

Investigations of the Y Chromosome, Male Founder Structure and YSTR Mutation Rates in the Old Order Amish

Toni I. Pollin^a Daniel J. McBride^a Richa Agarwala^b Alejandro A. Schäffer^b
Alan R. Shuldiner^{a, c} Braxton D. Mitchell^a Jeffrey R. O'Connell^a

^aDepartment of Medicine, University of Maryland School of Medicine, Baltimore, Md., ^bNational Center for Biotechnology Information, National Institutes of Health, Department of Health and Human Services, Bethesda, Md., ^cGeriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, Md., USA

Key Words

Amish · STR mutation rates · Y chromosomes STRs · Genealogy · Founder population

Abstract

Objectives: Using Y chromosome short tandem repeat (YSTR) genotypes, (1) evaluate the accuracy and completeness of the Lancaster County Old Order Amish (OOA) genealogical records and (2) estimate YSTR mutation rates. **Methods:** Nine YSTR markers were genotyped in 739 Old Order Amish males who participated in several ongoing genetic studies of complex traits and could be connected into one of 28 all-male lineage pedigrees constructed using the Anabaptist Genealogy Database and the query software PedHunter. A putative founder YSTR haplotype was constructed for each pedigree, and observed and inferred father-son transmissions were used to estimate YSTR mutation rates. **Results:** We inferred 27 distinct founder Y chromosome haplotypes in the 28 male lineages, which encompassed 27 surnames accounting for 98% of Lancaster OOA households. Nearly all deviations from founder haplotypes were consistent with mutation events rather than errors. The estimated marker-specific mutation rates ranged from 0 to 1.09% (average 0.33% using up to 283 observed meioses

only and 0.28% using up to 1,232 observed and inferred meioses combined). **Conclusions:** These data confirm the accuracy and completeness of the male lineage portion of the Anabaptist Genealogy Database and contribute mutation rate estimates for several commonly used Y chromosome STR markers.

Copyright © 2007 S. Karger AG, Basel

The Old Order Amish (OOA) of Lancaster County, Pennsylvania, are a closed founder population numbering approximately 30,000–50,000, nearly all of whom can trace their ancestors back to a small number of individuals who immigrated to the United States in the mid- to late 1700s [1, 2]. An additional group of OOA immigrated during this period to Ohio and Indiana, and later some Lancaster OOA migrated westward as well. The OOA have a strong interest in their ancestry, and their genealogical relationships are well-documented [3, 4]. These attributes make the OOA an attractive population for genetic studies [5], and indeed they have been subjects of study of the genetics of both single gene disorders for over 40 years [reviewed in 6] and of complex traits for almost as long [7], but particularly in the last 15 years by our group [8–16] and others [17, 18]. In recent years, the Anabaptist Genealogy Database [1, 19, 20] has been developed

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2007 S. Karger AG, Basel
0001-5652/08/0652-0091\$24.50/0

Accessible online at:
www.karger.com/hhe

Toni I. Pollin, MS, PhD
660 West Redwood Street
Room 445C
Baltimore, MD 21201 (USA)
Tel. +1 410 706 1630, Fax +1 410 706 1622, E-Mail tpollin@medicine.umaryland.edu

as a computerized resource searchable using the Ped-Hunter query software by groups with Institutional Review Board approved protocols [1], which enhances the ability to construct pedigrees defining the relationships between Amish subjects of genetic studies. These tools have been particularly useful for projects involving large study samples such as that of the University of Maryland, which has recruited well over 3,500 Lancaster Old Order Amish individuals for several different studies, primarily of complex adult onset conditions including diabetes, osteoporosis and cardiovascular disease. Despite the multitude of genetic studies of the OOA, a rigorous analysis of the founder structure has yet to be reported. The work reported here focuses on the Lancaster, Pennsylvania OOA only. For convenience, throughout the rest of this report, 'Amish' and 'Old Order Amish' will be used to refer to the Lancaster, PA Old Order Amish population, which is a subject of study at the University of Maryland.

The recent availability of numerous short tandem repeat (STR) markers on the non-recombining region of the Y chromosome [21–24] has been applied to population genetics [25–45], genealogy [46–52] and forensics [53–59]. The rate and mechanism of mutation of an STR (whether autosomal or sex-linked) determines not only its degree of polymorphism but also its usefulness for a particular application. High mutation rates with both gain and loss of repeat elements are potential confounders in analyses requiring inferences in the presence of substantial missing data. For example, in linkage analysis hypermutable autosomal STR mutations that lead to Mendelian inconsistencies are often classified as genotyping errors.

While the high mutation rate of STR markers also makes using them to track long-term evolutionary patterns difficult [60], their polymorphic nature makes them useful for distinguishing lineages, helping to understand relationships between lineages [23] and clarifying recent demographic history [61, 62]. By comparing the Y chromosome haplotypes between male lineages, we would be able to confirm the accuracy of the genealogical records and also determine whether individuals with similar surnames came from common founders.

In addition to allowing us to confirm genealogical records and estimate the number of male founders, such a large number of observed meioses in large families, as are present in the Amish, allowed us to estimate mutation rates in the STR markers. Two approaches have been used to calculate mutation rates in Y chromosome STR markers: large pedigrees with males connected through com-

mon male lineages with observed/inferred meioses and father-son pairs. Heyer et al. [63] typed 42 males from 12 'deep rooting' Canadian pedigrees for 9 STRs and estimated individual marker mutation rates ranging from 0 to 0.94%, with an average of 0.21%. However, since almost all of the meiotic events leading to the apparent mutation events were unobserved, mutations could not be confidently distinguished from nonpaternity (although Jobling et al. [64] partially addressed this concern using the minisatellite marker MSY1), and there was insufficient data to infer the direction of the mutations. Subsequently, Kayser et al. evaluated 4,999 meioses in 15 loci in typed father/son pairs that had undergone paternity testing and observed 14 mutation events, for an overall mutation rate of 0.28% [65], in contrast to Bianchi et al., who found no mutations in 1,743 meioses in seven loci [66]. Dupuy et al. studied 1,766 confirmed father/son pairs and found an overall mutation rate of 0.23% [28]. Several other investigators studying father/son pairs [26, 29, 67–72] obtained similar estimates. More recently, Bonn -Tamir et al. revived Heyer's method in 74 male samples from the highly isolated Israeli Samaritan population, which has similarly detailed genealogical records as the Amish, to arrive at an estimated mutation rate of 0.42% [73]. Additionally similar estimates have been made using sperm samples [74]. Lower 'evolutionary' estimates have also been made using cross-population samples [32, 43, 75]. The large number of Amish males genotyped in our studies (739 males at up to nine different STR markers) coupled with extensive pedigrees enabled us to use a large number of genotyped father/son pairs as well as a larger number of transmissions inferred from the pedigree structure to evaluate mutation frequency.

Subjects and Methods

Subjects

As of April 2003, a total of 2,480 subjects, including 1,080 males, had been recruited for several University of Maryland studies. Subjects consented to these studies via protocols approved by the University of Maryland Institutional Review Board. The construction and usage of the Anabaptist Genealogy Database is covered by a human subjects protocol overseen by an Institutional Review Board at NIH. Of the total 2,480 subjects, 1,249 subjects (506 male) from the Amish Family Diabetes Study (AFDS) [12, 14], Amish Family Osteoporosis/Calcification Study [76] or Amish Osteogenesis Imperfecta Study were genotyped using DNA extracted from leukocytes by the NHLBI Mammalian Genotyping Service for 800 STR markers (5 cM scan) from sets 11 and 51 (NHLBI Mammalian Genotyping Service), including nine markers on the Y chromosome (see below). An additional 514 (233 male) subjects from the AFDS were genotyped in an earlier 10 cM

scan using 400 markers from set 11 only, which included seven of the nine Y chromosome markers. Y marker data were thus available on a total of 739 males.

Marker Genotyping

Markers typed by the NHLBI Mammalian Genotyping Service in all subjects included DYS393/395, DYS391, DYS389-I, DYS389-II, DYS388, DYS390, and DYS392. The additional markers DYS19 and GGAAT1B07 were typed only in the 506 males in the 5 cM scan. Because two lineages with no genealogical, historical or surname evidence of relatedness shared the same apparent nine marker founder haplotype, we sequenced in a subset of individuals in a subset of lineages three additional single copy markers which were chosen based on high diversity statistics calculated previously [22]: DYS449, DYS456 and DYS458. Because we evaluated these markers in only a small number of individuals, we elected to sequence rather than genotype these markers to assure accurate allele calls.

