The Genetic Structure of the Jewish People

Doron M Behar
Molecular Medicine Laboratory
Rambam Health Care Campus
Haifa, Israel

From Generation to Generation: The Genetics of Jewish Populations
Stanford, October 11, 2012
Outline

- Review of some genetics
- Principles of archaeogenetics
- Archaeogenetics of the Jewish people
  - Maternal ancestry
  - Paternal ancestry
  - Genome-wide structure
Why?  When?  Where?

Forensic Medicine

Pharmacogenomics Whole genome associations

Phylogeography Archaeogenetics Ancestry
Review of genetics

- Each human cell has 46 chromosomes plus multiple copies of mitochondrial DNA (mtDNA)
- 22 pairs of autosomes (chromosomes 1 to 22)
- One pair of sex chromosomes
  - XX female (X from both mom & dad)
  - XY male (X from mom, Y from dad)
- Numerous copies of mitochondrial DNA
Phenomena affecting our genome

- Mutations/ Recombination
- Drift (Founding Event/ Bottlenecks)
- Natural Selection (Positive/ Negative)
- Migration
  - Short
  - Long
**Drift**

- A change in the gene pool of a small population that takes place strictly by chance.
  - Founding Event
  - Bottleneck
Migration

Isolated small populations

Short migration and merging

Long migration

Population growth
Mutations

A change in the genetic material:

- **Mechanism:** SNP, Deletion, Insertion, Duplication

- **Functional:** silent polymorphisms vs. gain or loss of function

- **Phylogeographic:** Unique Event Polymorphism (UEP)/Repeated Event
Mutations

- SNP: One nucleotide is changed
- Deletion: One/ few nucleotides are deleted
Recombination

- Autosomal chromosomes recombine. Hence, the copy of (say) chromosome one you got from your mom is a mixture of the copy she got from her mom (your maternal grandmother) and her dad (your maternal grandfather).
Suppose she has three offspring. Each time she contributes a copy of chromosome one to a child, it is usually a mixture of (roughly) half of each of her two copies:

What about the X and the Y?
Chromosomes and Sex Determination

Mother

Father

Daughter
Son
Daughter
Son

Daughter
Son
Daughter
Son
Chromosomes and Sex Determination

Mother

Father

XXX

XY

XX

XY
The study of Ancestry

- Each parent contributes one copy of each autosomal chromosome
- Y chromosomes only pass through males, hence a marker for direct male lineages
- mtDNA is strictly maternally inherited
- Hence, if male
  - Half your autosomes came from each parent
  - Your Y came from your dad, your X from mom
  - Your mtDNA came from mom
- If female
  - Again half your autosomes came from each parent
  - You got an X from both dad and mom
  - Your mtDNA came from mom
Y-Chromosome Genes

- Addiction to violent movies
- Switching TV channels
- Overconfidence
- Remembering jokes
- Difficulty in expressing feelings
- Difficulty in remembering anniversaries
The mitochondrial DNA

- No recombination
- Uniparental mode of transmission
- Not prone to selection
- Prone to drift
- Therefore, it is possible to combine markers into unambiguous phylogenetic trees which are highly informative with respect to shared ancestry
mtDNA Haplogroup M distribution

- Indian (M2-M6, +)
- E-Asian (M7-M10)
- African (M1)
- Papuan (Q)
West-Eurasian mtDNA lineages (excluding Indian and Berber U and Arabian pre-HV)

East-Eurasian mtDNA lineages (M excludes M1)

N1, X, W, HV, TJ, U

M, A, N9, R9(F), B

African mtDNA lineages

L1, L2, L3

zones of admixture or source?
Global distribution of the uniploids

Y-Chromosome

Mitochondrial DNA

The data in this map is supposed to represent the situation before the recent European expansion beginning about 1000 AD. In some cases such as some Native American tribes and the Maori this can be done readily, because STR typing is done. In other cases, especially in America, this has not been done. In this map, the only language group is American. The map indicates that there is no difference between the results of the previous map and this one.

Copyright © 2005 J. D. McDonald

Specific times in locations are shown at left. Unidentified past are for general population in the area. All maps, American and especially Western European areas are very vague. The data in this chart is supposed to represent the uniploids that European expansion beginning about 1000 AD.

Assignments in Australia are somewhat rough.
...events just happen, one after another. They don’t care who knows.

but history... ahh, history is different. History has to be observed.

