We are all Africans: what mtDNA and Y chromosome mutations reveal about human origins and migrations

Gyan Bhanot
Rutgers University
Cancer Institute of New Jersey
Institute for Advanced Study, Princeton

gyanbhanot@gmail.com

University of Hawaii 03/20/2008; Brown University 03/31/2008, Bar Ilan Hebrew University 09/11/2008, Cornell University, 10/16/2009
Nothing in Biology Makes Sense Except in the Light of Evolution

Theodosius Dobzhansky (1900-1975)
Inferences from Fossils
Homo sapiens

chimps & bonobos

TMRCA = 5 MYA

Courtesy of Meave Leakey
Donald Johanson and “Lucy”

D. Johanson:
‘Lucy: The Beginnings of Humankind’

R. Leakey, R. Lewin:
‘People of the Lake: Mankind & Its Beginnings’
Every creature alive today had ancestors

Not all fossils had descendents
How Evolved are YOU?

• True or False?
  – The first living organisms were like bacteria
  – The Coelacanth is the living fossil of the first four limbed vertebrate
  – Humans evolved from chimpanzees
  – Modern hunter-gathers resemble humans before the advent of agriculture
  – The Basque are a paleolithic relict population
  – The Australian aboriginals are like the original people that migrated “Out of Africa” 50-70 kybp.

• ALL ARE FALSE!

• Fallacy of the Contemporary Ancestor:
  – Misguided anthropocentric view regards humans as the pinnacle of evolution. *Evolution has no direction or goal!*

• If evolution = genetic change, humans are *less* evolved than monkeys or mice - they have lower genetic diversity (*and perhaps “civilization” is driving them to phenotypic identity as well*)
Fallacy of Linear Evolution

• Examples of Stupid & Dangerous questions:
  – What is the genetic basis for low IQ?
  – Which languages/species are the “most evolved”?
  – What is the ancestral biological homeland of population X?
    • Genetic diversity cannot be traced to a single time and place. *Different segments of genome converge to different ancestors.*
    • Nordic people have Irish mtDNA. Are they Irish?
  – Where did my ancestors live, a thousand years ago?
    • After n generations, we have $2^n$ ancestors.
      – $n=30$ (600-900 ybp): Ancestors ~ $10^9$
      – $n=40$ (800-1200 ybp): Ancestors ~ $10^{12}$
    • **OUR ANCESTORS LIVED EVERYWHERE!**
Inferences from Genetics

Sequencing cost: 6 Giga Bases = 2 human genomes = $10,000
Biology 101

- Life starts as a single cell with 46 chromosomes, 23 from the female egg and 23 from the male sperm.
- The cell divides and differentiates into \( \sim 10^{13}-10^{14} \) cells (our bodies).
- We regenerate our bodies \( \sim 500 \) fold in a lifetime: \( 10^7 \) new cells per second!
- Cells multiply by Mitosis: demo
- Meiosis produces germ line cells: demo
The Human Y Chromosome

- ~58 MB, paternal inheritance, 90% non-recombining
- Few coding genes
- Palindrome, self-recombination in male meiosis.
- Few variants (~5000)
- Derived from X about 300,000,000 years ago.
mtDNA

- 16569 bp, non-recombinant, maternally inherited
- Bacterial Origin, Circular, Symbiosis ~ 2 billion YBP
- Encodes 37 genes
  - 22 tRNAs, 2 rRNAs,
  - 13 proteins (energy metabolism)
- 100-300,000 copies in cells
- mutatation rate 1/400 births
  10X higher than nDNA
Uniparental Inheritance of Y & mtDNA
18 haploid genomes

10 Generations

T = 0

T = 10

←18 haploid genomes→
Mutations mark the tree of evolution

Scientists try to read this record!

