3-4-2014

Pedigree Structure and Kinship Measurements of a Mid-Michigan Community: A New North American Population Isolate Identified

Joseph D. Bonner
Michigan State University, joedonbonner@gmail.com

Rachel Fisher
Michigan State University

James Klein
Michigan State University

Qing Lu
Michigan State University

Ellen Wilch
Michigan State University

See next page for additional authors

Recommended Citation

Authors

Joseph D. Bonner, Rachel Fisher, James Klein, Qing Lu, Ellen Wilch, Karen H. Friderici, Jill L. Elfenbein, Debra L. Schutte, and Brian C. Schutte

Pedigree Structure and Kinship Measurements of a Mid-Michigan Community: A New North American Population Isolate Identified

Joseph D. Bonner,[1]* Rachel Fisher,[1] James Klein, Qing Lu,[1] Ellen Wilch,[1] Karen H. Friderici,[1] Jill L. Elfenbein,[2] Debra L. Schutte,[2] and Brian C. Schutte[1]

[1]Michigan State University.

[2]Wayne State University.

*Correspondence to: Joseph D. Bonner. E-mail: joedonbonner@gmail.com.

Abstract. Previous studies identified a cluster of individuals with an autosomal recessive form of deafness that resides in a small region of mid-Michigan. We hypothesized that affected members from this community descend from a defined founder population. Using public records and personal interviews, we constructed a genealogical database that includes the affected individuals and their extended families as descendants of 461 settlers who emigrated from the Eifel region of Germany between 1836 and 1875. The genealogical database represents a 13-generation pedigree that includes 27,747 descendants of these settlers. Among

these descendants, 13,784 are presumed living. Many of the extant descendants reside in a 90 square mile area and 52% were born to parents who share at least one common ancestor. Among those born to related parents, the median kinship coefficient is $3.7*10^{-3}$. While the pedigree contains 2,510 founders, 344 of the 461 settlers accounted for 67% of the genome in the extant population. These data suggest that we identified a new population isolate in North America and, as demonstrated for congenital hearing loss, this rural mid-Michigan community is a new resource to discover heritable factors that contribute to common health-related conditions.

Population isolates are not strictly defined. Rather, they share a number of general characteristics, such as a small founder population, rapid outgrowth, a history of isolation, an endogamous genealogy, environmental and phenotypic homogeneity, and the likelihood of genetically recessive health conditions (Arcos-Burgos and Muenke 2002). Endogamy is the practice of marriage between individuals of a common ethnic, social or religious group. Within a few generations most individuals of the endogamous group can be related to each other, often through multiple genealogical lines (Arcos-Burgos and Muenke 2002). Repeated generations of endogamy increase the likelihood of homozygosity of recessive alleles (McQuillan et al. 2008). This homozygosity might contribute risk of health-related conditions (Macgregor et al. 2010) and common diseases, like

cardiac conditions (Ismail et al. 2004; Ushasree et al. 2009; Youhanna et al. 2010) and some cancers (Denic 2003; Levy-Lahad et al. 1995). While relatedness of parents is noted as a risk factor for common complex diseases, individuals within population isolates are at no greater risk for childhood mortality (Bittles and Neel 1994) and might share positive health attributes such as longevity (Altukhov Yu and Sheremet'eva 2000) and increased fertility (Helgason et al. 2008).

The basic characteristics of population isolates make them important resources for biomedical research (Arcos-Burgos and Muenke 2002). For example, isolated populations were instrumental in the discovery of genetic factors for early-onset Alzheimer disease (Levy-Lahad et al. 1995; Rogaev et al. 1995), Hirschsprung disease (Puffenberger et al. 1994), and schizophrenia (Stoll et al. 2013), and the mapping of candidate loci in other complex traits such as Type II diabetes (Hanson et al. 1998), and hypertension (Hegele et al. 1996). These gene discoveries provided an immediate molecular diagnostic and screening benefit for the isolated population, and also benefit the broader population by advancing our understanding of pathophysiological mechanisms and gene functions and by accounting directly for some of the heritability of common diseases (Jonsson et al. 2012b; Stoll et al. 2013).

Over the last two decades, researchers have successfully identified genetic factors that cause some rare diseases in families. They have also identified common genetic factors that contribute risk for common diseases in broader

populations. However, despite these successes, much of the heritability for most common disorders remains elusive (Manolio et al. 2009). One possible source of the missing heritability of common diseases is rare DNA variants of moderate-to-strong effect. Because they are rare, such DNA variants are refractory to discovery strategies that rely on linkage disequilibrium or deep sequencing. Approaches based on linkage disequilibrium have very low power to detect rare DNA variants, while approaches based on deep sequencing require *a priori* knowledge of pathophysiological mechanisms and gene functions (Manolio et al. 2009).