Otherwise,

It’s not history. It’s just...well, events happening one after another...

Terry Pratchett/ Small Gods
...mutations just happen, one after another. They don’t care who knows.

but evolution...ahh, evolution is different. Evolution has to be observed.

Otherwise,

It’s not evolution. It’s just...well, mutations happening one after another...

Modified, without permission, from Terry Pratchett/ Small Gods
Phylogenies & haplogroups

Root

...A...C...T

C → G

A → T

T → A

...A...C...T

...A...G...T

...T...G...T

...T...G...A

Ancestral

Descendant
Phylogeographic Polymorphism

Unique events

- Single Nucleotide Polymorphism - SNP
- Insertions, Deletions
- Ancient events (>10k years)
- Stable polymorphisms
- More likely to occur in coding regions

Multiple Events

- Biallelic hot spot site
- Genetic loci
  - Control region in the mtDNA
  - Short tandem repeats (microsatellites)
- Allow a “genetic clock”
- More likely to occur in “non-coding regions”
mtDNA 16569bp bp1 16383 00057
Control region

Reference sequence

Combination of HVS-II SNPs

Combination of Coding region SNPs

What is a Haplotype?

What is a Haplogroup?

917bp

359bp HVS-I

16024

16569bp Coding region
# Haplogroups and Haplotypes

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1c2</td>
<td>16129 16187 16189 16223 16265 16278 16286</td>
</tr>
<tr>
<td>L3f</td>
<td>16209 16223 16311</td>
</tr>
<tr>
<td>M1</td>
<td>16129 16189 16223 16249 16311</td>
</tr>
<tr>
<td>Z</td>
<td>16185 16223 16224 16260 16298</td>
</tr>
<tr>
<td>T1</td>
<td>16126 16163 16186 16189 16294</td>
</tr>
<tr>
<td>J2</td>
<td>16069 16126 16193</td>
</tr>
<tr>
<td>K</td>
<td>16224 16311</td>
</tr>
<tr>
<td>U5a1a</td>
<td>16256 16270 16399</td>
</tr>
</tbody>
</table>
Phylogenies and time scales

The Dawn of Human Matrilineal Diversity

Doton M. Behar,1,3,5,6 Richard Villems,3,5,6 Himla Soodyall,1,3 Jason Blue-Smith,4 Luipa Pereira,3,4,6 Ene Metspalu,7 Rosaria Scozzari,7 Heesen Makkan,7 Shay Tsur,3 David Comas,8 Jaume Retnanpeit,8 Lluís Quintana-Murci,9 Chris Tyler-Smith,3,6 E. Spencer Wells,4 Saharon Rosset,10,11 and The Genographic Consortium14

The American Journal of Human Genetics 82, 1–11, May 2008
The ability to set and time migration
~65,000 ybp
European Commission has just announced an agreement whereby English will be the official language of the European Union rather than German, which was the other possibility.

As part of the negotiations, the British Government conceded that English spelling had some room for improvement and has accepted a 5-year phase-in plan that would become known as "Euro-English".

In the first year, "s" will replace the soft "c". Certainly, this will make the sivil servants jump with joy.

The hard "c" will be dropped in favour of "k". This should klear up konfusion, and keyboards kan have one less letter.

There will be growing publik enthusiasm in the sekond year when the troublesome "ph" will be replaced with "f". This will make words like fotograf 20% shorter.
In the 3rd year, publik akseptanse of the new spelling kan be expekted to reach the stage where more komplikated changes are possible.

Governments will enkourage the removal of double letters which have always ben a deterent to akurate speling.

Also, al wil agree that the horibl mes of the silent "e" in the languag is disgrasful and it should go away.

By the 4th yer people wil be reseptiv to steps such as replasing "th" with "z" and "w" with "v".

During ze fifz yer, ze unesesary "o" kan be dropd from words kontaining "ou" and after ziz fifz yer, ve vil hav a reil sensibl riten styl.

Zer vil be no mor trubl or difikultis and evrivun vil find it ezi tu understand ech oza. Ze drem of a united urop vil finali kum tru.

Und efter ze fifz yer, ve vil al be speking German like zey vunted in ze forst plaś.

