Family History Technology

Growing the Family Tree:

The Power of DNA

in Reconstructing Family Relationships

Luke A. D. Hutchison

Natalie M. Myres

Scott R. Woodward

Sorenson Molecular Genealogy Foundation (smgf.org)

Our Genetic Identity

- ***Every living individual has a unique genetic identity***
- ***This identity is a formed as a combination of the genetic signatures of ancestors, and is passed on to become part of future generations***
- ***We are thus intrinsically linked to, and part of, our forebears and our descendants***

Family History Technology

No man is an island

- ***"No man is an island, entire of itself; every man is a piece of the continent, a part of the main. . . every man's death diminishes me, because I am involved in mankind."***
  [John Donne, Meditation XVII]

- ***Knowing our ancestors helps us know ourselves***

Family History Technology

Molecular Genealogy

- ***Molecular (or genetic) genealogy is the application of***

DNA analysis techniques,
statistical population genetics
algorithmic analysis

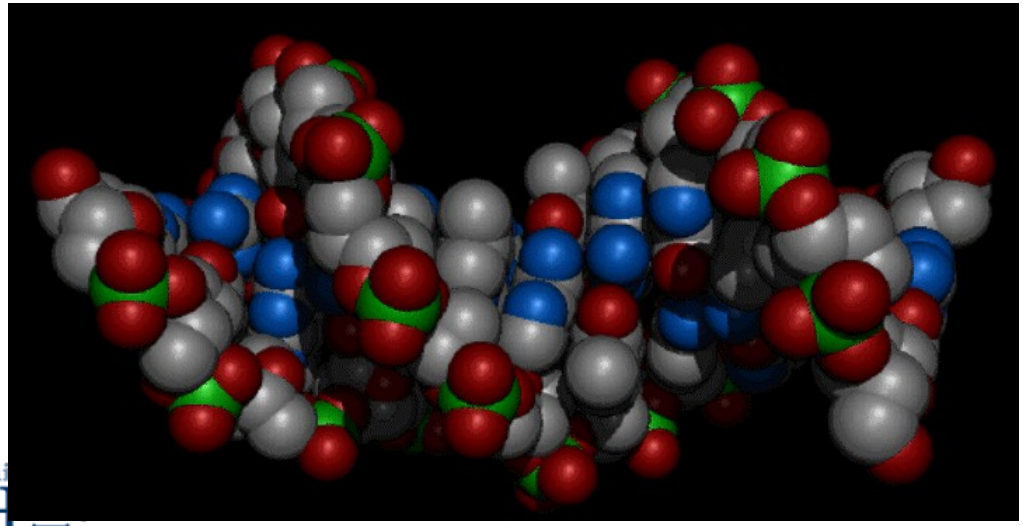
to the task of

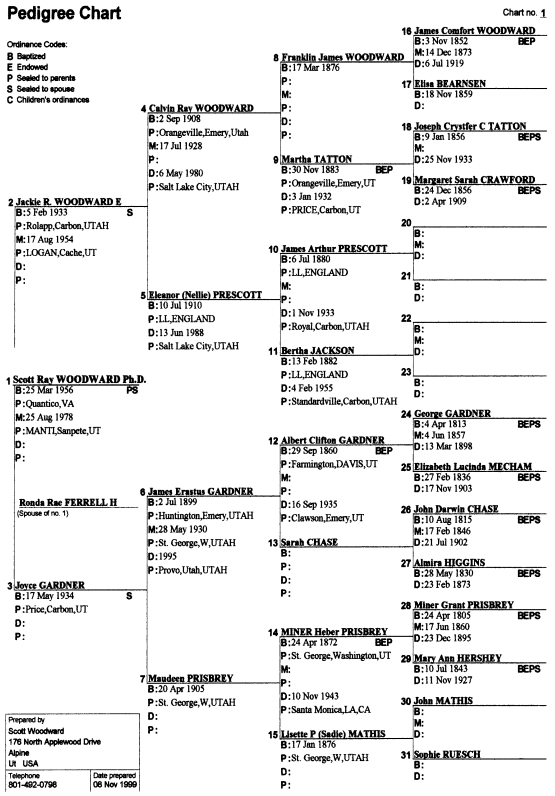
reconstructing unknown genealogies from the genetic and genealogical information of living individuals.

Family History Technology

Sorenson Molecular Genealogy Foundation

- ***The Sorenson Molecular Genealogy Foundation (www.smgf.org) is building the world's largest database of correlated genetic and genealogical information***





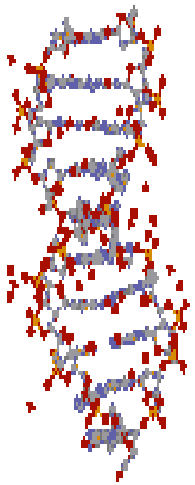
Family History Technology

Sorenson Molecular Genealogy Foundation

- ***Progress so far:***
  - **DNA and genealogies collected from over 47,000 volunteers**
  - **Up to 170 genetic markers analyzed**
  - **Pedigree charts extended as far as genealogical databases allow, to include over 1 million ancestral records**

Types of DNA

Y chromosome
(only males)



Autosomal DNA
(all individuals)

Mitochondrial DNA
(males and females; passed on female line)

Types of Genetic Data

- ***DNA sequence data: A, G, C, T***

- ***SNPs***

```
Reference Sequence: TAATCTGCCTTTACTTTTTCGGTACTGGAGAGCGTTTTTGTCCTATCCTCAGCAACTTCTAAGTTGTAATACGTAGAATT
Comparison Sequence: TAATCCGTCTTTACTTTTTCGGTGCTGGAGAGCGTTTTTGTCCTATCCTCAACGACTTCTGAGTTGTAATATGTAAAATT
```

- ***STRs / Microsatellite loci***

```
CTATTCATTCAATCATACACCCA ... TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA ... CAGACCCACCACAAAGAATC
Forward Primer Region          Repeat Region (Repeat Unit = TCTA; Number of Repeats = 10)          Reverse Primer Region
```

