Mitochondrial DNA Variation in North Dakota Residents of European Ancestry

Katelyn Kjelland
Department of Biology and Forensic Science Program
University of North Dakota
Grand Forks, ND 58202, USA.

Dr. Igor Ovchinnikov
Associate Professor
Department of Biology and Forensic Science Program
University of North Dakota
10 Cornell St, Starcher Hall, Stop 9019
Grand Forks, ND 58202, USA.

Abstract

This study aims to examine mtDNA variation in residents of European ancestry living in North Dakota. This state was settled by European migrants of distinct origins who originally formed isolated ethnic enclaves. MtDNA hypervariable segments HVS-I and HVS-II of 97 unrelated individuals revealed high diversity as well as admixture, inferring extensive postsettlement gene flow. A total of 106 polymorphisms in the data set defined 88 different haplotypes belonging to haplogroups of western Eurasian origin. Population statistics of North Dakota demonstrated no reduction in mtDNA diversity. Close genetic proximity was observed with Germans, Slavs, and Scandinavians. AMOVA tests provided no significant evidence of genetic structure within the population and suggested less stratification than is observed in European source populations. Comparison of genetic data with genealogical record indicated that emigration from Europe led to increased gene flow and, consequently, a more homogenous genetic structure due to maternal ancestry in North Dakota.

Keywords: mitochondrial DNA, hypervariable regions, North Dakota, European US Americans, population genetics

1. Introduction

North Dakota lies in the northernmost part of the Great Plains region of the United States (Figure 1). Two key dispersal events comprise the principal era of European settlement in the area. Attention was first brought to the region at the beginning of the 19th century by sequential speculations of transportation, ranching, and farming. These ventures continued for decades in what was, at the time, the Dakota Territory. The second phase began in the 1880s with the arrival of waves of immigrant forerunners from Europe and Canada. Promotions of land and opportunity attracted homesteaders of various ethnic backgrounds. During this period lasting until the First World War, hundreds of thousands of emigrants predominantly from northern Europe had flocked to the area. Being that most newcomers emigrated directly from their native countries, familiar cultural surroundings – especially in language – eased the transition. These influences ultimately led to a pattern of colonization across the state in which peoples with social commonalities clustered together. Lack of shared language, culture, and religion hindered interactions between founding communities until later years, as means of transport and communication became more readily accessible and intermarriages between individuals of diverse ethnic backgrounds more common (Sherman et al., 1988).

Today, descendants of over 40 different national groups of European descent live in North Dakota (http://www.census.gov). According to the 2000 United States Census, these people make up 92.4% of the state's population, with families having emigrated from across all of Europe. Primary ethnic groups of European immigrants to the area include 1) Scandinavians, 2) Germans, 3) Slavs, and 4) people of the British Isles (Sherman et al., 1988; http://www.census.gov).

Intensive analysis of European source populations has revealed underlying genetic stratification within groups that had previously been considered uniform, exhibiting congruencies between evolved phylogenies and physical topography (Novembre et al., 2008; Richards et al., 2002). However, the extent to which migration from the Old World to North America, followed by several generations of intermarriage, changed the original genetic variation and its distribution in European-diaspora populations in the United States overall and in particular regions remains largely unknown.

Investigation of genetic variation and stratification in "genetically homogeneous" populations is crucial for elucidating historic networks of human migration and reconstructing relationships between populations, disease-association research and improving applications of forensic identification by gaining an accurate perspective about the region's colonizing founders and events. Though common in other countries, studies of genetic variation and structure rarely involve small populations in the United States and are mostly limited to heterogeneous continental categories of European American, African American, Asian American, Hispanic, and Native American, that are better described as mosaic groups of diverse ancestral origins with varying genetic and cultural histories. To date, these studies include investigation of the genetic variation, structure, and ancestry of New Hampshire's population, most of which depicts European origin (Sloan et al. 2009), the Basque of Idaho, California, and Nevada (Davis et al., 2011; Valverde et al., 2011), the Hutterites of South Dakota (Ober et al., 2001; Pichler et al., 2010), and Mennonite communities of Kansas and Nebraska (Melton et al., 2010).

North Dakota's population offers a unique situation in which genetic consequences imposed by human migration may be readily observed, as the state features substantial Old World influence, relatively recent settlement, and communities of remote nature. The history and cultural heritage of different ethnic groups living in North Dakota is described well (Sherman et al., 1988), but genetic variation and its distribution across the state has never been examined.

This study aims to survey the degree of variation in mtDNA hypervariable segments HVS-I and HVS-II of current residents of European ancestry and understand the position of North Dakota's recently derived, integrative population among European parent populations. In this study, we explore working hypotheses of whether emigration from Europe reduced genetic diversity and emphasized the genetic matrilineal structure in communities of European ancestry in North Dakota.

2. Materials and Methods

2.1 DNA Sampling and Data Collection

All study materials and procedures were approved by the Institutional Review Board at the University of North Dakota (Protocol № IRB-201202-281), and informed consent was obtained from each participant. Questionnaires were administered to gather information about subjects' ancestral geographic origin as well as birthplaces of preceding family members and to confirm that participants were unrelated for at least two generations as indicated by surname data. Buccal swabs (Whatman) from 97 eligible volunteers born and living in North Dakota were collected and given number codes corresponding to the separately stored questionnaires to assure confidentiality. Regional distribution of collected samples according to volunteers' birthplaces is shown in Figure 1.