(anonymous)
Levantine origin followed by a highly complex history during forming the Diaspora
The maternal ancestry of Ashkenazi Jews

<table>
<thead>
<tr>
<th></th>
<th>K</th>
<th>N1b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Med NJ</td>
<td>6%</td>
<td>2%</td>
</tr>
<tr>
<td>Eur NJ</td>
<td>6%</td>
<td>0</td>
</tr>
<tr>
<td>Ash</td>
<td>32%</td>
<td>10%</td>
</tr>
</tbody>
</table>
Global network of mtDNA lineages

- India
- Europe
- East Asia
- Africa
Establishing the variation
Complete sequences Based network of Hg K
Combining topology and geography

Three different Ashkenazi lineages dominate the Ashkenazi mtDNA K haplogroup. These three lineages are virtually restricted to Jews and have expanded due to drift imposed by a strong founder effect.
## The maternal ancestry of non-Ashkenazi Jews

<table>
<thead>
<tr>
<th>Community</th>
<th>Estimated size</th>
<th>N lineages</th>
<th>% in the community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashkenazi</td>
<td>8,000,000</td>
<td>4</td>
<td>40.3</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>100,000</td>
<td>1</td>
<td>58.6</td>
</tr>
<tr>
<td>Georgia</td>
<td>100,000</td>
<td>1</td>
<td>58.1</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>120,000</td>
<td>5</td>
<td>41.4</td>
</tr>
<tr>
<td>Bene Israel</td>
<td>65,000</td>
<td>4</td>
<td>67.6</td>
</tr>
<tr>
<td>Cochin</td>
<td>10,000</td>
<td>4</td>
<td>66.7</td>
</tr>
<tr>
<td>Iran</td>
<td>&gt;150,000</td>
<td>7</td>
<td>47.6</td>
</tr>
<tr>
<td>Iraq</td>
<td>&gt;250,000</td>
<td>5</td>
<td>50.4</td>
</tr>
<tr>
<td>Libya</td>
<td>&gt;40,000</td>
<td>2</td>
<td>57.8</td>
</tr>
<tr>
<td>Morocco</td>
<td>&gt;250,000</td>
<td>5</td>
<td>24.2</td>
</tr>
<tr>
<td>Tunisia</td>
<td>&gt;100,000</td>
<td>4</td>
<td>43.2</td>
</tr>
<tr>
<td>Belmonte</td>
<td>400</td>
<td>1</td>
<td>93.3</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>&gt;50,000</td>
<td>4</td>
<td>26.8</td>
</tr>
<tr>
<td>Turkey</td>
<td>&gt;500,000</td>
<td>4</td>
<td>17.1</td>
</tr>
<tr>
<td>Yemen</td>
<td>&gt;50,000</td>
<td>5</td>
<td>47.1</td>
</tr>
</tbody>
</table>
Counting the Founders: The Matrilineal Genetic Ancestry of the Jewish Diaspora

Doron M. Behar, Ene Metspalu, Toomas Kivisild, Saharon Rosset, Shay Tzur, Yarin Hadid, Guennady Yudkovsky, Dror Rosengarten, Luisa Pereira, Antonio Amorim, Ildus Kutuev, David Gurwitz, Batsheva Bonne-Tamir, Richard Villems, Karl Skorecki
Y-chromosome studies

Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes


*Laboratory of Molecular Systematics and Evolution, University of Arizona, Tucson, AZ 85721, U.S.A. †Department of Genetics, Università degli Studi di Pavia, Pavia 27100, Italy. ‡Wadsworth Medical School, Hebrew University of Jerusalem, Jerusalem 91120, Israel. †‡Department of Genetics, University of Leicester, Leicester LE1 9HF, England. †§SARIAH, University of Witwatersrand, Johannesburg, South Africa. §Department of Pediatrics, New York University Medical Center, New York, NY 10016, and Department of Human Genetics, Sackler School of Medicine, Ramat Aviv 92135, Israel.

Communicated by Arno G. Motulsky, University of Washington, Seattle, WA, March 16, 2000 (received for review November 17, 1999).

Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations

Multiple Origins of Ashkenazi Levites: Y Chromosome Evidence for Both Near Eastern and European Ancestries

Doron M. Behar,1,2,* Mark G. Thomas,2 Karl Skorecki,1 Michael F. Hammer,4 Ekaterina Bulygina,2 Dror Rosengarten,1 Abigail L. Jones,2 Karen Held,2 Vivian Moses,2 David Goldstein,3 Neil Bradman,2 and Michael E. Weale2

Distribution of Hg J1e-P58 (containing the CMH)
The paternal ancestry of Ashkenazi Jews

Haplogroup frequencies  Founding events

Diagram showing the paternal ancestry of Ashkenazi Jews with haplogroup frequencies and founding events. The diagram illustrates the Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations.