Genealogy Analysis

The entire set of 1,080 male individuals enrolled in our studies was used in a query of the Anabaptist Genealogy Database version 3 (AGDB3), a large searchable database including content from three Amish genealogy sources [1, 3, 4, 20, 77]. The query utilized the PedHunter software to connect all phenotyped males as far back as possible through male lineages only.

Statistical Analysis

Founder Allele/Haplotype Designation

We used the genotypes of the typed individuals to designate a founder allele for each marker in each lineage. The founder was defined as the most recent common male ancestor (MRCMA) for all genotyped individuals within a lineage. All but two lineages for one locus (DYS391 in both cases) had unique putative founder alleles at each locus. For these two lineages at DYS391 we designated the founder allele as the one of two possible alleles that maximized the number of distinct gene flows with the fewest number of mutations that fit the data (see *Results*; *Y Chromosome Haplotypes*, for details). The putative founder haplotype was the set of founder alleles inferred in this manner. We note that for the two lineages above, the founder haplotypes are distinct from all other founder haplotypes regardless of which DYS391 allele is chosen, implying that comparisons made below with regard to similarity of lineages are actually independent of the choice of founder allele.

Y Chromosome Haplotype Reference Data

We used two publicly available databases for reference data on European Y chromosome genotypes and haplotypes. The Y-STR Haplotype Reference Database (YHRD), and its sister site, the Y-STR Haplotype Reference Database for U.S. Populations (YSTR-US) [78] (now combined into a single YHRD database [79]) are freely searchable but restrict submission of genotype data to forensic laboratories that have passed a quality control exercise and are limited to a set of ten STRs (DYS19, DYS389-I, DYS389-II, DYS390, DYS391, DYS392, DYS393, DYS385ab, DYS438 and DYS439), which includes seven of the nine typed in our study. All samples submitted to STR/STR-US must minimally be typed for the first eight of the ten markers. In YHRD, any subset of the nine STRs can be searched for matching haplotypes, and such search-

es yield the worldwide prevalence of a given haplotype along with region- and ethnicity-specific prevalences. The YHRD (Release 21) currently includes 51,253 haplotypes in 447 populations. To facilitate systematic comparison of our data to YHRD data, we downloaded the subset of 12,727 haplotypes in 91 European populations available at the YHRD web site and used in a recent publication describing the use of YSTRs to describe European population history [61].

Another database, YBase: Genealogy by Numbers, allows unrestricted submission and searching for 49 individual STRs, including all nine typed in our study. However, haplotype searching in YBase requires at least eight markers and is of limited utility for estimating population prevalence since there is no particular minimal set of markers required for inclusion in the database, and denominators are not given for haplotype search results. YBase primarily provides surnames of matching haplotypes with limited geographic information. The distribution maps show that most of the samples are sent by individuals and families residing in the eastern United States, the United Kingdom, Germany and Switzerland. YBase provides periodically updated tables of allele frequencies for individual STRs.

Pedigree Errors and Mutation Analysis

We initially reasoned that individuals with apparent non-founder alleles at multiple loci were most likely to represent pedigree errors and investigated these cases further, including autosomal loci, to confirm or refute this suspicion. After exclusion of pedigree errors, for a given locus within a lineage, if three or more individuals shared the same non-founder allele and had a MRCMA who was not the root and furthermore all of this MRCMA's descendants possessed this same allele, this was considered confirmatory evidence of a mutation. If the same were true in a set of two individuals, this was considered preliminary evidence of a mutation; sequencing of both the individual(s) possessing the putative mutation and additional relatives if available was used to confirm the mutation. Similarly, sequencing was used to confirm an apparent mutation appearing in a single individual. Sequencing was also used to localize historical mutation events, even those appearing in clusters of three or more individuals, if DNA was available from the relevant individuals.

PCR product sizes were used with available sequence information (see below) to convert allele names to repeat lengths named according to standard nomenclature [21]. A special case is DYS389-II, which has the structure $[TCTG]_n[TCTA]_m[48bp]$ $[TCTG]_3[TCTA]_q$, of which the last portion, $[TCTG]_3[TCTA]_q$, defines DYS389-I [80]. We provide the repeat length for DYS389-I as 3+q (as is standard), and for DYS389-II, we provide in table 1, which lists founder haplotypes, repeat length in the format $n+m+3+q_{n+m}$, in order to preserve the YHRD nomenclature ($n+m+3+q$) while simultaneously providing the repeat length of the DYS389-II specific segment ($n+m$) of the marker, which is used in some population and evolutionary studies. This representation enables our data to be readily compared with other publications and databases, which vary in their formatting of this marker. For example, a founder with the genotype $[TCTG]_5[TCTA]_{12}[48 bp][TCTG]_3[TCTA]_9$ would be denoted as DYS389-I = 12 and DYS389-II = 29₁₇. In discussing specific alleles and mutations in the DYS389-II-specific segment in the text, we use the DYS389-II-specific format $n+m$ (DYS389-II = 17 in the preceding example).

Table 1. Putative founder Y STR haplotypes: Lineages are rank ordered by number of male individuals genotyped

Lineage	DYS393	DYS19	DYS391	DYS389-I	DYS389-II ^a	DYS388	DYS390	DYS392	GGAAT1B07	DYS458	N ^b	MRCMA ^c
1	13	14	11	13	29 ₁₆	12	24	13	10	17	158	1749
2	13	14	11	13	31 ₁₈	13	24	13	10		135	1757
3	12	14	10	13	30 ₁₇	14	24	11	11		79	1778
4	13	14	11	12	28 ₁₆	12	24	13	10		78	1757
5	13	15	10	12	29 ₁₇	12	22	11	11		66	1729
6	13	14	11	13	29 ₁₆	12	23	13	10	18	50	1740
7	13	14	11	14	31 ₁₇	12	24	13	10		41	1737
8	13	14	10	14	30 ₁₆	12	24	13	10		20	~1690 ^d
9	14	16	10	13	31 ₁₈	12	25	11	9	15	16	1866
10	13	14	11	13	29 ₁₆	12	23	13	10	17	15	1894
11	13	14	11	14	30 ₁₆	12	23	13	11		15	1797
12	13	14	10	12	28 ₁₆	14	22	11	10		10	1850
13	13	14	11	14	32 ₁₈	12	24	13	8		8	1839
14	13	14	10	13	29 ₁₆	12	22	13	10	18	8	1763
15	14	15	10	14	30 ₁₆	12	22	9	11		7	1771
16	13	14	11	12	27 ₁₅	14	23	11	11		5	1838
17	13	14	12	13	29 ₁₆	12	24	13	10		4	1869
18	12	14	11	14	30 ₁₆	15	23	11	11		4	1920
19	14	14	10	12	28 ₁₆	13	22	11	11		3	1919
20	14	16	10	12	29 ₁₇	13	23	12	11		3	1864
21	13	15	11	13	29 ₁₆	12	23	13	10	17	2	1928
22	13	14	11	13	29 ₁₆	12	24	13		18	2	1918
23	13	15	10	13	28 ₁₅	12	24	14	10		1	1964
24	14	16	10	14	33 ₁₉	13	23	12			1	1952
25	13	14	10	13	28 ₁₅	12	24	13	10		1	1960
26	13	15	10	13	29 ₁₆	12	25	15			1	1951
27	13	14		13	29 ₁₆	12	23	13			1	1946
28	13	13	10	13	28 ₁₅	12	22	15	11		1	1933

^a Subscript indicates length of the DYS389-II specific segment.

^b Number of individuals in lineage genotyped in initial genome scan; includes those with mutations but excludes apparent pedigree errors.

^c Birth year of most recent common male ancestor of putative founder haplotype.

^d Birth year estimated based on 1710 birth of first child.