2.2 Sample Processing

Extraction of DNA from saliva samples was performed with Chelex 100 (Bio-Rad) and proteinase K digestion (Walsh et al., 1991). Primer pairs F15879/R16545 and F16495/R389 were used in amplification and sequencing processes (Taylor et al., 2001). Polymerase chain reaction (PCR) was performed on DNA sections containing HVS-I and HVS-II using GoTaq Flexi DNA polymerase (Promega) according to manufacturer's instruction. Amplicons were sequenced in both forward and reverse directions using the BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies). Sequences were generated on an Applied Biosystems 3100 Genetic Analyzer. To reduce risk of contamination, each procedural step was performed in a separate flow hood or detached room. Negative controls were included in extraction and amplification phases to ensure no samples exhibited contaminants.

MEGA5 (Tamura et al., 2011) was used to align sequences with the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999). All sites deviating from the rCRS in segments corresponding to nucleotides 16000 - 16400 and 40 - 390 for HVS-I and HVS-II, respectively, were recorded and each mtDNA profile was subject to independent second-person confirmation of polymorphisms.

Insertions and deletions as well as heteroplasmic nucleotides noted during visual inspection of the sequences were omitted during population comparison and phylogenetic reconstruction processes. Obtained hypervariable region (HVR) sequences deduced membership of each subject to a particular haplogroup with the use of HaploGrep (http://haplogrep.uibk.ac.at), which equates mtDNA profiles to a comprehensive phylogenetic tree of global human mtDNA variation (Phylotree Build 15) (Kloss-Brandstätter et al., 2011; van Oven & Kayser, 2008). Sequences with haplogroup match percentages of 80 or lower were deemed undetermined, as the percentage reflects reliability of the assignment based on polymorphisms known to be associated with certain lineages.

2.3 Data Analysis

Population statistics including internal diversity indices as well as measurements of genetic distance between populations (F_{ST}) were calculated in accordance with the Tamura-Nei model of DNA sequence evolution, a transition to transversion ratio of 10:1, and an alpha value of 0.2 using Arlequin v3.5 (Excoffier & Lischer, 2010). Evaluation of samples' demographic properties included Tajima's D (Tajima, 1989) and Fu's F_S (Fu, 1997) neutrality statistics and Harpending's raggedness index based on mismatch distributions (Rogers & Harpending, 1992). Standard errors were calculated from 1000 bootstrap replicates and significance of measurements determined by 10,000 permutations. Analysis of molecular variance (AMOVA) was performed in Arlequin to assess mtDNA distribution within and among subdivisions of populations (Excoffier & Lischer, 2010).

2.4 Comparison of Population Data Sets

Demographic parameters of the North Dakota sample were equated to data on source European populations in previous studies including Brits (Piercy et al., 1993), Czechs (Vanecek et al., 2004), Croatians (http://www.ncbi.nlm.nih.gov), Danes (Mikkelsen et al., 2010), Finns (Hedman et al., 2007), French (Dubut et al., 2004), samples from northeastern Germany ("GermanNE"; Poetsch et al., 2003) and southern Germany ("GermanS"; Lutz et al., 1998), Icelanders (Helgason et al., 2000), Norwegians (Helgason et al., 2001), Poles (Malyarchuk et al., 2002), Russian samples from areas surrounding Moscow ("RussianM"; Orekhov et al., 1999) and south of Moscow ("RussianS"; Malyarchuk et al., 2002), Scots (Helgason et al., 2001), Slovaks (Lehocky et al., 2008), and Swedish samples both from the northern region of Norrbotten ("SwedishN"; Sajantila et al., 1995) and from throughout Sweden ("Swedish"; Tillmar et al., 2010). For consistency with European data sets, the North Dakota mtDNA sequences were trimmed to segments of nucleotides 16024 - 16383 of HVS-I and nucleotides 72 - 340 of HVS-II. Comparative data sets for HVS-I alone were restricted to nucleotides 16024 - 16383. A multidimensional scaling (MDS) scatterplot was constructed using STATISTICS software (StatSoft) based on genetic distances between populations.

3. Results

3.1 MtDNA diversity in North Dakota

Of the 97 individual North Dakota mtDNA HVR profiles sequenced, 88 were found to be distinct. Haplotypes of the sample set were defined by 106 substitutions in 102 different sites. Transitions accounted for 99 of the mutations while transversions comprised the remaining seven substitutions. Parallel substitutions observed at nucleotide sites 72, 16183, 16239, and 16294 were previously indicated at all these positions (Ruiz-Pesini et al., 2007). All but one sequence contained insertions in cytosine-rich stretches following nucleotides 309 (49 sequences; 50.5%), 315 (96 sequences; 99%), or 16193 (11 sequences; 11.3%).

HVR haplotypes of the North Dakota sample belonged to nine major haplogroups including HV0, H, J, K, N1, T, U, W, and X. Individual HVR profiles and assigned haplogroups are shown in Table 1. Five HVR profiles received haplogroup match confidences $\leq$ 80% and were subsequently deemed to have an undetermined haplogroup assignment. All identified haplogroups were descriptive of western Eurasian origin, falling under either macrohaplogroup N or R.

A unimodal, bell-shaped mismatch distribution of nucleotide site differences suggested recent growth of the population (data not shown).

Indices of demographic history coincide with this scenario, as calculations for Tajima's D statistic gave a negative value (-1.68046), Fu's F_S resulted in a large negative (-24.4995), and a statistically significant raggedness *r* value ($0.002819 < 0.05$) was observed.