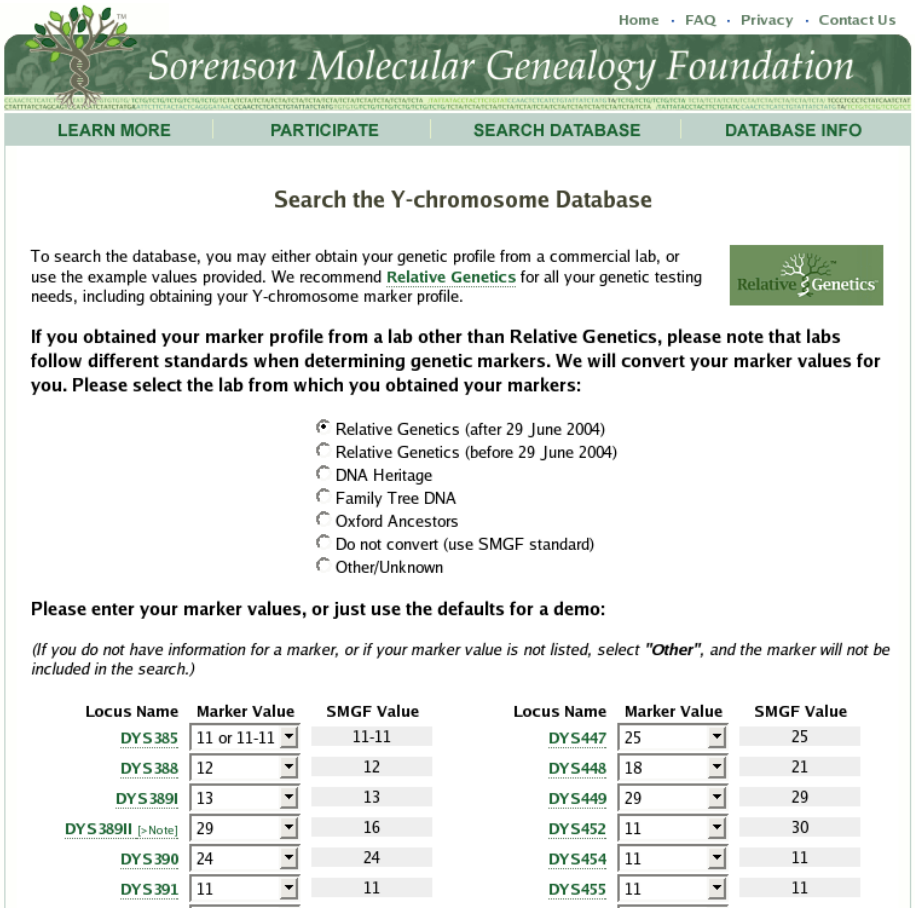
Family History Technology

Genetic Inheritance Models (Ycs)

- ***Y Chromosome (Ycs)***
  - Follows male (paternal) line
  - Single-stranded (haploid)
  - Same inheritance model as surname in many societies
  - Immediately useful to genealogists: correlation between Ycs patterns and surnames
  - Can search for similar Y chromosomes today on www.smgf.org

It's available now

- ***Search for potential paternal-line surnames and ancestors today***
- ***New smgf.org***



Home · FAQ · Privacy · Contact Us

Sorenson Molecular Genealogy Foundation

LEARN MORE | PARTICIPATE | SEARCH DATABASE | DATABASE INFO

Search the Y-chromosome Database

To search the database, you may either obtain your genetic profile from a commercial lab, or use the example values provided. We recommend Relative Genetics for all your genetic testing needs, including obtaining your Y-chromosome marker profile.

If you obtained your marker profile from a lab other than Relative Genetics, please note that labs follow different standards when determining genetic markers. We will convert your marker values for you. Please select the lab from which you obtained your markers:

- (•) Relative Genetics (after 29 June 2004)
- ( ) Relative Genetics (before 29 June 2004)
- ( ) DNA Heritage
- ( ) Family Tree DNA
- ( ) Oxford Ancestors
- ( ) Do not convert (use SMGF standard)
- ( ) Other/Unknown

Please enter your marker values, or just use the defaults for a demo:

*(If you do not have information for a marker, or if your marker value is not listed, select **"Other"**, and the marker will not be included in the search.)*

Locus Name	Marker Value	SMGF Value	Locus Name	Marker Value	SMGF Value
DYS385	11 or 11-11	11-11	DYS447	25	25
DYS388	12	12	DYS448	18	21
DYS389I	13	13	DYS449	29	29
DYS389II [>Note]	29	16	DYS452	11	30
DYS390	24	24	DYS454	11	11
DYS391	11	11	DYS455	11	11

Family History Technology

Matching Y Chromosome Profiles

Results

Congratulations! Genetic profiles similar to yours have been found within the SMGF database.

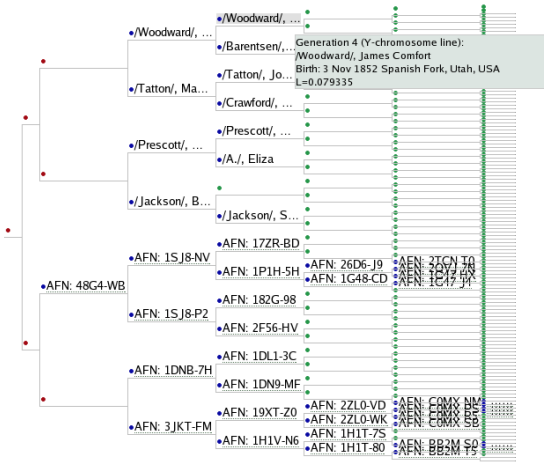
KEY: Locus Match | No Match | No Data

Click on the buttons to view the pedigree for each match result, and the buttons to view the likelihood of sharing a Most Recent Common Ancestor with the match result at each ancestral generation.

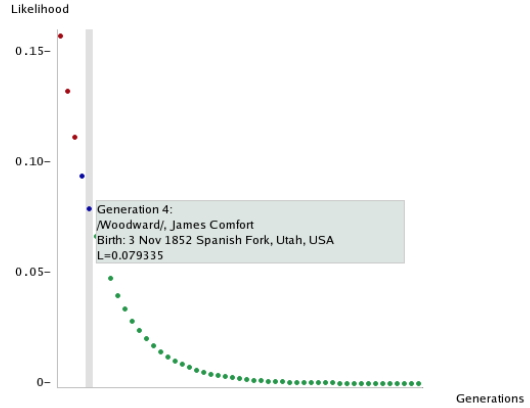
If you would like to re-run this search on a regular basis, you may bookmark this link.