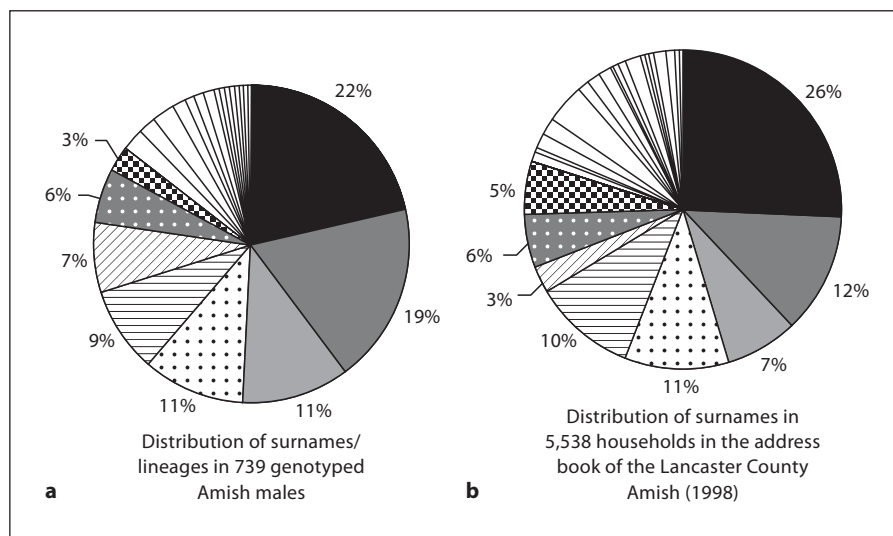
Two methods were used to estimate mutation rates. First, for each marker, the number of discordant typed father-son pairs was divided by the total number of typed father-son pairs (observed meioses). In the second method, which increased the sample size but also the number of assumptions made, we inferred as many genotypes as possible in each lineage using available genotypes along with our inferred founder genotypes. We then divided the total number of mutation events by the number of inferred/observed father-son transmissions for each marker. Binomial confidence intervals were calculated using the exact method as implemented in SAS Version 8.0 (Cary, NC).

STR Sequencing

Sequencing of STR markers, including previously typed loci to confirm and/or localize mutations and three additional loci to

distinguish lineages, was performed on an ABI 3700 DNA sequencer. In some cases primers were designed to amplify a PCR product larger than that originally detected by the NHLBI Mammalian Genotyping Service to guarantee readable sequence within the repeat region. For DYS389, the primer set used by the NHLBI Mammalian Genotyping Service and others, as a result of two binding sites of the forward primer, amplifies two products: the entire DYS389 region, classically denoted as DYS389-II, and the DYS389-I region contained within it. The DYS389 forward PCR primer was redesigned to bind to a unique site upstream of the original upstream binding site so that the entire sequence was only amplified once, allowing us to view within our sequencing result distinct sequences for DYS389-I and the DYS389-II-specific portion.

Fig. 1. Distribution of surnames and lineages in (A) our sample of 739 genotyped males and (B) the 5,538 households in the Address Book of the Lancaster County Amish. Surnames are listed in the same order in each figure; the 8 most common surnames in our studies are emphasized with distinct patterns.



Results

Genealogy Analysis

Querying AGDB with PedHunter for all 1,080 male subjects resulted in 30 male lineages. Two lineages comprised a total of three individuals with phenotype data, but no genotype data. Within each of the 28 genotyped lineages there was a unique surname after accounting for multiple spellings (e.g., Stoltzfus/Stoltzfoos, the most common Amish surname). However, the converse, that each surname corresponded to a unique lineage, was not true. Two surnames were each found in two separate lineages. The 739 males with Y chromosome STR markers genotyped could be traced to 28 of the founders, with each of the 28 founders having from one to 237 phenotyped descendants and from one to 159 descendants genotyped for all or some of the 9 STR markers. The distribution of founder descent of the genotyped individuals is shown in figure 1, along with the distribution of the corresponding surnames in the 1998 Address Book of the Lancaster County Amish. Seven founders accounted for 83% of these males, 14 for 95% and 21 for 99%.

Representativeness of Our Population Sample

To assess the representativeness of the general Old Order Amish population by our sample, we compared the number of individuals genotyped from each male lineage with the number of families with each corresponding surname as indicated in the 1998 Address Book of the Lancaster County Amish. Results are presented in figure 1. Our sampling of lineages appeared virtually com-

plete; the 27 surnames found in our collection of 739 Amish males accounted for 98% of all Lancaster County Old Order Amish households in the 1998 directory. The same eight surnames accounted for the majority of individuals in our sample (85%) and the majority of households (80%).

Y Chromosome Haplotypes

After exclusion of pedigree/genotyping errors (see below), putative founder Y haplotypes for the 28 lineages were inferred and assigned as described in *Methods* and are listed in table 1, rank ordered by the number of genotyped males. Some lineages had more than one allele at some markers. To assign a putative founder haplotype, for each marker we selected the configuration of the untyped individuals which minimized the number of mutation events and assigned the founder the allele designated in that configuration. The set of founder alleles designated in this manner was then assigned as the putative founder haplotype. In this manner the designation was made unambiguously for all markers in all lineages except in the cases of lineages #13 and 14 for one marker, DYS391. In these two lineages, there were three (equally likely) configurations minimizing the number of mutation events for DYS391; in each of those cases, the founder allele associated with the greatest number of these configurations (2 of 3 in both cases) was assigned as such. In these two lineages, the chosen allele was only marginally more likely than the alternative one; however, the uniqueness of the haplotype was independent of which of the two alleles was chosen. It should be noted that in lineage

#13, sequencing of a previously unstudied individual eliminated one possible configuration, leading to an equal likelihood of two different founder alleles. However, again, both possible founder haplotypes were unique among the Amish.

For 23 of the 28 lineages it was possible based on available data to assign a putative full 9-STR founder haplotype. Of the 23 lineages with data from all nine markers available, only two (#6 and #10) shared the same haplotype, although they have very distinct surnames and dates of entry into the population. Interestingly, one of these lineages entered the population in the mid-1800s, a rare event for the Old Order Amish. Of the five 'incomplete' putative founder haplotypes (#22 and #24–27), three (#24–26) could be distinguished from all others on the basis of available markers. One of these three (#25), whose haplotype was later completed by sequencing missing markers, had the same surname as #12, from which it differed at four loci, suggesting separate origins of that surname. Partial haplotype #27 matched #6 and #10, and partial haplotype #22 matched #1.

Sequencing three additional markers (DYS449, DYS456 and DYS458) in selected individuals enabled us to distinguish between haplotypes #6 and #10 and between #1 and #22. Interestingly, though, haplotypes #6 and #10 only differed at one of the three additional markers (for a total of one marker out of 12 genotyped), and only by one repeat unit, suggesting that the founders shared a relatively recent common male ancestor. Additional DNA was not available from the individual in lineage #27, so his partial haplotype remained indistinguishable from haplotypes #6 and #10. Six of the seven alleles available for the #27 individual are the most common alleles (and the seventh is the second most common) for their respective markers according to a large online Y chromosome genealogy database which allows unrestricted submission (Ybase). Also, in YHRD, the comparable 6-marker haplotype is present on three (including two which match haplotypes #6 and #10) of the 20 most common 8 marker haplotypes in European Americans. Thus it is not surprising that at such a low resolution we were unable to distinguish lineage #27 from the others by haplotype. In any case, we were able to establish that at least 27 of the 28 male founders of our study population had distinct Y chromosome haplotypes.

As noted previously, in addition to the 28 lineages comprising individuals included in the genome-wide scans, there were two additional male lineages. Lineage #29 was not pursued because the family was not practicing Amish and the males in the family were not of Amish