Somatic mutation – non-heritable
- **normal soma**: daughter cells carry mutation – ‘somatic mosaic’
- **cancer**: clonal expansion of mutant cells

Germ-line mutation – heritable
Mutation clues

T = 0

T = 10

↑10 Generations↓

eve1 eve2 eve3 eve4

←18 haploid genomes→
Two Main Models of Human Evolution

**MULTIREGIONAL HYPOTHESIS**

- African H. erectus
- European H. erectus
- Asian H. erectus
- Indonesian H. erectus
- Neanderthals
- Ngandong
- Modern Africans
- Modern Europeans
- Modern Asians
- Modern Australians

**OUT-OF-AFRICA HYPOTHESIS**

- African H. erectus
- European H. erectus
- Asian H. erectus
- Indonesian H. erectus
- Neanderthals
- Ngandong
- Modern Africans
- Modern Europeans
- Modern Asians
- Modern Australians
Distribution of Y Chr Groups
A LANDMARK STUDY OF THE HUMAN JOURNEY

Who was your first ancestor? New DNA studies say that all humans descended from an African ancestor who lived only 60,000 years ago. Uncover the specific paths that led from him to you—the ultimate human history, as written in our genes.

YOUR GENETIC JOURNEY

Explore your own genetic journey with Dr. Spencer Wells. DNA analysis includes a depiction of your ancient ancestors and an interactive map tracing your genetic lineage around the world and through the ages.

INTERESTED IN LEARNING MORE? FIND OUT HOW TO BECOME PART OF THE GENOGRAPHIC PROJECT AND DISCOVER YOUR OWN DEEP ANCESTRY.

GENETICS OVERVIEW

The human story is written in our genes. Explore the basics of genetics, from chromosomes and DNA to natural selection and genetic drift.

NEWS

- Global Gene Project to Trace Humanity’s Migrations
- Project Receives Approval From University Review Board

ALSO SEE

Video: Indigenous Representatives Talk About Their Migratory Histories (Download Windows Media)
- Deepest branches are exclusively sub-Saharan African

- TMRCA estimated at ~170,000 YBP

- TMRCA for the «out of Africa» group estimated at 50,000 YBP

Strong support to the recent African origin of modern humans


<table>
<thead>
<tr>
<th>Chr</th>
<th>Locus</th>
<th>Description</th>
<th>Samples</th>
<th>Origin</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>β Globin 2670 b</td>
<td>Hemoglobinopathies</td>
<td>326</td>
<td>African root</td>
<td>Harding et al, 1997</td>
</tr>
<tr>
<td>18</td>
<td>LPL 9734 b</td>
<td>Cardiovascular Diseases</td>
<td>71</td>
<td>Africa and elsewhere</td>
<td>Clark et al, 1998</td>
</tr>
<tr>
<td>16</td>
<td>MC1R 954 b</td>
<td>Skin pigment</td>
<td>356</td>
<td>Africa and elsewhere</td>
<td>Harding et al</td>
</tr>
<tr>
<td>X</td>
<td>PDHA1 4200 b</td>
<td>Neurological Diseases</td>
<td>35</td>
<td>Africa</td>
<td>Harris and Hey, 1999</td>
</tr>
</tbody>
</table>
B & F, which are groups found in Asia, are classified in the R sub-clade, which is mostly European.
Table 1. The ethnicities of the mtDNA data analyzed in this study

<table>
<thead>
<tr>
<th>Population</th>
<th>Ethnicity</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td></td>
<td>146</td>
</tr>
<tr>
<td>America</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Asia</td>
<td>China</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>674</td>
</tr>
<tr>
<td></td>
<td>Malaesia</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Other areas</td>
<td>86</td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Europe</td>
<td>Caucasian</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>Finland</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td>Italy</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Other areas</td>
<td>67</td>
</tr>
<tr>
<td>Unspecified areas</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1737</td>
</tr>
</tbody>
</table>

Rows = samples
Columns = mtDNA Loci
Green = 0 (no mutation)
Black = 1 (mutation)
900 Sequences in M and N
Out-group = L0/L1
Principal Component Analysis

