Molecular Genealogy—A Genetic Approach

Ugo A Perego, MS, Natalie M Myres, MS, Scott R Woodward, PhD, Jayne E Ekins, BS

Sorenson Molecular Genealogy Foundation

2511 South West Temple, Salt Lake City, Utah 84115 (USA)

INTRODUCTION

Everyone has ancestors, but not everyone knows who they are and where they came from. The majority of people currently living in the USA and in Australia descend from immigrants of foreign countries. Genealogical investigation has shown that through the process of immigration, or because of adoptions, illegitimacies, or other causes, genealogical records have often been changed, lost, destroyed, or never kept in the first place. As a result, many individuals cannot find a country of origin for one or more of their ancestors. What can be done to restore the link to their rightful heritage? The answer might be found in DNA.

Genetic research, with the purpose of tracing genealogies by using DNA, is currently underway at the Sorenson Molecular Genealogy Foundation. This study is known as the Molecular Genealogy Research Project (MGRP) and its main goal is to develop a very large and comprehensive database containing correlated genealogical and genetic data from all over the world. This database will then be used to develop applications that will assist family historians in their genealogical research.

WHAT IS DNA?

DNA (Deoxyribonucleic Acid) is the genetic material containing all of the genetic information necessary for living organisms and is the repository of hereditary traits. The human genome (DNA) is the complete genetic blueprint of a person, and it consists of almost four billion chemical pairs. It is found mainly in the nucleus of cells, in structures known as chromosomes: twenty-three received from the father and twenty-three received from the mother. In addition to nuclear DNA, there is also genetic material found inside mitochondria—the energy-producing organelles found in the cytoplasm. These genetic components contain all of the necessary information for the foundation and the sustaining of human life. The color of our eyes, our height, our predisposition to certain diseases are just a few examples of what is contained in the DNA of our cells.

WHAT IS MOLECULAR GENEALOGY?

DNA is transmitted from one generation to the next. Some parts are passed almost unchanged, while others experience a high rate of recombination. This mode of transmission from parents to children creates an unbreakable link between generations and it can be of great help in reconstructing family histories. Molecular genealogy is therefore a new way to do genealogy, where DNA is used in association with traditional written records. Since we have inherited our genetic material from our ancestors, our relatives share with us a portion of this information. The closer the relationship between two people, the more genetic information they share. This means that inside any family unit, the members of that unit share a greater quantity of genetic material than those outside of it. In other words, even though the entire DNA sequence of an individual is unique to that individual, similar genetic information can be found among those that descend from common ancestors. Through molecular genealogy it is possible to establish family links among individuals, families, tribes and populations by using the information encoded in DNA.

HOW DO YOU DO MOLECULAR GENEALOGY?