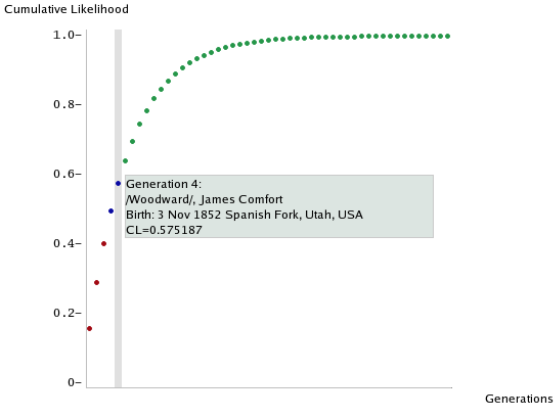
Result Number	Pedigree (CLICK)	MRCA (CLICK)	Num. matching loci	DYS385	DYS388	DYS389I	DYS389B [> Note]	DYS390	DYS391	DYS392	DYS393	DYS394/19	DYS426	DYS437	DYS438	DYS439	DYS441	DYS442	DYS444	DYS445	DYS446	DYS447	DYS448	DYS449	DYS452	DYS454	DYS455	DYS456	DYS458	DYS459	DYS460	DYS461	DYS462	DYS463	GGAAT1B07	YCAII	Y-GATA-A10	Y-GATA-C4	Y-GATA-H4
				11 11	12	13	16	24	11	13	13	15	12	14	12	11	13	12	12	13	25	21	29	30	11	11	15	16	9 10	11	11	11	24	10	19 23	12	23	11
1			36/36	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2			36/36	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3			32/36	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗
4			32/36	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓
5			31/36	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
6			31/36	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
7			31/36	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓

Pedigree of Match #1



Likelihood of Most Recent Common Ancestor (MRCA) Shared With Match #1

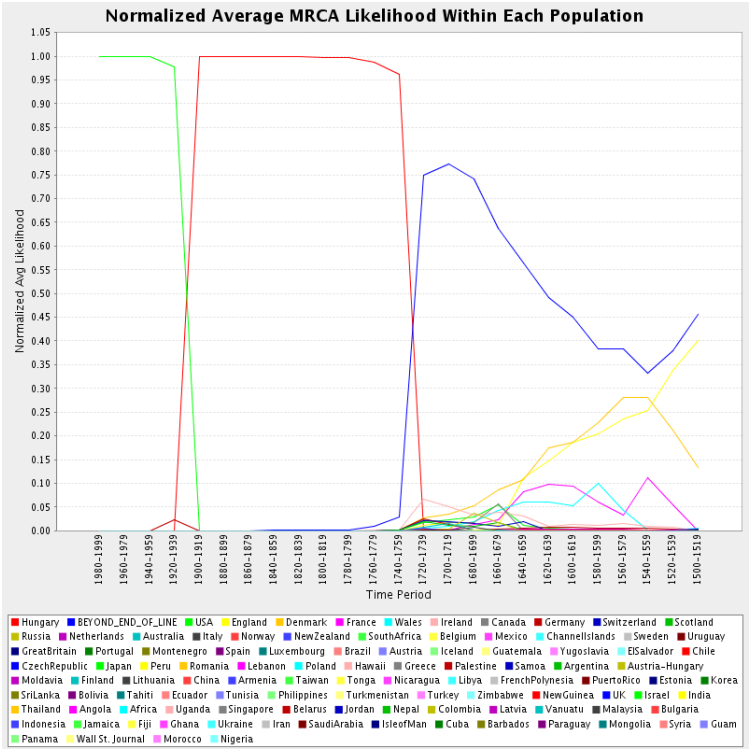




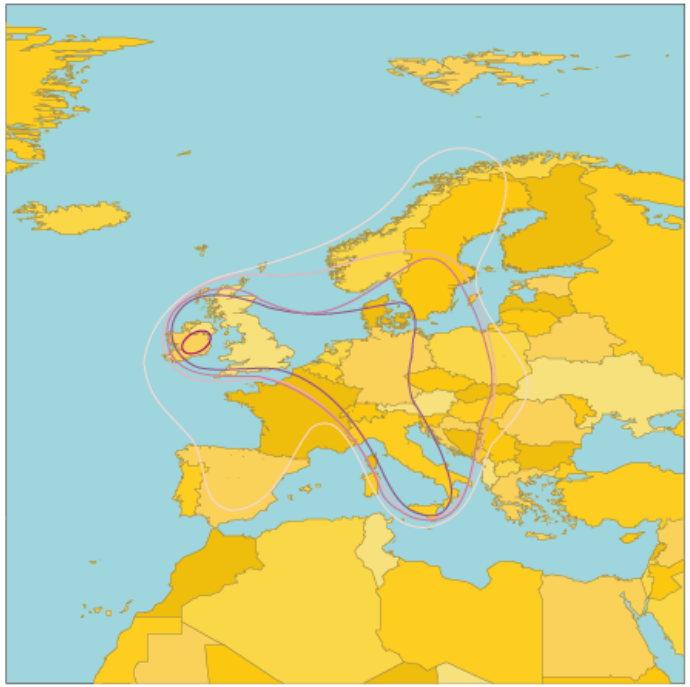
Family History Technology

Example for Surname: Anders

- ***Genetics show what the name does not intuitively show***



24-Locus Most Likely Geographic Assignment



Resolution	Likelihood of Assignment	Avg ttMRCA	StDev ttMRCA
1	0.999	140.307	34.474
2	0.566	132.387	36.765
3	0.459	128.705	38.794
4	0.357	117.732	34.523
5	0.963	64.200	20.921