In a population isolate the frequency of a DNA variant might be higher, thereby increasing the likelihood of finding a significant association with a common disease. For example, in the Icelandic population, two DNA variants, one in *TREM2* and one in *APP*, were found to be associated with late-onset Alzheimer disease (Jonsson et al. 2012a; Jonsson et al. 2012b). Both DNA variants were found at a lower frequency in North American and other European populations. In fact, the frequency of the *APP* variant in North American and other European populations was too low to be tested for an association. Thus, these discoveries exemplify how a research partnership with a population isolate can be used to identify low-frequency etiologic DNA variants. This example also highlights the potential impact of gene discoveries from population isolates. While both of these discoveries were highly significant because they provided

insights into the etiology of Alzheimer disease, the *TREM2* discovery had

additional immediate impact because the risk allele was found at a low, but

significant, frequency in other populations world-wide. Therefore, one way to

maximize the impact of genetic discoveries from population isolates is to

establish a research partnership with a population whose isolation was relatively

recent. In such populations, because the isolation was recent, the etiologic variant

would more likely be found in the broader population and account for disease risk

there, too.

North America is a challenging place to identify population isolates.

While such isolates exist, including the Hutterite communities (Steinberg AG

1967), Old Order Amish communities (Agarwala et al. 2003), and residents of

Tangier Island (Mathias et al. 2000), the relatively short time since initial

immigration, the high mobility of citizens, and the paucity of distant historical and

genealogical records make the identification of new population isolates difficult.

In this paper, we present the history and genealogy of a community in mid-

Michigan within the context of the features of a population isolate. This

community came to our attention through numerous cases of inherited deafness

within a narrow geographical region. Genealogical analysis elucidated a

community of highly related individuals with a defined set of founders. Our data

suggest that we identified a new North American population isolate of European

descent.

Materials and Methods

IRB and Informed Consent. All genealogical data for the community were collected as part of a study on congenital deafness (Rothrock et al. 2003; Wilch et al. 2010; Wilch et al. 2006). That study was approved by the Michigan State University (MSU) Institutional Review Board (IRB) and two community-based committees. The two committees (Research Advisory Committee and Research Ethics Committee) each have six to nine community members and approve all research protocols and scholarly products from within the community, including this manuscript.

Data Sources and Management. The genealogical data were ascertained from multiple sources: private family genealogical documents; a community history publication; genealogical websites that included the International Genealogy Index (IGI) from the Church of Jesus Christ of Latter Day Saints (LDS), searchmichigan.net website, individual family websites; cemetery searches in the community; verbal family histories; local historians; microfiche of early Michigan records at the local LDS Family History Library; and marriage banns, birth, death and anniversary announcements in the local newspaper. Ancestry.com© served as a secondary source to verify data from other sources or to find ancillary dates of

birth, immigration or death. Collection of genealogical data for a given individual ended when they married an unrelated spouse.

The genealogical data were kept in a large spreadsheet file of multiple worksheets. For the present analysis, the multiple worksheets were concatenated into a large text file of one row per unique individual. The concatenated file was used as input for extended family allocation and relatedness calculations. Each row for an individual contained an identification number, sex, birth date, death date and the names of parents and spouse(s).

Pedigree Structure and Relatedness Measurements. We used PedHunter2.0 to allocate individuals into extended families, identify the presumed living (extant) population and calculate measures of relatedness (Lee et al. 2010). To allocate individuals to extended families, we wrote a perl script to iteratively use individual identification numbers (ordered by birth year) as input to the "all_relatives" query with PedHunter2.0. This query returns a list of all individuals connected to the input individual through parents, spouses or offspring. The perl script accepted the output list of relatives from "all_relatives" and assigned the input individual identification number as the extended family identification number.

Founders are individuals who start the genealogic lines because they lack parental names in a pedigree database. Some historic community settlers are

pedigree founders. However, the genealogical database contains ancestors of some settlers thereby creating a distinction between settlers and founders. Founders who contributed only one child to the pedigree were not relevant to this study and were trimmed from the list of founders. We used the "trim_pedigree" utility of PedHunter2.0 to purge all non-relevant founders. Pedigree trimming logic was performed iteratively until no further genetically irrelevant individuals were identified.

