

Interview

With John Butler and Members of the NIST Human Identity Project Team

Interviews with people who play an important role in population genetics, forensic genetics, and genetic genealogy are presented in this feature. Suggestions for interview subjects are welcomed.

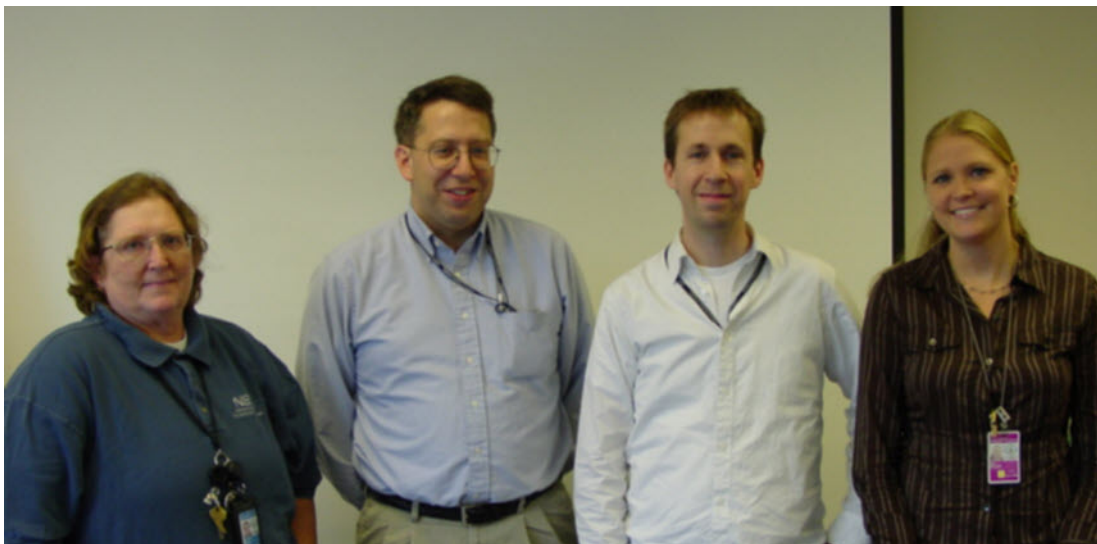
JoGG: Briefly, what is NIST and why is it important?

NIST: The National Institute of Standards and Technology (see <http://www.nist.gov>) is a non-regulatory federal agency within the U.S. Department of Commerce. NIST's mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life. The vision is to be the world's leader in creating critical measurement solutions and promoting equitable standards. With over 3,000 scientists spread primarily over two campuses in Gaithersburg, Maryland and Boulder, Colorado, NIST efforts center on stimulating innovation, fostering industrial competitiveness, and improving the quality of life. NIST aids quality assurance efforts and helps ensure compatible measurements by generating, certifying, and issuing Standards Reference Materials (SRMs). NIST defines time for the U.S. (see <http://www.time.gov>) and produces over 1100 different SRMs to aid measurements in a variety of fields.

JoGG: Why does NIST have a program in genetics?

NIST: Since the late 1980s, NIST has had scientists involved in DNA testing. Early concerns over measurement accuracy and issues with poor quality control of forensic DNA tests caused the Department of Justice to call upon NIST scientists to help with standards development and technology evaluation. Through Congressional funding, the National Institute of Justice (NIJ) has substantially funded the NIST and other nationwide efforts in improving forensic DNA testing. In 2000 with primary funding from NIJ, a Human Identity Project Team was established within the Biochemical Science Division of the Chemical Science and Technology Laboratory (CSTL) at NIST.

As of October 2008, the NIST Biochemical Science Division has been organized with efforts in Applied Genetics, DNA Science, Cell Systems Science, Protein Structure and Function, Bioassay Methods, and Multiplexed Biomolecular Science. There are efforts in our



NIST Human Identity Team, (L to R) Margaret Kline, John Butler, Peter Vallone, and Amy Decker

division to improve measurement capabilities for cell and tissue biological sciences for the biotech/pharmaceutical industry, DNA damage and repair, and data and standards applied to the fields of quantitative cell biology, proteomics, gene expression, and bioinformatics. Because our Division is tasked with many diverse measurement areas, and considering our limited staff and resources, we must prioritize our work and look at the impact in terms of needs for “standardization.” Historically, other agencies, such as the National Institute of Justice, have funded our work in the area of genetics.

JoGG: What are some of the activities of the Human Identity Project?

NIST: For the past several years, our Human Identity Project Team has been part of the DNA Measurements Group of the Biochemical Science Division. Beginning October 1, 2008, a new Applied Genetics Group has been formed to focus on forensic, clinical, and agricultural biotech genetic measurements. The Applied Genetics Group will continue to produce information, resources, and reference materials that should be able to benefit the genetic genealogy community as well although this is not our direct focus.