Resolution	Associated Surnames	Count	ttMRCA
5	GLENNON	1	47
5	WHITE	1	48
5	SMILEY	1	59
5	LINDSEY	1	69
5	MILES	1	98
4	SHAFER	1	62
4	CAPOZZA	1	74
4	CLARK	1	75
4	GUNTER	1	78
4	BAEST	1	79
4	BRECKON	1	85
4	HILLYARD	1	87
4	HAYES	1	88
4	CAPPS	1	89
4	JENSEN	1	90
4	ANDERSSON	1	91
4	HARVEY	1	92
4	GUEST	1	93
4	DAVIS	1	96
4	PLUMB	1	99

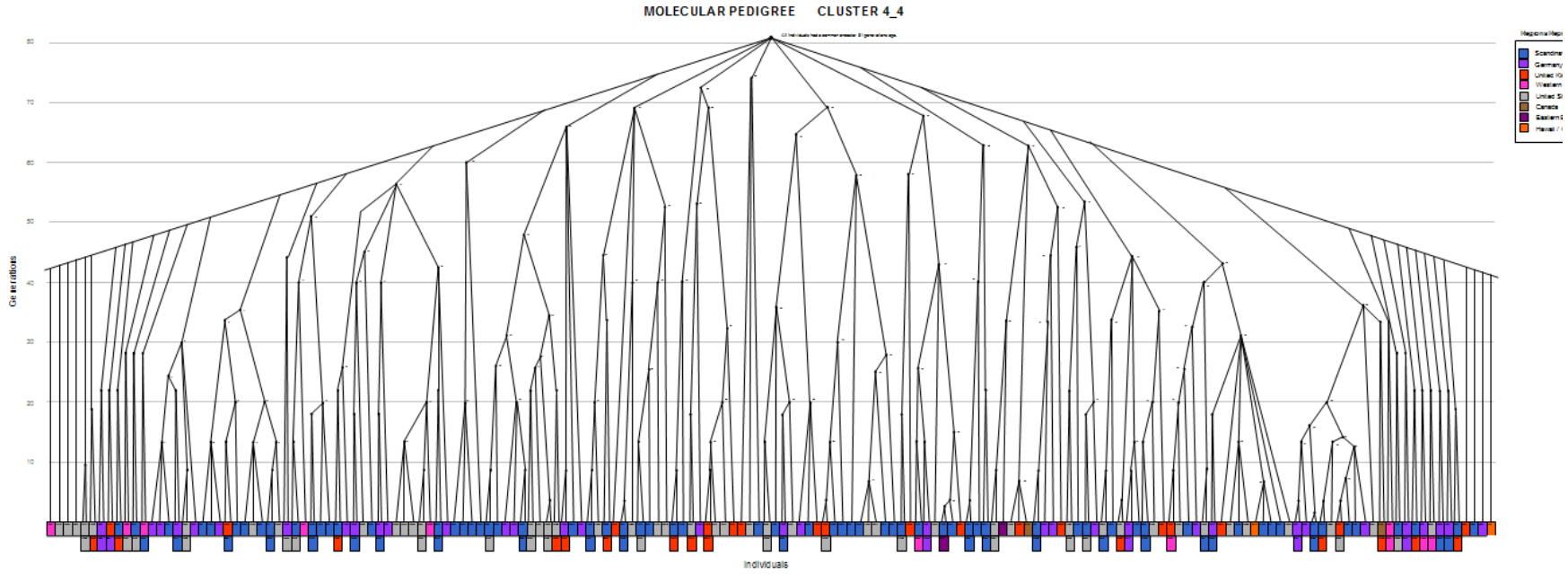
Genetic Inheritance Models (Ycs)

- ***Y Chromosome (Ycs)***
  - Forward through time: forms a tree structure
  - Backward through time: follows a single line

- ***Paternally-related populations***
  - No recombination of Ycs DNA (it is haploid)
  - Haploid populations behave differently from traditional populations
    - not affected by inbreeding
    - Population contractions are like slow expansions followed by fast expansions

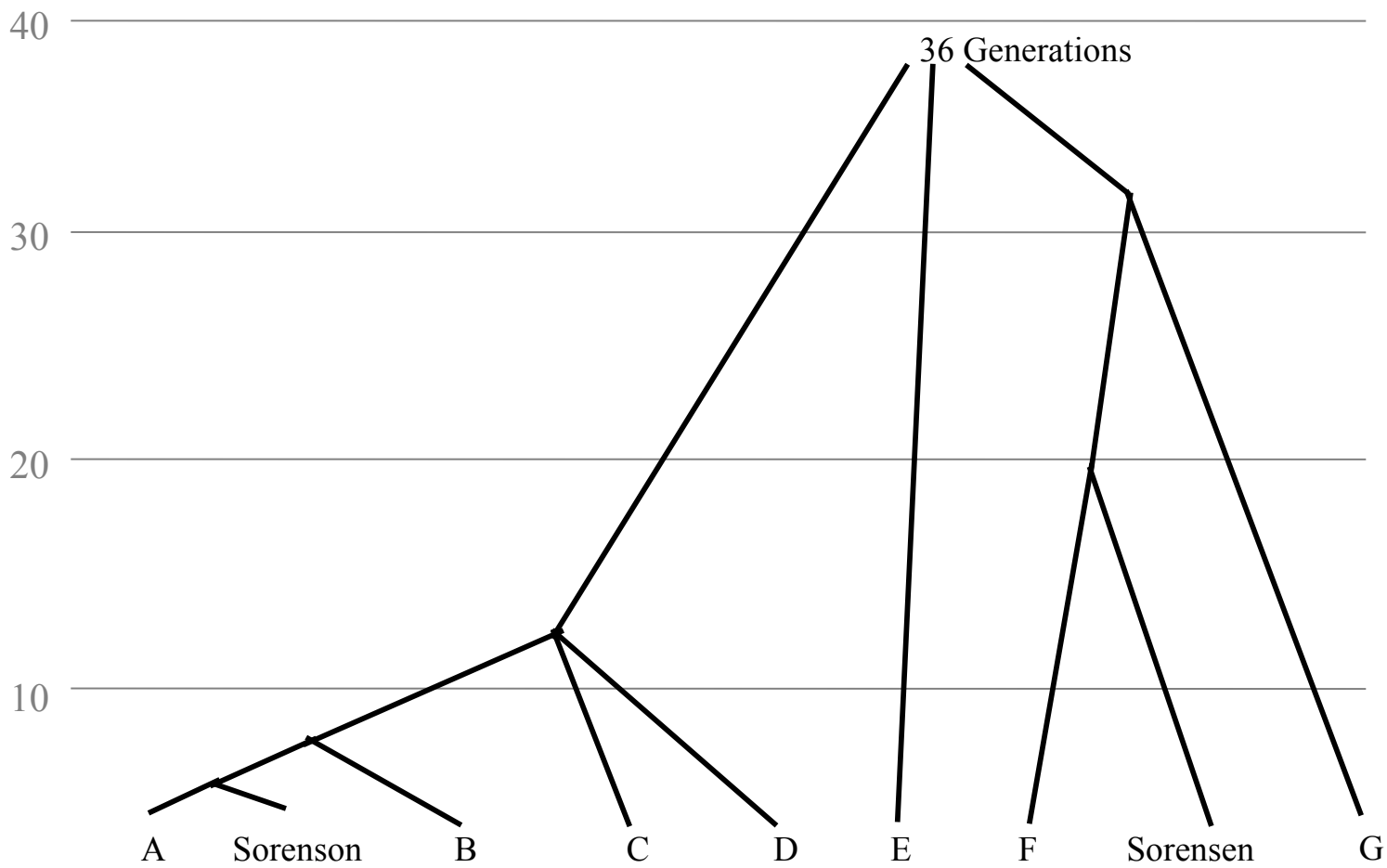
Phylogeny (Tree-Building)