Kinship coefficients represent the probability of an allele being shared by an ancestor common to both parents. PedHunter2.0 uses the algorithm of Weir (Weir 1996; Weir and Cockerham 1984) to calculate f-values as a kinship coefficient. We queried the database to produce a list of all mother-father combinations. We used this list as input to the "kinship" complex query. We used the output from "kinship" to assign to each child the kinship coefficient (f-value) of their parents.

Relative Founder Representation (RFR) or "r-value" is two times the kinship coefficient (f-value) and represents the proportion of a genome that is contributed by a specific ancestor. For example, an individual has an f-value of 0.125 with their grandparents. Conversely each grandparent contributes 25% of the genomes of their grandchildren, therefore the r-value between grandparent and grandchild is 0.25. When put into a broader pedigree context, the r-value each founder has with all their living descendants represents the relative representation

of each founder in the genome of the extant population. We generated the r-statistics for each relevant founder and their living descendants using the "calculate_r" and "average_r" complex queries of PedHunter2.0.

Plat Map. We obtained plat maps of property ownership in 1999 from the respective county governments. We scaled the maps to a common unit and highlighted the plots owned by individuals who have a surname in common with the settlers.

Results

Historical Context. In 1997, we were consulted about the case of a three-year-old child with congenital hearing loss. The team met with the extended family and received a copy of the child's pedigree going back six generations. The audiologist observed that the surname of an ancestor was the same as the surname of a child who was being counseled for post-cochlear implant speech therapy. With parental permission, this case joined the research study. A genealogist from the community took exhaustive family histories and established a genealogical link between the two families. In total, 15 cases of autosomal recessive deafness were identified in the community and linked within the pedigree. Genetic analyses of this extended pedigree showed that 11 cases were homozygous for the *35delG* mutation of *GJB2*, and four cases were compound heterozygous for *35delG* and a

novel deletion located upstream of *GJB2* (Rothrock et al. 2003; Wilch et al. 2010; Wilch et al. 2006).

Further genealogical analyses revealed that a preponderance of the community descended from a group of settlers whose initial arrival antedates Michigan statehood in 1837. In 1836, six individuals, including a Roman Catholic priest, emigrated from the Eifel region of Germany to mid-Michigan. These settlers cleared land, built homesteads, and established a parish church and school. Between 1836 and 1875, 456 settlers followed. Most settlers were other family members and acquaintances from Germany and were in their third or fourth decade of life. The demographics of the settlers were 46% female and 54% male and 350 were married to other settlers at the time of immigration. There were 288 distinct surnames among the 461 settlers. The most common surname was shared by 28 (6%) settlers.

Many of the living descendants of the settlers reside within a 90 square mile region. This region contains three villages. Much of the land surrounding the villages is owned by individuals whose surnames are the same as those of the settlers (Figure 1). Although the community is not isolated geographically, the German Catholic settlers remained isolated by language and religion from their predominantly English-speaking, non-Catholic neighbors. For example, classes were taught in both German and English in the early years of the settlement. In addition, the school in the primary settlement was exclusively parochial until

1956, and the Catholic churches in each village remain a central social institution supporting the close-knit nature of the population (Dykstra-Goris et al. 2014; Norris 1950).

Genealogical Database. The primary sources of the genealogical database were family history publications, accounting for about 40% of the primary records. Internet searches and newspaper publications contributed another 30% and 12% of the records, respectively. The remaining sources (local history publications, cemetery searches, verbal family histories, local historians, early Michigan records and parish marriage banns) each contributed 5% or less of the records. As of October 1, 2009, the genealogical database contained 28,256 ancestor or descendants of the original settlers. The birth decades of these individuals ranged from 1650 to 2000 (Figure 2A). The number of births in the genealogical database rises rapidly in the early 1800s and corresponds with the birth decades of the original settlers (Figure 2B). During the 20th century, the birth rates in the genealogical database precisely parallel the changes in the broader US population, with a sharp decrease in 1930s, a sharp increase in the 1950s and a sharp decrease in the 1960s. These data suggest that this community is strongly integrated into the broader society economically and culturally. The precipitous drop in the number of births after 1970 reflects missing data.