Because many comparison populations included only HVS-I data, a separate analysis was done for this segment alone with the North Dakota sample likened to a wider range of source populations. HVS-I contained 63 transitions and six transversions in 66 segregating sites, constituting the majority of polymorphisms found in the North Dakota mtDNA HVR profiles examined. HVS-I displayed 72 unique haplotypes, and the data set again illustrated population growth through values for Tajima's D (-1.6457), Fu's F_S (-25.0760), and a raggedness *r* value of 0.01096 < 0.05.

Of the compared HVR data sets, the North Dakota sample had greatest values for both nucleotide diversity (0.0127 ± 0.0066) and mean pairwise differences (7.98 ± 3.74) relative to eight European source populations, demonstrating maximal genetic variation (Table 2). The Czech population followed, with a nucleotide diversity of 0.0124 and mean pairwise differences at 7.80. Nominal genetic variation was found in the French sample, having a nucleotide diversity of 0.0104 and mean pairwise difference value of 6.57. For all data sets, Tajima's D, ranging from -1.6476 to -2.1046, and Fu's F_S, from -23.9577 to -24.4995, implied population growth.

As for HVS-I data, populations showed nucleotide diversity oscillating between 0.0105 and 0.0137, with the Czech sample exhibiting the highest variation and the GermanS sample displaying the least (Table 3). Mean pairwise differences presented a similar pattern, as the value for Czechs (4.93) inferred much variability relative to the GermanS sample (3.76). Comparatively, North Dakota was still among the most diverse, harboring a nucleotide diversity of 0.0131 ± 0.0072 and mean pairwise difference of 4.71 ± 2.33, each of which were only topped by 4 of the 17 comparative groups. All populations had demographic parameters characteristic of population expansion, with statistics for Tajima's D ranging from -1.3762 to -2.1416 and Fu's F_S between -24.3760 and -25.6635.

3.2 Genealogical Record of Maternal Ancestry in North Dakota

Collected genealogical data showed that people came to North Dakota from a number of existing European countries as well as some countries that dissolved from the political map during the 20th century. Ancestral origins of North Dakota donors were inferred from questionnaire sections describing from where in the Old World their maternal relatives had emigrated. Self-reported genealogical backgrounds, after being adjusted to coincide with modern political boundaries, chiefly included German (36.1%), Norwegian (33.0%), Swedish (5.3%), and Czech and English (4.3% each). Irish, Scottish, and French heritages were represented at 3.1% apiece, and the remainder of European ties was covered by 1.1% each of Dutch, Finnish, Greek, Luxembourgian, Polish, Slovak, or Ukrainian ancestry. From information provided by volunteers, the North Dakota population represents a product of immigration, settlement, and admixture of Europeans of diverse ethnic background, notably of German or Norwegian ancestry.

3.3 North Dakota mtDNA Pool among European Source Population Data

Pairwise F_{ST} values were calculated to understand the position of the North Dakota mtDNA pool among European populations. Overall, the North Dakota sample is similar to source populations of Slavonic, German, and Scandinavian origins. Statistically significant genetic distance was demonstrated with only three populations: Finns (F_{ST} = 0.00615; *P* = 0.03188), Icelanders (F_{ST} = 0.01283; *P* = 0.00059), and Scots (F_{ST} = 0.01064; *P* = 0.00366).

In order to place the North Dakota sample within the European mtDNA framework, MDS analysis was carried out, resulting in the two-dimensional scatterplot of populations with the stress value of 0.135 (Figure 2). The scatterplot demonstrated non-arbitrary clustering of North Dakotans together with Russians, Germans from southern Germany, and Swedes. Populations aggregating around the first group included Poles, Germans from northeastern Germany, Czechs, Brits, and Slovaks. Norwegians were more distant from the cluster than all other populations excluding Finns and Icelanders, who were most isolated in the first dimension, and the Scots who were secluded in the second dimension. We cannot identify factors that contributed to each of the two MDS dimensions. However, when considered separately, each provides a different story. In the first dimension, Swedes and Poles have the lowest distances to the North Dakota population, while in the second dimension lowest distances to the North Dakota sample are seen with Icelanders and Norwegians.

3.4 Genetic Structure in North Dakota

Several geographical and ethnic arrangements of putative subpopulations were explored, with divisions based on historical, economic, and ethnic factors that guided the settlement of North Dakota (Sherman et al., 1988).

First, population subdivision was examined based on participants' self-reported maternal ancestry by comparing those of German versus Scandinavian (Norwegian and Swedish) background in one test and groups of German, Scandinavian, British, or Slavonic ethnicity in another. Second, North Dakotans were separated according to their places of birth. Participants born in towns with populations over 5,000 were considered "urban", and those born in towns with less than 5,000 residents were grouped as "rural". Finally, the North Dakota population was divided into six regional subpopulations (West, Center, North, East, Southeast, or South) as depicted in Figure 1, again based on birthplace.

The AMOVA test demonstrated genetic homogeneity for the different arrangements of population structure in North Dakota. Approximately 100% of the total genetic variance is expressed within the ethnic groups of different ancestry as well as within the "urban" and "rural" groups. Minor but statistically non-significant differentiation was revealed between the six geographical regions of North Dakota, with 0.75% of the genetic variance ($P = 0.19159$) explained by differences between subpopulations (Table 4).