- ***Phylogeny programs (e.g. PAUP) can be used to rebuild possible inheritance trees***



Family History Technology

Discovering Previously-Unknown Relationships



Family History Technology

Problems with Phylogeny

- ***Many difficulties: size of problem space (intractability); significant difference in results between runs; IBS matches; inability to properly handle the inheritance topology of recombining DNA***

- ***Phylogeny results should be treated as informative but not authoritative***

Genetic Inheritance Models (mtDNA)

- ***Mitochondrial DNA (mtDNA)***
  - In mitochondria (energy units of cell) rather than in nucleus
  - Passed from mother to children (almost exclusively maternal-line DNA)
  - Usually mtDNA SNPs are used to trace deep genealogies (on an anthropological scale)
  - Haploid (single-stranded), so similar in population-genetic properties to Ycs DNA; phylogeny algorithms are applicable

Genetic Inheritance Models (Autosomes)

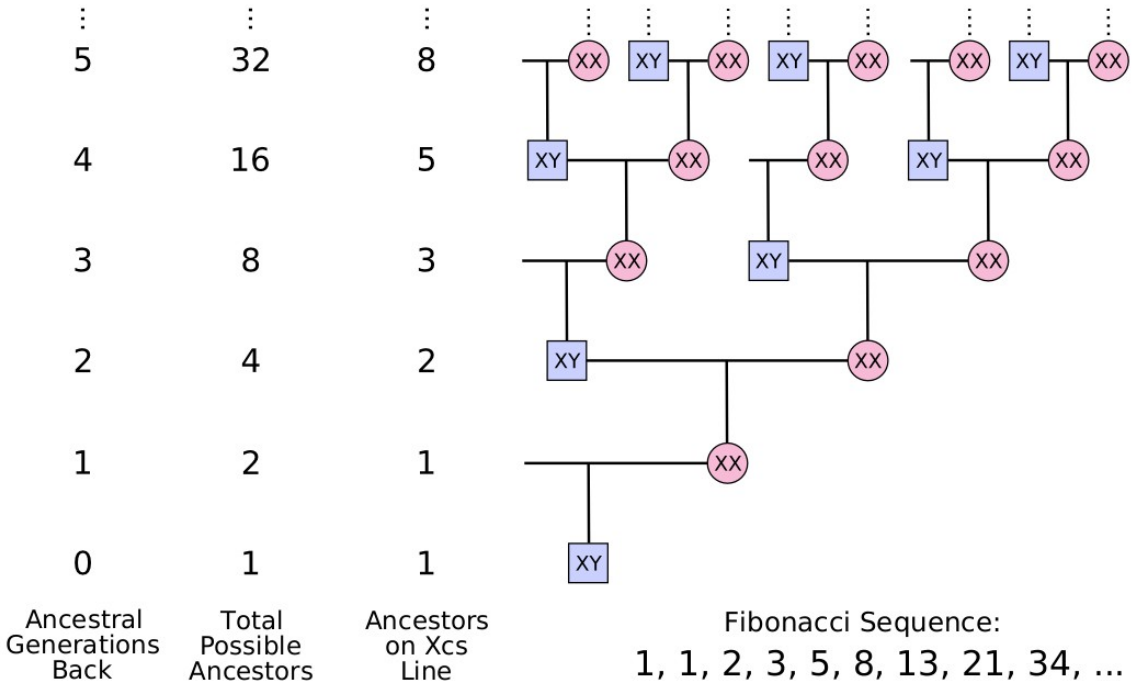
- ***Autosomal DNA***
  - The bulk of our nuclear DNA
  - Diploid (double stranded): pairs of homologous chromosomes
  - Recombining
  - We receive half of our autosomal DNA from each parent
  - Each parent only passes down half of their autosomal DNA to each child

Genetic Inheritance Models (Xcs)

- ***X chromosome (Xcs)***
  - Males: X-Y; Females: X-X
  - Any mother-daughter or father-son pair has exactly one X chromosome in common, allowing us to construct a phase-known set of haplotypes for testing haplotyping algorithms
  - Forward through time: X Passed from *father to all daughters*; *one* of mother's X chromosomes passed to *each child; X not* passed from father to son

Genetic Inheritance Models (Xcs)