DNA is extracted from samples donated by volunteers

Figure 1

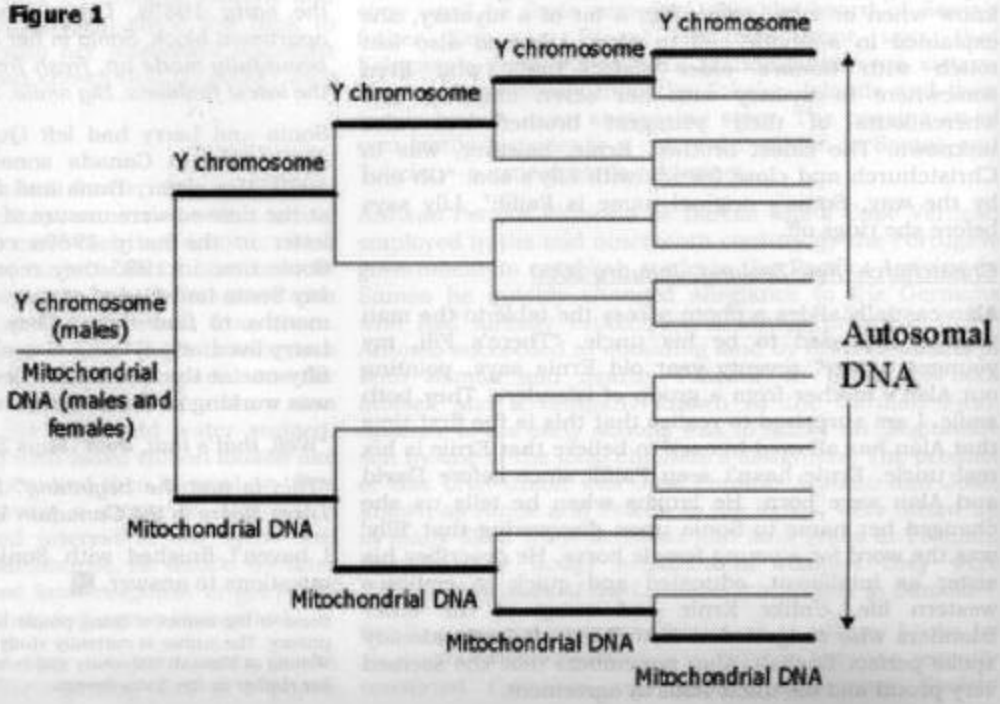
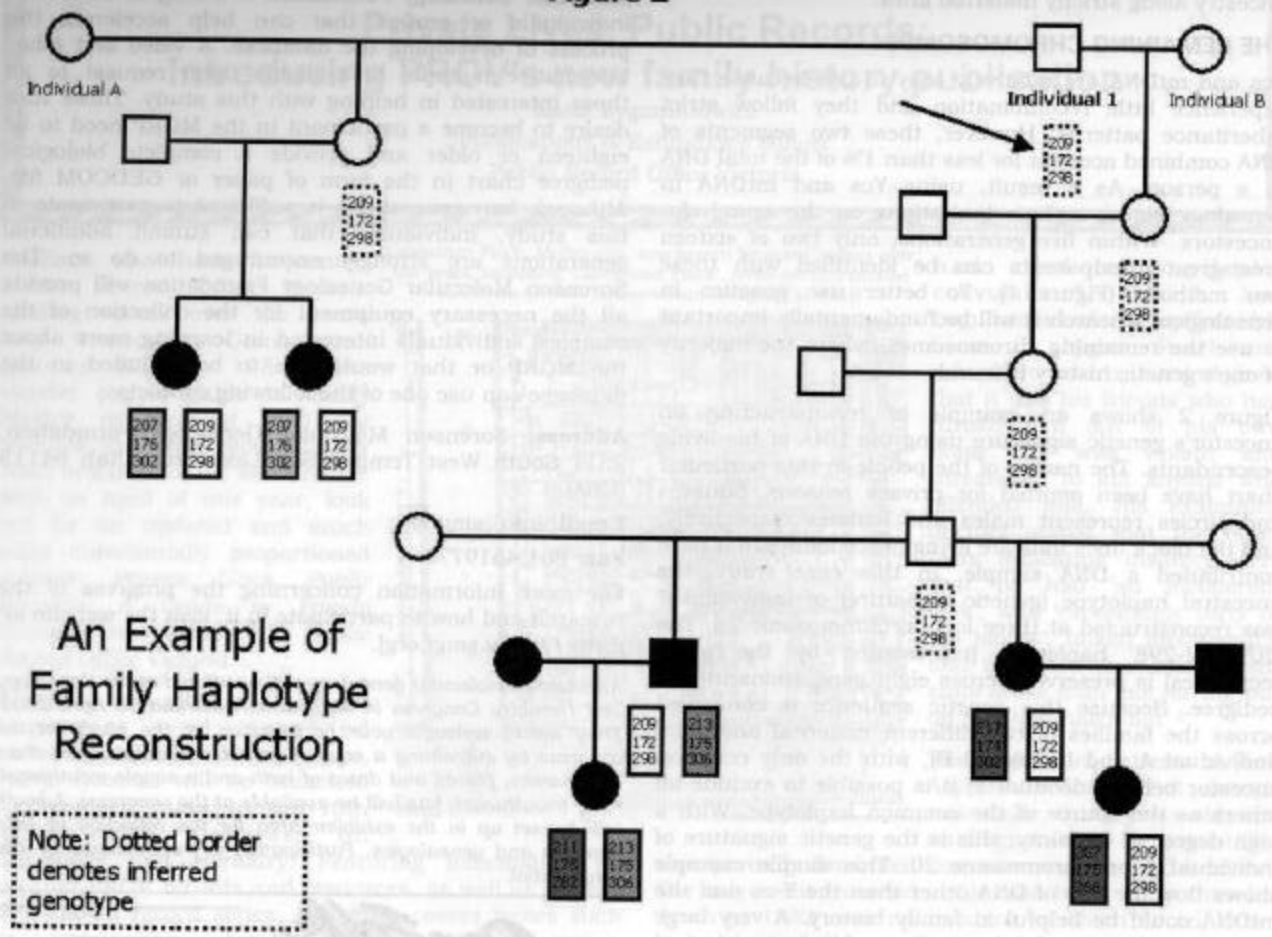


