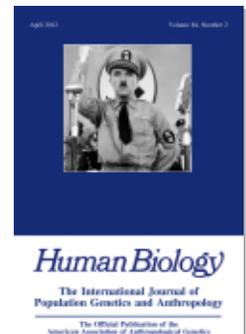# The Family Name as Socio-Cultural Feature and Genetic Metaphor: From Concepts to Methods

Pierre Darlu
Gerrit Bloothooft
Alessio Boattini
Leendert Brouwer

The Family Name as Socio-Cultural Feature and Genetic Metaphor: From Concepts to Methods

PIERRE DARLU,[1]* GERRIT BLOOTHOOFT,[2,3,4] ALESSIO BOATTINI,[5]
LEENDERT BROUWER,[3] MATTHIJS BROUWER,[3] GUY BRUNET,[6] PASCAL CHAREILLE,[7]
JAMES CHESHIRE,[12] RICHARD COATES,[8] KATHRIN DRÄGER,[9] BERTRAND DESJARDINS,[10]
PATRICK HANKS,[8] PAUL LONGLEY,[12] KEES MANDEMAKERS,[4] PABLO MATEOS,[12]
DAVIDE PETTENER,[5] ANTONELLA USELI,[5,11] AND FRANZ MANNI[1]

Abstract A recent workshop entitled ''The Family Name as Socio-Cultural Feature and Genetic Metaphor: From Concepts to Methods'' was held in Paris in December 2010, sponsored by the French National Centre for Scientific Research (CNRS) and by the journal Human Biology. This workshop was intended to foster a debate on questions related to the family names and to compare different multidisciplinary approaches involving geneticists, historians, geographers, sociologists and social anthropologists. This collective paper presents a collection of selected communications.

In 1983, *Human Biology* published a special May issue (volume 55, issue 2) devoted to surnames as tools to evaluate average consanguinity, to assess population isolation and structure, and to estimate the intensity and directionality of migrations. Major contributions written by scholars gave a special relevance to this special issue that remained, for many years, a reference (for a review see Lasker 1985; Colantonio et al. 2011).

Since that time, many surname studies have focused on extending knowledge on population structure, isonymy, and migration (for an exhaustive synthesis see Colantonio et al. 2003) and, to date, surname methodologies have

[1]UMR7206, CNRS, Muséum National d'Histoire Naturelle, Université Paris 7 Paris.
[2]Utrecht University, Utrecht institute of Linguistics.
[3]Meertens Institute KNAW, Amsterdam.
[4]International Institute for Social History KNAW, Amsterdam.
[5]Dipartimento di Biologia E.S., Area di Antropologia, Università di Bologna.
[6]UMR CNRS 5190 – Université Lyon 2.
[7]University of Tours, France, Centre d'Études Supérieures de la Renaissance (CESR).
[8]University of the West of England, Bristol.
[9]Deutsches Seminar, Albert-Ludwigs-Universität, Freiburg im Breisgau.
[10]Département de Démographie, Université de Montréal.
[11]Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari.
[12]Department of Geography/Center for Advanced Spatial Analysis, University College London (UCL).
*Correspondence to: Dr. Pierre Darlu, UMR 7206, National Museum of Natural History; 61, Rue Buffon, 75005 Paris, France. E-mail: darlu@mnhn.fr.

KEY WORDS: SURNAMES, POPULATION GENETICS, DATABASE, MIGRATION, SPACE-TIME EVOLUTION, Y CHROMOSOME, GEOGRAPHICAL PATTERNS, HISTORICAL DEMOGRAPHY.

been applied to about thirty societies all around the world with a geographic scale that ranges from a household or village, to a whole continent. Further and quite recent research put forward a spectrum of methods to analyze Y-chromosome DNA polymorphisms, thus allowing the examination of the degree of co-segregation of family names and Y-chromosome haplotypes, at least in patrilineal naming practice.

The workshop *The Family Name as Socio-Cultural Feature and Genetic Metaphor: From Concepts to Methods* (Paris, France, 5– 6 December, 2010) was organized to go further and, even if some presentations were focused towards more classical research, to pinpoint some particularly innovative aspects of current surname research. This summary article is meant to be a synthesis of the papers presented during the workshop; there are two main strands.

The first research direction relies on the use of surname databases that are increasingly exhaustive and easy to analyse thanks to the spread of digital techniques. In this respect, Pablo Mateos, James Cheshire and Paul Longley's UCL Worldnames database (which includes about 6 million surnames registered in 26 different countries, *http://worldnames.publicprofiler.org/*), constitutes an impressive quantity of information and an exciting tool for future research (Cross-Disciplinary Perspectives in the Spatial Analysis of Names, this article). Unfortunately, this large collection of data comes from different sources, such as national electoral registers or telephone directories, and problems of homogenization and representativeness need to be discussed further as they could not be addressed at the workshop. In the same way, long distance comparisons between stocks of surnames with very different historical and linguistic origins are also a challenge and deserve particular attention. The corpus of family names described by Kathrin Dräger (The German Surname Atlas Project. Computer-Based Surname Geography, this article) is based on the 2005 telephone directory of the Federal Republic of Germany and contains a set of one million different surname types accounting for about thirty million entries. This corpus can be organized according to linguistic properties (lexis, phonology, graphemics, and morphology) and to surname type (derived from place names, professions, nicknames, first names), thus allowing a quite deep exploration of the regional variation of surnames. In this way, from the current distribution of surnames it is sometimes possible to trace back ancient migratory movements. In the same vein, Gerrit Bloothooft et al. (Data Mining in Dutch Historical Civil Registration, 1811-Present, this article) presented a modern set of 16 million family names, collected from the Civil Registration, accounting for the entire Dutch population and including 314,000 different surnames. While their spatial distribution is fully available online, etymological and onomastic information is available for a subset of 100,000 names. To conclude the review of the presentations concerning large contemporary databases, the paper of Richard Coates and Patrick Hanks (The New Family Names of the United Kingdom Project (FaNUK), this article) is different from previous ones as also ancient sources were used. Overall, their *Family Names of the United Kingdom Project* is an assemblage of data coming

historical or recent dictionaries, primary sources of many kinds, and lists of surnames already published in England, Wales, and Scotland. The aim is to reconstruct the etymology of British surnames and to explain their morphological variations through space and time.

The second research direction implies the use of historical data. We note that this approach has a special relevance as, in the frame of accurate descriptions of population dynamics, surnames can be available backwards in time for consecutive generations, while comparable genetic data (Y-chromosome) can not. The research of Gerrit Bloothooft and co-workers we described in the previous section also included information on ancient data as the life cycles of 76,000 persons born between 1811 and 1922 were analysed. Similarly, Guy Brunet, Pierre Darlu and Bernard Desjardins (Writing the History of the Québec Populations Using Surname Frequencies, this article) used an almost exhaustive list of about 400,000 baptisms recorded in Québec from 1600 to 1800; Pascal Chareille (A Long-Term Perspective on Anthroponymic Corpora, this article) studied the surnames reported in the Normandy (France) currency tax rolls between 1383 to 1515 and exploited the household census in Burgundy (France) between 1376 and 1610. Finally Alessio Boattini, Antonella Useli and Davide Pettener (Reconstructing Past Genetic Structures in Recently Transformed Populations: Surnames and Y-Chromosomes in the Upper Savio Valley (Central Appenines, Italy), this article) analyzed the conscription list of individuals born between 1808 and 1987 in the Upper Savio Valley (Italy).

The large increase of the available data, both in time and space, has led to the development of new methods and analytical tools. Among them, and now widely used, are automatic geographic representations of surname diversity, which plot on a map either the variations of frequency of a given name or of a set of names sharing some phonetic or grammatical features (see figure in Bloothooft and Dräger contributions). Some innovative applications to surname data of available statistical methods were also presented, for example a Bayesian approach to infer the origins of migrants (Brunet et al.), Self-Organizing Maps to identify cluster of surnames sharing a same geographical origin (Boattini et al.), or naming network clustering into ethno-cultural groups (Mateos et al.).

Surnames are efficient markers for tracing the movements of people, and therefore most presentations focus on migration. Gerrit Bloothooft compares the distribution of birth places of current inhabitants of a given town and the corresponding distribution for their great-grandfathers. Guy Brunet discusses the origins of migrants who settled in various regions Québec (Canada) between the beginning and the end of the 18th century. Pascal Chareille extracts from the household census (14th century, Burgundy, France) annotations indicating movements of people around the city of Dijon. Finally, Kathrin Dräger, Richard Coates and Patrick Hanks, thanks to their databases providing etymological information on names, were able to localize the most likely geographic origin of given surnames.

In our opinion, the future of surname studies lies more in the rich information provided by the set of data preserved through the generations (one of

the oldest, which include 8,500 names, comes from the 9th century (Chareille and Darlu 2010) and in well-defined communities, than in the accumulation of surnames on a wider geographical scale. Moreover, the large amounts of time- and geo-referenced data that will be gathered will require new statistical methods taking into account the inescapable problems of sampling and lemmatization (the grouping together of related surnames).

In the past, surname studies were popular because they were considered as a cheaper alternative to Y-chromosome typing but, in our opinion, they have a special interest that molecular data do not provide: first they are much more variable and, second, they carry social and economic information that deserve inclusion in many interdisciplinary approaches. Historians, linguists, and geographers, as exemplified by our workshop, can play as active a role as biologists, in surname studies and population analyses. For the future, the trend should be to expand our traditional western-centered field of investigation, in order to study other modes of naming in other countries that have both different cultural traditions and large amounts of data readily available.

