



Research through DNA?

Sorenson Molecular Genealogy Foundation
Releases Y Chromosome Database by Ugo A. Perego, MSc

In March 2000, a group of university students, headed by molecular biologist Dr. Scott R. **Woodward**, began working on a large-scale genetic study called the Molecular Genealogy Research Project (MGRP). A few months later, *Everton's Family History Magazine* published a story about **Woodward's** research efforts.¹ The MGRP's main objective is to build a large and comprehensive database of correlated genealogical and genetic information. Currently, about 40,000 individuals from around the world have contributed DNA samples and copies of their pedigree charts to the study.

During the past four years, **Woodward's** team has made significant progress. Included in that was moving all of the project's operations from Brigham **Young** University in Provo, Utah, to the **Sorenson** Molecular Genealogy Foundation (SMGF), a non-profit organization located in Salt Lake City. By consolidating the project's resources under one roof, work has advanced at a more consistent and efficient pace. Another important milestone in the history of the MGRP was the release of the first version of the Y-chromosome database to the public in February.

DNA and family history

In recent years, DNA has become a familiar acronym among family historians. Lectures about genetics and genealogy are now a recurring theme at many genealogical conferences, and more people are using some form of DNA testing to complement traditional genealogical resources for finding lost branches of their family trees.² Among genetic tests available to family historians, the most common is Y-chromosome testing. The Y chromosome is a portion of DNA found only in males that is inherited strictly from the father, following a direct paternal line (just like the surname). Although females do not inherit a Y chromosome from their fathers, they can still take advantage of Y-chromosome

testing by asking a male relative sharing their same paternal line (surname) to contribute a DNA sample in their behalf.

This test is particularly useful for solving questions about relatedness among individuals sharing the same or similar surnames³, or in verifying biological relationships between individuals descending from alleged common paternal ancestors.⁴ Even though the MGRP's main focus is to use autosomal DNA — the DNA found across the whole human genome — in reconstructing family histories, a subset of data collected during the past four years has been compiled in the form of a Y-chromosome database. This new resource has recently become available to the public free of charge and can be accessed on the **Sorenson** Molecular Genealogy Foundation Web site at <www.smgf.org>.

Collection of genetic and genealogical data

During the first two years of the study, DNA samples were obtained through small blood samples collected exclusively by trained medical personnel. For this reason, participation in the study was possible only by attending a MGRP collection event, usually hosted by local genealogical societies and family history centers. In July 2002, a new technique for collecting DNA samples was implemented. The

new method, called GenetiRinse, involves the use of mouthwash for the collection of cheek cells. Individuals interested in participating in the MGRP can request a GenetiRinse kit through the mail and return their sample using a postage-paid envelope provided by the SMGF.

With the DNA sample, each participant also provides a copy of his or her pedigree chart with a minimum of four complete generations. Individuals who have done extensive research on their family lines and can submit more than four generations are strongly encouraged to do so. Each pedigree chart and DNA sample donated to the MGRP is barcoded upon submission by the collection personnel to protect the participant's privacy. Only genealogical data for ancestors born before 1900 are used in the compilation of the public database.

Development of the Y-chromosome database

For the development of the Y-chromosome database, genealogists at the **Sorenson** Molecular Genealogy Foundation have used the Ancestral File database and other available genealogical resources to extend the paternal lines of participants' pedigrees and obtain up to 14 generations of available genealogical information. Using customized PAF filters, the extended pedigrees were screened for possible errors, such as discrepancies among birth, death and marriage dates. After questionable genealogical information was manually accepted or rejected, the finalized pedigrees were imported into the genealogical component of the Y-chromosome database.

While the genealogists worked on the pedigree charts to prepare them for the database, Relative Genetics, an affiliated company specializing in genetic testing for genealogical purposes, extracted DNA from mouthwash samples using standard laboratory procedures. After extraction, the Y chromosomes of male participants were amplified and analyzed at 24 genetic

markers (loci) designated by SMGF. These markers were selected based on the high level of information they provide in genealogical studies. Taken together, they constitute a genetic signature, or haplotype, for the male donor and for his paternal ancestors. The 24-marker Y-chromosome profiles were imported into the genetic component of the database and linked to their corresponding extended pedigree charts.

Software used for tracking of genealogical and genetic data and for extending pedigree information, as well as algorithms used for genetic matching, was developed by UNIconnect of Salt Lake City.

Using the Y-chromosome database

Individuals querying the Sorenson Molecular Genealogy Foundation's Y-chromosome database must know their paternal lineage's Y-chromosome haplotype. This information can be easily obtained by submitting a biological sample to an independent laboratory offering this type of service, such as Relative Genetics. Females wanting to participate can ask a male relative sharing their paternal line to submit a DNA sample for the Y-chromosome testing.