Group 1 = N
Europe, Eurasia, N&S America

Group 2 = L0 / L1

Group 3 = L2

Group 4 = L3

Group 5 = M
Asia, China, Oceania, Australia

1737 mtDNA sequences
~20% variation
Haplogroups within Clades

- Data each clade
- 10-Fold Cross-Validation
  - $k=2, \ldots, k_{opt}$
  - k-level Ensemble Consensus Clustering
- Agreement Matrix
- Robust pattern identification
  - Tra/Test Protocol

Example: Partitioning / relocation

Agglomerative
Clustering Reveals tree

1. Sample equal size sub sets from k robust haplogroups
2. Repeat ensemble consensus clustering
3. Consensus tree
4. Root migration tree based on Poisson dynamics
5. Validate haplogroup assignment
Divide the Data into 2, 3, 4… $k_{opt}$ clusters
Major mtDNA CLADES

mtDNA ‘EVE’

150-200K Years

- **Africa**
- **Europe, Eurasia, N. and S. America**
- **Asia, Eurasia, Oceania, Australia**

---

Alexe et al, 2008. submitted
1-2-3 PCs for N Clade. Note location of B and A/J/T/U/H

Figure 5a. N-Clade PCA projections. The mnemonic in the middle assigns colors to the haplogroup labels for the samples in this clade. Note that B and A are much closer than B and any of J/T/H/U/V. The inference is that B and A separated later than B and J/T/H/U/V.
N Clade Tree
Time depth ~ 60,000 years

[Diagram of the N Clade Tree with clades and migration routes:
- Eastern Migration routes to Asia, N. & S. America
- Western Migration routes to Europe, Eurasia Americas]
Haplogroup JT → J and T

- **Haplogroup T** originated in Mesopotamia ~ 10,000 ybp and moved northwards. High concentrations around the eastern Baltic Sea.

- **Haplogroup J**: Defined by a mutation ~ 45,000 ybp in the DNA of a woman who lived in the Caucasus region. Further mutations in the J line are identified as J1a1 (27,000 yrs ago), J2a (19,000 yrs ago), J2b2 (16,000 years ago), J2b3 (5,800 yrs ago), etc.

- **Haplogroup J and T** associated with the spread of farming and herding in Europe during the Neolithic Era (8,000-10,000 yrs ago). All other West Eurasian-origin groups (H, V, U, K, W, I, X) were previously given to hunting and gathering.
M Clade Tree
Time depth ~ 60,000 years

SNP 5417, M7a/M7b, homoplasy
Jomon/Yayoi

Central and South Asia, Oceania, Australia, N. & S. America

M Clade Founder
Divergence of haplogroup M7 into subhaplogroups M7a, M7b, and M7c.

SNP 6455 and 9824 in M7a and M7b.
How Long Ago Did ‘mtDNA Eve’ Live?

\[ \frac{T_{\text{H-C}}}{T_{\text{H-H}}} = 24.3 \pm 1.7 \Rightarrow T_{\text{H-H}} = 206K \pm 14K \text{ if } T_{\text{H-C}} = 10^6 \]
M Clade
Migration

N Clade
Migration

L0, L1, L2
L3 Clades
in Africa

39 KYA

40 KYA

47 KYA

50 KYA

M Clade
Migration
mtDNA and migrations into Europe

(a) Neolithic 10–5 KYA

(b) Mesolithic glacial refugia
Across the Bering Straits into N. & S. America 11,000-15,000 YBP

Ocean Levels were lower during Ice-Age
SCANDINAVIANS HAVE SCOT AND IRISH mtDNA
Y-Chromosomes, Slavery and Colonization

~0.5 million Portuguese (1500–1808)

~2.4 million Native Americans (in situ)

~6 million Europeans (1808–)

~4 million Africans (1550–1850)
European Gypsies came from Rajasthan, India.
mtDNA and the Jewish Migrations out of Palestine
Where can I learn more?