The German Surname Atlas Project. Computer-Based Surname Geography [Kathrin Dräger]

German surnames preserve linguistic material which is up to 900 years old, from Middle High German, Middle Low German and Early New High German (Kunze 2004). This enables us to draw conclusions regarding medieval dialectal variations, writing traditions and cultural life by using the current surname distributions.

The high degree of territorial variation of the German surname system is now being made accessible by the German Surname Atlas Project (*Deutscher Familiennamenatlas*; begun 2005), a cooperation between the University of Freiburg and the University of Mainz under the direction of Prof. Dr. Konrad Kunze and Prof. Dr. Damaris Nübling.

The most frequent and impressive examples are selected from the ~1 million different surnames in Germany to address lexical (e.g., *Schröder/Schneider*, both surnames derived from the profession of tailor) as well as phonological (e.g., *Hauser/Häuser/Heuser*, *Walter/Walther*) and morphological (e.g., patronyms such as *Petersen/Peters/Peter*) questions. The database consists of all of the landline telephone connections in the Federal Republic of Germany in the year 2005 as provided by Deutsche Telekom AG. To estimate the number of people who bear a specific name, one multiplies the number of telephone connections by 2.9. In Germany, telephone connections are the only comprehensive database available. They are arranged by postal code districts comprising five digits each.

The atlas will contain two parts: one grammatical and one lexical. The first part, comprising phonology, graphics and morphology, will be published in three volumes: (1) vowels, (2) consonants, (3) morphology and syntax. The second part of the atlas will be divided into three volumes based on the five surname types: (4) provenance and residence names, (5)

profession names and nicknames, (6) patronyms. Volume 1 was published in 2009 (Kunze and Nübling 2009), volume 2 in 2011 (Kunze and Nübling 2011), volumes 3 in 2012 (Kunze and Nübling 2012) and 4 will follow also in 2012. The final two volumes are scheduled for 2015.

Each surname map in the atlas is accompanied by a commentary containing six sections: (i) the topic being illustrated, and why this special case has been chosen. Usually, very frequent names are selected which are preferably etymologically unambiguous; (ii) the quantitative database for the map, with the regular expression applied, the output types, and the frequencies of the different types; (iii) etymological information regarding the names; (iv) further details about the map and auxiliary maps, which contain details from the main map or illustrate the same topic with other examples; (v) historical forms of the names. The *German Surname Atlas* is the first linguistic atlas which takes data from both present and past, reaching as far back as the Middle Ages, into consideration; and (vi) bibliographical references, cross-references and further information; e.g., the frequency and distribution of names in neighboring countries.

The following case studies are taken from volume 4 of the *German Surname Atlas*. With surnames derived from the provenance of recently arrived persons, we can illustrate ancient migratory movements because surnames emerged in a time characterized by a large degree of migration within the country.

The example of *Westphal*, which is concentrated in Schleswig-Holstein and Mecklenburg-Vorpommern (see Figure 1), illustrates the migration of Westphalian settlers in the context of the German eastward expansion (*mittelalterliche deutsche Ostsiedlung*) of the 9th to the 14th centuries, in which Germans from modern-day western and central Germany settled less-populated regions of eastern Central and Eastern Europe, formerly inhabited mostly by Slavic and Baltic peoples. As this example shows, Westphalian settlers must have participated in the German eastward expansion to a major extent. This is supported by both historical evidence showing that a large part of the population in today's Mecklenburg-Vorpommern has its roots in the western low German area, as well as by linguistic similarities between dialects in Westphalia and in Mecklenburg-Vorpommern (Schmuck 2009).

Surnames such as *Unger* and *Hunger,* which refer to *Hungary,* are concentrated in Saxony and in the eastern part of Thuringia. The surnames *Böhm* and *Böhme* agglomerate not only in Saxony and Thuringia, but also in northern Bavaria, so that the latter surnames can be found in a curve around Bohemia in today's Czech Republic. According to Walther (1993, p. 498), the surnames *Unger*, *Hunger*, as well as *Böhm* and *Böhme* reflect the fact that Saxonian miners often moved to Bohemian and Hungarian mining sites. After they returned home, they were named after their former places of work.

Figure 2 shows the distribution of the name *Schweizer*. The varieties with *z* exist mainly in Baden-Württemberg, while those with *tz* are attached in the north, mainly in Rheinland-Pfalz and Hessen. These surnames also appear in France in

Figure 1. Relative distribution of the surname *Westphal*.

about 3,500 births between 1891 and 1990 (www.notrefamille.com, 28.09.2011), as well as in Switzerland, with about 4,500 telephone connections (www.verwandt.ch, 28.09.2011). The reason why *Schweizer* and its variants appear quite often in Switzerland itself is that during the time when surnames arose, *Schweizer* originally referred to the village Schwyz and the surrounding canton. The name of the village and canton *Schwyz* was applied to the entire Swiss confederation only from the 14th century on. Diphthongization led to the standard German form *Schweiz*. Mainly after the Thirty Years War, many people from the village and canton of Schwyz and from the whole Swiss confederation settled in today's southwestern Germany.
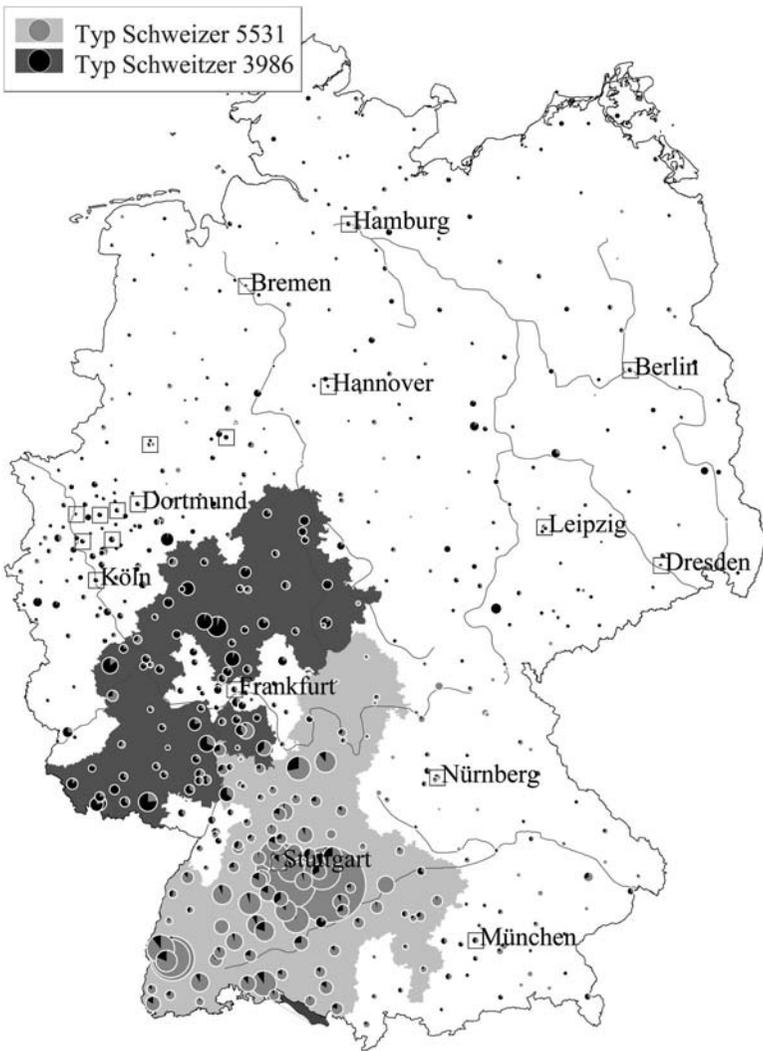
Figure 2. Relative distribution of surname types *Schweizer* and *Schweitzer*.

Figure 3 gathers surnames which refer to the names of the low mountain ranges Westerwald, Odenwald and the region of Bergstraße, which is part of the Odenwald. The surnames which trace back to the toponym *Westerwald* are located around the corresponding low mountain range: *Westerwald* is concentrated around Frankfurt, *Westerweller*, with assimilation of *ld* to *ll*, in the northeast of Frankfurt and the eastern part of the Ruhrgebiet, while *Westerwelle* is found in the area of Bielefeld and in the eastern part of the Ruhrgebiet. The surnames which trace back to the toponym *Odenwald* (*Odenwald*, *Odenwälder*, *Odenweller*, *Odenwäller*, *Ottenwälder*, *Ottenweller*) are located in southern Hessen, northwestern Bavaria
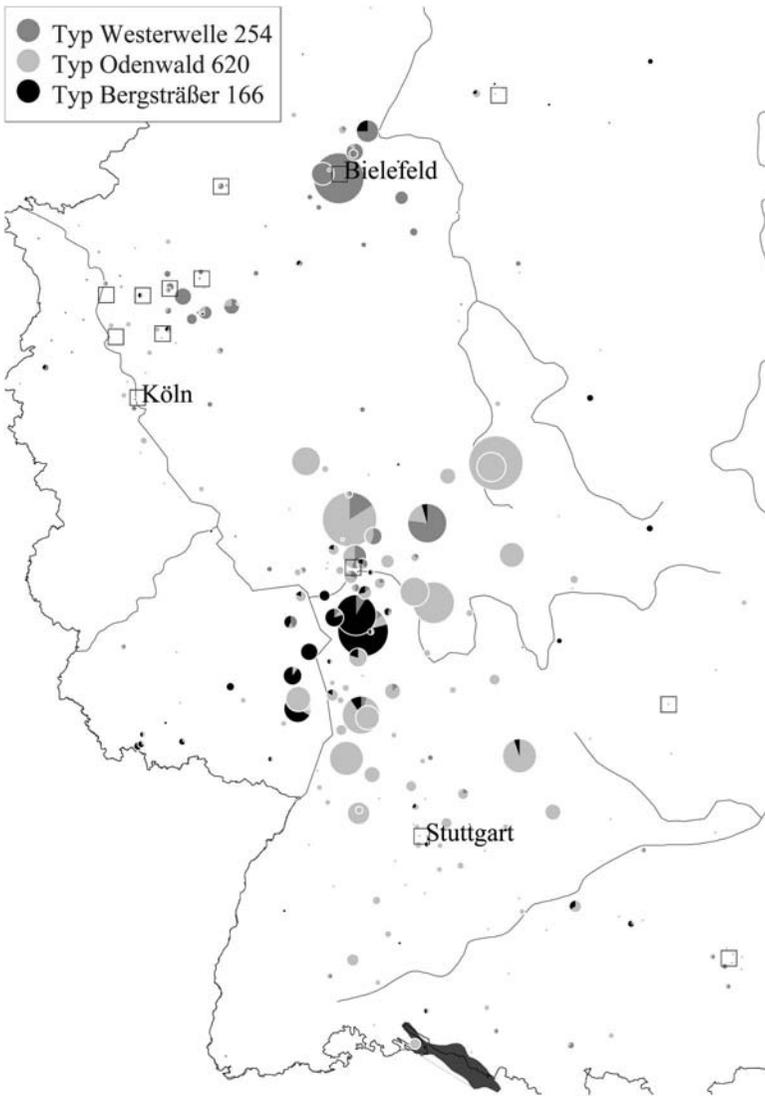
Figure 3. Absolute distribution of surname types *Westerwelle*, *Odenwald*, and ***Bergsträßer*** in Western and Southwestern Germany.