AMOVA suggested less population stratification among North Dakota resident subdivisions than in European populaces used for the comparison study (Table 4). Tests performed with HVS-I source population data similarly split into three groups labeled Slavonic (Czech, Polish, RussianM, RussianS, and Slovak data), Scandinavian (Danish, Swedish, SwedishN, and Norwegian data), and Germans (GermanNE and GermanS data) also calculated little variation (0.16%) among populations, though this value was deemed significant ($P = 0.00782$) unlike results in the North Dakota AMOVA tests. The same scenario was reached when a fourth group, Icelanders, was added, increasing variation among populations to 0.34% ($F_{ST} = 0.00343$).

4. Discussion

Studies of genetic variation and structure in communities and states throughout the United States remain fundamental for medical and forensic genetics as well as for reconstructing and understanding the history of human migration and colonization. A population genetics study recently conducted in New Hampshire demonstrated that genetic structure due to influx of ethnically distinct immigrants from eastern Europe exists in the state whose population was previously considered ancestrally uniform and highly admixed, with 96% of residents having European ancestry (Sloan et al., 2009). The New Hampshire study was carried out using single nucleotide polymorphisms in the nuclear genome, but mtDNA analysis is also capable of detecting micro-differentiation within small groups both in North America and Europe, e.g., in Mennonite communities from the Midwest United States (Melton et al., 2010).

This study evaluated variation of mtDNA in the North Dakota population, which is also regarded as having a homogeneous European background. The sample's mtDNA data was then equated to self-reported maternal ancestry. The first hypothesis tested was whether emigration from Europe to North Dakota has reduced the genetic diversity in communities of European ancestry. Previous studies have demonstrated different effects on the mtDNA diversity of European communities in North America due to recent migrations. The Basque population living in Idaho, when compared to Basques in Spain, displayed a loss of genetic diversity as a result of founder effects (Davis et al., 2011). Conversely, the Quebec population of primarily French descent does not appear significantly different in the degree of mtDNA variation compared to European populations (Moreau et al., 2007). Analysis of the North Dakota sample included parameters such as the number of different haplotypes along with haplogroup each defines, nucleotide diversity, the mean number of pairwise nucleotide differences, as well as Tajima's D and Fu's F_S statistics. These considerations were all unable to detect reduction in mtDNA diversity or impacts by founder events in the North Dakota population. Even more, slightly higher mtDNA diversity is seen in North Dakota than in many European populations. The mechanism behind this change is not completely clear but can likely be explained by two processes. First, immigrants came to North Dakota from several European countries and constituted over 40 ethnic groups (Sherman et al., 1988; http://www.census.gov). The influx of immigrants from across such a broad territory brought mtDNA lineages present in virtually every part of Europe.

Second, because many volunteers' predecessors came to North Dakota before or at the turn of the 20th century, it may be postulated that the increased mtDNA variation found in North Dakota reflects a higher level of mtDNA diversity which had existed in Europe prior to the First and Second World Wars and local conflicts that led to drastic shifts in national borderlines in the last century.

Effects these 20th century divergences may have had on mtDNA diversity in Europe are unknown; however, consequences of political processes on Y-chromosome variation in Poland and Germany have been verified (Kayser et al., 2005; Rebala et al., 2013).

Next, the hypothesis was tested of whether North Dakota's mtDNA data correlates with self-reported maternal ancestries of residents. From mtDNA, it can be shown that either strong correlations exist between genetic structure and self-reported ancestry, like was observed in Australia (Byrne et al., 2008) or that such correspondence is lacking, as was seen in admixed Colombian populations (Salas et al., 2008). The majority of donors in North Dakota reported German or Norwegian maternal ancestry. Analysis of genetic distance and the two-dimensional MDS scatterplot both placed the North Dakota sample centroid to many European populations, with closest proximity to Germans, Russians, and Swedes. In general, this correlates with the German-Slavonic-Scandinavian background of donors and many communities in North Dakota. The Norwegian group's more remote location in the MDS scatterplot somewhat contradicts volunteers' self-reported maternal ancestry. However, it is the second closest population to North Dakotans in the second dimension, after Icelanders. Positioning of the North Dakota sample may be attributed to the complex history of settlement in North Dakota followed by increased gene flow between once-isolated ethnic communities of Scandinavian, German, and Slavonic origins over the decades. The close proximity between Russians and North Dakotans may be rationalized by the many German-speaking migrants that came to North Dakota not from within contemporary German borders but from other countries, especially Russia and Ukraine (Sherman et al., 1988), thus carrying with them mtDNA lineages descriptive of the eastern European genetic landscape. The substantial distance between populations of North Dakota and Norway versus that between North Dakota and Sweden may be a result of the union amongst Norway and Sweden in 1814 - 1905 and subsequent changes in the distribution of mtDNA lineages throughout recent history. The closer proximity of North Dakota with Denmark than with Norway is similarly supported. Both Denmark and Norway have an extensive history of unity from 1380 - 1814 (Rying, 1988) that would have theoretically enhanced gene flow between the countries. Additionally, records of fused ethnic backgrounds may have been misinterpreted or forgotten upon emigration to the United States, and such relationships are therefore inaccurate or not reflected in self-declared genealogies of current residents at all.

Finally, the third hypothesis was tested for whether genetic matrilineal structure exists between different ethnic groups, urban and rural communities, or separate geographic regions of North Dakota. It appears that genetic stratification originally observed among European founder populations has been eradicated in the state's postsettlement era. Demographic data of our donors shows that many families in North Dakota have accounts of intermarriage between ancestors of particular European origins in previous generations. It can be concluded that the decades of mixed marriages between immigrants has led to increased gene flow between once-isolated ethnic communities and, therefore, depleted much of the genetic structure that may have existed among initial settlements.

