

Genetic Genealogy: A New Tool for Mennonite Genealogists

by Tim Janzen MD

Portland, Oregon

tjanzen@comcast.net

503-761-8781

MHSS Mennonite Genealogy Workshop

Herbert, Saskatchewan

August 4, 2007

Historical background

- DNA structure described by James Watson and Francis Crick in 1953
- First worldwide study of mitochondrial DNA published by Rebecca Cann in 1987
- Detailed worldwide study of Y chromosome DNA published in 2000 by Peter Underhill
- Commercial DNA testing for genealogists became available in 2000
- Human Genome Project completed in 2003

Biological Background

- Humans have 46 chromosomes; of these 44 are autosomal chromosomes and 2 are sex chromosomes
- About 3.2 billion base pairs in the human genome, of which 95-98% don't code for anything
- 99.9% of human autosomal DNA is identical
- Mitochondrial DNA has 16,569 base pairs, of which 1449 base pairs don't code for anything

The four nucleotides (bases)

- Adenine (A)
- Cytosine (C)
- Guanine (G)
- Thymine (T)
- Adenine pairs with Thymine
- Cytosine pairs with Guanine

Major Types of DNA

- Y chromosome: found only in males and passed from father to son; only 26 million base pairs sequenced thus far out of about 60 million
- Mitochondrial DNA: found in both males and females, but passed on only by the mother to her children; 16,569 base pairs in a circle
- Autosomal DNA: 44 chromosomes; each parent contributes one half of the DNA to their children
- X chromosome: 2 in females and 1 in males; sons receive one chromosome from their mother and daughters receive one chromosome from each parent

Common Terms in DNA research

- Marker: an identifiable locus (location) on a chromosome
- Haplotype: a set of values for a group of genetic markers that is inherited as a unit
- Haplogroup: a group of similar patterned and related haplotypes that share a common ancestor due to a specific mutation

More DNA terms

- Single Nucleotide Polymorphism (SNP): common variations in the allele value at a specific nucleotide position
- Short Tandem Repeats (STR): Patterns in DNA sequences that repeat over and over again in tandem right after each other. For example GATAGATAGATAGATA is a pattern where 4 nucleotides are repeated 4 times.

Y chromosome research

- Used to determine the relative degree to which two males are related to each other on their paternal lines
- Used to determine a male's haplogroup
- Usually at least 12 markers are tested; at least 130 markers currently available for testing
- Values given as a total of the number of STRs for a particular marker
- The more markers two males have that match the more likely they are to be closely related to each other

Mitochondrial DNA

- Used to determine the relative degree to which two people are related to each other on their maternal lines
- Used to determine one's mitochondrial DNA haplogroup
- Values are given as differences to the Cambridge Reference Sequence, which was the first sequence completed for the human mitochondrial DNA molecule
- 3 hypervariable regions: HVR1 (15841-16569; 729 bases), HVR 2 (00001-00437; 437 bases), HVR 3 (00438-00720; 283 bases); 1449 in total

X chromosome

- Exact value in DNA testing is still being researched
- Research is complicated by the fact that the chromosome recombines at conception
- Likely will be shown to have significant value when the results are used in conjunction with other DNA tests

Autosomal DNA

- May be used to determine the relative degree to which 2 people are related to each other on any line of descent if they share at least one common ancestor in the recent past
- Has significant potential for genealogical researchers since there are a total of 44 chromosomes that can be tested
- Limited by the fact that the chromosomes recombine at conception and thus one half of each parent's markers is not passed to a specific child
- Of most benefit in determining relationships between people who share a common ancestor within 6 to 8 generations

Mutation Rates

- Y chromosome STR mutation rates vary depending on the marker. Some are more stable than others. Generally, the higher the allele value, the higher the mutation rate. Average mutation rate is about $4/1000$. Thus if 37 markers are checked then chances are about $1/7$ that one marker will have mutated in any one transmission.
- Mitochondrial DNA: $3/100,000$ per base in the hypervariable regions or $3/100$ for the entire mitochondrial DNA sequence per transmission
- SNPs: $1/50,000,000$ per transmission

DNA testing labs

- Family Tree DNA www.familytreedna.com
- Sorenson Molecular Genealogy Foundation
<http://smgf.org>
- National Geographic Society Genographic Project
www5.nationalgeographic.com/genographic
- RelativeGenetics www.relativegenetics.com
- Oxford Ancestors www.oxfordancestors.com