and northern Baden-Württemberg. Right in the middle, around the homonymous region, *Bergsträßer* and *Bergsträsser* are to be found.

In the Middle Ages, German towns flourished and attracted rural populations, and the newcomers were often named after their places of origin. So with the surnames derived from the provenance of recently arrived persons which relate to single settlements, we can reconstruct where the migrants came from and where they settled down.

Onomasticians such as Grünert (1958, pp. 537–553, map 1–9), Hellfritzsch (2007, pp. 525–539, maps 1–4), Neumann (1970, pp. 182–187, map 2), and Neumann (1981, pp. 276–283, maps 1–4) collected historical documents regarding surnames related to single settlements and then mapped them. Thus they found out that the medieval catchment areas of smaller towns had a radius of barely 100 kilometers.

Conversely, the distribution of surnames can also illustrate where former citizens of a certain town or village moved, because newcomers were often named after their place of origin. In many cases, most persons who bear a specific name based on a small town or village still live within a radius of about 50 kilometers around the eponymous settlement (cf. the contribution of Pascal Chareille in this article). Figure 4 illustrates this with the example of the surname *Rothenbucher*, with umlaut *Rothenbücher*. Here, the ancestor was named after the small village of Rothenbuch in the Spessart.

In addition to the Middle Ages and the early modern period, the database of the *German Surname Atlas* also opens up possibilities to reconstruct migratory movements during the 20th century because it contains not only German but also foreign surnames. This provides a broad field of research in which linguists, historians, human geographers and geneticists can collaborate.

Data Mining in the Dutch (Historical) Civil Registration (1811–Present) [Gerrit Bloothooft, Kees Mandemakers, Leendert Brouwer, and Matthijs Brouwer]

Names identify individual persons. As such, names are central in research dealing with individuals, and groups defined by properties of these individuals – such as families. In the latter, generations also come into play, carrying the dimension of time and historical developments in society. The spatial dimension also influences groups: members migrate and interact. For studies of subjects including genetics, health, demography and sociology, the identification of groups and knowledge of their dispersion in time and space are valuable if not essential information.

In Dutch and other modern civil registrations, people are identified not only by name but also by a persistent ID. By having the parents' IDs in the record of every individual, and a complete and accurate digital registration, all family relations in society are basically known, at least for a couple of generations. In these systems, names are no longer essential to demonstrate relations between people. However, for older registrations, no IDs were used, and reconstructions of relations between people depend strongly on their names and the description of relationships in certificates of birth, marriage and decease. Accuracy of these archives is often problematic, completeness is rare, and full digitization is a long-term goal only.

Available Data and Major Ongoing Projects in The Netherlands.

Modern Civil Registration.   In 2000, a new law on the Civil Registration (CR) opened the possibility to acquire data for scientific research. This opportunity

Figure 4. Absolute distribution of *Rothenbucher* and *Rothenbücher* in Northern Bavaria.

was used by Utrecht University and the Meertens Institute to request two selections of data; one centered around first names and the other around family names. Full-population data were acquired for all first names of 21 million persons (5 million deceased). As well as all first names, the (internal) ID, the first names and IDs of the parents, and the date, place and country of birth of all individuals, were provided. This constitutes a full-population genealogy for several generations—but with only the first name known. The data describe the full-population born after 1930. They become gradually less complete for earlier

years of birth but still provide a 30% sample of all persons born in 1880. All in all, these data entailed 500,000 unique first names which were made public in June 2010 on *www.meertens.knaw.nl/nvb*. For the family names, full-population data were acquired for the 16 million persons alive in 2007 with information about the following attributes: the family name, date, place and country of birth, and the current place of residence (compare Cheshire et al. 2011; Dräger, this article; Coates and Hanks, this article). These data were linked to data from the 1947 census. The 16-million persons carried 314,000 unique surnames. The website presenting the surnames was launched in December 2009 on *www.meertens.knaw.nl/nfb*.

Historic Civil Registration.   Hundreds of volunteers are digitizing historical registers of birth, marriage and decease from the civil-registration system that started in 1811, based on Napoleonic law. Currently about half of the job is done. There are now over 16-million registers digitized, containing information on about 70-million (not unique) persons (see *www.genlias.nl*). Automatic reconstruction of families from these data is now in progress in the LINKS project (*Link*ing *s*ystem for historical family reconstruction). Ideally, the goal of LINKS is to identify all individuals mentioned in the certificates uniquely, and, just like the modern CR, to tag them with a persistent ID and the IDs of their parents. It is possible to link this historical "population registration" with the modern one, provided privacy concerns do not prevent this.

Historical Sample of The Netherlands.   The Historical Sample of The Netherlands is a project that started in 1991, with the aim to reconstruct life cycles for an unbiased random sample of an eventual 78,000 persons (born 1812–1922) sampled manually from birth certificates. In addition to standard personal data, religious affiliation, occupation, household composition, literacy, social network, and migration history are also collected from the civil certificates and population registers (Mandemakers 2000). More information can be found on *www.iisg.nl/hsn*.

Data Mining, Considerations, Tools and Examples.

Geographic Spread.   Current geographic spread of a family name can be shown immediately on the website of the Dutch Family Name Corpus at the municipality level. By providing an online possibility to search by *regular expression*, properties of all kinds of *sets* of surnames can be shown as well—see the example in Figure 5. These properties may include all kinds of spelling variation, or may require the presence of certain morphemic properties which may be typical for some language or dialect. The same options exist for the first-names website.

Migration.   A complete (historical) civil registration would allow for migration studies by tracing the places of births of subsequent generations. On the basis of
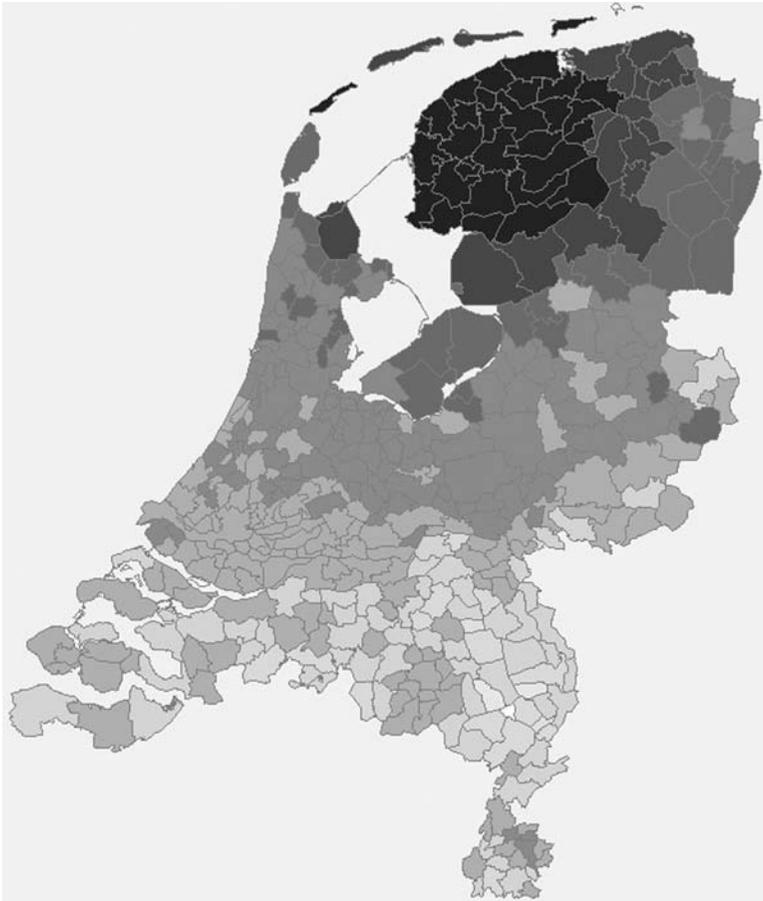
Figure 5. Geographic distribution in the Netherlands of all surnames that fulfil the regular expression "stra$," implying 483 names ending with –stra, in percentage per municipality. This is a typical Frisian name ending, expressing "coming from". The map shows the province of Friesland (the darker area) with more than 5% -stra names, the circular shape of the decrease of the presence of the name in the North, a relative sharp boundary with the Catholic south of the country — with exceptions in areas of industrial development (in the coal mines of Limburg, around Eindhoven (Philips company) and the textile factories in the eastern part). The 10 gray shades follow a logarithmic scale from over 5% (dark) to less than 0.01% (light).

