MOLECULAR GENEALOGY

at the

Brigham Young University

By Ugo A. Perego

PR for the Molecular Genealogy Research Group
Born in Milan, Italy

Introduction

A couple of weeks ago I talked to a man who spent many years traveling to the most remote regions of Peru microfilming records to be stored in Family History Centers for the benefit of genealogy. He told me that one night he arrived in a rural church and as he walked in, he saw the pastor burning some of the records to keep himself warm! This is just one sad example of how precious written materials are destroyed and the information they contain lost forever. What can be done when the records no longer exist? How is it possible to find names of our ancestors where a fire, a flood, or a careless keeper destroyed the only source available? What about the name you may be looking for that may not have been recorded in the first place?

The July-August 2000 issue of the *Everton's Genealogical Helper*, page 42, had a short article about the Molecular Genealogy Research Group (MGRG) at Brigham Young University (BYU) in Provo, Utah. This full text article intends to provide more details about this study and explain how genealogy work can be enhanced and extended through the use of molecular genetic information.

We are all familiar with the Human Genome Project (HGP), which aims are to decode the entire DNA sequence of a human being. The implications of this research are many. Pharmaceutical companies and health care providers are already talking about individualized medicine, where patients may be treated with a medication suitable to their genetic makeup or where "defective" genes may be replaced with healthy ones. We are all excited about the benefits that the HGP findings will bring to the medical field, but many of us may not be aware of other applications that can be made through this work. The information obtained from the HGP has been crucial to identify the specific genetic markers that can be used to characterize world populations. As the project continues to develop and the gaps are "filled in" we will have more information available for the genetic study of populations and families.

What is DNA?

The complete set of genetic instructions for making an organism is called its genome. It contains the master blueprint for all cellular structures and activities for the lifetime of the cell or organism. Found mainly in the nucleus of a person's many trillions of cells, the human genome consists of tightly coiled threads of deoxyribonucleic acid (DNA) and associated protein molecules, organized into structures called chromosomes. If unwound and tied together, the strands of DNA would stretch more

than 5 feet but would be only 50 trillionths of an inch wide. For each organism, the components of these slender threads encode all the information necessary for building and maintaining life, from simple bacteria to remarkably complex human beings. The color of our eyes, our height, and how our body functions are just some of the traits coded by the three billion pieces of information in every person's DNA. We inherit our genetic compliment from our parents, which they received from their parents, and so on. This process results in a genetic family history, which can tell a great deal about our lineage and our relationship to other individuals.

What is Molecular Genealogy?

Our ancestry and history is contained within each of our cells, thus creating a molecular lineage. Molecular Genealogy is therefore a new type of genealogy based on DNA rather than strictly relying on written records. Since we receive this information from those who came before us, we share portions of our history and genetics with our closest relatives. In other words, even though every individual posseses a very unique genetic makeup, two brothers will have more information in common, compared to their first degree cousins, which will share more than their second degree cousins, and so on. In summary, Molecular Genealogy is the unique identification of individuals, families, clans or tribal groups, and populations using the information encoded in the DNA of an individual and the linking of these individuals and groups together in "family trees" in ways not possible with traditional methods.

Why do Molecular Genealogy?

For some, genealogy is a hobby, while for others it is part of finding out who they are. Throughout the world there is intense interest in the origins and histories of people. Some of this information is transmitted through oral or written histories. Civic and religious records have also documented the histories of families and communities. Unfortunately, the histories of some peoples and communities have been lost or destroyed through time. In such instances the written documents are uninformative or simply don't exist. This can present a significant obstacle for individuals who are trying to trace their "roots." By utilizing the genetic record that each individual retains of his / her past, it is possible to reveal important clues as to the origin and relationship of any individual to other persons or populations.

What makes Molecular Genealogy possible?