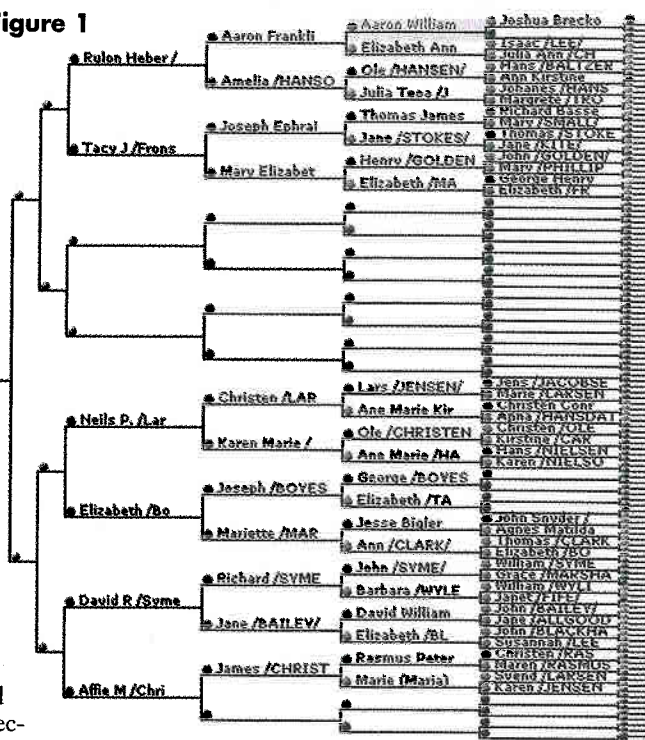
Table 1 (below) shows an example of the type of report an individual would expect to receive from an independent laboratory providing Y-chromosome testing. In this example, the 24 markers used by the foundation are listed in the first column of the table and the individual's genetic value for each marker is reported in the second column as number of repeats — the standard unit for measuring genotype values. Knowing a family's Y-chromosome signature in molecular genealogy is as important as knowing a family's surname when querying a traditional name-based database.

With a family Y-chromosome haplotype available, queries for the Y-chromosome database can be performed. When a haplotype is submitted to the database, a search for similar haplotypes takes place among the thousands in the database. Results are reported for each haplotype within the database that matches the submission by at least 21 markers. Results are reported this way because two individuals who have matching Y haplotypes, at a minimum of 21 markers, usually share a common paternal ancestor in the recent past. Genealogical information associated with each matching haplotype also is reported in the form of a pedigree chart and a most-recent-common-ancestor (MRCA) graph.

Figure 1 (above) shows an example of a pedigree chart given by the database. Only information for individuals born before 1900 is reported in order to protect the privacy of living individuals. Figure 2 (right) is an example of a MRCA graph. Each dot on the graph represents an individual on the paternal line of the corresponding pedigree chart. Vital statistics for each ancestor born before 1900 are available from the database. The curved pattern of the dots shows the likelihood that the submitted haplotype and the match identified in the database share a common paternal ancestor in any number of past generations. The dot or ancestor at the peak of the graph (designated MLE) is estimated to be the most likely common paternal ancestor for a pair of haplotypes. Although genealogical information contained in the database is screened for obvious errors, inaccuracies still can be present. Therefore, the pedigrees and MRCA graphs are meant to be used only as a guide for identifying possible paternal ancestors and not as conclusive evidence.

DNA samples submitted by volunteers to the Sorenson Molecular Genealogy Foundation for constructing the database cannot be used to generate individual genetic profiles for personal database queries. Because of federal regulations, these samples have been stripped of any personal identifiers to protect the confidentiality of the participants and

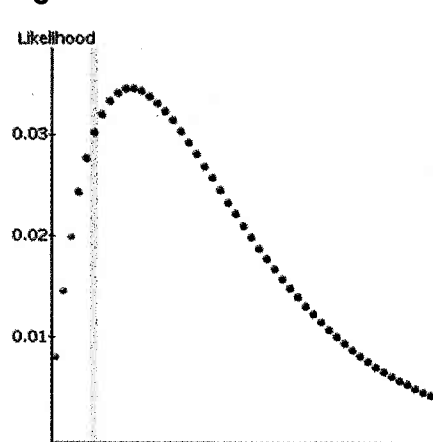
Figure 1



Match Number 1

| | | |
|-------------|-------------------------|------------------------|
| Name | Aaron William /BRACKEN/ | |
| Generation | 6 | MRCA [0.0301 |
| Birth | 26 Jan 1843 | Hartland, Niagara, NY |
| Christening | | |
| Marriage | 1 Apr 1865 | Tooele, Tooele, UT |
| Death | 12 Jul 1874 | Liberty, Bear Lake, ID |
| Burial | 12 Jul 1874 | Tooele, Tooele, Utah |

Figure 2



Match Number 1

| | | |
|-------------|-------------------------|------------------------|
| Name | Aaron William /BRACKEN/ | |
| Generation | 6 | MRCA [0.0301 |
| Birth | 26 Jan 1843 | Hartland, Niagara, NY |
| Christening | | |
| Marriage | 1 Apr 1865 | Tooele, Tooele, UT |
| Death | 12 Jul 1874 | Liberty, Bear Lake, ID |
| Burial | 12 Jul 1874 | Tooele, Tooele, Utah |