our first-name corpus from the modern civil registration, we identified grandparents and their grandchildren, and computed the distance between their places of residence in 2006 (Figure 6). When the grandchildren are young they live with their parents at an average distance of a stable 22.5 km. Between the ages of 20 and 30, the grandchildren settle themselves, and the average distance increases to 34 km, which remains stable again in further life. Distances do not sum over generations since on average grandchildren randomly move in all directions.
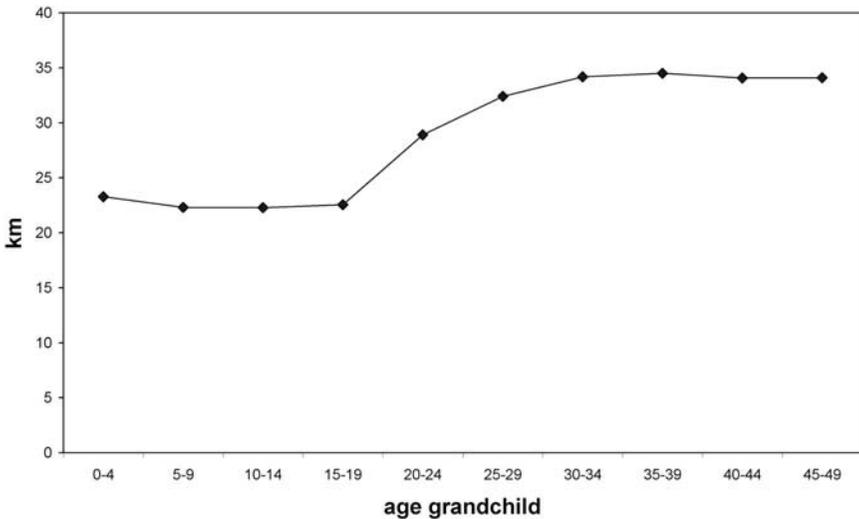
Figure 6. Distance between places of living of grandparents and their grandchildren in 2006.

Another analysis of geo-distributional nature, and related to migration, can be done for surnames. Given a limited migration some surnames may still be found in the region where the ancestor adopted the name, often many centuries ago. We determined for which surnames 50% of the bearers nowadays live within 30 km of a center municipality. Subsequently, we computed per municipality the percentage of the population with such a "regional" name. Results are shown in Figure 7. Rural areas and closed communities such as fishing villages can have up to 43% of the population with a regional name—and a high percentage of consanguinity. Larger towns and newly reclaimed polders are a melting pot of families and obviously have much lower percentages (Bloothooft 2011).

Co-Variation.　An important property of the data in the civil registration (and reconstructed life courses) is that on the basis of known family relations, studies within families and across generations can be performed, thus informing on the social strata of the population. We explored this in a study of modern first names. The assumption was that parents do not choose names for their children at random, but (largely unconsciously) on the basis of what is fashionable or expected in their social environment. This would imply that the names of children in the same family convey some of this fashion. Traditional parents may name their children with old Dutch names like *Willem* and *Dirk*, and this combination of names will appear in such families more frequently than can be expected on the basis of individual probabilities of the names. By analyzing the names of millions of children in families with more than one child, we could cluster the names in such a way that names within a cluster have a higher probability to be found in a single family than across clusters (Bloothooft and de Groot 2008). For modern naming, fifteen clusters or name groups gave a fair
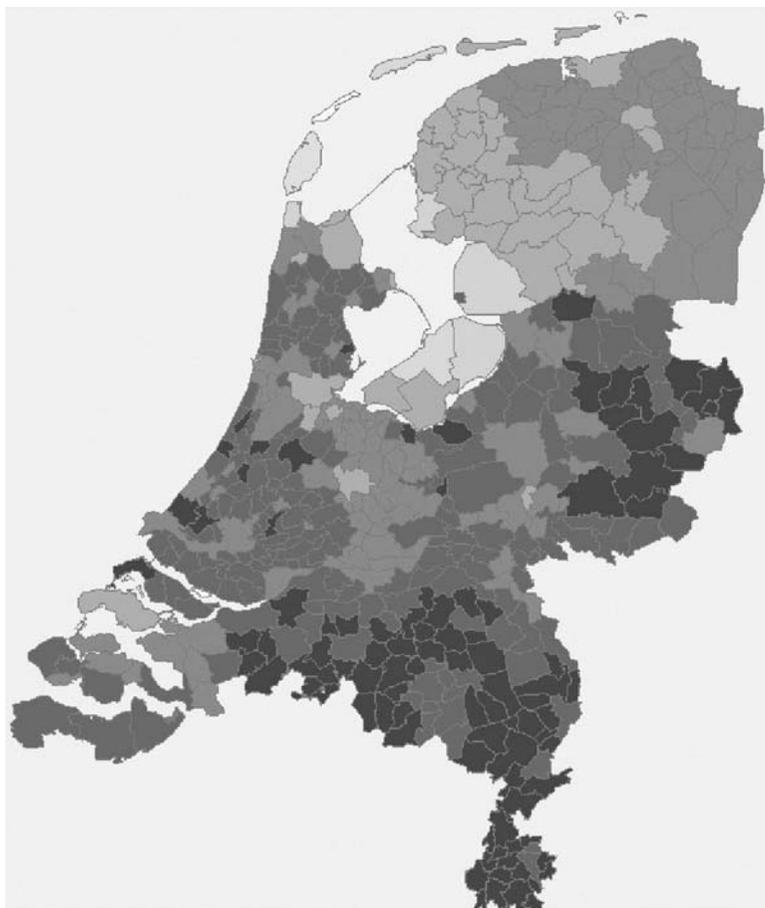
Figure 7. Density of regional surnames in The Netherlands. The five gray-shades indicate 1–2% (light), 2–5%, 5–10%, 10–20%, and 20–50% (dark) of inhabitants in a municipality with a regional name.

description of the 1,409 most frequent names (naming 75% of all children). These are (1) traditional Latinized names [*Johannes, Maria*]; (2) Dutch traditional names *Trijntje*]; (3) Hebrew names [*David, Esther*]; (4) Frisian names [*Jelle, Nienke*]; (5) longer premodern Dutch names (popular before 1990) [*Wouter, Suzanne*]; (6) short international names (popular before 2000) [*Mark, Laura*]; (7) English names [*Kevin, Samantha*]; (8) short modern Dutch names [*Tim, Anne*]; (9) other modern names [*Milan, Lara*]; (10) Nordic and French names [*Niels, Anouk*]; (11) elite names [*Floris, Amber*]; (12) French names [*Jules, Dominique*]; (13) Italian and Spanish names [*Lorenzo, Felicia*]; (14) Arabic names [*Mohamed, Samira*]; and (15) Turkish names [*Hakan, Meryem*].

The geographic spread of each name group has significant features across the country, as shown in Figure 8 for traditional Dutch names, which mainly
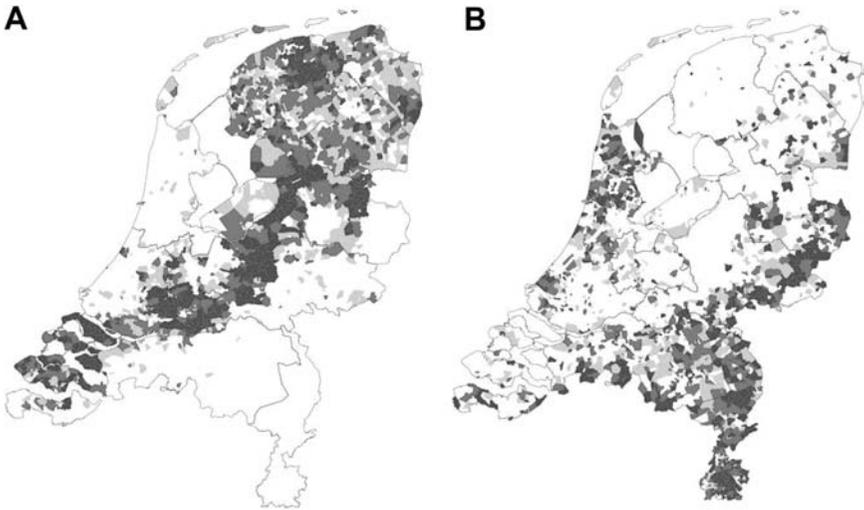
Figure 8. Geographic spread of Dutch traditional first names (A) and short English names (B).

follow the Dutch Bible Belt—a narrow region of conservative Protestantism from the southwest to the middle of the country—and ends more widely distributed in the Northern provinces, while short English names are preferred in the areas of Catholic dominance, which earlier chose traditional Latinized names.

In a subsequent study, we had available diverse socio-economic data from about 281,751 households, including the names of the children in the households. This allowed us to investigate the relation between socio-economic parameters, such as educational level and income of the parents, and the name groups. We also had lifestyle profiles of the households (summarizing all data), and could map the name groups on major lifestyle dimensions related to them (Bloothooft and Onland 2011). Results are shown in
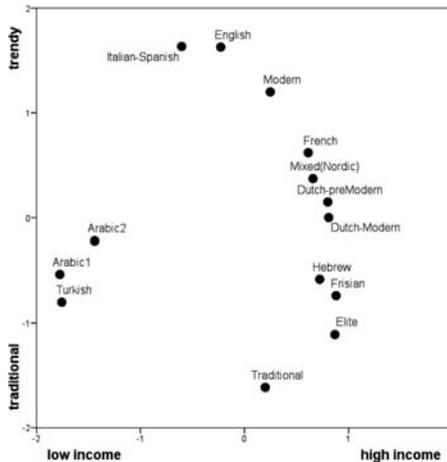

Figure 9. Name groups and lifestyle dimensions.