Informative DNA testing web sites

- Charles Kerchner's DNA Info and Resources Page
www.kerchner.com/dna-info.htm
- EthoAncestry <http://www.ethnoancestry.com/dna.htm>
- ISOGG <http://www.isogg.org>
- World Families Network
<http://worldfamilies.net/y-haplogroups.htm>
- Dr. Whit Athey's utility for prediction of haplogroup from a Y chromosome haplotype
<https://home.comcast.net/~hapest5/index.html>
- Dean McGee's Y DNA comparison utility
<http://www.mymcgee.com/tools/yutility.html>
- Dr. John McEwan's "How to" guide for Y chromosome haplotypes www.geocities.com/mcewanjc/howto.htm

Informative books about DNA testing

- Trace Your Roots with DNA by Megan Smolenyak and Ann Turner, 2004.
- DNA and Family History: How Genetic Testing Can Advance Your Genealogical Research by Chris Pomery, 2004.
- Unlocking Your Genetic History by Thomas Shawker, 2004.
- The History and Geography of Human Genes by L. Luca Cavalli-Sforza, Paola Menozzi, and Alberto Piazza, 1994.
- The Journey of Man by Spencer Wells, 2002.
- Deep Ancestry by Spencer Wells, 2006.
- The Seven Daughters of Eve by Bryan Sykes, 2001.

Genetic Genealogy Discussion Groups

- RootsWeb Genealogy-DNA list
<http://archiver.rootsweb.com/th/index/GENEALOGY-DNA>
; most active list with an average of about 2000 messages per month
- FTDNA <http://www.familytreedna.com/forum>; very active
- ISOGG Newbies; about 350 messages per month
<http://groups.yahoo.com/group/DNA-NEWBIE>
- www.Genealogy.com DNA forum;
<http://genforum.genealogy.com/dna>; about 50 messages per month
- World Family Network www.wfnforum.net; minimally active

History of the Sorenson Molecular Genealogy Foundation

- Began from a project called the Molecular Genealogy Research Project (MGRP) started in 1999 by Dr. Scott Woodward at Brigham Young University
- SMGF founded in 2002
- Funded by philanthropist James Sorenson
- Now comprised of about 40 staff members
- About 80,000 samples collected thus far

Sorenson Molecular Genealogy Foundation

- Only lab that offers DNA testing for free
- Kits available at http://www.smgf.org/pages/request_kit.jsp
- Tests 37 Y chromosome markers (43 alleles)
- Tests the 3 hypervariable regions of mitochondrial DNA
- Tests X chromosome markers
- Tests 300 autosomal DNA markers
- Test for 50,000 SNPs to be done in the future
- A pedigree chart is submitted along with the DNA test kit

Sorenson Molecular Genealogy Foundation

- Results take at least 12 months to be placed in the databases
- Results must be manually retrieved from the databases
- Y chromosome database searchable by surname and haplotype
- Mitochondrial DNA database released in July 2006 and now searchable by surname
- Autosomal DNA and X chromosome databases will not be released for at least 6 months

Sorenson Molecular Genealogy Foundation Databases

- Y chromosome database now has 19,113 haplotypes with 7 or more markers; 14,720 haplotypes with 34 or more markers
- Mitochondrial DNA database has results for 25,104 people from the 3 Hypervariable Regions
- Databases updated about every 8 weeks
- About 5000 new mitochondrial DNA results and about 1500 new Y chromosome results in each new release

Autosomal DNA testing basics

- Siblings share 50% of the same markers
- First cousins share 12.5% of the same markers
- Second cousins share 3.125% of the same markers
- Third cousins share 0.78% of the same markers
- Fourth cousins share 0.195% of the same markers
- Fifth cousins share 0.049% of the same markers
- Sixth cousins share 0.0122% of the same markers

Autosomal DNA statistics if SMGF tests 300 markers

- Siblings would have 150 markers in common
- First cousins would have 37.5 markers in common
- Second cousins would have 9.375 markers in common
- Third cousins would have 2.34 markers in common
- Fourth cousins would have .586 markers in common
- Fifth cousins would have .146 markers in common

Inheritance of Autosomal DNA markers if SMGF tests 300 markers

- Each person inherits about 150 markers from each parent
- Each person inherits about 75 markers from each grandparent
- Each person inherits about 37.5 markers from each great grandparent
- Each person inherits about 18.75 markers from each great great grandparent
- Each person inherits about 9 markers from each great great great grandparent
- Each person inherits about 4 or 5 markers from each great great great great grandparent