DOI 10.1007/s00439-003-1072-7

Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations

Doreen M. Behar · Daniel Garrigan · Matthew E. Kaplan
Zahra Mobasher · Bror Rosengarten · Tatiana M. Karalet
Luis Quintana-Murci · Harry Oster · Kurt Skorecki
Michael F. Hammer
## Distribution of Hg R1a1-M17/M198 among Ashkenazi Levites

### Haplogroup Frequencies across Six Jewish and Four Non-Jewish Sample Sets

<table>
<thead>
<tr>
<th>HAPLOGROUP</th>
<th>AC (n = 76)</th>
<th>SC (n = 69)</th>
<th>AL (n = 60)</th>
<th>SL (n = 31)</th>
<th>AI (n = 100)</th>
<th>SI (n = 63)</th>
<th>Ger (n = 88)</th>
<th>Nor (n = 83)</th>
<th>Sorb (n = 112)</th>
<th>Bel (n = 306)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y*(xBR,A3b2)</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.159</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>BR*(xDE3,JR)</td>
<td>0.0263</td>
<td>0.0290</td>
<td>0.0333</td>
<td>0.1935</td>
<td>0.1400</td>
<td>0.0794</td>
<td>0.3295</td>
<td>0.4458</td>
<td>0.1786</td>
<td>0.2582</td>
</tr>
<tr>
<td>E*(xE3a)</td>
<td>0.0395</td>
<td>0.0435</td>
<td>0.2000</td>
<td>0.0968</td>
<td>0.2200</td>
<td>0.1905</td>
<td>0.0341</td>
<td>0.0120</td>
<td>0.0625</td>
<td>0.0458</td>
</tr>
<tr>
<td>J</td>
<td>0.8684</td>
<td>0.7536</td>
<td>0.1000</td>
<td>0.3226</td>
<td>0.3700</td>
<td>0.3651</td>
<td>0.0341</td>
<td>0.0241</td>
<td>0.0268</td>
<td>0.0327</td>
</tr>
<tr>
<td>K*(xL,N3,O2b,P)</td>
<td>0.0263</td>
<td>0.0290</td>
<td>0.0333</td>
<td>0.2258</td>
<td>0.0800</td>
<td>0.1270</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>L</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0323</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0114</td>
<td>0.0361</td>
<td>0.0089</td>
<td>0.1046</td>
<td>...</td>
</tr>
<tr>
<td>P*(xR1a)</td>
<td>0.0263</td>
<td>0.0870</td>
<td>0.1167</td>
<td>0.0968</td>
<td>0.1500</td>
<td>0.2063</td>
<td>0.4659</td>
<td>0.2651</td>
<td>0.0893</td>
<td>0.0425</td>
</tr>
<tr>
<td>R1a*</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>R1a1</td>
<td>0.0132</td>
<td>0.0580</td>
<td>0.5167</td>
<td>0.0323</td>
<td>0.0400</td>
<td>0.0159</td>
<td>0.1250</td>
<td>0.2169</td>
<td>0.6339</td>
<td>0.5098</td>
</tr>
</tbody>
</table>

* Population sample labels: AC = Ashkenazi Cohen; SC = Sephardi Cohen; AL = Ashkenazi Levite; SL = Sephardi Levite; AI = Ashkenazi Israeli; SI = Sephardi Israeli; Ger = German; Sorb = Sorb; Bel = Belarusian; Nor = Norwegian.

---

**Multiple Origins of Ashkenazi Levites: Y Chromosome Evidence for Both Near Eastern and European Ancestries**


1. Bar-Ilan University; 2. The Center for Population Genetics and Human Health, Department of Biology, University College London, London; and 3. Division of Biotechnology, University of Arizona, Tucson, AZ.
Global distribution of Hg R1a1

- > 2,000 R1a1 samples
- No good internal structure
- STR haplotypes are not informative
- Marker M458 reaches frequencies of 30-70% in East Europe
Origin of Ashkenazi Levites

- A strong founding event within the last 2,000 years
- >50% of contemporary Ashkenazi Levite
- Close relatedness to non-Jewish groups
- Unresolved origin
Direct Parental Ancestry