JoGG: It appears that most of your group’s efforts go to support the forensics community—is that a fair assessment?

NIST: Yes, our project team is funded through the National Institute of Justice to conduct research that benefits the human identity testing community and to create tools that enable forensic DNA laboratories to be more effective in analyzing DNA. We certify standard reference materials, conduct interlaboratory studies, produce new assays to enable improved recovery of information from degraded DNA, evaluate new loci for potential future use in human identity applications, and generate standard information and training materials that are made available on the NIST STRBase website: <http://www.cstl.nist.gov/biotech/strbase>.

JoGG: There seems to be a disconnect between the forensics and genetic genealogy communities concerning the value (usefulness) of Y haplogroups. The forensics community seems to have no interest in Y haplogroups (why is that?). Genetic genealogists seem to be completely puzzled as to why forensics journals publish so many articles on allele frequencies in particular populations, when those distributions are over mixed haplogroups. On the other hand, allele frequency distributions within haplogroups are of great interest to genetic genealogists. Surely one or the other of the two communities is missing something?

NIST: The forensic community is typically focused on using genetic markers that exhibit the highest possible

power of discrimination. Thus, autosomal DNA testing is preferred over lineage marker analysis since paternal (in the case of Y chromosome) or maternal (in the case of mtDNA) relatives would be expected to match. In the case of Y-chromosome testing, Y-STR haplotypes show more variability than Y-SNP haplogroups. Forensic analysts are looking for the “biggest bang for their buck,” since they are usually dealing with low amounts of available sample. Depending on the specific case, it probably will not be very helpful to know that a sample came from an individual that is R1b (or even R1b1b2a) because there may be hundreds of people from a region with potential access to the crime scene that would be considered suspects based on a Y haplogroup.

Although probably not as significant of a reason, Y-SNPs haplogroups may also be linked to ethnicity and thus may be avoided by some labs due to the potential perception of “racial profiling.” Of course, Y haplogroups have shown their value in human migration “deep ancestry” studies because they have a much lower mutation rate than Y-STR haplotypes.

Important as well is the fact that there is no common SNP typing assay platform, core Y-SNP markers, and widely-accepted commercial kits for Y-SNP analysis. Thus, it would be difficult to prepare reasonable reference materials or methods.

The bottom line is that each community is asking different questions. Forensic labs would like to distinguish suspects who are not the perpetrator (and have a greater chance that a matching result is to the true person) while genetic genealogists are looking for matches that link relatives and thus do not want so much of a genetic separation with their DNA test that they are individualized and cannot be grouped with related people. In simple terms, forensic scientists are primarily looking for exclusions while genealogists are looking for inclusions.

Forensic laboratories, in order to simplify quality control of their DNA testing reagents, utilize commercially available kits. Genetic genealogy test providers on the other hand use proprietary, in-house assays to analyze many more Y-STR markers than are examined by the forensic community. The currently available Y-STR kits used by forensic DNA labs are PowerPlex Y (Promega Corporation) and Yfiler (Applied Biosystems) that examine 12 or 17 Y-STRs, respectively. Both of these kits examine a core set of 11 Y-STR loci that were agreed upon in 2003 by a group of scientists as part of the FBI’s Scientific Working Group on DNA Analysis Methods (SWGDM).

JoGG: Your group maintains a Y-STR database with 11 markers. Why don’t you determine the haplogroup for each haplotype? Wouldn’t that help determine if the

haplogroup would have value for the forensics community? Why isn't this database searchable on all of the Y-STR data that you have available (many of your samples have been tested on many more than 11 markers)?

NIST: Actually, Y haplogroups generated with 42 Y-SNPs have been available on a subset of our population samples since July 2004 (Vallone and Butler, 2004). We found only 18 different haplogroups in 229 individuals tested and therefore decided that Y-SNPs were not as useful as Y-STRs in terms of separating unrelated males from one another. Since that work, our Y-chromosome studies have been focused almost entirely on Y-STRs. It is worth noting that a major challenge exists in terms of standardizing Y haplogroup measurements against a constantly expanding phylogenetic tree.

More recently our entire set of samples has been categorized into 19 Y haplogroups (with a set of 19 Y-SNPs) through collaboration with scientists in the Netherlands. This data will be forthcoming as part of an effort to understand ethnicity estimation capabilities and SNP variation across autosomal, Y-chromosome, and mitochondrial DNA compared to autosomal STRs and Y-STRs in the same sample set.