Table 1

| Y-Chromosome Haplotype | |
|------------------------|-------------------|
| Locus | Number of Repeats |
| DYS385 | 11,12 |
| DYS388 | 13 |
| DYS389I | 13 |
| DYS389BII | 25 |
| DYS390 | 25 |
| DYS391 | 9 |
| DYS392 | 11 |
| DYS393 | 12 |
| DYS394/19 | 15 |
| DYS437 | 14 |
| DYS438 | 10 |
| DYS439 | 12 |
| DYS448 | 18 |
| DYS449 | 28 |
| DYS454 | 11 |
| DYS458 | 19 |
| DYS459 | 8,8 |
| DYS460 | 7 |
| DYS461 | 13 |
| DYS462 | 13 |
| GGAAT1B07 | 9 |
| YGATAA10 | 14 |
| YGATAC4 | 23 |
| YGATAH4 | 12 |

cannot be shared with anyone, including the participants themselves.

Conclusion

The **Sorenson** Molecular Genealogy Foundation Y-chromosome database is the newest tool available to genealogists. It provides the opportunity to search for paternal ancestors when two individuals share the same or similar Y-chromosome haplotypes. To enhance database usefulness, the **Sorenson** Molecular Genealogy Foundation is adding new Y-chromosome haplotypes and genealogical information regularly. As the database grows, so will the chances of finding matches to individual queries. Additionally, the number of markers used to construct Y-haplotypes also will increase in order to provide more accuracy in MRCA calculations.

It is important to recognize that even though Y-chromosome testing is a popular method for identifying common ancestors, its use is limited to the paternal line. The Y-chromosome accounts for less than 1 percent of a male's total DNA. For this reason, the **Sorenson** Molecular Genealogy Foundation continues to develop ways in which all of the chromosomes in the human genome can be used effectively in reconstructing and verifying genealogies. Researchers at the foundation are developing

a more comprehensive database consisting of information from all human chromosomes. Updates, scientific publications, and progress of the MGRP can be monitored on the foundation Web site at <www.smgf.org>.

Individuals who would like to contribute a DNA sample and a pedigree chart to future versions of the Y-chromosome database and for the development of the autosomal database are invited to contact the **Sorenson** Molecular Genealogy Foundation and request a free GenetiRinse kit. Both male and female participants who can provide a minimum of four complete generations of pedigree information are welcome to participate. Donation of DNA samples and genealogical information to the MGRP is free, confidential and voluntary. GenetiRinse kits can be requested by visiting the project Web site or by writing to: **Sorenson** Molecular Genealogy Foundation, 2511 S. West Temple, Salt Lake City, UT 84115, or by sending e-mail to info@smgf.org.

More Contacts:

Relative Genetics, 2495 S. West Temple, Salt Lake City, UT 84115; Phone: 801.461.9760; Fax: 801.461.9761; E-mail: info@relativegenetics.com

Endnotes

- 1 Ugo A. **Perego**, Scott R. **Woodward**. 2000. *Molecular Genealogy*. Everton's Genealogical Helper. Vol. 54 No. 5: 22-25.
- 2 Kevin **Duerinck**. 2004. *Y-chromosome DNA Surname Projects* [online]. Available from the World Wide Web: <www.duerinck.com/surname.html>
- 3 Mark A. **Jobling**. 2001. *In the Name of the Father: surnames and genetics*. *TRENDS in Genetics*. Vol. 17 No. 6: 353-357.
- 4 Ugo A. **Perego**, Natalie M. **Myres**, Scott R. **Woodward**. 2004. *Reconstructing the Y-chromosome of Joseph Smith Jr. and Genealogical Applications*. BYU Studies.

Ugo A. Perego, MSc, is the senior project administrator for the Molecular Genealogy Research Project, **Sorenson** Molecular Genealogy Foundation. He was aided in writing this article by Natalie M. **Myres** and Scott R.

Woodward. Natalie M. **Myres**, MSc, is co-principal investigator for the Molecular Genealogy Research Project, **Sorenson** Molecular Genealogy Foundation. Scott R. **Woodward**, Ph.D., is the principal investigator for the Molecular Genealogy Research Project, **Sorenson** Molecular Genealogy Foundation, and is professor of molecular biology at Brigham **Young** University.

PUBLISH YOUR MANUSCRIPT

We'll Show You How!

Quality Books - Reasonable Prices
Acid Free Paper - Professional Design
Hardback or Paperback



HERE'S WHAT SOME OF OUR
AUTHORS HAVE TO SAY:

"You made a new experience
simple and pleasant."

"...no better first step than
getting in touch with you."



Call Ann Hughes at (410) 837-8271
or Email: ahughes@gatewaypress.com

Send for our free brochure "Guide for Authors"

www.gatewaypress.com

1001 North Calvert Street, Baltimore, MD 21202-3897

CIRCLE #179 ON READER RESPONSE CARD FOR MORE INFORMATION