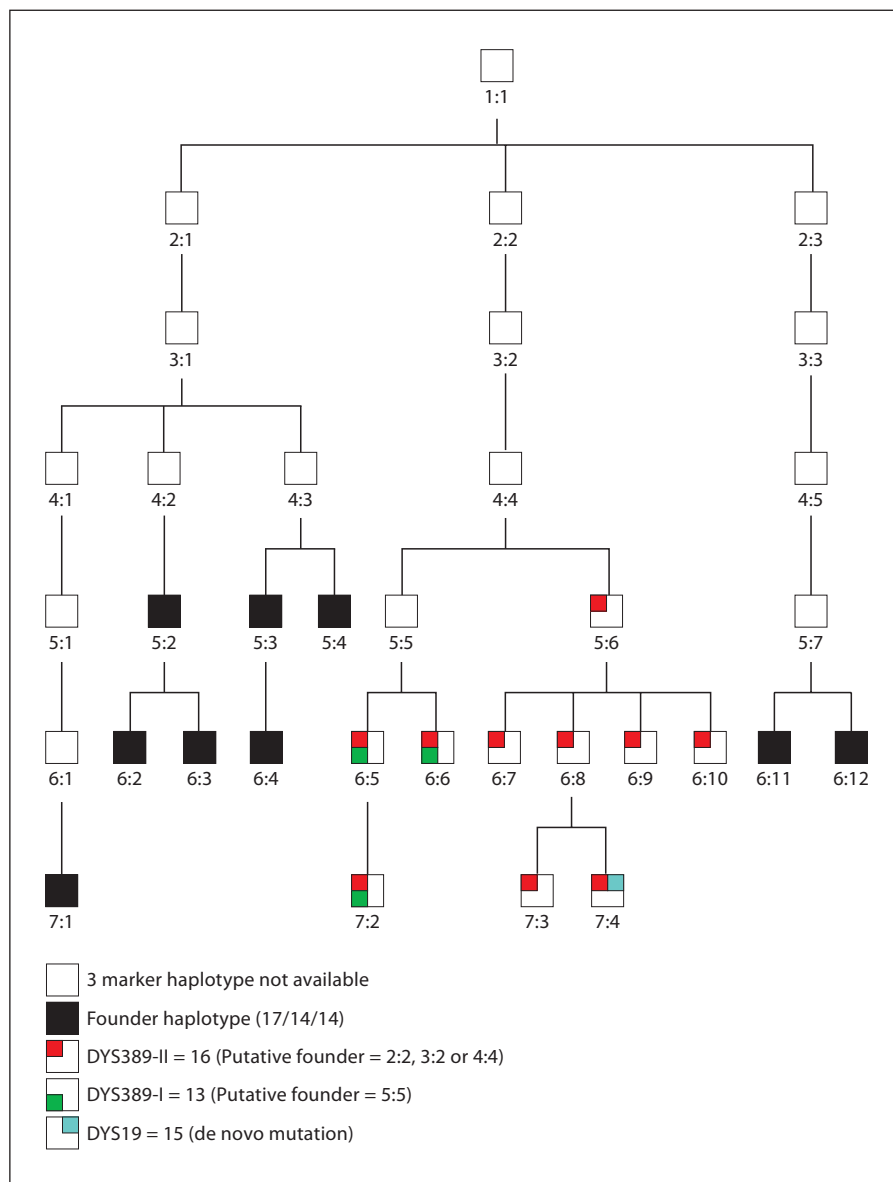
descent. Lineage #30 comprised one individual with the same unusual surname as lineage #20 who could not be connected to this lineage via the Anabaptist Genealogy Database (AGDB). However, successful sequencing of three previously typed STR markers (DYS19, DYS390 and DYS392) in the limited DNA available on this subject revealed that he matched the family on all three alleles. Only one other family, #24, which had a different surname, possessed this same three marker haplotype, which was rare in the both the European (0.33% of 25,904 haplotypes) and worldwide (0.43% of 51,253 haplotypes) samples as well as in the Pennsylvania European-American (0.00% of 67 haplotypes) and national European-American (0.84% of 359 haplotypes) samples in the YHRD (Release 21) [79]. It is therefore highly likely that the additional individual was either (1) descended from the founder of lineage #20 via a path omitted from the genealogy or (2) shared with founder #20 a recent common male ancestor who lived prior to the immigration.

Relationship of the Old Order Amish to Other Populations

The Lancaster Old Order Amish reportedly immigrated from Western Europe, specifically from South Germany and Alsace, where they had originally fled from Bern, Switzerland, to escape religious persecution [4, 5]. We used YHRD data to attempt to corroborate this historical account. For the seven loci genotyped in both our data and the YHRD, our Amish sample contains 27 complete haplotypes, with 25 of these haplotypes being unique. Twenty-two of these Amish founders' haplotypes (20 unique haplotypes) are found among the 12,727 individuals in the pan-European YHRD subset of 91 populations used by Roewer et al. [61] in an analysis of recent European historical events. Nineteen founders' haplotypes (17 unique haplotypes) are found in a subset of three German (Freiburg, Münster and Leipzig) populations and the Bern, Switzerland population. Fifteen of the 19 (13 of 17 unique) haplotypes are over-represented in this subset of four populations, that is, found at a greater frequency in this subset of 1,293 German/Swiss individuals than in the entire set of 12,727, with frequency ratios ranging from 1.11 to 9.84. To include the remaining three unique haplotypes, it is necessary to add one Polish (Bydgoszcz) and one Austrian (Tyrol) population.

If the set of Amish founder haplotypes is expanded to include haplotypes differing by one repeat at one locus ('one step neighbors') to allow for frequent mutation, then all Amish haplotypes except one can be located in the Roewer subset. Using this same expanded definition,

Fig. 2. Portion of a pedigree providing a rational explanation for the existence of two non-founder alleles in a single individual. Individuals shown in black have been genotyped and possess the putative founder allele for all three of DYS389-II, DYS389-I and DYS19. A cluster of 10 individuals with DYS389-II = 16 (vs. 17 in founder haplotype) is consistent with a 17→16 mutation in individual 2:2, 3:2 or 4:4. A cluster of 3 individuals with a DYS389-I = 13 allele is consistent with an historic DYS389-I 14→13 mutation in individual 5:5, a descendant of the putative DYS389-II 17→16 founder. Another such descendant, 7:4, has a de novo DYS19 14→15 mutation.



the set of four German/Swiss populations described above includes 25 Amish founders' haplotypes (23 unique haplotypes), and including the Polish and Austrian populations brings the total founders' haplotypes found to 26 of 27 (24 of 25 unique haplotypes), the most that can be found in the Roewer data. Thus even this expanded haplotype definition fails to locate one of the 27 founder haplotypes in one of the European populations analyzed by Roewer. Interestingly, the founder of this excluded lineage (also not found in the entire YHRD database, with one one step neighbor found in a non-European population), was reported to be kidnapped and brought to the

United States as a young boy to work as an indentured servant; thus his family origins are obscure.

The single marker genotypes among the founders exhibited similar frequencies to those in both the YHRD and YBASE online databases (data not shown).

Distinguishing Mutations and Pedigree Errors

In tabulating differences between individual marker genotypes and putative founder alleles to assess marker mutation rates, we first needed to rule out cases of pedigree and/or sample errors. We reasoned that multiple alleles discrepant from respective founder alleles in a single

Table 2. Deviations from Familial Single Marker Genotypes observed in genome scan data and confirmed by sequencing or clustering in 3 or more individuals

Marker	# genotyped	# with non-founder allele	# apparent mutation events	Gains	Losses
DYS393/395	728	0	0	0	0
DYS19	470	4	3	2	1
DYS391 ^a	671	7	5	3 (2)	2 (1)
DYS389-I	698	9	3	1	2
DYS389-II ^b	699	24	12	2	10
DYS388	713	0	0	0	0
DYS390	703	23	4	1	3
DYS392	701	1	1	1	0
GGAAT1B07	411	0	0	0	0
Total ^a	5,794	68	28	10 (9)	18 (17)

^a Numbers in parentheses indicate number of gains or losses if the two cases of ambiguous DYS391 founder alleles (lineages #13 and #14, see text and fig. 2) are excluded.

^b Excludes those resulting from DYS389-I mutations.

individual, while possibly resulting from multiple mutations, would have the greatest likelihood of resulting from errors. Three samples with multiple mismatches (two individuals each with a one-step mismatch at one locus and a two-step mismatch at another locus, and one individual with a one-step mismatch at each of two loci and a two-step mismatch at a third locus) were confirmed to be pedigree errors based on Mendelian inconsistent autosomal data. A fourth individual had only one one-step mismatch with his putative founder haplotype, but was investigated because he was successfully genotyped for only four of the nine markers. He was subsequently found to be discrepant with his founder at two of the three additional markers sequenced (DYS449 and DYS458, mismatched by one step each). He was Mendelian consistent at all autosomal loci with his two sisters and his mother, but a comparison of his and his two sisters' chromosome 1 haplotypes (inferred using a Markov Chain Monte Carlo algorithm as implemented in Simwalk 2 [81]) with other lineage members revealed that he did not appear to share any haplotype segment with these distant cousins, consistent with a paternity error in a recent ancestor. Notably, this individual and his immediate family were not registered members of the Amish Church.

