Advanced Topics at Family Tree DNA
Part 1: Y-DNA

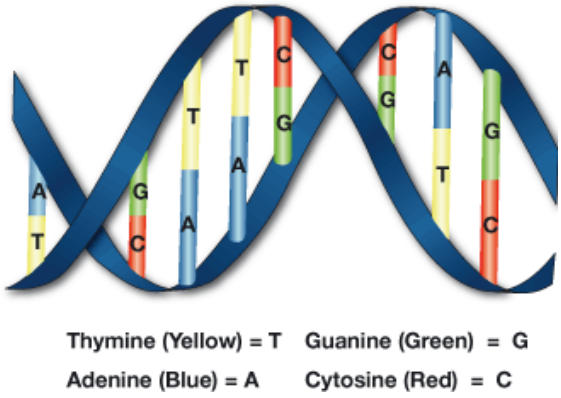
Presented By
Elise Friedman, Relative Roots

Take your genetic genealogy knowledge to the next level! This presentation further details the usage & genetics of Y-DNA. Topics include: NIST Standards, Compound Markers, Palindromic Region, Multi-Copy Markers, Micro-Alleles, Genetic Distance Models, Modal values, Haplogroups, SNP Discovery, Recurrent SNPs.

DNA Building Blocks

DNA is made up of four (4) bases (nucleotides):

- Adenine
- Cytosine
- Guanine
- Thymine



Mutations (changes) in the DNA happen over time.

Mutations happen randomly, but always between a parent/child pair. Mutations are what enable us to do genealogy and ancestry studies. They differentiate us from people who we're not closely related to, while matching us with people who we are closely related to. Genetic genealogy uses two types of mutations: STR and SNP.

Y-STR Results

Family Tree DNA provides the STR allele values on the *Standard Y-STR Results* page of your MyFTDNA account:

PANEL 1 (1-12)

Marker	DYS393	DYS390	DYS19*	DYS391	DYS385	DYS426	DYS388	DYS439	DYS389I	DYS392	DYS389II
Value	12	23	14	10	13-17	11	16	11	13	11	30

How to Read the Results

DYS393=12, DYS390=23, DYS19=14, etc.

Example: DYS393 = 12

- At marker DYS393
- Which has the repeat structure AGAT [This info isn't shown on the FTDNA pages]
- "AGAT" is repeated 12 times:

  AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT

The NIST website has details about many of the markers that Family Tree DNA tests, including the repeat structure for each one:

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

STR Naming Standards

DYS390, DYF411

- **D** stands for DNA.
- **Y** stands for Y-chromosome.
- **S, Z, or F** stands for the complexity of the repeat segment.
  - **S** is a unique segment.
  - **Z** is a number of repetitive segments at one site.
  - **F** is a segment that has multiple copies on the Y-chromosome.
- The number is a unique identification number.
- Some older STRs don't follow these standards.

Primers

5'-GTGGTCTTCTACTTGTGTCAATAC-3'

5'-AACTCAAGTCCAAAAAATGAGG-3'

- Primers are used to isolate the DNA fragment to be tested.
- Length of a primer is usually not more than 30 nucleotides (commonly 18-24)
- Primers for each marker can be found on the NIST website. For example:

http://www.cstl.nist.gov/biotech/strbase/str_y393.htm

NIST Standards / Nomenclature

- National Institute of Science & Technology (U.S.)
- Developed standards for many Y-STR markers.
- All companies adopting NIST standards.
    - Companies using Sorenson lab have already converted.
    - FTDNA still has changes to make, including full display of micro-alleles. HUGE project.

Websites about the nomenclature standards:

http://www.cstl.nist.gov/strbase/

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

https://www-s.nist.gov/srmors/view_detail.cfm?srm=2395

http://www.cstl.nist.gov/biotech/strbase/YSTRs/H4_nomenclature.htm

Conversion Between Companies

Until NIST standards are adopted by all companies,
conversions are required on several markers for compatibility across companies.

Marker	Conversion
DYS441	FTDNA = Sorenson – 1
DYS442	FTDNA = Sorenson – 5
Y-GATA-A10	FTDNA = Sorenson – 2
Y-GATA-H4	FTDNA = Sorenson – 1

Compound Marker: DYS389

DYS389I	DYS392	DYS389II
13	11	30

DYS389 has two parts: a & b

DYS389I = a = 17 in the example to the right

DYS389II = a + b = 30 in the example to the right

So b = 13 (30-17)

When DYS389I mutates, the DYS389II value will also change since DYS389I is part of DYS389II. Keep this in mind when doing Genetic Distance calculations.