Acknowledgements

This study was supported by the North Dakota EPSCoR, UND faculty start-up awards to Igor Ovchinnikov, and a UND Academic Programs and Student Awards Committee award to Katelyn Kjelland. We thank Olga Kholina, Kenneth Drees and Kenneth Flagg for technical assistance, Daniel Bradley, Minttu Hedman, Miroslav Slip, Andreas Tillmar, and Tomas Vanecek for providing us with mtDNA data sets of European populations. We express gratitude to all volunteers who contributed buccal swabs to this study as well as to the Minnkota Genealogical Society, its President Mike Swanson, and members for help in advertising our North Dakota Genetic Ancestry Project.

References

Andrews, R.M., Kubacka, I., Chinnery P.F., et al. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics, 23, 147.

Byrne, E., McRae, A., Zhao, Z., et al. (2008). The use of common mitochondrial variants to detect and characterise population structure in the Australian population: implications for genome wide association studies. European Journal of Human Genetics, 16, 1396-1403.

Davis, M.C., Novak, S.J., & Hampikian, G. (2011). Mitochondrial DNA analysis of an immigrant Basque population: loss of diversity due to founder effects. American Journal of Physical Anthropology, 144, 516-525.

Dubut, V., Chollet, L., Murail, P., et al. (2004). MtDNA polymorphisms in five French groups: importance of regional sampling. European Journal of Human Genetics, 12, 293-300.

Excoffier, L., & Lischer, H.E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology, 10, 564-567.

Fu, Y. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, 147, 915-925.

Hedman, M., Brandstätter, A., Pimenoff, V., et al. (2007). Finnish mitochondrial DNA HVS-I and HVS-II population data. Forensic Science International, 172, 171-178.

Helgason, A., Hickey, E., Goodacre, S., et al. (2001). MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. American Journal of Human Genetics, 68, 723-737.

Helgason, A., Sigurethardottir, S., Gulcher, J., et al. (2000). MtDNA and the origin of the Icelanders: deciphering signals of recent population history. American Journal of Human Genetics, 66, 999 -1016.

Kayser, M., Lao, O., Anslinger, K., et al. (2005). Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y chromosome analysis. Human Genetics, 117, 428-443.

Kloss-Brandstätter, A., Pacher, D., Schoenherr, S., et al. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Human Mutation, 32, 25-32.

Lehocky, I., Baldovic, M., Kadasi, L., et al. (2008). A database of mitochondrial DNA hypervariable regions I and II sequences of individuals from Slovakia. Forensic Science International: Genetics, 2, e53-59.

Lutz, S., Weisser, H., Heizmann, J., et al. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. International Journal of Legal Medicine, 111, 67-77.

Malyarchuk, B.A., Grzybowski, T., Derenko, M., et al. (2002). Mitochondrial DNA variability in Poles and Russians. Annals of Human Genetics, 66, 261-283.

Melton, P.E., Mosher, M.J., Rubicz, R., et al. (2010). Mitochondrial DNA diversity in Mennonite communities from the Midwestern United States. Human Biology, 82, 267-289.

Mikkelsen, M., Sorensen, E., Rasmussen, E., et al. (2010). Mitochondrial DNA HV1 and HV2 variation in Danes. Forensic Science International: Genetics, 4, e87-88.

Moreau, C., Vezina, H., & Labuda, D. (2007). Effets fondateurs et variabilite genetique au Quebec. Medicine/Sciences (Paris), 23, 1008-1013.

Novembre, J., Johnson, T., Bryc, K., et al. (2008). Genes mirror geography within Europe. Nature, 456, 98-101.

Ober, C., Abney, M., & McPeek, M.S. (2001). The genetic dissection of complex traits in a founder population. American Journal of Human Genetics, 69, 1068-1079.

Orekhov, V., Poltoraus, A., Zhivotovsky, L., et al. (1999). Mitochondrial DNA sequence diversity in Russians. FEBS Letters, 445, 197-201.

Pichler, I., Fuchsberger, C., Platzer, C., et al. (2010). Drawing the history of the Hutterite population on a genetic landscape: inference from Y-chromosome and mtDNA genotypes. European Journal of Human Genetics, 18, 463-470.

Piercy, R., Sullivan, K., Benson, N., et al. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. International Journal of Legal Medicine, 106, 85 -90.

Poetsch, M., Wittig, H., Krause, D., et al. (2003). Mitochondrial diversity of a northeast German population sample. Forensic Science International, 137, 125-132.

Rebala, K., Martinez-Cruz, B., Tonjes, A., et al. (2013). Contemporary paternal genetic landscape of Polish and German populations: from early medieval Slavic expansion to post-World War II resettlements. European Journal of Human Genetics, 21, 415-422.

Richards, M.B., Macaulay, V.A., Torroni, A., et al. (2002). In search of geographical patterns in European mitochondrial DNA. American Journal of Human Genetics, 71, 1168-1174.

Rogers, A.R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution, 9, 552-569.

Ruiz-Pesini, E., Lott, M., Procaccio, V., et al. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Research, 35, D823-828.

Rying, B. (1988). Danish in the South and the North: Denmark history, vII. Copenhagen, Denmark: The Royal Danish Ministry of Foreign Affairs Press and Cultural Relations Department.