There were four additional individuals who each had two mismatches with their founder haplotype; however, these cases were consistent with true mutations occurring in multiple generations. In lineage #7 (see table 1), 10

individuals possessed a non-founder allele in DYS389-II consistent with an historic 17→16 mutation in the DYS389-II-specific segment (fig. 2). A cluster of three of these 10 individuals revealed an apparent DYS389-I 14→13 mutation in a descendant (5:5) of the putative DYS389-II 17→16 founder, with the end result being that these three individuals (6:5, 6:6 and 7:2 in the figure) each had two non-founder alleles. In addition, in this same lineage (apparently by coincidence), one individual with the 17 allele (7:4 in fig. 2) had a non-founder allele at DYS19, consistent with a new 14→15 mutation. Of course, even finding evidence of two mutations occurring in the same individual would also not have been surprising statistically, as observed in a study of 9 STRs in 415 father-son pairs [65].

Confirmation of Apparent Mutations

Those non-founder alleles found in three or more family members in a pattern consistent with a single historical mutation event were considered to be confirmed evidence of a mutation event not requiring further molecular investigation. To increase confidence that apparent mutations manifesting in only one or two family members were real and not results of genotype errors, we sequenced relevant markers in putative mutation carriers and discordant 'normal' fathers and/or brothers (or the closest relative(s) available) along with previously ungenotyped sons and/or other relatives expected to share the mutation to confirm those mutations. Of the 31 initially observed or inferred mutations, five were confirmed by their presence in clusters of three to 16 individuals, 23 were confirmed by sequencing of individuals possessing the mutation and/or close relatives, and three were refuted by sequencing. In lineage #5, sequencing an additional individual to confirm and localize a DYS389-II 17→16 mutation in addition revealed a new or recently inherited DYS389-I 12→13 mutation.

Mutation Rate Analysis

Prior to estimating mutation rates, we excluded the 4 individuals representing pedigree errors, the three mutations refuted by sequencing, and the newly identified DYS389-I mutation in lineage #5.

All markers except DYS393/395 and GGAAT1B07 had at least one apparent mutation event. A summary of these mutation events is shown in table 2. Of 5,794 genotypes, 68 differed from the expected family genotype. Pedigree analysis revealed that several of these differences could be attributed to a total 28 putative historical and de novo mutation events.

Table 3. Apparent Y STR mutation events

Marker	Observed meiotic mutation events (typed father-son pairs)				All meiotic/mutation events (entire pedigrees)			
	meioses	mutations	mutation rate, %	95% exact CI, %	meioses	mutations	mutation rate, %	95% exact CI, %
DYS393/395	283	0	0.00	0.00–1.30	1,232	0	0.00 ^b	0.00–0.30
DYS19	155	1	0.65	0.02–3.54	906	3	0.33	0.07–0.96
DYS391	230	0	0.00	0.00–1.59	1,191	5	0.42	0.14–0.98
DYS389-I	274	0	0.00	0.00–0.13	1,189	3	0.25	0.05–0.74
DYS389-II ^a	274	3	1.09	0.40–3.70	1,178	12	1.02 ^c	0.53–1.77
DYS388	271	0	0.00	0.00–1.35	1,217	0	0.00 ^b	0.00–0.33
DYS390	263	2	0.76	0.09–2.72	1,215	4	0.33	0.09–0.84
DYS392	257	1	0.39	0.01–2.15	1,209	1	0.08	0.00–0.46
GGAAT1B07	119	0	0.00	0.00–3.05	833	0	0.00	0.00–0.44
Total	2,126	7	0.33	0.13–0.68	10,170	28	0.28	0.18–0.40
Tetranucleotide only	1,478	6	0.41	0.15–0.88	6,911	27	0.39	0.26–0.57

^a Excludes those resulting from DYS389-I mutations.

^b Fisher's exact $p = 0.067$ vs. all markers and $p = 0.026$ vs. tetranucleotide markers.

^c $p < 0.0001$ vs. all markers and $p = 0.0096$ vs. tetranucleotide markers.

We used two methods to estimate mutation rates in our Y chromosome markers. We first restricted our analysis to typed father-son pairs (table 3) to generate results that could be compared with a previous study by Kayser et al. [65] and other studies of father-son pairs [26, 28, 29, 67, 72]. For the nine markers, the number of such pairs available ranged from 119 to 283. Mutation rates, calculated as the proportion of discordant father-son pairs over the total number typed for each marker, ranged from 0 to 1.09%, with an overall mutation rate of 0.33% (0.41% for tetranucleotide repeats only). These rates are similar to those calculated previously [26, 28, 29, 65, 67–72].

The second method traces haplotypes and mutation events back to putative founders and considers all meioses, observed or inferred, as a denominator. This method was used previously [63] with a much smaller sample size (42 individuals from 12 pedigrees) than ours (739 individuals from 28 pedigrees). Mutation rates (table 3) were similar to those we calculated using father-son pairs: an overall mutation rate of 0.28% (0.39% for tetranucleotides only). This method also resulted in a sample size sufficient to detect significant departures from the overall and/or tetranucleotide marker mutation rate in three markers, DYS393 (mutation rate = 0%, Fisher's exact p value = 0.067 versus all markers and $p = 0.026$ versus tetranucleotide markers), DYS388 (same mutation rate and p values as DYS393) and DYS389-II (mutation rate = 1.02%, Fisher's exact p value < 0.0001 versus all markers

and $p = 0.0096$ versus all tetranucleotide markers). By Fisher's exact test, no significant differences between mutation rates calculated by the two methods were observed for any of the nine markers.

In two families, the marker with the highest mutation rate, DYS389-II, showed evidence of multiple independent mutation events. Lineage #2 showed evidence of five independent occurrences of the same mutation in DYS389-II (18→17). Furthermore, in this same lineage, there were two additional alleles for DYS389-II, both resulting from de novo mutations as evidenced by fathers possessing the founder allele: 18→20 (the only two step mutation observed) and 18→19. There was also evidence of three independent occurrences of a 17→16 mutation in this same marker in lineage #5.

Discussion

By examining the genotypes at several STR markers on the Y chromosome in several hundred Amish study volunteers, we have confirmed the historical accuracy of the genealogical records of the ancestors that connect individuals in our current pedigrees recruited for the study of complex phenotypes. The combination of genealogical records and Y chromosome genotypes indicates that virtually every surname in the Amish represents a unique founder. Comparison of putative Amish founder Y chro-

Table 4. Comparison of single locus mutation rates observed in this study to previously published and online mutation rates

Study	Method*	DYS19	DYS389-I	DYS389-II	DYS390	DYS391	DYS392	DYS393/395
Kayser et al. (2000) [65]	pairs	0.20 (2/996)	0.24 (1/425)	0.47 (2/425)	0.86 (4/466)	0.48 (2/415)	0.00 (0/415)	0.00 (0/415)
Bianchi et al. (1998) [66]	pairs	0.00 (0/249)	0.00 (0/249)	0.00 (0/249)	0.00 (0/249)	0.00 (0/249)	0.00 (0/249)	0.00 (0/249)
Heyer et al. (1997) [63]	pedigrees	0.00 (0/213)	–	–	0.00 (0/213)	0.00 (0/213)	0.47 (1/213)	0.00 (0/213)
Kurihara et al. (2004) [29]	pairs	0.00 (0/161)	0.62 (1/161)	0.62 (1/161)	0.00 (0/161)	0.62 (1/161)	0.00 (0/161)	0.00 (0/161)
Dupuy et al. (2004) [28]	pairs	0.17 (3/1,766)	0.23 (4/1,766)	0.23 (4/1,766)	0.45 (8/1,766)	0.45 (8/1,766)	0.00 (0/1,766)	0.06 (1/1,766)
Ballard et al. (2005) [67]	pairs	0.41 (1/245)	0.41 (1/247)	0.81 (2/246)	0.00 (0/248)	0.81 (2/248)	0.00 (0/226)	0.00 (0/248)
Budowle et al. (2005) [71]	pairs	0.29 (2/692)	0.14 (1/692)	0.14 (1/692)	0.00 (0/692)	0.14 (1/692)	0.00 (0/692)	0.14 (1/692)
Gusmão et al. (2005) [72]	pairs	0.14 (4/2,807)	0.11 (2/1,793)	0.11 (2/1,781)	0.11 (3/2,816)	0.32 (9/2,815)	0.11 (3/2,803)	0.13 (2/1,569)
Hohoff et al. (2006) [69]	pairs	0.58 (6/1,027)	0.10 (1/1,027)	0.49 (5/1,027)	0.20 (2/1,027)	0.20 (2/1,028)	0.00 (0/1,026)	0.10 (1/1,027)
Lee et al. (2007) [70]	pairs	0.54 (2/369)	0.27 (1/369)	0.54 (2/369)	0.27 (1/369)	0.00 (0/369)	0.00 (0/369)	0.27 (1/369)
Domingues et al. (2007) [68]	pairs	0.74 (1/135)	0.00 (0/135)	0.00 (0/135)	0.00 (0/135)	0.00 (0/135)	0.00 (0/135)	0.00 (0/135)
YHRD pooled (includes all of the above plus 2 un-published studies)	mixed	0.25 (22/8,944)	0.18 (13/7,148)	0.27 (19/7,135)	0.24 (20/8,426)	0.30 (25/8,375)	0.05 (4/8,339)	0.08 (6/7,128)
Bonné-Tamir et al. (2003) [73]	pedigrees	1.45 (2/138)	1.44 (2/139)	0.00 (0/139)	0.00 (0/138)	0.72 (1/138)	0.00 (0/139)	0.00 (0/139)
Present study	pairs	0.65 (1/155)	0.00 (0/274)	1.09 (3/274)	0.76 (2/263)	0.00 (0/230)	0.39 (1/257)	0.00 (0/283)
Present study	pedigrees	0.33 (3/906)	0.25 (3/1,189)	1.02 (12/1,178)	0.33 (4/1,215)	0.42 (5/1,191)	0.08 (1/1,209)	0.00 (0/1,232)