As explained earlier, each individual carries within their DNA a record of who they are and how they are related to all other people. The genome of an individual can be divided into three regions namely, genes, regulatory regions, and spacer DNA. Depending on which region is analyzed, information is found that can identify every person as a unique individual, as belonging to a family group, and belonging to a larger population. Pedigrees based on these genetic markers can reveal relationships not detectable in genealogies based only on names, written records, or oral traditions. There may be a number of individuals named John Smith, but a genetic identification is unique and can even discriminate between closely related individuals or those sharing the same name.

How do you do Molecular Genealogy?

DNA is isolated by simple techniques from volunteer participants. Blood is usually the preferred way to collect the genetic material, since a small amount will provide the large quantity and quality of DNA necessary for database construction. DNA can also be extracted from saliva or hair samples. Giving the DNA samples coded identification numbers protects the identity of participants. Samples are genotyped in a laboratory on the BYU campus. Once the genetic information is available, we link it to the genealogies, including birth date and place of the donor and all other genealogies in the database. In this way a large family tree is built with the genetic markers correlated with the appropriate written genealogies. Once a genealogical question is posed, the database can be used to answer the question. For example, it is not uncommon for information about a grandparent to be missing or unknown. In such cases, the written records can only be used to provide a portion of an individual's family history. However, despite the absence of written records, genetic clues are available as to the identification of this "missing person." By testing the DNA of individuals in the pedigree, we can identify genetic markers inherited from ancestors and compare those with information in the database in order to find a close match. Additionally, it will be possible to identify family groups and populations of origins for people in the event that genealogical information is not known.

Y Chromosome Inheritance

Genetic tools have also been used to help establish family relationship in instances where families share the same surname. The sex of an individual is determined by the combination of the sex chromosomes that are designated X and Y. Males contain an X and Y chromosome. Females will have two copies (one received from each parent) of the X chromosome. While it is estimated that there are 3.2 billion units of DNA in the entire genome, the Y chromosome (Ycs) is relatively small, having only about 60 million units of genetic information. The Ycs is passed intact from father to son at each generation. If we were looking at a family tree, the Ycs inheritance pattern follows the paternal line to the left-outermost part of the chart. Individuals with the same surname (or similar last names) may not be able to prove an existing relationship by using traditional written records, but through the analysis of the Ycs it may be possible to discover if there was a shared male ancestor. Fig. 1 shows the Ycs inheritance pattern indicated by the black squares on the left side. Females are represented with circles. Male 4, 8, 11 and 15 all share the same Ycs, inherited from individual 1. Note that even though male 15 and male 18 have the same set of great-great grandparents, they possess a different Ycs.

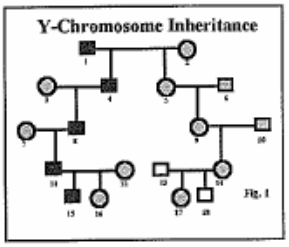

Fig. 1

Mitochondrial Inheritance

While the majority of genetic material is contained within the nucleus of every cell, a small quantity of DNA is found in the cellular mitochondria (the organelle responsible for the production of cellular energy). This genetic material is called mitochondrial DNA (mtDNA) and it is inherited by all children (both male and female) from their biological mother, but only passed on through daughters. In this case, if we were looking at a family tree, the mtDNA inheritance pattern follows the right-outermost part of the chart. Again, this information is very useful in the search for common ancestors, where written records are not available, or to explore an existing relationship. In Fig. 2, the black squares and circles represent the mitochondrial DNA inheritance pattern. Individuals 15-16 and individuals 17-18 share the same set of great-great grandparents, but inherited mtDNA from two different lines.

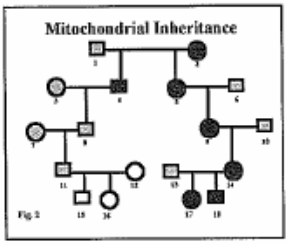

Fig. 2

Nuclear Inheritance

When passed to the children, the Y-chromosome and the mitochondrial DNA experience limited or negligible recombination. However, the remaining 22 pairs of autosomal chromosomes found in the nucleus are subject to shuffling or recombination from one generation to the next one and are not necessarily inherited intact. This characteristic property of genetics introduces the diversity found among peoples and is responsible for the unique genetic identity that defines an individual. By looking at specific locations on the nuclear DNA, it is possible to identify markers that will reveal the degree of relationship between any two individuals.