Figure 9, with the horizontal axis related to household income or highest education of the parents (low-high), and the vertical axis related to affinity to tradition versus fashion. Major features are the tendency for well-educated and somewhat traditional parents to choose Dutch, Hebrew or Frisian names, while the medium-educated and trendy parents favor foreign or fancy modern names.

This type of analysis could be done for surnames as well on the basis of known family relations and data from sources external to the civil registration, such as family income, education level, occupation, or ethnicity. This would underpin relationships between surnames and cultural, ethnic, and linguistic (CEL) parameters (Mateos et al. 2007).

The New Family Names of the United Kingdom Project (FaNUK) [Richard Coates and Patrick Hanks]

The major new research project called Family Names of the United Kingdom (FaNUK) began on April 1, 2010, and will run for four years, based at the Bristol Centre for Linguistics in the University of the West of England, Bristol. It receives funding from the Arts and Humanities Research Council, and has an attached doctoral studentship. Some 5,000 UK family names have no accepted etymological explanation; many others have been wrongly explained. FaNUK's goal is to make good these deficiencies through the creation of a database of family names containing an evidence-based account of the linguistic and geographical origins, history, and demography of at least the 43,000 most frequent extant names.

Research Context. Public interest in the origins, history, and demography of family names is attested by the vast amount of amateur work and media interest in genealogy. This is poorly served by existing literature, not radically improved since work done in the 1950s (Reaney 1958, last revised in 1991). Many seemingly plausible earlier explanations are incompatible with new facts about name history and geographical distribution. Misperceptions have arisen because county-based research by medievalists lacks a national framework. Reliable new resources are needed which are accessible to an increasingly sophisticated public.

Family-name research is interdisciplinary. New resources from history, family history, place-name study, official statistics, and genetics include collections and editions of medieval evidence, machine-readable census data, and new statistical methods for correlating family names and locations (cf. the contribution to this article by Pascal Chareille). Geneticists have begun working with local historians on the relationship between the distribution of individual family names and their origin. Such work needs bringing together, allowing existing accounts of family-name origins and history to be evaluated, corrected, and supplemented, and allowing a satisfactory multidisciplinary framework to be created. FaNUK will emphasize family names as linguistic and historical entities, rather than focus on genealogy and family history. But it will systematically take

account of the work of genealogists and family historians—especially the Guild of One-Name Studies (http://www.one-name.org/) "128" to ensure maximum credibility for a resource of which they represent the major likely consumers.

Although there is reliable smaller-scale work exists (e.g., the best one-name studies, and surveys of seven counties dealing with medieval family names), no current resource brings together medieval evidence for comparison with distributional evidence derived from modern online geodemographic tools. FaNUK prepares the ground for detailed genealogical work which will eventually secure the connections across time. When all this material is brought together, critical assessment of previous etymological and historical claims about names and their alleged continuity will be possible, new patterns in their historical demography will appear, and new etymologies for problematic names will be facilitated through direct comparison of the data set. Research on this scale is entirely new in the UK. The proposed product will be by far the most wide-ranging, complete, and reliable source of relevant information. No competing online resource exists, and FaNUK will counterbalance much misinformation on amateur web-sites (often taken from existing literature).

The standard work on **English** surnames is Reaney (1958, and last revised 1991; R&W). Its defects are now apparent. For example, comparison with 1881 census data reveals no entry for common names such as *Alderson* (northern England), *Blair* (Scotland), and *Critchley* (Lancashire) and over 20,000 other family names with more than 100 modern bearers. Being essentially a dictory of medieval surnames without declaring this in the title, it includes over 3000 defunct surnames e.g., some derived from obsolete nicknames such as *Ballox*, *Barebone*, *Beardless*, etc. It takes little account of geographical distribution or local sources, explaining *Broadhead* as a nickname and *Gawkrodger* as "awkward Roger"; both are in fact from minor place-names. Reaney's links between medieval evidence and modern surnames are often demonstrably untenable, and some other etymologies are unreliable or misleading. Other previous English-oriented works include: Cottle (1967, 1978, 2009), and the nine counties of the English Surnames Series (ESS), based on McKinley's discontinued program at Leicester University. A major critique of Reaney's methodology is Redmonds (2002). He and Hey (2000) have shown the need to integrate the study of family history with local history. Hanks and Hodges (1988; H&H), like its successor Hanks (2003; DAFN), is a general resource containing much material relevant to the UK and foreshadowing FaNUK in that its data set has a broad ethnic and etymological scope, but the etymologies mostly lack medieval evidence.

Despite our reservations about these predecessors, they are usable as a foundation for FaNUK. They offer systematic hypotheses for confirmation or correction, in the light of new evidence. We are therefore grateful to the publishers and copyright owners who have made the material in R&W, H&H, and DAFN available to FaNUK in electronic form.

The best resource for **Welsh** surnames is Morgan and Morgan (1985). However, the headwords are Welsh personal-name forms, not surnames. References are regularly to undated secondary sources, not to dated primary documents. It is therefore not user-friendly for a non-Welsh-speaking public, and potentially misleading for unwary users. For **Scottish** surnames, the standard work is Black (1946), a fine collection of historical data where, as with R&W, names are selected from pre-modern evidence rather than a modern inventory, and the etymologies need systematic revision. The main **Irish** resources (de Woulfe 1906; MacLysaght 1985) are based on old work, though we now have de Bhulbh (1997). Both H&H and DAFN include reliable etymological information on Irish surnames, but none of these works provides evidence for early bearers of Irish names. Such evidence exists, for example, in the Tudor Fiants (Nicholls 1994), authorizations to the Court of Chancery in Ireland for the issue of letters patent under the Great Seal of English monarchs in the 16th and 17th centuries, which show surnames in transition from their Irish to their anglicized forms. FaNUK will include, for each Irish family name, evidence from such sources. While the Republic of Ireland is not part of the UK, we cannot omit Irish names, both because of the mass Irish immigration into Britain, and because the northeastern six counties of Ireland still form part of the UK.

On the basis of such previous work, FaNUK prepares the ground for a history of family names *in* the UK. Most academic effort will be directed at names of insular origin. However, the UK's multiethnic character will be addressed by including most **immigrant** names (principally **Huguenot** and **Jewish**, and those more recent arrivals having up to 100 current bearers), making FaNUK's range unique. The focus will be on (a) linguistic source (culturally important to those with foreign genealogy), (b) cultural and religious associations, and (c) how and when each name reached the UK, rather than its entire remote history elsewhere. For well-represented cultures, this will lead to projects beyond the end of FaNUK.

UK surname research lags far behind that in many other European countries. In The Netherlands, two institutions are building large surname databases: Meertens Instituut in Amsterdam (www.meertens.knaw.nl/nfb) and the Central Bureau of Genealogy in The Hague (www.cbg.nl). In Poland, scholars at Pracownia Antroponimiczna (Anthroponymic Research Group, www.ijp-pan.krakow.pl/en/struktura-organizacyjna/zaklad-onomastyki/), Kraków, are researching a comprehensive historical dictionary of Polish surnames whose first volume appeared in 2007. Current UK family-name research also compares unfavorably with funded allied areas like English and Scottish place-names. FaNUK seeks to redress the balance.

Research Methods and Project Outcomes.   We intend to address the research lacunae mentioned above by creating a database using data from the range of sources provided by copyright owners and consultants, gathered by the investigators, and screened, explained, and commented on by the investigators in conjunction with consultants. Machine-readable versions of R&W and H&H

were successfully loaded into an experimental prototype database with the active collaboration of the publishers. Before the project began, we audited the availability of other reference sources for possible addition to the database. We also have lists of relevant historical resources containing many individuals' surnames, and where such resources exist in e-form, permission is sought for electronic links between these data and the FaNUK database. Where they are not yet available, we are actively exploring with project leaders and copyright owners the potential for digitization to our mutual benefit. As a last resort, FaNUK mines documents conventionally.

The database will also establish the inventory of surnames in the post-1880 UK, accompanied by their geographical distribution and frequency. Surname distributions have been derived computationally by current collaborators from electoral rolls and the 1881 census, both now publicly available online.

FaNUK requires many consultants with various specialisms, philological and computational; we do not have space to mention them all here. As the project has progressed, we have benefited considerably from the cooperation of Steven Archer, who has created mappings of the frequency and geographical distribution of surnames recorded in the 1881 national census. A surname whose association with a particular locality is statistically significant may in many cases have originated there, and this possibility needs to be exhaustively investigated before other possibilities are considered. We say this with confidence, because although people can move around, ample evidence exists that a large number of surnames still cluster around a point of origin. Because of this phenomenon, we have been able to resolve some issues about the original distribution and source of some surnames deriving from place-names which are recorded from medieval times but wrongly explained in R&W, (e.g., that the surname *Harmison* originates in Hermiston in Roxburghshire rather than in Harmston in Lincolnshire). Place-names are comparatively stable, both linguistically and geographically. Surnames are not. Families and individual bearers move around; competing spellings are commonplace; people adopt other surnames; surnames are not necessarily transmitted as counterparts of the Y-chromosome; and surnames die out. Archer's work (2003, 2011), has confirmed the essential correctness of H. P. Guppy's hypothesis of a significant relation between many surnames and locations, though many such relations remain unexplained. The association between *Fazackerley* and Lancashire is obvious because a place exists in Lancashire called *Fazakerley*; no place anywhere else has this name, or with a name remotely like it. Elsewhere associations exists between variants of names and particular places, as with *Pardoe* and *Pardey*, which share a linguistic origin, but have no known genealogical connection or shared source; one may be waiting to be discovered through statistical work on distributions.