Rates given as percents with number of mutations over number of meioses shown in parentheses.

* Methods: 'Pairs' refers to studies using only typed, confirmed father/son pairs. 'Pedigrees' refers to studies using pedigrees that included untyped but inferred transmissions in the calculations.

mosome haplotypes with online European Y chromosome haplotype data support the reported Western/Central European origin of the Amish.

Our Amish data afforded us the opportunity to use two different but complementary approaches for estimating Y STR mutation rates. A number of studies have evaluated mutation rates in Y STR markers, which are impor-

tant for forensic applications [82]. Observation of frequent mutations in YSTRs is also a reminder of the general mutability of STRs, which in autosomes can lead to Mendelian discrepancies that may be mistaken for genotype errors when using the markers for linkage analysis. Heyer et al. [63] and Jobling et al. [64] used information from 42 individuals in 12 'deep rooting' pedigrees

and were able to use as a denominator 213 to 248 primarily unobserved transmissions [63]. The disadvantage of this approach, as pointed out by others [65, 83], is that paternity cannot be completely resolved. In fact, Heyer et al. used three different scenarios to estimate Y chromosome mutation frequencies because they could not distinguish multiple apparent mutations in one individual from nonpaternity. The minisatellite MSY1 genotyping applied by Jobling et al. [64] to the same pedigrees provided evidence but did not prove definitively that the single marker differences were true mutations and the multiple marker differences represented instances of nonpaternity. To remove nonpaternity concerns from mutation rate estimation, Kayser et al. [65] and later others [26, 28, 29, 67, 72] studied father-son pairs in conjunction with autosomal genotyping. In our Amish data, we were able to use both approaches, which yielded similar results and complemented each other well. The father-son pairs, which included 2,126 meiotic events, led to a 0.33% estimate of the mutation rate, very similar to that calculated by Kayser et al. Since these individuals were genotyped for approximately 400 (in the case of the AFDS) or 800 (in the case of the AFOS) autosomal and X-linked STRs, we were able to rule out nonpaternity in these pairs to an even greater certainty than Kayser et al., who used only 11–13 autosomal markers to confirm paternity.

The estimated overall mutation rate in the Amish was similar using both methods (0.33% with father-son pairs and 0.28% with the whole pedigrees). A pedigree based study in the Samaritan population similarly estimated a mutation rate of 0.42% [73], also consistent with our findings. Since the landmark study by Kayser et al. [65], similar mutation rates have been estimated using father-son pairs in several other populations, including 0.22% in 161 Japanese pairs typed for 14 YSTRs [29], 0.20% in 3,026 Spanish and Portuguese pairs typed for 17 YSTRs [72], 0.46% in up to 249 mixed UK pairs typed for 13 YSTRs [67], 0.31% in 109 Taiwanese pairs typed for nine YSTRs [26], 0.16% in 692 North American pairs typed for 12 YSTRs [71], 0.21% in 1,029 German pairs typed for 15 YSTRs [69], 0.39% in 369 Korean pairs typed for 22 YSTRs [70] and 0.18% in 135 'Afro-Brazilian' pairs typed for 12 YSTRs [68]. Similarly, an estimated mutation rate of 0.18–21% for repeat gains (losses could not be evaluated due to the methodology used) was calculated using 2 STRs in sperm samples from three donors; overall mutation rate was estimated at 0.4% based on the assumption of equilibrium between gains and losses [74]. These latter results should be interpreted with caution due to the technical limitations of the small-pool PCR and fluo-

rescence-based fragment-length analysis methods used, including the inability to detect repeat losses. Previously published mutation rates for the seven commonly typed loci are shown in table 4 (study using sperm sample excluded because of its technical limitations) in conjunction with the two sets of mutation rates estimated in the present study.

In addition, 'evolutionary' mutation rate estimates of 0.026% per 20 years [75], 0.069% per 25 years [32] and 0.027% per generation [43] have been reported. The discrepancy between these estimates and those based on father-son pairs and pedigree analysis appears to be attributable to several factors, including assumptions about the age of the population [32, 75, 84], the specific characteristics of the markers and alleles evaluated [85], and possibly haplogroup-based selection effects [86]. The consistency of the mutation rates estimated using the pedigree and father-son pair methods in the present study and previous studies suggests that for the purpose of forensic and genetic epidemiology quality control applications, the mutation rate for YSTRs is between 0 and 1% of meioses per marker, varying by specific marker.

One marker, DYS389-II, showed a significant departure from the overall mutation rate when evaluated using the whole pedigree method. This marker along with two others (tetranucleotide DYS393/395 and trinucleotide DYS388) showed significant departure from the overall tetranucleotide marker mutation rate. The marker with the highest mutation rate, DYS389-II, showed a virtually identical rate between the two methods (1.09% in father-son pairs and 1.02% in the whole pedigrees). This marker also had the greatest number of alleles in our study, exemplifying the advantages and disadvantages of STRs: (autosomal) markers with high diversity are useful for linkage analysis but increase the possibility of mutations to contribute to the overall 'error' rate. Not surprisingly marker DYS393/395, which has no observed mutations in either our data or the data of Kayser et al. and Heyer et al., and few mutations in other studies, has very low diversity, with a single allele accounting for over 70% of Amish founder alleles as well as alleles in YHRD and YBase.

Summary

In summary, our genotype analysis of Y chromosome STR markers in our Amish study subjects has (1) confirmed the accuracy of the male lineage portion of the genealogy and completeness of the Anabaptist Genealogy

Database; (2) showed that Lancaster Amish founder Y chromosomes exhibit diversity similar to the general Caucasian population, reinforcing that the surnames delineate fairly distinct founders, and (3) added to existing data on mutation rate estimates for several commonly used Y chromosome STR markers.

Acknowledgements

The authors wish to acknowledge Drs. Elizabeth Streeten and Jay Shapiro, the Amish Research Clinic staff and Amish Liaisons for their energetic recruitment of subjects into the Amish Family Studies, Jack Shelton for STR marker sequencing and the NHLBI Mammalian Genotyping Service for STR genotyping. This work would not have been possible without the outstanding

cooperation of the Amish community. This research was supported in part by the intramural research program of the NIH, National Library of Medicine the American Diabetes Association, and NIH grants R01-DK64621, R01-AR4638 and R01-HL69313.