Sajantila, A., Lahermo, P., Anttinen, T., et al. (1995). Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Research, 5, 42-52.

Salas, A., Acosta, A., Alvarez-Iglesias, V., et al. (2008). The mtDNA ancestry of admixed Colombian populations. American Journal of Human Biology, 20, 584-591.

Sherman, W.C., Thorson, P.V., Henke, W.A., et al. (1988). Plains Folk: North Dakota's Ethnic History. Fargo, ND: North Dakota Institute for Regional Studies.

Sloan, C.D., Andrew, A.D., Duell, E.J., et al. (2009). Genetic population structure analysis in New Hampshire reveals eastern European ancestry. PLoS ONE, 4, e6928.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585-595.

Tamura, K., Peterson, D., Peterson, N., et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28, 2731-2739.

Taylor, R.W., Taylor, G.A., Durham, S.E., et al. (2001). The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations. Nucleic Acids Research, 29, e74.

Tillmar, A., Coble, M., Wallerstrom, T., et al. (2010). Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations. International Journal of Legal Medicine, 124, 91-98.

Valverde, L., Rosique, M., Kohnemann, S., et al. (2011). Y-STR variation in the Basque diaspora in the western USA: evolutionary and forensic perspectives. International Journal of Legal Medicine, 126, 293-298.

van Oven, M., & Kayser, M. (2008). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation, 30, e386-394.

Vanecek, T., Vorel, F., & Slip, M. (2004). Mitochondrial DNA D-loop hypervariable regions: Czech population data. International Journal of Legal Medicine, 118, 14-18.

Walsh, S., Metzger, D., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques, 10, 506-513.

Table 1: Individual HVR Profiles of the North Dakota Sample with Assigned Haplogroups. Samples Arranged by Haplogroup Association, Then by Motif Similarity, Giving HVS-I Priority

Sample	HVR Motif [a]		Haplogroup assignment
	HVS-I (16000 - 16400) [b]	HVS-II (40 - 390)	
103	51, 162	73, 152, 263, 309.1C, 315.1C	H1
58	92, 263	263, 315.1C	H1
35	162, 172, 209	73, 263, 315.1C	H1
34	172, 183C, 189, 193.1C, 356, 362	263, 315.1C	H1
97	189, 356	263, 309.1C, 309.2C, 315.1C	H1
32	263	263, 315.1C	H1
55	263	263, 315.1C	H1
92	270	263, 309.1C, 315.1C	H1
36	rCRS	152, 263, 315.1C	H2
41	rCRS	152, 263, 315.1C	H2
42	rCRS	263, 309.1C, 315.1C	H2
65	rCRS	263, 309.1C, 315.1C	H2
101	rCRS	263, 309.1C, 315.1C	H2
25	rCRS	263, 315.1C	H2
38	rCRS	263, 315.1C	H2
47	rCRS	263, 315.1C	H2
91	rCRS	263, 315.1C	H2
99	rCRS	263, 315.1C	H2
4	183	263, 309.1C, 315.1C	H2
67	354	263, 315.1C	H2
40	239G	152, 263, 309.1C, 315.1C	H3
75	294A, 304	263, 315.1C	H5
102	304	263, 309.1C, 309.2C, 315.1C	H5
44	304	263, 315.1C	H5
105	rCRS	239, 263, 309.1C, 315.1C	H6
3	92C, 362C	239, 263, 309.1C, 315.1C	H6
21	93, 221	263, 309.1 C, 315.1 C	H10
82	293, 311	143, 195, 263, 309.1C, 315.1C	H11
111	311, 390	195, 263, 309.1C, 309.2C, 315.1C	H11
39	227	263, 309.1C, 315.1C	H22
73	319	72G, 146, 195, 263, 309.1C, 315.1C	H31
30	189, 193.1C, 298	72, 152, 195, 263, 309.1C, 309.2C, 315.1C	HV0
74	298	72, 263, 309.1C, 315.1C	HV0
76	298	72, 263, 309.1C, 315.1C	HV0
23	298	72, 263, 310	HV0
2	51, 86, 129, 214, 223, 391	73, 152, 199, 204, 207, 239, 250, 263, 309.1C, 315.1C	I3
45	69, 93, 126, 145, 172, 222, 261	73, 146, 234, 242, 263, 295, 315.1C	J1
72	69, 126	73, 150, 185, 188, 228, 263, 295, 315.1C	J1
110	69, 126	73, 152, 185, 189, 263, 295, 309.1C, 315.1C	J1
71	69, 126	73, 185, 188, 228, 263, 295, 309.1C, 315.1C	J1
53	69, 126	73, 185, 188, 228, 263, 295, 315.1C	J1
14	69, 126	73, 185, 228, 263, 295, 315.1C	J1
54	69, 126, 148	41, 73, 185, 188, 228, 263, 295, 309.1C, 315.1C	J1
79	69, 126, 319	73, 185, 228, 263, 295, 309.1C, 315.1C	J1
63	69, 126, 145, 231, 261	73, 150, 152, 195, 215, 263, 295, 315.1C, 319C	J2
93	224, 311	73, 263, 309.1C, 315.1C	K
7	93, 224, 311	73, 263, 315.1C	K1
94	224, 239, 311, 320	73, 146, 152, 263, 315.1C	K1
51	224, 242, 311	73, 195, 263, 315.1C	K1
37	224, 245, 311	73, 146, 263, 309.1C, 315.1C	K1