Pedigree Analysis. The family allocation logic assembled all individuals from this genealogical database into 310 extended families. One of the extended families included 27,747 (98%) individuals, spanning 13 generations. The extended family within the database contained 2,754 founders. We defined founders as individuals that lacked parental information in the genealogical database. However, we trimmed 244 (9%) non-significant founders who only contributed one descendent to the pedigree, leaving 2,510 founders. While the birth decade of founders ranged from 1650 to 1982, there was a rapid rise in the number of founders at the beginning of the 19th century (Figure 2B). As expected, the birth decade of these early founders coincided with the birth decade for the settlers. In fact, the 461 settlers accounted for 344 of the founders in the pedigree. From 1850 to 1910, the rate of new founders declined or was steady (Figure 2B). However, since the number of births increased during this same period, these data suggest a period of high endogamous growth. The rate of new founders increased after 1920 suggesting an increased rate of exogamous marriages. Like the decline in the total number of births, the decline in the number of founders after the 1970s reflects incompleteness of the genealogical data.

There are 14,756 (53%) individuals in the extended family who were born after 1930. Of these 972 (3%) have a recorded death date. Thus, the genealogic database contains 13,784 living descendants of the 461 settlers in equal numbers of males and females. Among the 461 settlers, 389 (84%) have living descendants

among the presumed living population. One settler couple has over 8,000 living descendants.

The extant descendants carry 963 surnames, 116 of which are shared with the settlers. The shared surnames are used by 70% (9,531/13,784) of the extant descendants. One surname is shared by 8% (1,047/13,784) of the living descendants.

Measures of Relatedness. We measured relatedness in two ways: 1) the frequency of children born to parents who share at least one common ancestor, and 2) the median f-value (kinship coefficient) between all children born to related parents. For a closed population, both of these values should increase with time, and following completion of the immigration phase, all children should be born to related parents. In this population we observed that the frequency of related parents peaked at 68% in 1970 and after an initial modest rise, the kinship coefficient decreased (Figure 2C). Again, these data suggest that while the community was isolated, it was not closed. However, the median f-value (kinship coefficient) was $3.7*10^{-3}$ (range $2.0*10^{-6} – 3.4*10^{-2}$) demonstrating the presence of measurable amounts of homozygosity within the extant population.

The extant genome represents the genetic contributions of the full set of founders. To estimate the settler's contribution to the extant genome, we calculated the relative founder representation (RFR) for all founders. One

founder, born in 1811, contributed 1% of the extant genome. The settlers

collectively contributed 67% of the extant genome (Figure 3). In comparison, the

Old Order Amish of Lancaster, PA (a closed population) (Lee et al. 2010) has

fewer founders, and thus each founder has a greater contribution to the extant

genome (Figure 3).

Discussion

The purpose of this study was to consider whether a community in mid-Michigan

settled by a few immigrants in 1836 constitutes a North American population

isolate. While there is no strict definition for a population isolate, we can compare

the characteristics of this community with three known population isolates of

European descent in North America: the Old Order Amish, the Hutterites of South

Dakota, and Tangier Island (Table 1).

The origin of the isolated community was well-documented, beginning

with 461 settlers who immigrated primarily from the Eifel region of Germany. In

comparison, the Old Order Amish, Hutterites and Tangier Island populations

settled in North America a century earlier, indicating a more recent isolation in

our partner community. The Old Order Amish and Hutterite communities are

closed, while the present community is only isolated with 52% of the extant

members born to parents that have a common ancestor. Additionally the median

kinship coefficient value ($f = 3.7*10^{-3}$) indicates that the related individuals share

a common ancestor within the past six to seven generations. This value is much lower than other North American isolates, yet considerably higher than the broader population estimates of 99 generations (Lachance 2009) and is consistent with a population isolate.

The pedigree database at the core of this work was assembled by community members and the authors, with the goal of elucidating the biological relationships among the descendants of the settlers. Thus, the founders' descendants were tracked until an individual married someone who was not a descendant of the settlers. While this focus would tend to increase the proportion of individuals born to related parents, the focus does not inflate the degree of relatedness between parents.

Another characteristic of a population isolate is the occurrence of recessive disorders. We previously identified a number of cases of autosomal recessive deafness in the community (Rothrock et al. 2003) caused by homozygosity or compound heterozygosity for *GJB2* mutations. We discovered a novel mutant allele that bears a deletion of distant *GJB2*-regulatory sequence (Wilch et al. 2010; Wilch et al. 2006). This allele was found in compound heterozygosity with a common mutation of *GJB2* in four deaf individuals in separate sibships. These four individuals and the other deletion carriers in our study population share a set of four common ancestors born in Germany between 1702 and 1723, suggesting its presence among German descendants of these four

individuals. However, this allele has not been found in other populations (Wilch et al. 2010) and is likely to be rare. Although this rare mutation has not itself contributed to diagnostic yield in genetic testing for hearing loss outside of Michigan, it has provided strong evidence for the location of critical GJB2-regulatory elements.