Why Genotype the World?

Every individual living today has (or had) two parents, four grandparents, and eight great grandparents. The number of progenitors doubles at each generation. This means that at a level of 30 generations (approximately 750 years) a single person would have one billion ancestors (Fig. 3). At the same time it is believed that the population of the world in 1250 AD was only about 400 million. The way to explain this discrepancy is that at some point in the past we each share one or more common ancestors (Fig. 4). Therefore the number of actual ancestors is much smaller than the number of possible ancestors (Fig. 5). This means that we are all related to varying degrees. These family bonds are continually lost through time and very few individuals are capable of keeping in touch

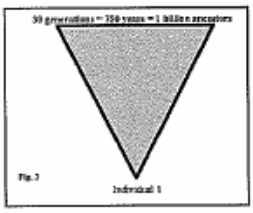

Fig. 3

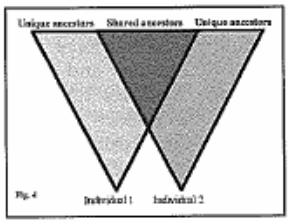

Fig. 4

with their extended families. This study aims to identify genetic links between populations and to construct a "family tree" of the people of the world. Moreover, we will be able to generate genealogical information in populations with no written records by using the samples collected in other areas. For example, we have collected several thousand samples in Peru, even in the tiniest villages in the mountains and jungles. Often these people barely knew who their grandparents were and sometimes have no written records about their past. We also collected samples in the big cities of Peru, such as Lima. Almost no one from Lima is originally from Lima, but have ancestors that moved there from other parts of the country. By matching the DNA of the people in the big cities with those in the small villages, we have been able to identify places of origins and migration patterns. This information can be used to help people "know where to look" for evidence of an unknown ancestor.

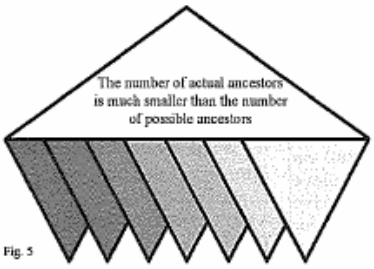

Fig. 5

How do we genotype the world?

Since March 2000, over 2500 DNA samples have been collected for this database primarily at Brigham Young University. Gradually this study will expand to other groups throughout the country and throughout the world. The ultimate goal for this study is to collect at least 100,000 samples representing 500 different populations across the globe. As the samples are collected, information regarding the place of origin of the donors and their ancestors is represented on a world map to give an idea of the provenience of the samples. We expect that to build such a database may require up to 5 years. However, we believe that after the first 10,000 samples are compiled in the database, it should be possible to generate some helpful insights that can be used in genealogical research. The key to make this project successful is to complete the database. The researchers are also interested in looking at genetic polymorphisms among groups of people who share the same geographic background. For example, it is possible to collect samples from an Italian population in the United States without having to travel abroad. Such individuals will retain genetic markers from their respective population of origin.

Conclusion

Family trees with blocked genealogies, cases of adoptions or illegitimacies, records showing individuals with the same name, etc. are all situations in which the Molecular Genealogy Research can help in providing answers. This is the reason why the Molecular Genealogy Research Group is very interested in people with known genealogies that would be willing to participate in this study and help in building this database. In order to participate volunteers must be 18 years or older and provide a four-generation pedigree chart. Those interested can write to the Molecular Genealogy Research Lab to find out when samples will be collected in a location near to you. Also, several groups have expressed interest in participating in this study. Groups with 100 or more people can contact the MGRG to schedule a visit at the earliest possible time. Participant confidentiality is maintained throughout the whole study. For more information, write to the Molecular Genealogy Research Group, 788 WIDB - BYU, Provo, Utah 84602 (USA), or send an e-mail to: molecular-genealogy@email.byu.edu. You can also access the website at: http://molecular-genealogy.byu.edu.