Example:

Compare 13-30 with 14-31

Although this looks like GD=2 because both numbers are different, it's actually GD=1.

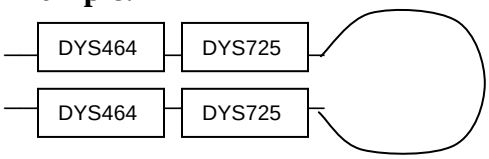
13-30: a = 13, b = 17

14-31: a = 14, b = 17 – so only "a" has changed.

Palindromic Region

A region of the Y chromosome where sections loop around like a hairpin. The markers in these sections have multiple copies, and they sit opposite each other on the long edges of the hairpin.

Example:



See Thomas Krahn's presentation about the Palindromic Region for a complete map of this region:

http://www.dna-fingerprint.com/static/PalindromicPres.pdf

Multi-Copy Markers

- Multi-copy markers can have the same or different values for each copy – they mutate independently.
- Standard STR tests cannot distinguish which value belongs to each position on the palindrome.
- Multi-copy marker values are always listed from low value to high value.

Example: DYS464 = 12-15-16-17

FTDNA Standard Marker Panels	FTDNA Advanced Orders
DYS385	DYF371
DYS459	DYS385 Kittler
DYS464	DYF387
YCAII	DYF397
CDY (aka DYS724)	DYF399
DYF395S1	DYF401
DYS413	DYF408
	DYF411
	DYS464X
	DYS725

Multi-Copy Markers: DYS385

- One pair on P4 (see the Palindromic Map)
- "a" copy on bottom of hairpin
- "b" copy on top of hairpin
- Kittler test can determine which value is in each position!

So DYS385 = 11-14 could really be DYS385=14-11, where a=14 (bottom copy), b=11 (top copy)

Multi-Copy Markers: DYS464

- Usually 4 copies:
  - One pair on P1
  - One pair on P2
- Can be up to 8 copies, perhaps even more
- DYS464x test sometimes help to determine which values are in each position
  - Two possible marker types:
    - G-Type
    - C-Type
  - Most haplogroups have only G-Types
  - Haplogroup R1b usually has three C-Types and one G-Type (though a SNP mutation can change this)
  - In haplogroup R1b, the single G-Type is on P2

Multi-Copy Markers: DYF371

- Usually 4 copies
  - One pair on P1
  - One pair on P5
- Three C-Types, one T-Type
- DYS425 is the DYF371 T-Type
  - DYS425 = NULL if the T-Type doesn't exist (due to mutation)

Multi-Copy Markers: DYF399

- 3 copies
  - One pair on P1
  - Single copy inserted between P2 & P3
- P1 has one C-Type and one T-Type
- Insertion between P2 & P3 is usually a T-Type

Recombinational Loss of Heterozygosity (RecLOH)

- One arm of a palindrome can folder over onto the other arm, which overwrites the the value on one arm with the value on the other arm.
- Results in duplicated values in multi-copy markers
- Not every duplicated value is due RecLOH. Standard mutations can cause two markers that previously had close values (ie, 17-18) to now have the same values (17-17 or 18-18).

Example:

- Haplogroup R1b normally has DYS385 = 11-14
- RecLOH can result in DYS385 = 11-11 or 14-14

Micro-Alleles (Partial Repeats)

cccgtgggaaaGTACGTACGTACGTACGTgacacccccgtca

- GTAC repeated 4 times
- Extra "GT" at the end – not a full repeat
- 4.2 repeats, rounded to 5
- Round up or down depending on how many bases are in the partial repeat in relation to the expected repeat motif:
  - Down: Partial repeat has less than half the nucleotides in the repeat motif.
  - Up: Partial repeat has half or more of the nucleotides in the repeat motif.

Another Example:

- DYS464 = $[CCTT]_n$
- If you have CCTT repeated 15 times, then CCT at the end:
  - 15.3
  - Rounds to 16 (3 nucleotides is more than half of the expected 4 nucleotides)

Genetic Distance Models

- Stepwise
  - Count every difference in marker values
    ie: 28 vs 30: GD=2
- Infinite Alleles
  - Count one difference even if the values differ by more than one.
    ie: 28 vs 30: GD=2

Family Tree DNA uses a mix of stepwise and infinite alleles, depending on the marker.