• Take a Genetics Class

• Popular Books:
  – Brian Sykes: ‘The Seven Daughters of Eve’, ‘Adam’s Curse’

• Text Books:
  – Hartl, Clark: ‘Principals of Population Genetics’
  – J. Gillespie: ‘Population Genetics, a concise guide’
  – Jobling et al: ‘Evolutionary Pop. Genetics and Disease’
  – Ewens, ‘Mathematical Population Genetics’
How can I find my ancestral mtDNA and Y haplogroup?

How can I trace my ancestry?

- http://www.familytreedna.com/
HVS1 Sequence
Haplogroup: M*
16111T, 16223T, 16311C, 16519C

ATTCTAATTAAACTATTCTCTGTCTTTCTTTCAATGGGGGAAGCAGATTTGGGTA
CCACCCAAGTATTGACTCACCCCATCAACAACCGCTATGTATTTTCGTACATT
ACTGCCAGTTCACCATGAATATTGTACGGGTACCATAAATACTTGGACCACCTG
TAGTACATAAAAAACCAATCCACATCAAAAACCCCCCTCCCCCATGCTTACAAG
CAAGTACAGCAATCAACCGCTTCAACTATCACACATCAACTGCAACTCTCAAAG
CCACCCCTCACCCTACAGGATACCCAACAAACCTACCACCCACCCCTTAACAGTAC
ATAGCACATAAAGCCATTTCACCCTACATAGCACCATTAGTCAAATCCCTT
CTCGTCCCCATGGATGACCCCCCTCTAGATAGGGGTCCCTTGACCACCATCC
TCCGTGAAATCAATATCCGCACAAGAGTCTACTCTCTCCTCGCTCCGGGCC
CATAAACACTTGGGGGTAGCTAAAGGTGTAACACTGATTCCGACATCTGGTTCCA
CTTCAGGGCCATAAGGCCTAAATAGGCCACACGTCTCCCTTAATAGACACT
TCACGATG
Y-Chromosome Report

Haplogroup: R (M173)

STRs:

DYS393: 13, DYS439: 12
DYS388: 12, DYS385a: 12
DYS19: 15 DYS389-1: 14
DYS390: 23 DYS385b: 13
DYS391: 10 DYS389-2: 16
DYS426: 12 DYS392: 10
Typical Reports

- Maternal Lineage based on 10 SNPs plus HVS1 sequence
- Paternal Lineage based on Y chromosome STRs + SNPs
- mtDNA report 1, mtDNA report 2
- Y-Chromosome report
Open Questions
12 Major Language Groups (Greenberg) into which the 6000 languages in the world can be classified
Questions from fossils and archaeology

• Did early hominids (H. heidelbergensis, H. neanderthalensis) contribute to the modern gene pool?

• What were the relative contributions of Paleolithic and Neolithic settlers?

~800 KYA
heidelbergensis (antecessor)

~250 – 28 KYA
neanderthalensis

~40 KYA
Modern humans

~20 KYA
Glacial maximum

~10 KYA
Beginning of Neolithic transition
Were Neanderthals a Different Species or do we still carry their DNA?
How 50% of humanity views EVOLUTION!
And then there are the doubters
The Real Promise of Genetics

- Individualized therapy: markers for disease risk, prognosis and treatment efficacy.

- Understanding and treating complex diseases: Cancer, Diabetes, Heart disease, Parkinson’s Disease, Alzheimers, Obesity, Hypercholesterolemia

- Understanding Complex Traits: Longevity, Athleticism, Height, Musical Ability

- Making effective Vaccines: HPV, FLU, HIV/AIDS, Hepatitis
Physics is what physicists do at night.

R. Feynman
Physics is what physicists do at night.

Who is this physicist?
Venki Ramakrishnan 2009 Nobel Prize for Chemistry
TIME TO LEARN A NEW GAME