Web Resources

The URLs for data presented herein are as follows:
 NHLBI Mammalian Genotyping Service, <http://www.marshfieldclinic.org/research/genetics>
 PedHunter, <http://www.ncbi.nlm.nih.gov/CBBresearch/Schaffler/pedhunter.html>
 Ybase: genealogy by numbers, <http://www.ybase.org>
 The Y-STR Haplotype Reference Database (YHRD), <http://www.yhrd.org>

References

- Agarwala R, Biesecker LG, Hopkins KA, Francomano CA, Schäffer AA: Software for constructing and verifying pedigrees within large genealogies and an application to the Old Order Amish of Lancaster County. *Genome Res* 1998;8:211–221.
- Cross HE: Population studies and the Old Order Amish. *Nature* 1976;262:17–20.
- Beiler K: *Fisher Family History*. Eby's Quality Publishing, 1988.
- Gingerich HF, Kreider RW: *Amish and Amish Mennonite Genealogies*. 2nd Printing ed. Gordonville, PA: Pequea Publishers, 2002.
- McKusick VA, Hostetler JA, Egeland JA: Genetic studies of the Amish, background and potentialities. *Bull Johns Hopkins Hosp* 1964;115:203–222.
- Francomano CA, McKusick VA, Biesecker LG: Medical genetic studies in the Amish: historical perspective. *Am J Med Genet* 2003;121C:1–4.
- Rimoin DL: Ethnic variability in glucose tolerance and insulin secretion. *Arch Intern Med* 1969;124:695–700.
- Hsueh W-C, Mitchell BD, Schneider JL, St Jean PL, Pollin TI, Ehm MG, Wagner MJ, Burns DK, Sakul H, Bell CJ, Shuldiner AR: Genome-wide scan of obesity in the Old Order Amish. *J Clin Endocrinol Metab* 2001;86:1199–1205.
- Pollin TI, Tanner K, O'Connell JR, Ott SH, Damcott CM, Shuldiner AR, McLenithan JC, Mitchell BD: Linkage of plasma adiponectin levels to 3q27 explained by association with variation in the APM1 gene. *Diabetes* 2005;54:268–274.
- Sorkin J, Post W, Pollin TI, O'Connell JR, Mitchell BD, Shuldiner AR: Exploring the genetics of longevity in the Old Order Amish. *Mech Ageing Dev* 2005;126:347–350.
- Steinle NI, Hsueh WC, Snitker S, Pollin TI, Sakul H, St Jean PL, Bell CJ, Mitchell BD, Shuldiner AR: Eating behavior in the Old Order Amish: heritability analysis and a genome-wide linkage analysis. *Am J Clin Nutr* 2002;75:1098–1106.
- Hsueh W-C, Mitchell BD, Aburomia R, Pollin T, Sakul H, Gelder EM, Michelsen BK, Wagner MJ, St Jean PL, Knowler WC, Burns DK, Bell CJ, Shuldiner AR: Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* 2000;23:595–601.
- Hsueh W-C, Mitchell BD, Schneider JL, Wagner MJ, Bell CJ, Nanthakumar E, Shuldiner AR: QTL influencing blood pressure maps to the region of PPH1 on chromosome 2q31-34 in Old Order Amish. *Circulation* 2000;101:2810–2816.
- Hsueh W-C, St Jean PL, Mitchell BD, Pollin TI, Knowler WC, Ehm MG, Bell CJ, Sakul H, Wagner MJ, Burns DK, Shuldiner AR: Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: Evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes* 2003;52:550–557.
- Pollin TI, Hsueh W-C, Steinle NI, Snitker S, Shuldiner AR, Mitchell BD: A genome-wide scan of serum lipid levels in the Old Order Amish. *Atherosclerosis* 2004;173:89–96.
- Mitchell BD, Hsueh W-C, King TM, Pollin TI, Sorkin J, Agarwala R, Schäffer AA, Shuldiner AR: Heritability of life span in the Old Order Amish. *Am J Med Genet* 2001;102:346–352.
- Ginns EI, St Jean P, Philibert RA, Galdzicka M, Damschroder-Williams P, Thiel B, Long RT, Ingraham LJ, Dalwaldi H, Murray MA, Ehler M, Paul S, Remortel BG, Patel AP, Anderson MC, Shaio C, Lau E, Dymarskaia I, Martin BM, Stubblefield B, Falls KM, Carulli JP, Keith TP, Fann CS, Lacy LG, Allen CR, Hostetter AM, Elston RC, Schork NJ, Egeland JA, Paul SM: A genome-wide search for chromosomal loci linked to mental health wellness in relatives at high risk for bipolar affective disorder among the Old Order Amish. *Proc Natl Acad Sci USA* 1998;95:15531–15536.
- Platte P, Papanicolaou GJ, Johnston J, Klein CM, Doheny KF, Pugh EW, Roy-Gagnon M-H, Stunkard AJ, Francomano CA, Wilson AF: A study of linkage and association of body mass index in the Old Order Amish. *Am J Med Genet* 2003;121C:71–80.
- Agarwala R, Biesecker LG, Tomlin JF, Schäffer AA: Towards a complete North American Anabaptist genealogy: A systematic approach to merging partially overlapping genealogy resources. *Am J Med Genet* 1999;86:156–161.
- Agarwala R, Biesecker LG, Schäffer AA: Anabaptist genealogy database. *Am J Med Genet* 2003;121C:32–37.
- Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF: A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int* 2002;129:10–24.
- Redd AJ, Agellon AB, Kearney VA, Contreas VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF: Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int* 2002;130:97–111.

- 23 Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Qasim Mehdi S, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C: A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet* 2004;74:1183–1197.
- 24 Jobling MA, Tyler-Smith C: The human Y chromosome: An evolutionary marker comes of age. *Nat Rev Genet* 2003;4:598–612.
- 25 Immel U-D, Krawczak M, Udolph J, Richter A, Rodig H, Kleiber M, Klitschar M: Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population. *Eur J Hum Genet* 2006;14:577–582.
- 26 Tsai L-C, Yuen T-Y, Hsieh H-M, Lin M, Tzeng C-H, Huang N-E, Linacre A, Lee JCI: Haplotype frequencies of nine Y-chromosome STR loci in the Taiwanese Han population. *Int J Legal Med* 2002;116:179–183.
- 27 Pestoni C, Cal ML, Lareu MV, Rodríguez-Calvo MS, Carracedo A: Y chromosome STR haplotypes: Genetic and sequencing data of the Galician population (NW Spain). *Int J Legal Med* 1999;112:15–21.
- 28 Dupuy BM, Stenersen M, Egeland T, Olaisen B: Y-chromosomal microsatellite mutation rates: Differences in mutation rate between and within loci. *Hum Mutat* 2004;23:117–124.
- 29 Kurihara R, Yamamoto T, Uchihi R, Li S-L, Yoshimoto T, Ohtaki H, Kamiyama K, Katsumata Y: Mutations in 14 Y-STR loci among Japanese father-son haplotypes. *Int J Legal Med* 2004;118:125–131.
- 30 Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrila L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarinov V, Mammi C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofaneli S, Malaspina P, Novelletto A: Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* 2004;115:357–371.
- 31 Chaix R, Austerlitz F, Morar B, Kalaydjieva L, Heyer E: Vlach Roma history: what do coalescent-based methods tell us? *Eur J Hum Genet* 2004;12:285–292.
- 32 Zhitovskiy LA, Underhill PA, Cinnioglu, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisoli G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004;74:50–61.
- 33 Berger B, Niederstatter H, Brandstätter A, Parson W: Molecular characterization and Austrian Caucasian population data of the multi-copy Y-chromosomal STR *DYS464*. *Forensic Sci Int* 2003;137:221–230.
- 34 Zei G, Lisa A, Fiorani O, Magri C, Quintana-Murci L, Semino O, Santachiara-Benerecetti AS: From surnames to the history of Y chromosomes: The Sardinian population as a paradigm. *Eur J Hum Genet* 2003;11:802–807.
- 35 Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinkas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L: Origins and divergence of the Roma (gypsies). *Am J Hum Genet* 2001;69:1314–1331.
- 36 Walsh B: Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics* 2001;158:897–912.
- 37 Dupuy BM, Andreassen R, Flønes AG, Tomassen K, Egeland T, Brion M, Carracedo A, Olaisen B: Y-chromosome variation in a Norwegian population sample. *Forensic Sci Int* 2001;117:163–173.
- 38 Nebel A, Filon D, Weiss DA, Weale M, Faerman M, Oppenheim A, Thomas MG: High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. *Hum Genet* 2000;107:630–641.
- 39 Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA: The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 2000;66:979–988.
- 40 Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW: Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol Biol Evol* 1999;16:1791–1798.
- 41 Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC: Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. *Am J Phys Anthropol* 1999;108:381–399.
- 42 Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC: Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 1998;62:1171–1179.
- 43 Caglià A, Novelletto A, Dobosz M, Malaspina P, Ciminelli BM, Pascali VL: Y-chromosome STR loci in Sardinia and continental Italy reveal islander-specific haplotypes. *Eur J Hum Genet* 1997;5:288–292.
- 44 el Khil HK, Marrakchi RT, Loueslati BY, Langaney A, Fellous M, Elgaaied BA: Distribution of Y chromosome lineages in Jerba island population. *Forensic Sci Int* 2005;148:211–218.
- 45 Behar DM, Garrigan D, Kaplan ME, Mombashar Z, Rosengarten D, Karafet TM, Quintana-Murci L, Ostrer H, Skorecki K, Hammer MF: Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum Genet* 2004;114:354–365.
- 46 McEvoy B, Bradley DG: Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum Genet* 2006;119:212–219.
- 47 Jobling MA: In the name of the father: surnames and genetics. *Trends Genet* 2001;17:353–357.
- 48 Sykes B, Irven C: Surnames and the Y chromosome. *Am J Hum Genet* 2000;66:1417–1419.
- 49 Boster JS, Hudson RR, Gaulin SJC: High paternity certainties of Jewish priests. *Am Anthropologist* 1998;100:967–971.
- 50 Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB: Origins of Old Testament priests. *Nature* 1998;394:138–140.
- 51 Trumme T, Herrmann B, Hummel S: Genetics in genealogical research-reconstruction of a family tree by means of Y-haplotyping. *Anthropol Anz* 2004;62:379–386.
- 52 King TE, Ballereau SJ, Schürer KE, Jobling MA: Genetic signatures of coancestry within surnames. *Curr Biol* 2006;16:384–388.
- 53 Betz A, Bässler G, Diel G, Steil X, Weyermann G, Pflug W: *DYS* STR analysis with epithelial cells in a rape case. *Forensic Sci Int* 2001;118:126–130.
- 54 Sibille I, Duverneuil C, Lorin dlG, Guerrouache K, Teissière F, Durigon M, de Mazancourt P: Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa. *Forensic Sci Int* 2002;125:212–216.
- 55 Parson W, Niederstätter H, Köchl S, Steinlechner M, Berger B: When autosomal short tandem repeats fail: optimized primer and reaction design for Y-chromosome short tandem repeat analysis in forensic casework. *Croat Med J* 2001;42:285–287.
- 56 Sinha SK, Budowle B, Chakraborty R, Pounovic A, Guidry RD, Larsen C, Lal A, Shaffer M, Pineda G, Sinha SK, Schneida E, Nasir H, Shewale JG: Utility of the Y-STR typing systems Y-PLEX 6 and Y-PLEX 5 in forensic casework and 11 Y-STR haplotype database for three major population groups in the United States. *J Forensic Sci* 2004;49:691–700.
- 57 Shewale JG, Nasir H, Schneida E, Gross AM, Budowle B, Sinha SK: Y-chromosome STR system, Y-PLEX 12, for forensic casework: Development and validation. *J Forensic Sci* 2004;49:1278–1290.
- 58 Dettlaff-Kakol A, Pawlowski R: First Polish DNA ‘manhunt’ – an application of Y-chromosome STRs. *Int J Legal Med* 2002;116:289–291.
- 59 Cerri N, Ricci U, Sani I, Verzeletti A, De Ferrari F: Mixed stains from sexual assault cases: autosomal or Y-chromosome short tandem repeats? *Croat Med J* 2003;44:289–292.
- 60 Gusmão L, Sánchez-Diz P, Alves C, Beleza S, Lopes A, Carracedo A, Amorim A: Grouping of Y-STR haplotypes discloses European geographic clines. *Forensic Sci Int* 2003;134:172–179.