Figure 2



using basic laboratory techniques. Although there are many ways to obtain genetic material, the Sorenson Molecular Genealogy Foundation is currently using a simple mouthwash rinse. Each sample is analyzed at approximately 250 loci (sections of DNA) across the entire genome. This genetic information is then correlated to the genealogical data provided by the donors. The only genealogical information used is places and dates of birth for individuals born within the last one hundred years. After one hundred years, personal identifiers, such as names, are considered public domain and can be added to the database to be used for genealogical purposes. Names of donors are substituted with codes to protect participants' confidentiality. During the process of correlating the genetic markers to the genealogical information, individuals with similar ancestry and genetics will fall together into clusters representing specific populations. This procedure will allow the MGRP to reconstruct a genetic map of the world. Individuals with blocked genealogies could use the genetic map to trace the migration pattern of their ancestors in time and space. This objective will be developed in steps. Initially, the data will be useful to reconstruct genotypes found in large populations. As the database continues to grow and more data is collected, analyzed and correlated, populations will eventually be divided in sub-populations, then in clans and tribes, and eventually in family groups.

THE Y CHROMOSOME

While many DNA inheritance patterns are complex and difficult to interpret, two methods of genetic analysis known as Y-chromosome and mitochondrial DNA testing are relatively straightforward and more easily understood. These types of analysis are currently available to genealogists and are reliable for establishing certain types of family relationships.

Y chromosome (Y-cs) analysis is based on a fragment of DNA that is only found in males. It follows a strict inheritance pattern since it is always passed from father to son. Additionally, the Y-cs remains relatively unchanged from one generation to the next. Because of these inheritance properties, Y-cs testing can be a valuable tool for surname studies. There are hundreds of these studies currently underway, with the number growing each week. Additionally, Y-cs testing was used to support the existence of familial relationships in the highly publicized 1998 Jefferson-Hemings case, and the Jewish priestly class of Cohen study.

THE MITOCHONDRIAL DNA

Similar to the Y-cs, mitochondrial DNA (mtDNA) also follows a strict inheritance pattern, and remains relatively unchanged from one generation to the next. In contrast to the Y-cs, mtDNA follows a maternal inheritance pattern with a mother passing her mtDNA to all of her children, with only the daughters passing it to the subsequent generation. Genetic testing based on mtDNA analysis can be used to reveal the existence of a common female ancestor between two or more individuals. For example, mtDNA analysis was used to confirm the identity of buried remains thought to be Nicholas II, Czar of Russia, and members of his family, who were killed in 1918. mtDNA extracted from the remains were compared to that of living relatives sharing a common maternal line and supported a familial relationship. Additionally, mtDNA testing has recently

been offered to verify the presence of Native American ancestry along strictly maternal lines.