Summary Review of Targets and Plans for Dissemination.   FaNUK's primary target is to create reliable explanations for the approximately 43,000 long-established or traditional insular surnames in the UK with more than 100

current bearers. A secondary target is to add explanations for unproblematic names of lower frequency. A tertiary target is to add entries for about 3,000 names of recent immigrant origin, indicating where they came from, what (if anything) is known about their meaning, and giving information relevant to their UK status, such as date of arrival. Data from recent electoral rolls and censuses show that over 370,000 different surnames are present in Britain today, but the vast majority of them are extremely rare, being borne by only a handful of people. Surprisingly, over 300,000 are the names of recent immigrants from a vast number of countries including, but by no means restricted to, the countries of the former British Empire. That leaves the 43,000 surnames referred to above.

The principal output of FaNUK, its publicly accessible database, will be valuable to genealogists, geneticists, local historians, historical demographers, historians of the English and Celtic languages, other philologists, and place-name scholars.

Writing the History of the Québec Populations Using Surname Frequencies [Guy Brunet, Pierre Darlu, and Bernard Desjardins]

The study of the geographical distributions of surnames obtained from various registers has already demonstrated its efficiency to infer migration of people, either by applying statistical models when surnames are recorded only once at a given time, using Fst statistics or probabilistic models (Karlin and McGregor 1967; Yasuda et al. 1974; Zei et al. 1983), or by comparing surname frequencies recorded at least twice at the same location (Wijsman et al. 1984; Degioanni and Darlu 2001; Darlu et al. 2011). This second strategy has been less frequently used because it requires historical records. These are now more abundantly available, thanks to the efforts of historians, as exemplified by the present article showing original analysis of migration in Québec.

The arrival of French immigrants in Québec during the 17th century was the starting point for the growth of the French Canadian population, which increased from 18,000 inhabitants in 1700 to 200,000 in 1800 with a corresponding geographic dispersal. On their arrival, the pioneers colonized a strip of land along the Saint-Laurent River expanding first from the two main poles of settlement (Montréal and Québec). During the 18th century, northern and southern parts of the river were progressively occupied, as well as the places between Montréal and Québec.

From the very beginning, baptisms, marriages, and deaths were systematically recorded in parish registers, allowing the reconstruction of the temporal and spatial evolution of the European population. Data on the native Americans were insufficient to allow a similar analysis. The onomastic information drawn from these records was analyzed to infer the demographic growth of this population, its renewal, migration, and geographic expansion.

The present work is based on 392,998 baptism records noted between 1608 and 1799. For each of them, corresponding to a baptized child, the

surname, the birth date, and the birth place (in term of parish and county) were noted. Although the question of lemmatization of the surname variants is far less difficult in Québec than in the situation described by Chareille in the case of the 14th and 15th century documentations (Chareille, this volume), surnames had to be first standardized to allow for orthographic variations. Then their frequency was studied by parish and County for four successive periods of time: P1: 1700–1724; P2:1725–1749; P3:1750–1774; P4:1775–1799.

Global Dynamic of the Population.   The set of surnames, already largely diversified before 1700 (1349 surnames) was relatively stable in the first part of the 18th century, because of the reduced number of immigrants. The number of baptisms increased fourfold between the first (P1) and the fourth (P4) period. The proportion of surname per baptisms (S/N) was rather high before 1725, and progressively decreased during the rest of the century, indicating that there were new arrivals of migrants with new surnames. This is also stated by the evolution of S′ and S″ (See Table 1). Indeed, the number of new surnames arriving at the end of the period (P4, S = 4266) was four times higher than those arriving during the previous period (P3, S = 923). The turnover between the surnames disappearing (S″) and those arriving (S′) leads to a positive although weak balance of 239 surnames in P2, larger in P3 (1,947), and in P4 (2,679). The burst of growth occurred in the middle of the century, with the arrival of many surnames superimposed upon the maintenance of a core of surnames brought by the first settlers. The proportions of singletons (name occurring only once) confirm this point.

The two main towns (Québec, Montréal) show a larger diversity of surnames than the parishes or the counties, obviously following a linear relation with the population, as shown in Figure 10. However, one can also show that Montréal and Québec display an excess of surnames compared to the other places. This excess is larger for the P4 than for the P1 period, meaning that the immigrants are preferentially arriving in these two largest towns, particularly at the end of the century. Actually, the proportion of singletons is respectively 50% and 49% in the parishes of Montréal and Québec (and the weight of the three most frequent surnames is 3.6% and 2.6%) whereas the proportion of singletons is only 29% (and the three most frequent surnames account for 6%) in a typical parish like Saint-Eustache, where 3,500 baptisms were recorded.

Such a contrast between large and small populations has long been reported (Zei et al., 1983). The largest towns attract first the immigrants that have heterogeneous origins and consequently have a larger diversity of surnames.

Surname Resemblance, Tree Representation, and its Geographic Projection.   To specify the geographic structure of the surname distributions in Québec, we calculated the pairwise surname distances between counties, using

Table 1. Distribution of the Numbers of Baptisms (N), Surnames (S) and of Their Ratio According to Periods. H/S is the Proportion of Singletons (Hapax), S' is the Number of New Surname Arriving at a Given Period and Still Found in All Next Periods, S'' is the Number of Surnames Already Present or Arriving at a Given Period and Disappearing at the Next Periods

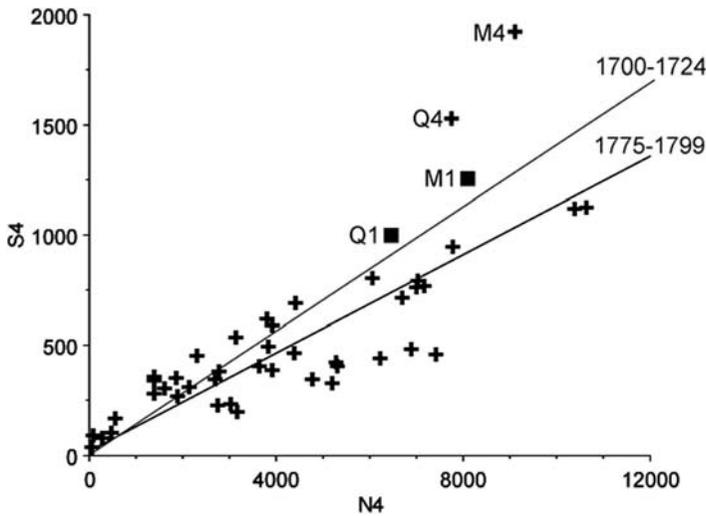|	Number N of Baptisms	Number S of Surnames	Proportion (%) of Surnames Among the Baptisms	Proportion (H/S %) of Singletons Among the Surnames	Number S' of Newly Arriving Surnames	Number S'' of Surnames Disappearing Next Period
Before 1700	41759	1349				
1700–1724 P1	44857	2709	6.0	8.7	1704	544
1725–1749 P2	56246	2768	4.9	1.9	411	348
1750–1774 P3	107919	4798	4.4	13.4	923	1587
1775–1800 P4	183961	7571	4.1	19.2	4266	

Figure 10. Regression of the number of surnames (S4) on the number of baptisms (N4) observed in 43 Counties for the period 1775–1799 (P4). The line of regression for the period 1700–1724 periods (P1), is also drawn for comparison, and is identical for the 1725–1749 (P2) and 1750–1774 (P3). Montréal and Québec are plotted for the P1 and P4 periods (M1, M4, and Q1,Q4 respectively), to show the larger than expected increase of the number of surnames between these two periods of time (P4 versus P1) whereas the trend is stable or even inverted for the other towns.

the classical Nei's distance, as used first by Chen and Cavalli-Sforza (1983). The idea is that two counties sharing close surname frequencies were exchanging people in the past more intensively than two counties that show a large surname distance.

Once the surname-distance matrix was obtained, trees were constructed by the neighbor-joining method (Saitou and Nei 1987), with bootstrap resampling (Felsenstein, 1985) to estimate robustness at nodes of the tree. The consensus tree can be projected on a geographic map, connecting surfaces being clustered together with a given level of bootstrap proportion.

Figure 11 shows that the surname resemblances are clearly high between neighboring Counties, which can exchange individuals readily, and an absence of noticeable division between the two banks of the Saint-Laurent River both near the Montréal and the Québec counties. Moreover, no significant structure distinguishes the area around Montréal from that around Québec. In fact, few strong structures exists except those plotted in Figure 11, suggesting that the dispersion of people (and surnames) was already well-advanced on a large scale at the beginning of the 18th century.