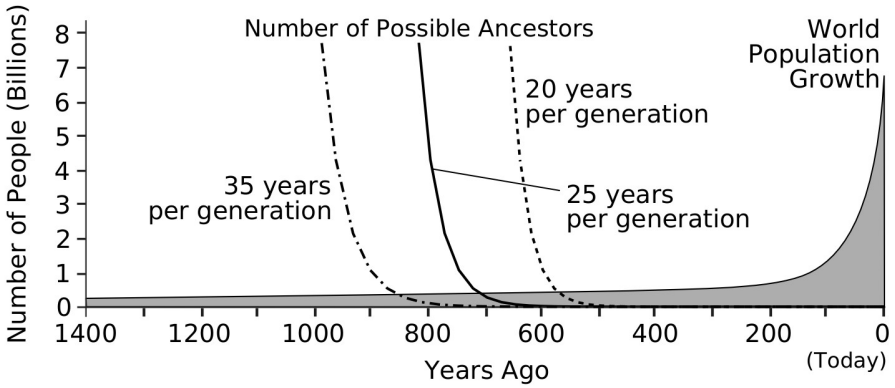
- Backward through time: number of possible Xcs ancestors follows the Fibonacci Sequence:



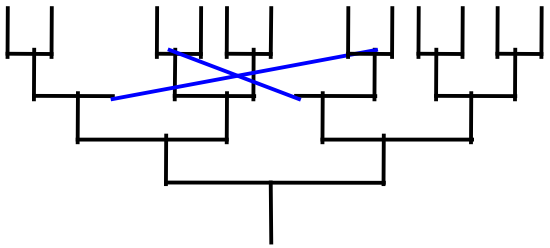
Family History Technology

Population growth through time

- ***Number of possible (autosomal) ancestors quickly outstrips world population size***

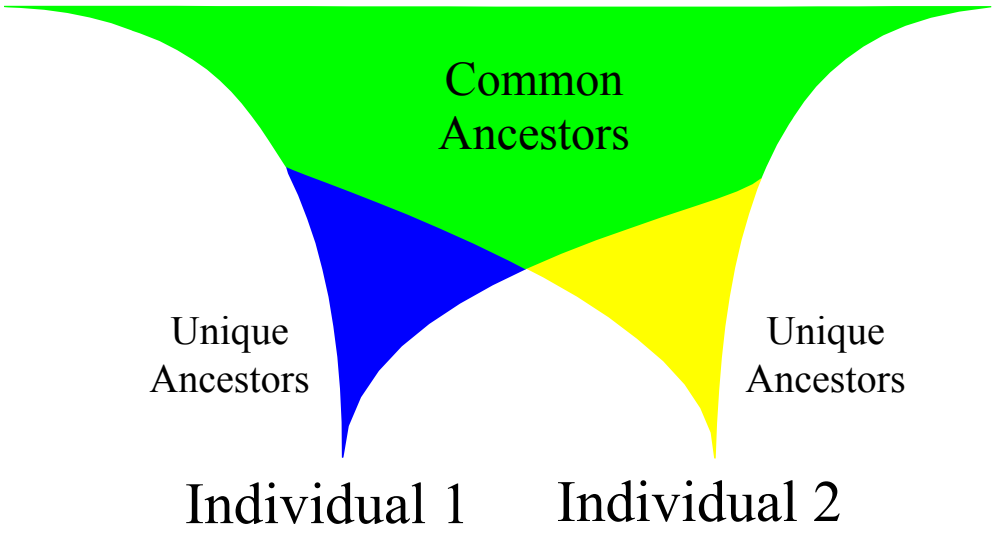


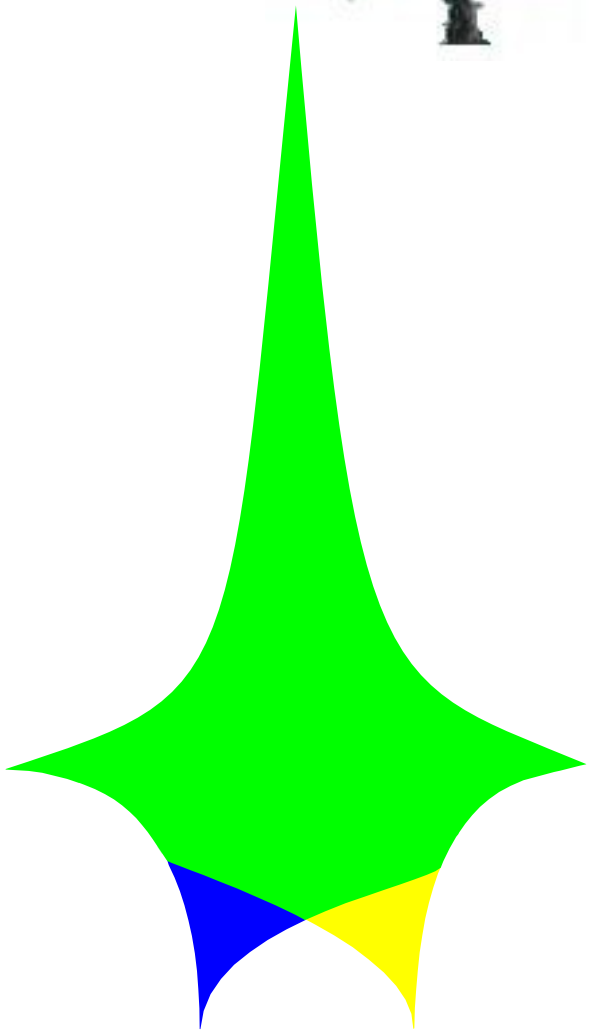
- ***Genealogi then coale***



Coalescence

- ***Two individuals theoretically share all their ancestors at a very recent point in time***





Family History Technology

Collaboration

- ***We are seeking collaborators***

- ***Help us build the tools to reunite living individuals with their ancestors through their DNA ...***

  ... or help us build the database – contribute your DNA and your genealogy!

www.smgf.org

Family History Technology

Conclusions

- ***Molecular Genealogy allows for DNA to be used in combination with pedigree data to fill in unknown genealogy***
- ***New field, many exciting problems***
- ***Several useful analysis techniques already exist, e.g. Y chromosome surname search***
- ***Much work still needs to be done, particularly in the areas of algorithm design and statistical analysis***

Family History Technology

Family History Technology

Family History Technology

Additional Slides

(included for informational purposes, will probably not be covered in the presentation)

Family History Technology

Goals of Molecular Genealogy

- ***To create a comprehensive database of the peoples of the world, using correlated genealogical and genetic information***

- ***To provide tools to reconstruct genealogies using DNA, to reunite us with our ancestors***

- ***To change the way that we think about each other, and hopefully the way we act towards each other, by showing that we are really one great human family***

Family History Technology

Why Family History?