4th

3rd

2nd

1st
Ancestry at the population level

1st

2nd

3rd

4th
Autosomal signals

- If two individuals share a common ancestor k generations back, then the chance they share the same allele from that ancestor is \((1/2)^{2k-1}\):
  - For \(k = 1\), this is 50%
  - For \(k = 2\), this is 12.5%
  - For \(k = 3\) this is 3.1%
  - For \(k = 4\), this is 0.78%
  - For \(k = 5\), this is 0.19%
  - For \(k = 6\), this is 0.05%
  - For \(k = 7\), this is 0.01%

- However, Individuals belonging to the same population share hundreds of parental ancestor
19 (14) Jewish communities  N=184 (121)
118 (69) Old World populations  N=1936 (1166)
(incl. HGDP, Hapmap Gujaratis)
ca. 600 000 SNP (Illumina)
Admixture mapping $K=8$
Three Zooms

Zoom 1

Zoom 2

Zoom 3

- Arabia
- Caucasus
- Central Asia
- Iran
- East Asia
- Europe
- Near East
- North & East Africa
- Siberia
- South Asia
- Sub-Saharan Africa
- Jews
European and North African Jews

Yemenite Jews

Middle East and Caucasus Jews
Khazaria and Ashkenazi Jews

Caucasus

Ashkenazi Jews

24 Caucasus populations
214 Samples
Illumina arrays
Yunushbaev et al. 2012
Bene Israel (Jews of Mumbai)

Bene Israel girls (circa 1913).

Y-chromosome – has a link to Levant

mtDNA – mostly autochthonous Indian

Haplogroup J, J1e-P58 (20%) in particular (incl. Cohen Modal Haplotype) provides a link between Levant and Bene Israel.
...Support for a Jewish contribution to the Lemba gene pool is, nevertheless, found in the presence, at high frequency in the Lemba, of the CMH (.088 of the entire population and .135 of UEP group 1)...Of particular interest is the Buba clan...
**Scientific Correspondence**

*Y chromosomes of Jewish priests*

Karl Skorecki, Sara Selig, Shraga Blazer, Robert Bradman, Neil Bradman, P. J. Waburton, Monica Ismailowicz & Michael F. Hammer

**Origins of Old Testament priests**

Mark G. Thomas, Karl Skorecki, Haim Ben-Amid, Tudor Parfitt, Neil Bradman & David B. Goldstein

---

**Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood**

Michael F. Hammer, Doron M. Behar, Tatiana M. Karafet, Fernando L. Menezes, Brian Hallmark, Tamar Erez, Lev A. Zhivotovsky, Saharon Rosset & Karl Skorecki
Uniparental ancestry among Lemba

Y-chromosome (N=164)

mtDNA (N=164)

• J1c3d: 7/12 among Buba
• Buba: 7/14 are J1c3d
Uniparental ancestry among Lemba

Y-chromosome

- African: 33%
- Middle Eastern: 67%

mtDNA

- African: 100%
- Middle Eastern: 0%
Most Jewish samples form a remarkably tight subcluster that overlies Druze and Cypriot samples but not samples from other Levantine populations or paired Diaspora host populations. These results cast light on the variegated genetic architecture of the Middle East, and trace the origins of most Jewish Diaspora communities to the Levant.
Acknowledgements

Tartu
Estonian Biocentre
Mait Metspalu
Bayazit Yunushbaev
Gyaneshwer Chaubey
Monika Karmin
Jüri Parik
Tuuli Reisberg
Mari Järve
Siiri Rootsi
Toomas Kivisild
Ene Metspalu
Richard Villems

Barcelona
David Comas
Porto
Luisa Pereira

Cambridge
Toomas Kivisild

Ufa
Elza K. Khusnudtinova

Moscow
Oleg Balanovsky

Thessaloniki
Costas Triantaphyllidis
Anastasia Kouvatsi

Paris
Lluis Quintana-Murci

Tucson
Michael F. Hammer
Tanya Karafet

Tel Aviv
Saharon Rosset
David Gurwitz
Batsheva Bonne-Tamir

Jerusalem
Sergio DellaPergola

Haifa
Shay Tzur
Guennady Yudkovsky
Sara Selig
Karl Skorecki

London
Neil Bradman
Tudor Parfitt

Pavia
Ornella Semino
Antonio Torroni