Probability of Geographic Origin (pgo): A Bayesian Approach. Since migration of people involves migration of their surnames (or at least the surnames of their children quoted in the birth registers), the movement of
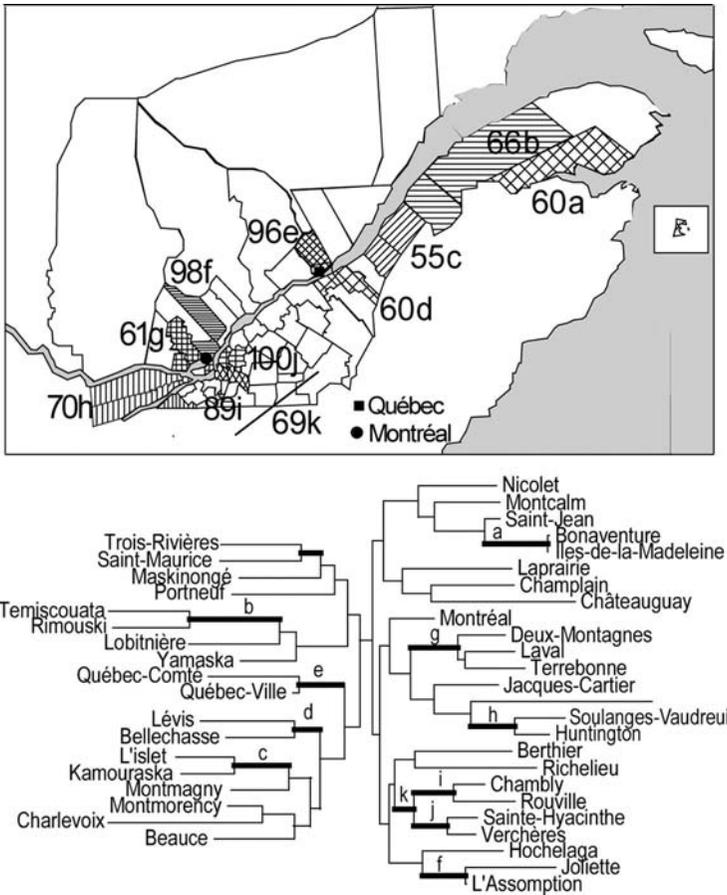
Figure 11. Projection of the clusters defined by bootstrap proportion larger than 55% in the unrooted tree reconstructed by Neighbor-Joining from the Nei's pairwise surname distances between the 43 Counties (P3 and P4 pooled). Numbers in the map are the bootstrap proportions (%) attached to the branches labeled with the corresponding italic letters.

people—usually the males because surnames are paternally transmitted—can be reasonably inferred from the movements of their surnames, although with some limitations (Darlu and Degioanni, 2007, Darlu et al. 2011; Degioanni et al. 2001). A Bayesian approach can be applied, as detailed elsewhere (Chareille and Darlu 2011; Darlu and Degioanni 2007; Degioanni and Darlu 2011).

For the area under investigation (here a county), called the "recipient area," the probability that the surname s_k which is present at time $t+1$ and absent at time t originated from another area, a_i called the "source area" i, is, according to Bayes' Theorem:
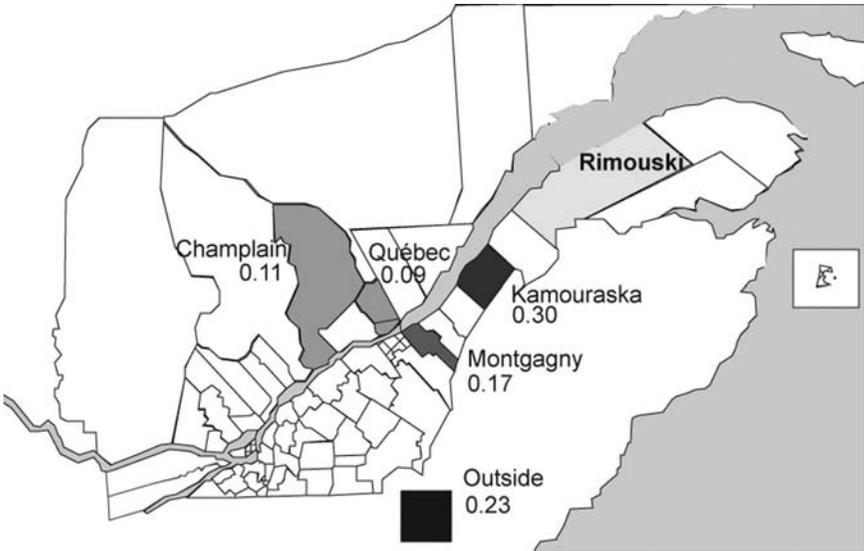
Figure 12. Probabilities of geographic origin of migrants newly arriving at P4 (1775–1799) in the Rimouski county from other Counties of the Province of Québec, or from elsewhere (Outside) (e.g., 30% of the migrants arriving in Rimouski at P4 came from Kamouraska).

Table 2. Probabilities of Geographic Origins of Migrants Coming From Montréal, Trois-Rivières, Québec, and From Outside to These Cities, Between 1725–1775 (P2+P3) and 1775–1799 (P4)

| | | *From* | | | |
		Montreal	*Trois-Rivieres*	*Quebec*	*Outside*
	Montreal		0.01	0.19	0.50
To	Trois-Rivieres	0.04		0.28	0.21
	Quebec	0.06	0.00		0.66

$$p(a_i \mid s_k) \; = \; \frac{\pi(a_i) \, p(s_k \mid a_i)}{\sum_i \pi(a_i) \, p(s_k \mid a_i)}$$

where $p(s_k \, a_i)$ is the probability of observing the surname s_k within the a_i-th area. This probability can be estimated by the observed frequency of the kth surname in the a_i-th area; $\pi(a_i)$ is the a priori probability of emigration from the geographic area a_i to any other area, whatever the surname. The sum is over all-considered geographic areas.

As this probability of origin of surnames is estimated for each surname s_k, one obtains a more accurate estimate by summing all surnames and then by calculating the weighted mean probability of geographic origin, pgo_i, of any surname newly arriving between two periods in a given recipient area i as:

$$pgo_i \;=\; \frac{1}{\sum_k \omega_k} \sum_k \omega_k\, p(a_i \mid s_k)$$

where ω_k is a weight taking into account the fact that several persons could share the same surname. Once these probabilities are obtained, they are used as a new estimate of the a priori probability $\pi(a_i)$ and are replaced into the Bayesian formula which is recalculated. This iterative process is carried on until a convergence criterion is met (for extensive discussion, see Degioanni and Darlu 2001).

Figure 12 shows the probability of geographic origin of newly arriving immigrants at Rimouski. Most of them did not come from the 43 counties (outside: 23%). Most of them came from the neighboring counties (Kamouraska, 30%; Montgagny, 17%). Clearly, the settlement in this part of Québec was done from place to place at short distance. A large town like Québec did not participate much in this process of migration.

The same method was applied to the migrations between the three main towns, Montréal, Trois-Rivières, and Québec. Table 2 shows the probability of geographic origin for each town. Most of the immigrants were coming from "outside," $p = 0.44$ and 0.57 for Montréal and Québec respectively, much more than for Trois-Rivières ($p = 20$). If some migrants to Montréal came from Québec ($p = 0.18$) the reverse is not true ($p = 0.06$). Trois-Rivières received its immigrants mainly from Québec.

Conclusions.   As demonstrated by the example of the Province of Québec for which accurate and exhaustive data are available for a long period of time, the use of surname frequencies in a geographic and historical context allows inferences on the peopling and on the spatial population structuring. The few methods used in this paper (analysis of surname distribution, calculation of the surname distances between places, use of agglomerative procedures to estimate robustness of surname proximities and their geographic representation, estimation of the probabilities of origin of migrants) allow us to conclude that the various Canadian parishes in Québec were, at the end of the 18th century, not very strongly structured, reflecting the dispersal of the previous generations, but nevertheless maintaining exchanges and migrations at short distances between neighboring places, and retaining the Saint-Laurent River and the two main centers of population (Montréal and Québec) as the most important delineating geographical elements.

A Long-Term Perspective on Anthroponymic Corpora
[Pascal Chareille]

It was the 11th century that saw the emergence of the two-element naming system still in use in France today. While this system was certainly not initially patronymic, the transmission of the surname—although not systematic before the

18th century—probably became "usual" as early as the 13th century. In the written sources used by historians, names provide abundant material for study. In France, since the Revolution and the establishment of a civil status register, potentially exhaustive nominative data for the whole territory are available, strengthening a system of registration which had existed since the early 16th century. The vicissitudes of archival conservation, however, are such that not all these documents have come down to us. Indeed, they are even relatively rare for the 16th century. And the further back in time one goes, the less the data are spatially exhaustive. The documents which predate the parish registers never contain the whole population. Thus the tax rolls from the 14th and 15th centuries, some admirable regional series of which have survived, only name the head of the household, and almost never the other members. In these documents, in which men are over-represented, the mode of designation of individuals already very broadly associates a name (or forename) with a surname (either individual, family or patronymic), and hence it is possible to envisage a study of anthroponymic stocks, in particular stocks of surnames, over a long duration (15th to 20th centuries).

The exploitation of medieval sources in this perspective, however, remains a perilous exercise: identifying individuals, and hence anthroponyms, may be uncertain: is the *hug*[*ue*]*s boy laigue* thus designated in a census of households in Dijon in 1376 the same person as the *hug*[*ue*]*s boilleaux* identified a year later in the same street? Examples of this type are legion, and it is often not simple to decide, since the transcription of names was largely phonetic at a period when writing was not yet in general use and spelling still inconsistent. Numerous criteria (orthographic, linguistic, phonetic, etc.) can be involved in the differentiation of variants, and the choice whether to group the latter together or treat them separately is obviously decisive for the constitution of such historical corpora. The differentiation of names such as *Fabre*, *Favre*, *Febvre*, *Fèvre*, *Lefebvre*, *Lefèvre*, *Lefébure*, etc. or *Gauthier*, *Gautier*, *Galtier*, *Vautier*, *Vaultier*, etc., which goes uncontested in present-day lists of patronyms, is not necessarily pertinent for the Middle Ages. Lemmatization is therefore a necessary and unavoidable stage in the anthroponymist's task. In practice, it leads to the establishment of separate corpora depending on the level of lemmatization adopted, either only grouping together the minor spellings and/or variants ("weak lemmatization"), or else associating, in a common "root form," all the related forms ("strong lemmatization").