69	224, 311	73, 146, 152, 263, 309.1C, 315.1C	K1
10	224, 311	73, 146, 195, 263, 315.1C	K1
106	145, 176G, 223, 390	73, 152, 263, 309.1C, 315.1C	N1
66	126, 294	73, 263, 315.1C	T
5	126, 163, 186, 189, 294	73, 152, 195, 263, 309.1C, 315.1C	T1
11	126, 163, 186, 189, 294	73, 152, 263, 309.1C, 315.1C	T1
90	126, 186, 189, 294	73, 152, 195, 263, 309.1C, 315.1C	T1
26	104, 126, 294, 304	73, 152, 263, 309.1C, 315.1C	T2
62	126, 153, 294, 296	73, 150, 263, 309.1C, 315.1C	T2
6	126, 189, 193.1C, 294, 296, 311	73, 263, 309.1C, 309.2C, 315.1C	T2
31	126, 294, 296, 304	41, 73, 263, 315.1C, 319C	T2
108	126, 294, 296, 304	73, 146, 263, 309.1C, 315.1C	T2
96	126, 294, 304, 309	73, 263, 309.1C, 315.1C	T2
56	168, 192, 343	73, 150, 263, 315.1C	U3
1	189, 193.1C, 343, 390	73, 150, 263, 309.1C, 315.1C	U3
22	193, 249, 343	73, 150, 263, 309.1C, 315.1C	U3
49	134, 356	73, 152, 195, 263, 309.1C, 309.2C, 315.1C	U4
64	179, 356	73, 195, 263, 315.1C	U4
18	193, 342, 356	73, 195, 263, 309.1C, 315.1C	U4
57	265, 356, 362	73, 195, 247, 263, 315.1C	U4
20	265, 356, 362	73, 195, 247, 263, 315.1C, 379G	U4
50	356	73, 152, 195, 263, 315.1C	U4
46	51, 256, 270, 399	73, 263, 309.1C, 315.1C	U5
17	93, 192, 248, 270, 304	73, 150, 228, 263, 315.1C	U5
60	114A, 192, 256, 270, 294	73, 152, 263, 315.1C	U5
77	144, 148, 183C, 189, 193.1C, 270, 335	73, 150, 263, 309.1, 309.2C, 315.1C	U5
68	144, 189, 193.1C, 270	73, 150, 263, 315.1C	U5
80	144, 189, 270	73, 150, 263, 315.1C	U5
29	172, 192, 256, 270, 291, 399	73, 263, 315.1C	U5
61	189, 192, 270	73, 125, 127, 150, 217, 263, 309.1C, 315.1C	U5
9	189, 270	73, 150, 217, 263, 309.1C, 315.1C	U5
70	192, 256, 270, 311, 399	73, 199, 263, 315.1C	U5
15	256, 270, 320, 399	73, 195, 263, 315.1C	U5
13	256, 270, 399	73, 152, 263, 309.1C, 315.1C	U5
24	256, 270, 399	73, 263, 315.1C	U5
48	179, 189, 193.1C	73, 263, 282, 315.1C	U8
16	93, 183C, 189, 193.1C, 213, 223, 278	73, 143, 153, 195, 200, 263, 309.1C, 315.1C	X2
8	93, 186, 189, 223, 278	73, 153, 195, 225, 226, 263, 309.1C, 315.1C	X2
59	183C, 189, 193.1C, 213, 223, 278	73, 143, 153, 195, 200, 263, 298, 309.1C, 315.1C	X2
33	189, 193.1C, 193.2C, 223, 278	73, 153, 195, 225, 226, 263, 309.1C, 315.1C	X2
81	104, 223, 292	73, 189, 194, 195, 204, 207, 263, 315.1C	W
27	223	73, 143, 189, 192, 194, 195, 196, 204, 207, 263, 315.1C	W4
107	rCRS	150, 248, 263, 315.1C	und.[c]
28	86, 147A, 223, 248, 320, 355	73, 152, 189, 199, 204, 207, 263, 315.1C	und.
43	92, 301, 311	195, 263, 315.1C	und.
104	183C, 193.1C, 271	73, 150, 263, 315.1C	und.
52	242, 319, 320	195, 263, 315.1C	und.

[a] All substitutions are transitions unless otherwise noted.
[b] HVS-I substitutions are listed as the nucleotide position minus 16000.
[c] Undetermined
rCRS means that the HVS-I sequence is identical to the revised Cambridge Reference Sequence (Andrews et al., 1999).

Table 2: Molecular Indices for HVR Data of North Dakota Sample and Eight European Populations

Population	Sample size	Nucleotide diversity	Mean number of pairwise differences
North Dakota	97	0.012694 ± 0.006590	7.984612 ± 3.742457
Czech	177	0.012405 ± 0.006422	7.802867 ± 3.649835
Danish	201	0.012306 ± 0.006370	7.740695 ± 3.621076
French	209	0.010442 ± 0.005482	6.568265 ± 3.116320
GermanNE	297	0.011049 ± 0.005762	6.949524 ± 3.276519
Polish	436	0.012126 ± 0.006268	7.627239 ± 3.564420
RussianS	201	0.012162 ± 0.006302	7.650165 ± 3.582162
Slovak	367	0.011089 ± 0.005778	6.975276 ± 3.285845
Swedish	296	0.012001 ± 0.006215	7.548467 ± 3.533784