- 61 Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M: Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* 2005;116:279–291.
- 62 Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziwska M, Roewer L, Dobosz T, Ploski R: Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet* 2005;117:428–443.
- 63 Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P: Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 1997;6:799–803.
- 64 Jobling MA, Heyer E, Dieltjes P, de Knijff P: Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1. *Hum Mol Genet* 1999;8:2117–2120.
- 65 Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A: Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000;66:1580–1588.
- 66 Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS: Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* 1998;63:1862–1871.
- 67 Ballard DJ, Phillips C, Wright G, Thacker CR, Robson C, Revoir AP, Court DS: A study of mutation rates and the characterisation of intermediate, null and duplicated alleles for 13 Y chromosome STRs. *Forensic Sci Int* 2005;155:65–70.
- 68 Domingues PM, Gusmao L, da Silva DA, Amorim A, Pereira RW, de Carvalho EF: Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci. *Int J Legal Med* 2007.
- 69 Hohoff C, Dewa K, Sibbing U, Hoppe K, Forster P, Brinkmann B: Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *Int J Legal Med* 2006;Oct. 26 [Epub ahead of print].
- 70 Lee HY, Park MJ, Chung U, Lee HY, Yang WI, Cho SH, Shin KJ: Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs. *Int J Legal Med* 2007;121:128–135.
- 71 Budowle B, Adamowicz M, Aranda XG, Barina C, Chakraborty R, Cheswick D, Dafoe B, Eisenberg A, Frappier R, Gross AM, Ladd C, Lee H-S, Milne SC, Meyers C, Prinz M, Richard ML, Saldanha G, Tierney AA, Viculis L, Krenke BE: Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Forensic Sci Int* 2005;150:1–15.
- 72 Gusmao L, Sánchez-Diz P, Calafell F, Martín P, Alonso CA, Álvarez-Fernández F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Lonardo AM, Espinheira R, Fagundes de Carvalho E, Farfán MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza Góes AC, Santapa OA, Sumita DR, Vallejo G, Rioja LV, Vide MC, Vieira da Silva CI, Whittle MR, Zabala W, Zarrabeitia MT, Alonso A, Caracedo A, Amorim A: Mutation rates at Y chromosome specific microsatellites. *Hum Mutat* 2005;26:520–528.
- 73 Bonnè-Tamir B, Korostishevsky M, Redd AJ, Pel-Or Y, Kaplan ME, Hammer MF: Maternal and paternal lineages of the Samaritan isolate: Mutation rates and time to most recent common male ancestor. *Ann Hum Genet* 2003;67:153–164.
- 74 Holtkemper U, Rolf B, Hohoff C, Forster P, Brinkmann B: Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. *Hum Mol Genet* 2001;10:629–633.
- 75 Forster P, Röhl A, Lünemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B: A short tandem repeat-based phylogeny for the human Y chromosome. *Am J Hum Genet* 2000;67:182–196.
- 76 Streeeten EA, McBride DJ, Lodge AL, Pollin TI, Stinchcomb DG, Agarwala R, Schäffer AA, Shapiro JR, Shuldiner AR, Mitchell BD: Reduced incidence of hip fracture in the Old Order Amish. *J Bone Miner Res* 2004;19:308–313.
- 77 Agarwala R, Schäffer AA, Tomlin JF: Towards a complete North American Anabaptist genealogy II: analysis of inbreeding. *Hum Biol* 2001;533–545.
- 78 Kayser M, Brauer S, Willuweit S, Schädlich H, Batzer MA, Zawacki J, Prinz M, Roewer L, Stoneking M: Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. *J Forensic Sci* 2002;47:513–519.
- 79 Willuweit S, Roewer L, on behalf of the International Forensic Y Chromosome User Group: Y chromosome haplotype reference database (YHRD): Update. *Forensic Science International: Genetics* 2007;1:83–87.
- 80 Rolf B, Meyer E, Brinkmann B, de Knijff P: Polymorphism at the tetranucleotide repeat locus *DYS389* in 10 populations reveals strong geographic clustering. *Eur J Hum Genet* 1998;6:583–588.
- 81 Sobel E, Lange K: Descent graphs in pedigree analysis: Applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 1996;58:1323–1337.
- 82 Kayser M, de Knijff P, Dieltjes P, Krawczak M, Nagy M, Zerjal T, Pandya A, Tyler-Smith C, Roewer L: Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 1997;18:1602–1607.
- 83 Kayser M, Sajantila A: Mutations at Y-STR loci: Implications for paternity testing and forensic analysis. *Forensic Sci Int* 2001;118:116–121.
- 84 Zhivotovsky LA, Underhill PA: On the evolutionary mutation rate at Y-chromosome STRs: Comments on paper by Di Giacomo et al. (2004). *Hum Genet* 2005;116:529–532.
- 85 Carvalho-Silva DR, Santos FR, Hutz MH, Salzano FM, Pena SDJ: Divergent human Y-chromosome microsatellite evolution rates. *J Mol Evol* 1999;49:204–214.
- 86 Zhivotovsky LA, Underhill PA, Feldman MW: Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. *Mol Biol Evol* 2006;23:2268–2270.