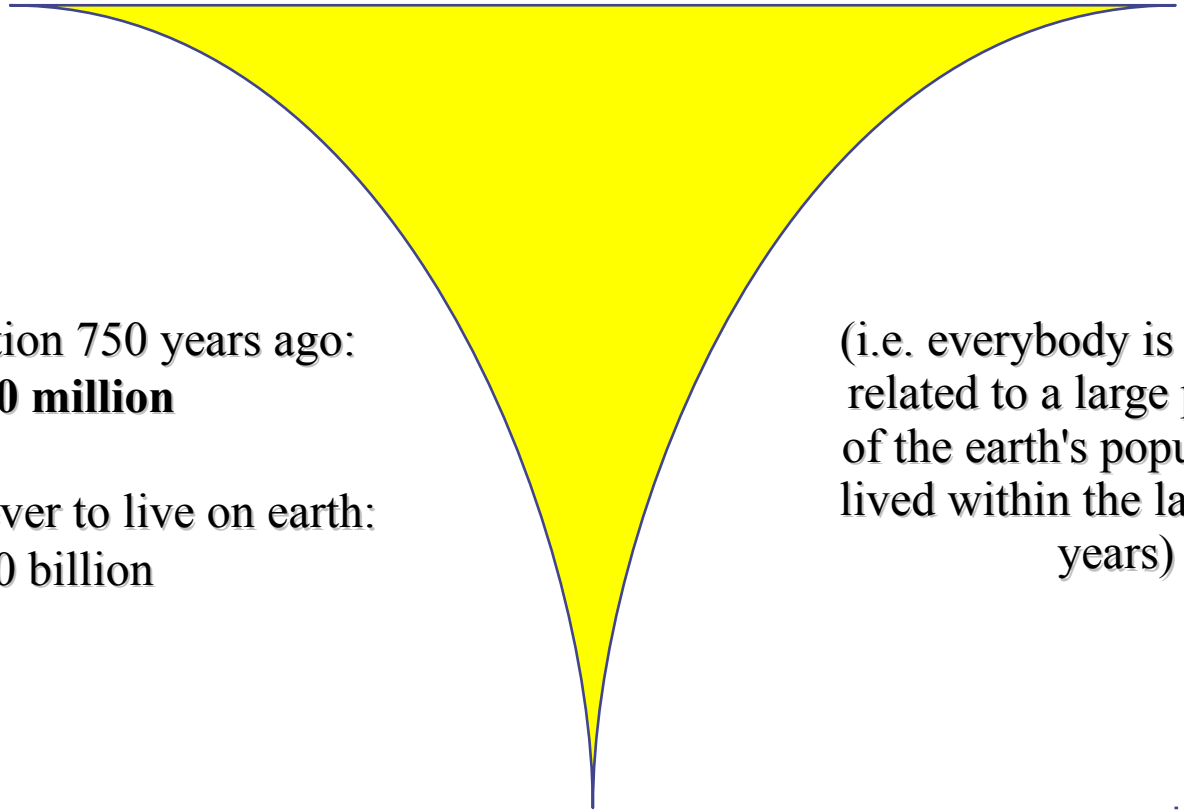
- ***Ask a genealogist!***

- ***"No man is an island"***
  - Our family is part of our identity and purpose
  - We cannot fully know ourselves without knowing those through whom we came
  - We all have a responsibility to search out our ancestors

Family History Technology

Problems with the numbers

30 generations = 750 years = **1 billion** possible ancestors



World population 750 years ago:
~ **450 million**

Total humans ever to live on earth:
~ 70 billion

(i.e. everybody is potentially related to a large proportion of the earth's population that lived within the last 500-750 years)

Living Individual

Family History Technology

The Basis of Molecular Genealogy

- ***Each individual carries within their DNA a record of who they are and how they are related to all other people.***

- ***Specific regions of DNA have properties that can:***
  - Identify an individual
  - Link them to a family
  - Identify extended family groups
  - Tie the individual to their ancestral populations

Family History Technology

The DNA Paradox

- ***Almost 4 billion pieces of information***
- ***Can identify you as a unique individual***
- ***All humans share many regions exactly***
- ***The level of sharing is directly related to the degree of relationship***

- ***DNA is what makes us different***
- ***DNA is what makes us the same***

Translating the Language of DNA

- ***Unique approach***
  - We focus specifically on using DNA to accelerate the work of ***family history***.

- ***We extract and interpret information in DNA to:***
  - ***identify*** individuals who lived in the past, and
  - ***link*** them to individuals living today.

Family History Technology

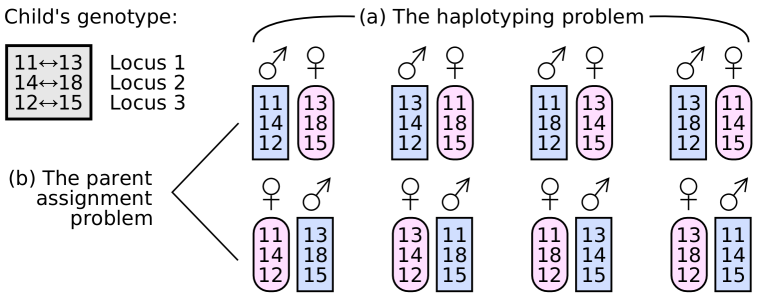
We are one family

"[…] the word generosity has the same derivation as the word genealogy, both coming from the Latin genus, meaning of the same birth or kind, the same family or gender. We will always find it easier to be generous when we remember that this person being favored is truly one of our own."