THE REMAINING CHROMOSOMES

Ycs and mtDNA are relatively easy to use because they experience little recombination and they follow strict inheritance patterns. However, these two segments of DNA combined account for less than 1% of the total DNA in a person. As a result, using Ycs and mtDNA in genealogy places serious limitations on the search for ancestors. Within five generations, only two of sixteen great-great-grandparents can be identified with these two methods (Figure 1). To better use genetics in genealogical research it will be fundamentally important to use the remaining chromosomes, where the majority of one's genetic history is found.

Figure 2 shows an example of reconstructing an ancestor's genetic signature using the DNA of his living descendants. The names of the people in this particular chart have been omitted for privacy reasons. Squares and circles represent males and females respectively, and the black ones indicate living descendants that have contributed a DNA sample. In this case study, the ancestral haplotype (genetic signature) of Individual 1 was reconstructed at three loci on chromosome 20. The '209-172-298' haplotype (represented by the white rectangles) is preserved across eight generations in this pedigree. Because this genetic sequence is conserved across the families of two different maternal ancestors (Individual A and Individual B), with the only common ancestor being Individual 1, it is possible to exclude all others as the source of the common haplotype. With a high degree of certainty, this is the genetic signature of Individual 1 on chromosome 20. This simple example shows how the use of DNA other than the Y-cs and the mtDNA could be helpful in family history. A very large and comprehensive database of correlated genealogical and genetic data such as the one being built by the MGRP could provide the necessary frame of reference wherein unique haplotypes could be associated to specific populations or family groups.

THE DATABASE

In the past three years, tens of thousand of people have volunteered to participate in the MGRP by donating a small biological sample and a copy of their extended pedigree chart. These samples have come mainly from people that live in the United States. Other sample collections have taken place in Oceania, Europe, South America and the Middle East. The first objective of the MGRP is to create a database of correlated genetic and genealogical information representative of worldwide populations. To accomplish this goal, the Sorenson Molecular Genealogy Foundation will collect hundreds of thousands more samples from all over the world. The database will continue to grow until every population of the earth will be properly represented, both genealogically and genetically. Due to the size and scope of this project, it is expected to take several years. Many findings and the related applications to genealogical research are expected from the study on a regular basis for the years to come. Participants in the first stage of the project will not receive any information back because of confidentiality issues.

CONCLUSION

Blocked genealogies, adoptions, and records that are missing or unreliable are all situations in which the MGRP might provide some help in the future. The success of this project is based on the participation of individuals with known genealogies from all over the world in the construction of the database. The Sorenson Molecular Genealogy Foundation is willing to work with individuals or groups that can help accelerate the process of developing the database. A video and other informative material is available upon request to all those interested in helping with this study. Those that desire to become a participant in the MGRP need to be eighteen or older and provide a complete biological pedigree chart in the form of paper or GEDCOM file. Although four generations is sufficient to participate in this study, individuals that can submit additional generations are strongly encouraged to do so. The Sorenson Molecular Genealogy Foundation will provide all the necessary equipment for the collection of the samples. Individuals interested in learning more about the MGRP or that would like to be included in the database can use one of the following contacts:

Address: Sorenson Molecular Genealogy Foundation, 2511 South West Temple, Salt Lake City, Utah 84115 (USA)

Email: info@smgf.org

Fax: 801 4619775

For more information concerning the progress of the research and how to participate in it, visit the website at: [http://www.smgf.org].

A lecture on molecular genealogy will be offered at the Genealogy and Heraldry Congress in Melbourne, Australia 25 April 2003. They would welcome your participation in the study at the congress by submitting a copy of your complete pedigree chart with names, places and dates of birth and a simple mouthwash rinse (mouthwash kits will be available at the congress). A booth will be set up in the exhibitor area for the collection of DNA samples and genealogies. Participation is free, voluntarily and confidential.