Lastly, isolated populations share a relatively homogenous environment. Although this study did not directly study the environment of our partner community, the plat map provides a very broad overview. Both the rural setting and the close proximity in which descendants of the settlers currently reside are consistent with a shared and homogenous environment (Igl et al. 2010).

In summary, although the extent of isolation is somewhat lower than in other well-characterized North American isolates, the historical, genealogical and geographical data are consistent with the conclusion that the community herein presented represents a newly identified North American population isolate.

Perspectives

We describe herein a newly identified population isolate. Importantly, the community in which the descendants reside has demonstrated an ongoing willingness to participate in biomedical research projects, which were focused initially on the genetic causes of congenital deafness. A community health needs assessment and some early environmental descriptions are underway to guide an

expanded research focus to examine the genetic and environmental causes of common complex diseases. The interests and priorities of the community will set the directives. We hypothesize that these processes, coupled with the unique community characteristics, provide a powerful tool to aid in searching for the missing heritability in common complex diseases.

Literature Cited

Agarwala, R., L. G. Biesecker, and A. A. Schaffer. 2003. Anabaptist genealogy database. *Am. J. Med. Genet. C Semin. Med. Genet.* 121C:32<N>37.

Altukhov Yu, P., and V. A. Sheremet'eva. 2000. Genomic heterozygosity and human longevity. *Dokl. Biol. Sci.* 371:197<N>199.

Arcos-Burgos, M., and M. Muenke. 2002. Genetics of population isolates. *Clin. Genet.* 61:233<N>247.

Bittles, A. H., and J. V. Neel. 1994. The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.* 8:117<N>121.

Bowen, P., and G. J. Conradi. 1976. Syndrome of skeletal and genitourinary anomalies with unusual facies and failure to thrive in Hutterite sibs. *Birth Defects Original Article Series* 12:101.

Denic, S. 2003. Consanguinity as risk factor for cervical carcinoma. *Med. Hypotheses* 60:321<N>324.

Dykstra-Goris, E., D. L. Schutte, J. Rivard et al. 2014. Community leader perceptions of the health needs of older adults. *Western Journal of Nursing Research* In Press.

Hanson, R. L., M. G. Ehm, D. J. Pettitt et al. 1998. An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.* 63:1,130<N>1,138.

Hegele, R. A., J. H. Brunt, and P. W. Connelly. 1996. Genetic and biochemical factors associated with variation in blood pressure in a genetic isolate. *Hypertension* 27:308<N>312.

Helgason, A., S. Palsson, D. F. Gudbjartsson et al. 2008. An association between the kinship and fertility of human couples. *Science* 319:813<N>816.

Igl, W., A. Johansson, and U. Gyllensten. 2010. The Northern Swedish Population Health Study (NSPHS)--A paradigmatic study in a rural population combining community health and basic research. *Rural Remote Health* 10:1,363.

Ismail, J., T. H. Jafar, F. H. Jafary et al. 2004. Risk factors for non-fatal myocardial infarction in young South Asian adults. *Heart* 90:259<N>263.

Jonsson, T., J. K. Atwal, S. Steinberg et al. 2012a. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488:96<N>99.

Jonsson, T., H. Stefansson, D. S. Ph et al. 2012b. Variant of TREM2 Associated with the Risk of Alzheimer's Disease. *N. Engl. J. Med*. 368:107<N>116.

Lachance, J. 2009. Inbreeding, pedigree size, and the most recent common ancestor of humanity. *J. Theor. Biol.* 261:238<N>247.

Lee, W. J., T. I. Pollin, J. R. O'Connell et al. 2010. PedHunter 2.0 and its usage to characterize the founder structure of the Old Order Amish of Lancaster County. *BMC Med. Genet.* 11:68.

Levy-Lahad, E., E. M. Wijsman, E. Nemens et al. 1995. A familial Alzheimer's disease locus on chromosome 1. *Science* 269:970<N>973.

Macgregor, S., C. Bellis, R. A. Lea et al. 2010. Legacy of mutiny on the Bounty: founder effect and admixture on Norfolk Island. *Eur. J. Hum. Genet.* 18:67<N>72.

Manolio, T. A., F. S. Collins, N. J. Cox et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747<N>753.

Mathias, R. A., C. A. Bickel, T. H. Beaty et al. 2000. A study of contemporary levels and temporal trends in inbreeding in the Tangier Island, Virginia, population using pedigree data and isonymy. *Am. J. Phys. Anthropol.* 112:29<N>38.