Mennonite DNA Project

www.mennonitedna.com

- Began in July 2004
- Coordinated by Glenn Penner (gpenner@uoguelph.ca) and Amelia Reimer (nutmeg@centurytel.net)
- At least 750 people who have been tested thus far
- FTDNA: 102 males with Y chromosome (11 also did mitochondrial DNA) and 4 females
- SMGF: 365 males (8 also in FTDNA) and 262 females (1 also in FTDNA) as well as 50 Mennonites in Mexico

Goals of the project

- Determine the number of male progenitors for each Mennonite surname, of which there are about 300 of Prussian/Dutch origin
- Determine the number of female Mennonite progenitors
- Determine the haplogroups (deep ancestry) of each progenitor
- Use the DNA results to complement traditional genealogical research in determining relationships among various ancestors

Penner Y chromosome data

- Results from 26 previously unconnected Penners show that all but two descend from a common ancestor who likely lived about 400-600 years ago.
- Glenn Penner continues to recruit Penners to be tested and will pay for testing of male Penners who are not more closely related than being a third cousin of some other Penner who has already tested.

Other Mennonite Y chromosome data

- Froese, Hiebert, Janzen, Penner, Peters, Schroeder, and Wieler surnames each have at least two progenitors
- Descendents of other surnames tested thus far share a common ancestor
- Haplogroups represented thus far: E3b, I, G, J2, R1a, R1b, R1b

Haplogroup I Surname Analysis

- I1a; Anglo-Saxon in origin, suggesting that they were originally from the Netherlands or NW Germany: Heinrich Bock (b. ca 1784) GM#44124, Benjamin Fehr (b. 1733) GM#196504; Friesen; Harder; Peter Jantz (b. 1650) GM#39121; Salomon Neufeld (b. 1701) GM#265863; Peter Siemens (b. ca 1790) GM#58879; Peter Thiessen (b. ca 1717) GM#95226; Wall
- I1b2a; Continental type 1; common in the Netherlands and NW Germany: Braun; Peter Wolf (b. 1756) GM#196568
- I1b2a*: Root type 3; found from Iberia and Italy through Denmark: Wiens

Haplogroup R1b Surname Analysis

- R1b Frisian (DYS390=23, DYS391=11): Epp, Hiebert, Paul Janzen (b. 1704) GM #11942, Jacob Loewen (b. ca 1794) GM#51579, Wiebe, Jacob Klaas Wieler (b. 1794) GM#55032, Engbrecht
- R1b Ubiquitous: Gerhard Peters (b. 1772) GM#18759, Paul Schellenberg (b. 1634) GM#56777, Julius Toews (b. 1741) GM#187161, Karl Winter (b. 1810) GM#222126
- R1b North/South: Heinrich Bartsch (b. 1826) GM #32380, Dyck, Johann Reimer (b. 1815) GM #180814
- R1b Atlantic Modal Haplotype: Ratzlaff, Isaac Schroeder (b. 1738) GM#222095, Peter Schroeder (b. 1718) GM #694669, Johann Wieler (b. 1771) GM#127055, Zacharias, Isaak, Flaming

Mitochondrial DNA results

- Results from only 21 people of Prussian/Dutch Mennonite ancestry tested thus far
- 16 haplotypes have been found thus far
- 4 haplotypes have at least one match and thus they share a common maternal ancestor in the past
- Haplogroups represented thus far: A, H, I, J, T, and U

Potential applications of autosomal DNA testing for genealogists

- Determine if two people are distant cousins by reviewing the number of autosomal markers that they share in common and the percentage of markers that they share in common out of the 300 autosomal markers that SMGF tests.
- Determine precisely which markers that a person has inherited from each of his ancestors; this can be done only after analyzing the results from many closely and distantly related people and then theorizing as to which markers each ancestor must have carried in their genome.

Long term autosomal DNA testing goals for Mennonite genealogy

- Determine multiple markers for recent ancestors who lived in the late 1800s and early 1900s
- Determine one or more markers for each ancestor who lived in the 1700s and early 1800s
- Attempt to determine precisely which markers living descendants inherited from each of their Mennonite ancestors

Caveats

- SMGF long term funding
- Privacy issues
- Test reliability
- “Skeletons in the closet”
- Care must be taken not to overreach when interpreting results
- Mitochondrial DNA heteroplasmy

Who should be tested?

- Anyone who wants to be tested!
- The most important people to be tested are the oldest living members of each family and anyone who has already had one or more of their parents die.
- The more people who are tested the more information that will be learned that will be potentially useful for Mennonite genealogists in the long term.