(Jeffrey R. Holland, SLC General Conference, April 2002)

Haplotyping

- ***Haplotyping or setting phase is the problem of determining which alleles (marker values) in a diploid genotype were located on the same chromosome strand***

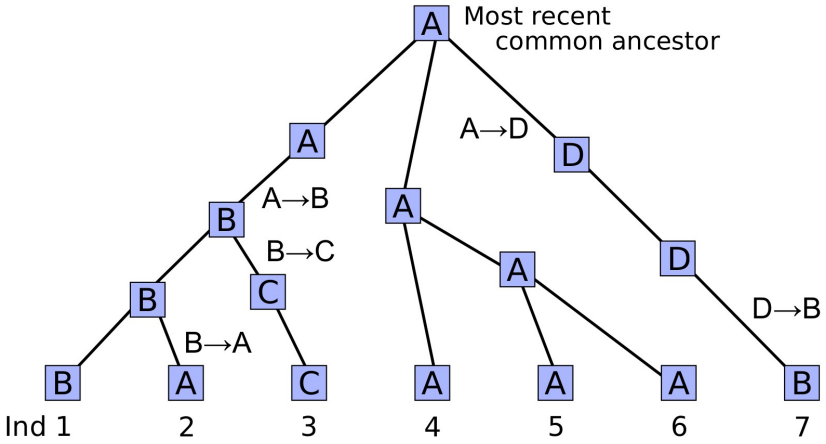
- ***Haplotypes are*** 

al alleles (less chance of IBS r

Family History Technology

IBD and IBS

- ***Genetic markers that match because they were passed down from a common ancestor are "identical by descent" (IBD)***
- ***Genetic markers that match after mutation are "identical by state" (IBS)***
- ***IBS Matches can be misleading***



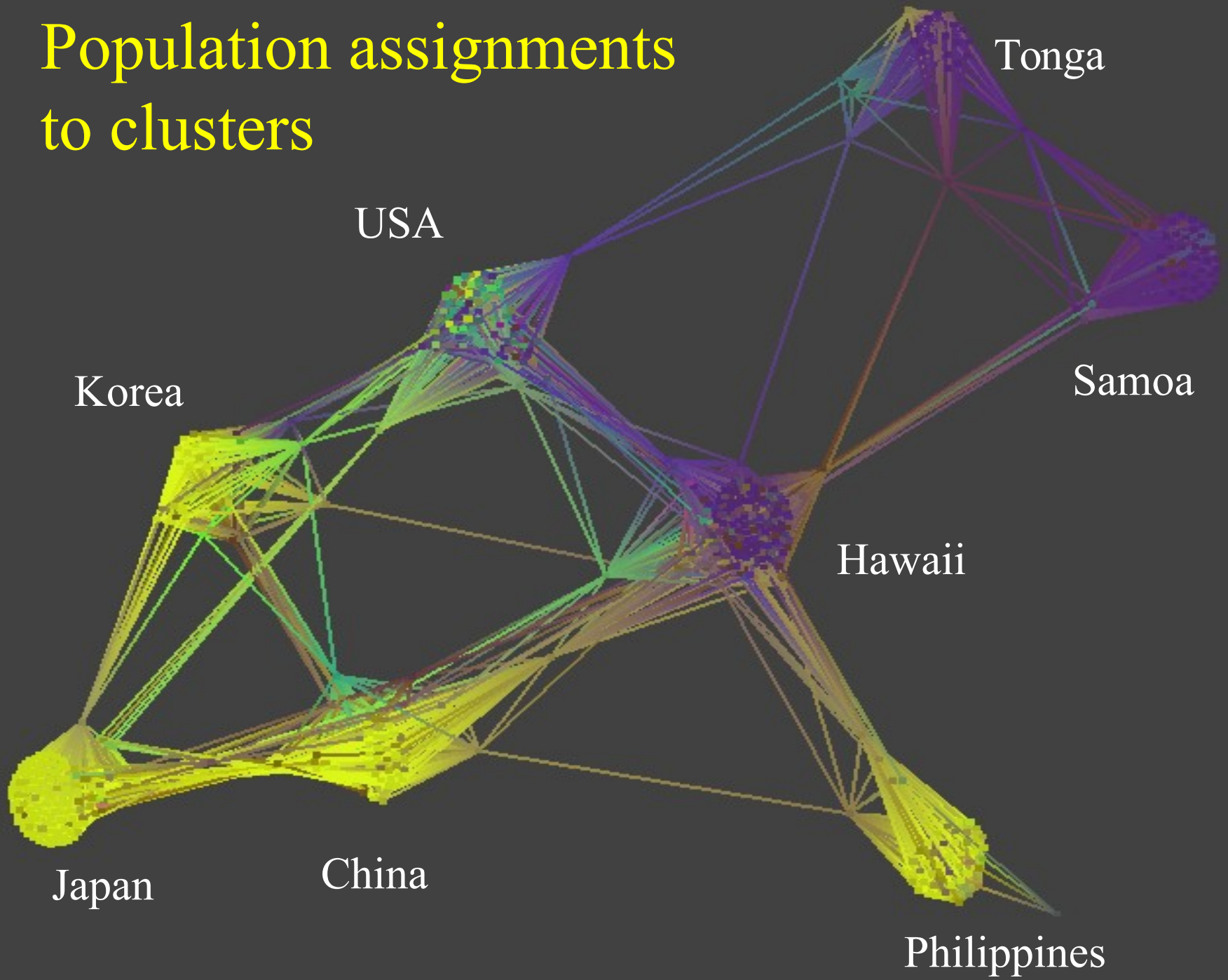
Mutation Models and Rates

- ***Mutation can happen between generations***
- ***Only approximate mutation models exist to explain mutational changes***
  - Stepwise Mutation Model (SMM)
  - Infinite Alleles Model
- ***Mutation rates have been estimated only approximately, e.g. 0.3%/STR locus/gen and 0.000002%/nucleotide/gen***

Clustering of Pacific Island Populations

- ***Collected 1500+ individuals from the Pacific Islands***
- ***Typed at 60+ autosomal loci***
- ***Clustered with STRUCTURE***
  - 682 individuals using 58 loci
  - Clustered into 8 pops
- ***Visualized with TULIP***

Population assignments to clusters



Other Issues

- ***Clustering***
  - Some success with autosomes
- ***Tracing of autosomal-line DNA***
  - "Goldmine" but harder to work with
- ***Statistical population genetics***
  - Gives us understanding of population dynamics, e.g. Hardy-Weinberg Equilibrium
- ***Accuracy of genealogical data***
  - Deep, accurate genealogies crucial to molecular genealogy [need common ancestors]