McQuillan, R., A. L. Leutenegger, R. Abdel-Rahman et al. 2008. Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83:359<N>372.

Norris, T. L. 1950. Acculturation in a German Catholic community. East Lansing: Michigan State University.

Puffenberger, E. G., K. Hosoda, S. S. Washington et al. 1994. A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* 79:1,257<N>1,266.

Rogaev, E. I., R. Sherrington, E. A. Rogaeva et al. 1995. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* 376:775<N>778.

Rothrock, C. R., A. Murgia, E. L. Sartorato et al. 2003. Connexin 26 35delG does not represent a mutational hotspot. *Hum. Genet.* 113:18<N>23.

Steinberg, A. G., H. K. Bleibtreu, T. W. Kurczynski, A. O. Martin, and E. M. Kurczynski. 1967. Genetic studies in an inbred human isolate. In *Proceedings of the Third International Congress of Human Genetics*, J. H. Crow and J. V. Neel, editors. Johns Hopkins University Press, Baltimore, 267<N>290.

Stoll, G., O. P. Pietilainen, B. Linder et al. 2013. Deletion of TOP3beta, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. *Nat. Neurosci.* 16:1,228<N>1,237.

Stone, D. L., R. Agarwala, A. A. Schaffer et al. 1998. Genetic and physical mapping of the McKusick-Kaufman syndrome. *Hum. Mol. Genet.* 7:475<N>481.

Ushasree, B., V. Shivani, A. Venkateshwari et al. 2009. Epidemiology and genetics of dilated cardiomyopathy in the Indian context. *Indian J. Med. Sci.* 63:288<N>296.

Weir, B. S. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Sunderland, MA: Sinauer.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 1984:1,358<N>1,370.

Wilch, E., H. Azaiez, R. A. Fisher et al. 2010. A novel DFNB1 deletion allele supports the existence of a distant cis-regulatory region that controls GJB2 and GJB6 expression. *Clin. Genet.* 78:267<N>274.

Wilch, E., M. Zhu, K. B. Burkhart et al. 2006. Expression of GJB2 and GJB6 is reduced in a novel DFNB1 allele. *Am. J. Hum. Genet.* 79:174<N>179.