There are searchable databases containing all of our Y-STR data. The 17-locus Yfiler haplotypes from over 650 NIST samples have been included as anonymous samples in both the US Y-STR Database¹ and the YHRD database.²

Our U.S. population samples³ were obtained as anonymous blood donors from multiple blood banks and thus cannot be linked to any particular family or individual. The condition of anonymity was required by our institutional review board at NIST that approves the use of biological samples. Thus, while these samples are valuable to study general marker variation, they are less helpful for genetic genealogical analysis that may be interested in linking a result to a specific individual or family surname.

It is also worth noting that studies from various forensic labs around the world have found that approximately 95% of samples examined can be differentiated from one another with the 17 Y-STR markers found in the Yfiler kit. So, in terms of forensics investigations, this is typically enough information especially since these labs are testing limited amounts of biological evidence and may not have enough material for more than one or two

multiplexed PCR amplifications. However, for research purposes, our laboratory at NIST and Jack Ballantyne's lab at the University of Central Florida have examined a number of additional Y-STR loci to see how well samples can be resolved from one another with additional haplotype information. Our groups have developed numerous Y-STR assays, sought to understand which Y-STR loci are optimal for resolving paternal lineages, and examined mutation rates in father/son pairs.

JoGG: What are some of the specific activities of your group that JoGG readers may be interested in?

NIST: The certificate for SRM 2395 Human Y-chromosome DNA Profiling Standard⁴ has been updated to include information on new Y-STR loci—again with a primary focus on markers used in forensics or showing a high degree of variability. We hope to contribute to future discussions regarding STR nomenclature issues and feel that our certified reference materials can play a role in aiding quality measurements and reporting by the genetic genealogy companies. We have also examined the impact of additional Y-STR markers in terms of resolving common types (see Decker et al., 2007; and presentations available on our web site⁵). A set of almost 400 father-son samples were examined at the 17 Yfiler loci to measure mutation rates (Decker et al., 2008).

JoGG: You have a table on your web site with information on about 80 Y-STR markers. If you count up all of the Y-STR markers offered for testing by all of the genetic genealogy labs, there are about 120 of them, depending on just how one counts. Are these all of the Y-STR markers that are likely to become available to the genetic genealogy community, or are there significantly more that could be developed?

NIST: The table you are referring to on our NIST website is the Y-STR Fact Sheet page.⁶ Yes, there are many Y-STR markers that are not yet included on this page. We intend to continue adding information to this page as there are over 400 Y-STR markers characterized on the Y-chromosome. However, many of the loci formally located on the Y-chromosome have significant homology (sequence similarity) with the X-chromosome making it very difficult to design a Y-chromosome-specific assay. In addition, some information has not been verified or cannot be found in our searches of available public genomic databases.

1 See <http://www.usystrdatabase.org/>

2 See <http://www.yhrd.org/>

3 See <http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm>

4 See <http://www.cstl.nist.gov/biotech/strbase/srm2395.htm>

5 See <http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm>

6 See http://www.cstl.nist.gov/biotech/strbase/ystr_fact.htm

Our list of Y-STRs contains 151 markers, considerably fewer than 400, because some are not submitted correctly (repeated or redundant- different names for the same Y-STR), and some can not be found in GenBank (not enough information found). Most of our studies have focused on 82 Y-STRs that we originally thought might be informative in U.S. populations. However, some of these Y-STRs only had one or two alleles at that locus, which does not help in discriminating between individuals. What we have done has been fairly comprehensive in terms of reviewing Y-STR loci; therefore, we probably won't come across a "new" set of loci that are better than what is already being used.

JoGG: How much of the Y chromosome has been searched for Y-STRs?

NIST: More than half of the human Y-chromosome has never been sequenced—and probably never will be. The Human Genome Project did not attempt to determine the complex repeat patterns found on the long arm of the Y-chromosome in what is commonly referred to as the "heterochromatic region." The remaining euchromatic region, which is approximately 23 million bases in length, was reported as the "finished" reference Y-chromosome published in the June 19, 2003 issue of *Nature* by researchers from the Whitehead Institute and Washington University.

About five years ago, Manfred Kayser, a German scientist now at Erasmus University in Rotterdam (the Netherlands), led a group in comprehensively characterizing Y-STR markers. Their efforts were published in the June 2004 issue of the *American Journal of Human Genetics* (Kayser, 2004). Kayser's team found 166 new Y-STRs of which 139 were polymorphic in a set of eight diverse samples.

Erin Hanson and Jack Ballantyne from the University of Central Florida published an article in the March 2006 issue of the journal *Legal Medicine* which describes at least 417 separate Y-STR markers based on information deposited in public databases. The figures and tables from their article are available at <http://ncfs.ucf.edu/ystar/ystar.html>. As they note in their Table 2 there are, unfortunately, many redundant Y-STR locus designations meaning that the true number of "real" Y-STRs is less than 400.