- Palindromic Markers
  - Infinite Allele method
  - Counts a difference for two types of changes:
    - a mismatch (values not the same)
    - a copy number change (more or fewer copies of the marker)

Examples:

(1)	***(2)***	***(3)***
14-15-15-16	14-15-16-17	14-14-15-16
14-15-18-18	14-15-15-16-16-17	14-15-15-16-16-17
GD = 1	GD = 1	GD = 2
Different values (16,18)	Same values	Different value (17)
Same # of copies	Different # of copies	Different # of copies

- NULL values
  - Single event, so GD=1
  - Exception: DYS439 (for historical reasons, hopefully this will be changed)
    - Null vs 12
    - GD=12

SNP Names

Designation	Research Lab
IMS-JST	Institute of Medical Science-Japan Science and Technology Agency, Japan
L	The Family Tree DNA Genomic Research Center, Houston, Texas, United States of America
M	Stanford University, California, United States of America
P	University of Arizona, Arizona, United States of America
PAGE/PAGES/PS	Whitehead Institute for Biomedical Research, Massachusetts, United States of America
PK	Biomedical and Genetic Engineering Laboratories, Islamabad, Pakistan
U	University of Central Florida, Florida, United States of America V La Sapienza, Rome, Italy
CTS	Welcome Trust Sanger Institute, Cambridge, UK
F	Fudan University, People's Republic of China
PF	Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, Italy
Z and DF	Various members of the Genetic Genealogy community, Worldwide

Source: Family Tree DNA FAQ

SNP Discovery

Company/Product	Description
FTDNA Walk The Y (WTY)	Tested DNA fragments using existing primers to search for new SNPs. No longer offered, may be replaced by something new.
Geno2.0	Genographic Project – Phase 2. Tests thousands of SNPs, those already on the Haplotree and a very significant number of new ones too.
1000 Genomes	First project to sequence the full genomes of a large number of people to provide a comprehensive resource on human genetic variation. Many new Y-DNA SNPs found through the results of this project.
23andMe	Tests nearly 4,000 Y-DNA SNPs, many already on the Haplotree, but there have been some significant new ones too.

Recurrent SNPs

- While it's not very common for the same position on the Y chromosome to mutate more than one, it can and does happen.
- When a SNP recurs, it's most often in different Haplogroups, making the SNP hierarchy necessary for understanding which haplogroup a person is really in when he tests positive for a recurrent SNP.
- In rare cases, the same position mutates more than once within the same haplogroup.

Example:

- M35 is the SNP that defines haplogroup E1b1b1, characterized by a mutation from G to C.
- One family within E1b1b1 tested negative for M35, so they were originally thought to not be in E1b1b1.
- However, analysis of the STRs led to the realization that this family very likely was in haplogroup E1b1b1.
- Additional testing was done to confirm. FTDNA discovered that the M35 position had mutated to a T! This happened many thousands of years later.
- The M35 SNP test did not detect the T because it just looks for the C. If there's not a C at that position, then the test is negative for the M35 SNP.
- The family was also tested for M243, which is currently considered to be equivalent to M35 on the Y-DNA tree, and was positive. And they tested positive for V13, which defines a subclade below E1b1b1.
- Original M35 re-named to M35.1; new mutation named M35.2

Websites

- Relative Roots
  - http://www.relativeroots.net
  - http://www.facebook.com/RelativeRoots
  - http://www.twitter.com/RelativeRoots
- Family Tree DNA
  - http://www.familytreedna.com/
  - http://www.familytreedna.com/faq/
  - http://www.facebook.com/FamilyTreeDNA
- International Society of Genetic Genealogy (ISOGG) – it's free to join!
  - http://www.isogg.org
- Genetic genealogy mailing lists and discussion forums:
  - http://www.yahoogroups.com/group/ISOGG
  - http://www.yahoogroups.com/group/DNA-Newbie
  - http://forums.familytreedna.com
  - http://www.dna-forums.org
  - http://lyris.jewishgen.org/ListManager (JewishGen)
  - http://lists.rootsweb.ancestry.com/index/other/DNA/GENEALOGY-DNA.html

Public Searchable Databases

- Ysearch http://www.ysearch.org
- SMGF http://www.smgf.org
- YHRD http://www.yhrd.org

Thank you for attending a *Relative Roots* webinar!

Please visit our website often to learn about additional webinars and/or join the Relative Roots Webinars Announcements email list at http://www.relativeroots.net/webinars/

Follow Relative Roots on Facebook and Twitter!

www.facebook.com/RelativeRoots

www.twitter.com/RelativeRoots