Youhanna, S., D. E. Platt, A. Rebeiz et al. 2010. Parental consanguinity and family history of coronary artery disease strongly predict early stenosis. *Atherosclerosis* 212:559<N>563.
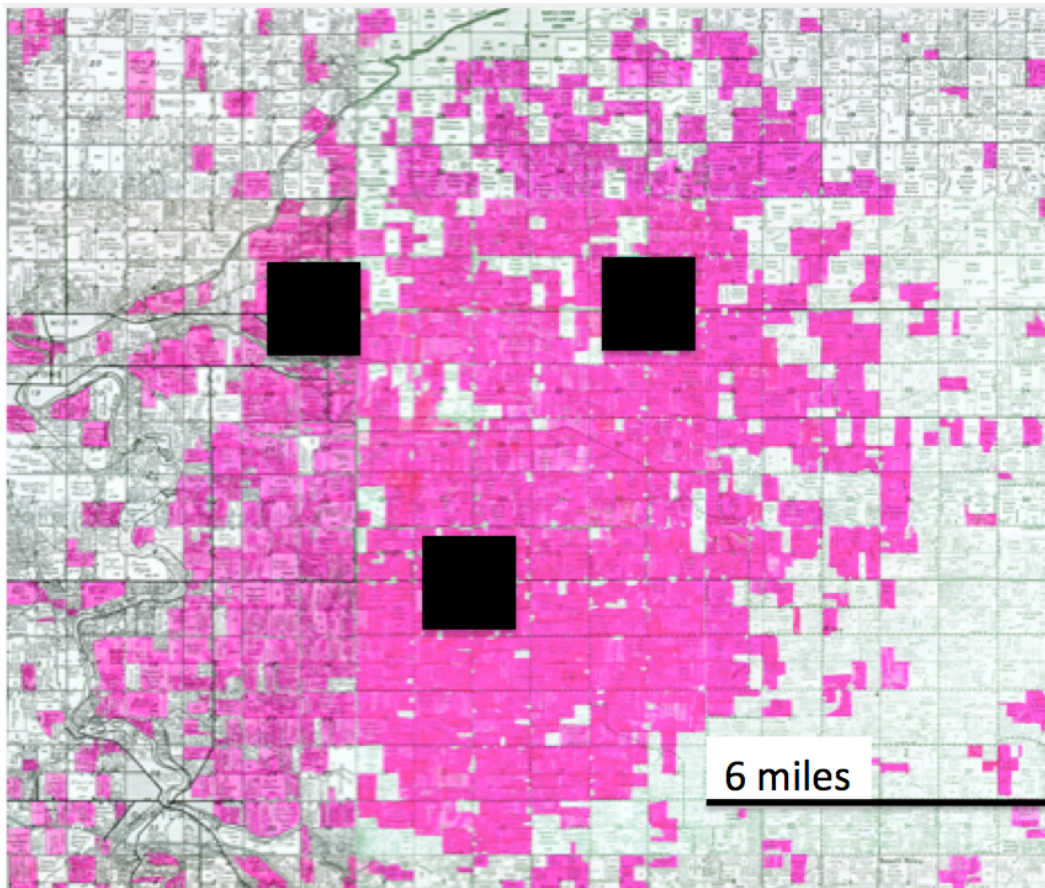
Figure 1. Plat map of mid-Michigan region. Black squares indicate the location of the three villages and pink squares indicate property owned by an individual whose surname is the same as one of the original settlers.
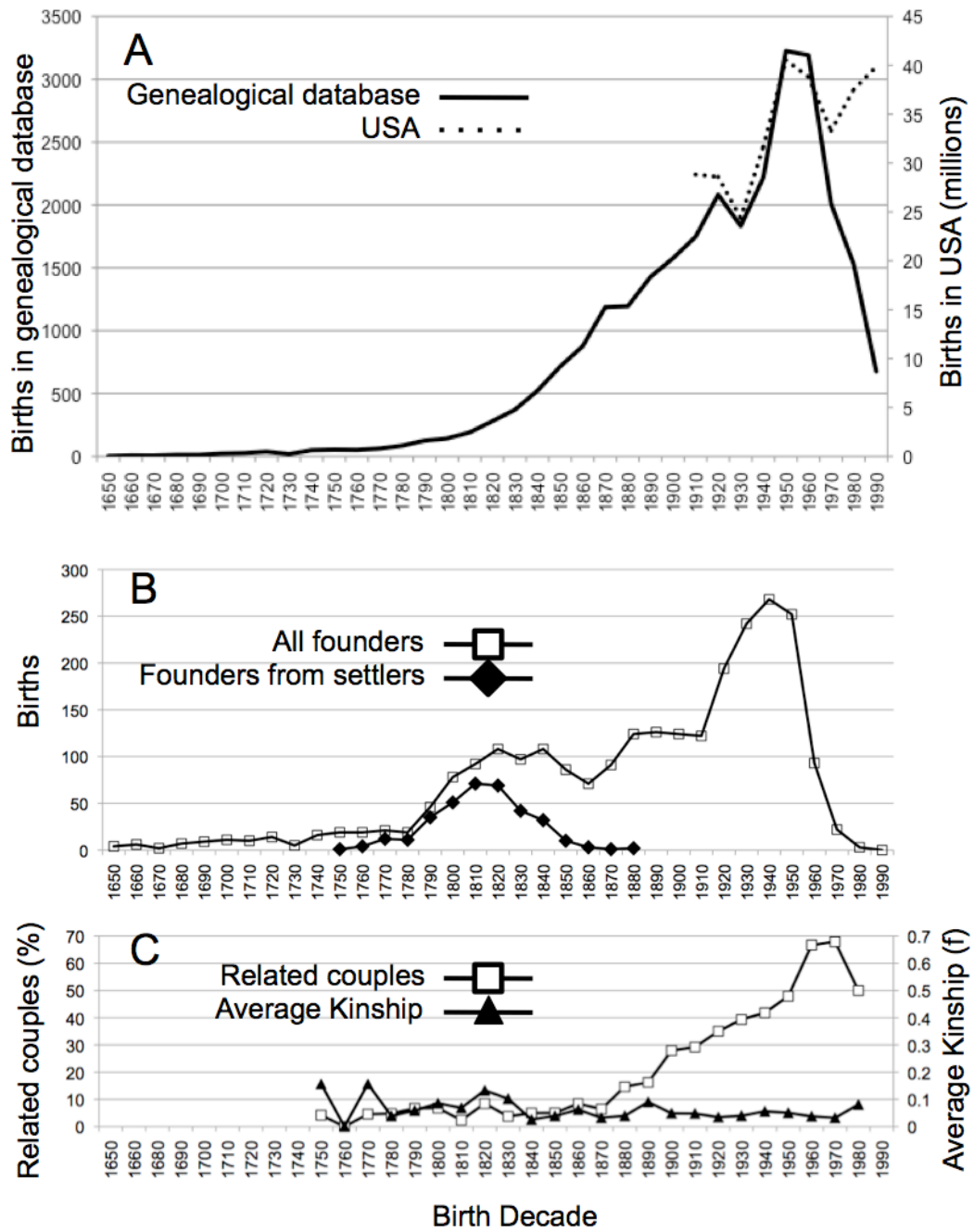
Figure 2. Genealogical analyses of mid-Michigan settler community. (A) The