It is important to realize that not all Y-STR markers are of equal value in terms of exhibiting variation between individuals or being able to be PCR-amplified as Y-chromosome specific. In order to examine variation in individual samples, Y-STR markers must also be put into multiplex assays that work reliably to order to make sample testing more manageable and cost effective. Thus, not every Y-STR marker that exists on the human Y-chromosome is going to be useful or used.

JoGG: There are now about one million SNPs tested simultaneously by "gene chips" such as the Illumina chip. How likely is it that someone will develop a chip that could measure hundreds (or thousands?) of Y-STRs at once?

NIST: The Illumina SNP chip systems, such as used by 23andMe, can generate an incredible amount of genetic information from bi-allelic SNP markers. Unfortunately, these hybridization arrays are unable to analyze multi-allelic STR markers. Thus, Y-STRs will continue to be processed for the foreseeable future with proven capillary electrophoresis technology.

JoGG: Your web site now has some references to genetic genealogy. What role would you like NIST to play in the genetic genealogy area?

NIST: While our focus has been on Y-STR markers of forensic interest due to funding in this area, we certainly would like to see measurements performed properly and consistently between genetic genealogy test providers. Hopefully our efforts in producing the Human Y-chromosome DNA Profiling Standard Reference Material (SRM 2395) – see:

<http://www.cstl.nist.gov/biotech/strbase/srm2395.htm>

will aid genetic genealogy as well as forensic DNA Y-chromosome analysis. Unfortunately, we probably will not be able to easily provide certified reference values for all Y-STR loci currently under examination by the genetic genealogy community due to limited resources at NIST. We have reviewed information on specific Y-STR loci per requests in the past, made nomenclature recommendations, and provided genetic diversity information on a number of Y-STR markers.

JoGG: One of the biggest problems in the genetic genealogy community is the different Y-STR nomenclature standards used by different labs. FTDNA stated at its last meeting in Houston that they had written to you and requested that NIST provide guidance on how each marker should be scored. Have you responded to this request? Can we expect NIST to make public some recommendations in this area, and when might that happen?

NIST: In October 2007, at the International Symposium on Human Identification, John Butler met with Matt Kaplan, Taylor Edwards, and Thomas Krahn from FamilyTree DNA and discussed a number of specific Y-STR nomenclature examples.

The DNA Commission of the International Society of Forensic Genetics (ISFG), in their latest update of recommendations on the use of Y-STRs in forensic analysis, has published eight nomenclature guidelines (Gusmão et

al., 2006). Specific guidance is given on over 70 Y-STRs (see their Tables 1 and 2). However, there are some additional Y-STRs run by progressive genetic genealogy companies that are not included in this article. Unfortunately, these recommendations do not necessarily capture all possibilities and some ambiguity can exist, especially with more complex repeat structures.

The five male sample components present in the NIST SRM 2395 Human Y-chromosome DNA Profiling reference material have been certified for some additional Y-STRs. We have prepared a paper (Butler, et al., 2008--see this issue of JoGG) to address some of the issues seen with Y-STRs we have reviewed in our lab and included on the SRM 2395 update.

JoGG: If a genetic genealogy company decides to introduce a new marker, would your group be willing to provide advice on scoring it?

NIST: We are open to discussing new loci, our response to whether or not we can assist in determining repeat motifs will depend on our focus at the time of the request. We will try to assist as time and resources permit. Knowledge of what loci are being investigated assist us in assessing standards needs that help formulate our future plans

JoGG: What can the genetic genealogy community do for your group?

NIST: Please continue to inform us about what information is of interest to the genetic genealogy community (such as consistent nomenclature with Y-STRs) and continue open communication with us on ideas and projects. Realize that our focus is where our funding is and therefore we cannot always respond to every desire of those who contact us. Although we are spread thin and working on other projects, we definitely would like to support progress and quality measurements in genetic genealogy.

Web Resources

<http://www.usystrdatabase.org/>

U. S. Y-STR Database

<http://www.yhrd.org/>

International YHRD Database

<http://www.cstl.nist.gov/biotech/strbase/srm2395.htm>

SRM 2395 Human Y-Chromosome Profiling Standard

References

- Butler JM, Kline MC, Decker AE (2008) Addressing Y-chromosome short tandem repeat (Y-STR) allele nomenclature. *J Genet Geneal*, 4:125-148.
- Decker AE, Kline MC, Redman JW, Reid TM, Butler JM (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *ESI Genetics*, 2(3):e31-e35.
- Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci Int*, 157:187-197.
- Hanson EK, Ballantyne J (2006) Comprehensive annotated STR physical map of the human Y chromosome: forensic implications. *Legal Med*, 8:110-120; see also <http://ncfs.ucf.edu/ystar/ystar.html>
- Kayser M, Kittler R, Ralf A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet*, 74(6):1183-1197.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 453:825-837.
- Vallone P, Butler J (2004) Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *J Forensic Sci*, 49(4):723-732.