Table 3: Molecular Indices for HVS-I Data of North Dakota Sample and 17 European Populations

Population	Sample size	Nucleotide diversity	Mean number of pairwise differences
North Dakota	97	0.013079 ± 0.007154	4.708448 ± 2.325117
British	89	0.012341 ± 0.006805	4.442640 ± 2.211359
Croatian	59	0.011418 ± 0.006398	4.110629 ± 2.076746
Czech	177	0.013704 ± 0.007419	4.933445 ± 2.413421
Danish	201	0.012987 ± 0.007071	4.675279 ± 2.300521
Finnish	200	0.012496 ± 0.006837	4.498726 ± 2.224217
French	209	0.011695 ± 0.006451	4.210188 ± 2.098947
GermanNE	297	0.011374 ± 0.006289	4.094713 ± 2.046534
GermanS	200	0.010455 ± 0.005858	3.763711 ± 1.905709
Icelandic	392	0.013241 ± 0.007185	4.713946 ± 2.312614
Norwegian	314	0.011261 ± 0.006233	4.054005 ± 2.028610
Polish	436	0.013204 ± 0.007156	4.753363 ± 2.329124
RussianM	103	0.011995 ± 0.006628	4.318377 ± 2.154555
RussianS	201	0.013211 ± 0.007179	4.756082 ± 2.335447
Scottish	179	0.012944 ± 0.007055	4.660017 ± 2.295081
Slovak	367	0.011875 ± 0.006524	4.275051 ± 2.123429
Swedish	296	0.012197 ± 0.006682	4.390952 ± 2.174676
SwedishN	32	0.013008 ± 0.007278	4.682993 ± 2.355441

Table 4: AMOVA Results of Genetic Variation Within and Among Grouped Arrangements of North Dakota Subpopulations and European Source Populations

Populations	% variation	*P*-value
European populations		
German vs Scandinavian vs Slavonic		
Among populations	0.16	0.00098 ± 0.00098
Within populations	99.84	
German vs Scandinavian vs Slavonic vs Icelandic[a]		
Among populations	0.34	0.00000 ± 0.00000
Within populations	99.66	
North Dakota subpopulations		
Rural vs Urban		
Among populations	-0.83	0.89932 ± 0.01162
Within populations	100.83	
German vs Scandinavian ancestry		
Among populations	-0.05	0.46823 ± 0.01310
Within populations	100.05	
German vs Scandinavian vs Slavonic vs British Isles ancestry		
Among populations	-0.06	0.49267 ± 0.01875
Within populations	100.06	
Six regional subpopulations		
Among populations	0.75	0.19159 ± 0.01351
Within populations	99.25	

[a] Sample size (n) is as follows for each group: "German" n = 497 (GermanNE, GermanS); "Scandinavian" n = 843 (Danish, Norwegian, Swedish, SwedishN), "Slavonic" n = 1284 (Czech, Polish, RussianM, RussianS, Slovak), and "Icelandic" n = 392.

Figure 1. Map of the United States in which North Dakota's location is shaded, with an enlarged map of North Dakota outlining the geographic divisions used to group residents into one of six regional subpopulations. Modified, with permission, from original National Atlas of the United States image (http://www.nationalatlas.gov).

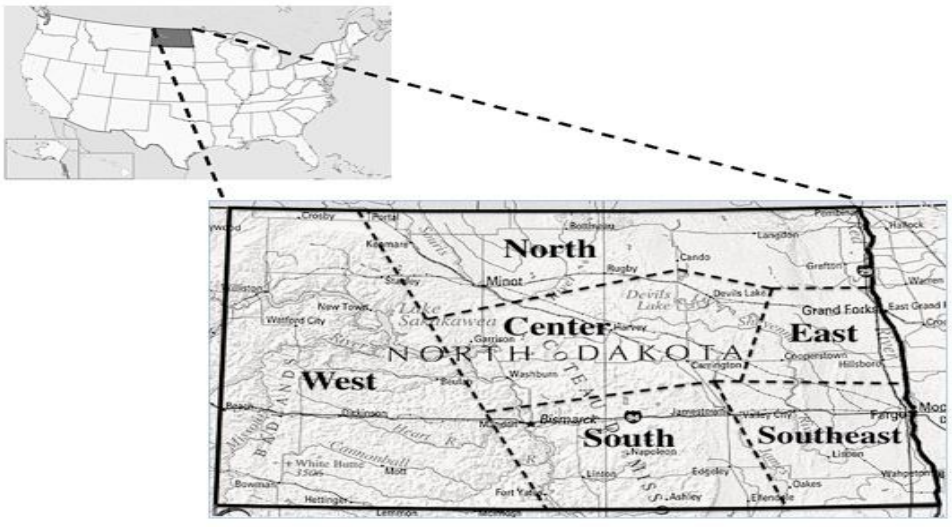

Figure 2. MDS scatterplot of the F_{ST} measurements calculated for HVS-I sequence data of North Dakota sample among 15 European source populations. ("Brit" = Brits, "Cro" = Croatians, "Cze" = Czechs, "Dan" = Danes, "Fin" = Finns, "Fre" = French, "GerNE" = Germans from northeastern Germany, "GerS" = Germans from southern Germany, "Ice" = Icelanders, "ND" = North Dakotans, "Nor" = Norwegians, "Pol" = Poles, "Rus" = Russians, "Scot" = Scots, "Slo" = Slovaks, "Swe" = Swedes).