number of births in the pedigree (solid line) increased rapidly from the time of the

original settlement in 1836 to the 1930s when economic and cultural factors affected the birth rate in the pedigree (solid line) and the US population (dashed line) similarly. The apparent decline in birth rate in the pedigree since 1970s reflects the end of data collection. (B) 344 of the original 461 settlers were founders and were born between 1780 and 1860 (filled diamonds). As expected, the birth decade of all founders (open squares) tracks about two decades behind the birth decade for the total number of members in the community (Panel A). (C) Relatedness of Settler community by birth decade. Fraction of couples who have a common ancestor (open squares) and average f-value (closed triangle) by birth decade of the female in each couple.
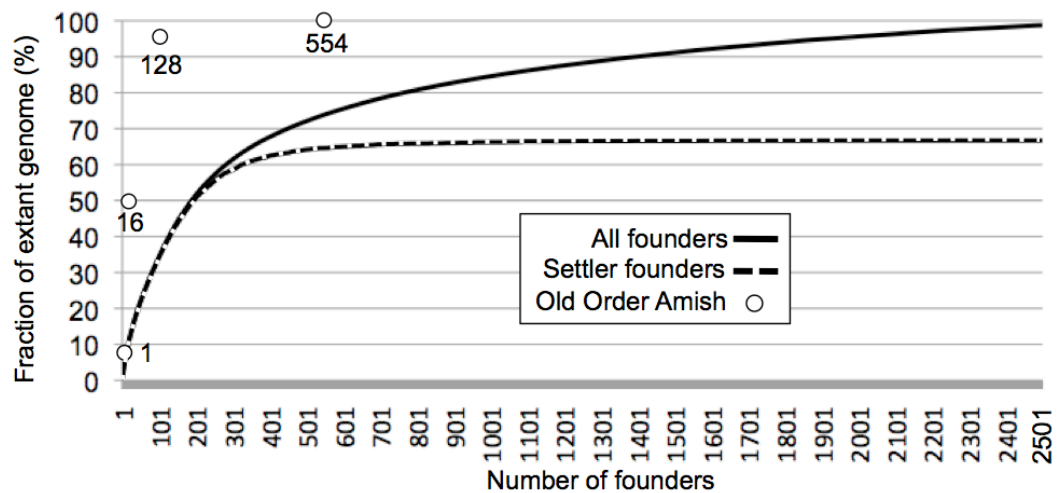

Figure 3. Cumulative mean relative founder representation (RFR) in genealogical database of our partner community. The RFR curves are organized from the

founder with the highest to lowest representation in living members. The genealogical database contains 13,874 living members who descended from 2500 total founders (solid line). Of the original settlers, 334 were founders and accounted for 67% of the extant genome (dashed line), including one original settler that contributed 1%. In comparison, the Old Order Amish community consists of about 34,160 individuals that descended from 554 total founders and one founder contributed 7% of the extant genome (open circles; data from Lee et al. 2010).

Table 1. Characteristics of Population Isolates of North America of European Descent

	German-Catholic	Old Order Amish (Lancaster, PA)	Hutterite (South Dakota)	Tangier Island
Date of origin	1836	1757	Early 1700s	1722
Founders	2,510	554	64	104
Contemporary cohort	~14,000	34,160	722	~200

Isolation	Language/ Religion	Religion	Religion	Geography
Mean F-value	0.0037[a]	0.015	0.034	0.018
Homogeneity	Geography 90 square mile region	Lifestyle	Communal lifestyle	Geography
Examples of recessive traits (not exhaustive)	Congenital hearing loss	McKusick-Kaufman syndrome	Bowen-Conradi Syndrome, Merkel Syndrome	Tangier disease
References	(Rothrock et al. 2003; Wilch et al. 2010; Wilch et al. 2006)	(Lee et al. 2010; Stone et al. 1998)	(Bowen and Conradi 1976; Steinberg AG 1967)	(Mathias et al. 2000)

[a] For children born to related parents.