Patronymic Stability: Normandy 1383 to 1515.   Normandy is one of the regions for which we have at our disposal a considerable historical corpus of 64,000 anthroponymic occurrences, concerning more than 55,000 individuals, drawn from the perusal of some 1400 *rôles du monnéage* [rolls of a currency stabilization tax], dating from 1383 to 1515, concerning nearly 550 parishes scattered over five viscountcies (Bayeux, Caen, Falaise, Vire and Orbec) (Angers and Chareille 2010). Nearly 13,000 different patronyms have been identified, a number which was reduced to 7600 after "strong lemmatization."

Despite this high level of lemmatization, nearly three out of every four patronyms is only attested in a single viscountcy, and less than 3.3% are present in all five. In 15th-century Normandy, then, the monophyletic character of patronyms is marked, suggesting an essentially local distribution of patronymic homonymy and a rooting of populations. It is, however, difficult to determine whether the high degree of micro-regional specificity in the 15th century is ascribable to low-population mobility or to the relatively recent adoption of patronyms, as the spatial dispersion of the hypothetical original corpora proves to be a slow process. Furthermore, the linguistic dimension of the problem, which is indisputable, still needs to be evaluated.

Despite these specificities proper to the above viscountcies, the most frequent patronyms are those which are also to be found in various places all over Normandy. None of the 100 most frequent patronyms in the whole set of corpora from 14th- and 15th-century Normandy is absent from more than two viscountcies.

The division of this corpus into four periods (P1 = 1383–1413; P2 = 1416–1449; P3 = 1452–1479; P4 = 1482–1515) makes it possible to examine its evolution over a long duration: *Lefebvre*, *Jehan*, *Hue*, *Martin* and *Hebert* are the five most frequent patronyms and, with the exception of *Hebert*, they always occupy one of the eight leading positions. The stability of these results over a very long duration is remarkable. The 25 most frequent patronyms in the 15th-century corpus are all, with the exceptions of *Regnault* and *Gueroult*, among the 150 most frequent today in the department of Calvados. This stability, however, only concerns the most frequent patronyms. In those parishes for which the documentation is continuous, less than 15% of these patronyms are attested over the total period (1383–1515), one which admittedly was particularly troubled. It is not an easy task to interpret this renewal, but the latter does not appear to be specific to either the period or the chosen analytical scale (see Darlu et al. 1997, for the period 1891–1940).

The Question of Migrations: The Example of the Dijonnais Region, 1376–1610. Historians, following the example of geneticists, use anthroponomy as one of the ways of tracing population movements, whether it be a matter of studying "long-distance" migrations within a vast territory or between one linguistic area and another, or of intraregional migrations.

A few rare documents allow a systematic count of instances of explicit extra-urban mobility. This is the case with a household census carried out in Dijon during 1376–1377 (see the extract in Figure 13): the origin (parish and street) and destination of the known migrants are often clearly mentioned (Beck and Chareille 1998).

In the absence of direct information, the study of migrations can also be envisaged on the basis of the count of surnames corresponding to place-names. We are aware that the method is imperfect and questionable (Emery 1952, 1955; Kedar 1973), but its application to the above enumeration concerning the Dijonnais region allows the construction of a map (Figure 14) which is perhaps
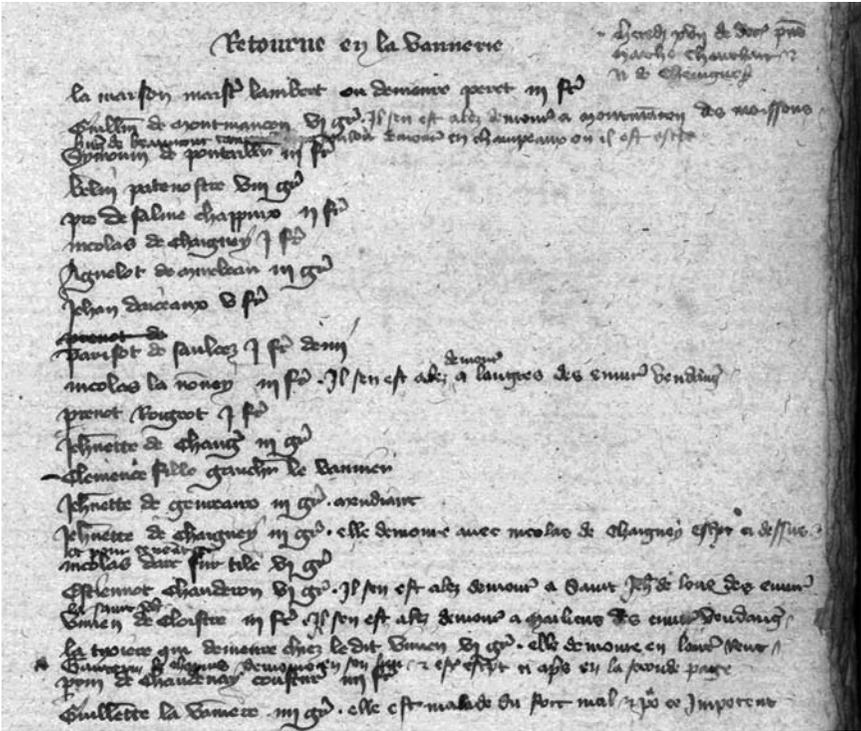
Figure 13. Annotated household census (1376–1377) [*dénombrement des feux*] for Dijon, available at: http://archivesenligne.cotedor.fr/console/ir_ead_visu_lien.php?ir=630& id=73969140(FRAD_021_B_11574_0109, Chambre des Comptes de Bourgogne Dijonnais). In this extract, concerning a street known as "Retourne en la Vannerie," the annotations mention that, for instance, "Guill[em]in de Montmancon" (entry 2) "left to live in Montmançon at harvest-time" [*Guill[em]in de Montmancon sen est alez demour[é] à montma[n]con des moissons*], and that "Nicolas la Monney" (entry 12) "left to live in Langres around the time of the grape harvest" [*nicolas la mon[n]ey sen est alez demour[é] a langres des envir[ons] vendang[es]*], etc.

less indicative of the main axes of migration toward Dijon than of a perception of the surrounding space (Beck and Chareille 1997).

The application to historical corpora of tools developed for the study of population genetics is not impossible and, moreover, enables an approach to the question of mobilities (Bourin and Sopena 2010; Darlu et al. 2011). Their use can, however, be difficult, constrained as it is by the limitations of the documentation: the absence of exhaustivity in the corpus, and the relative uncertainty as to both the hereditary nature of surnames and the extent to which they were fixed, which was certainly the norm in the 14th century but was by no means an exclusive rule. Nominative lists do not, except in exceptional cases, make it possible to identify a migrant who might have given up his former surname in favor of another recording his provenance or,
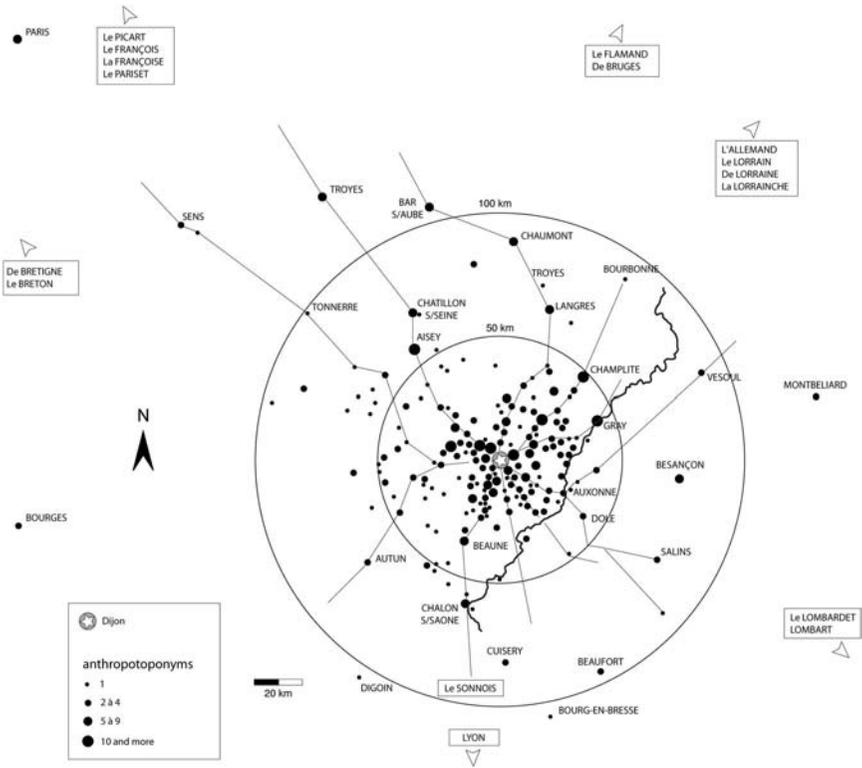
Figure 14. Surnames with place-name elements (or "anthropotoponyms") at Dijon in 1376–1377. This map, which is visibly articulated along the main routes from or toward Dijon (the strategic, political and economic routes of Burgundy in the period of the Valois dukes), is probably a fair reflection of both a large proportion of the migratory realities of the time and also, indirectly, of the perception of their surrounding space by late 14th-century inhabitants of Dijon.

on the contrary, sealing his adhesion to a new community through the use of local sound patterns in place of "exotic sonorities." And we do not know the possible extent of this phenomenon, which is attested in various places.

Despite these difficulties, the diachronic analysis of spatial distributions allows the—however fragmentary—reconstitution of the histories of certain patronyms, and hence possibly of families, and thereby makes it possible to formulate hypotheses on migrations.

Phylogenetic methods make it possible to evaluate the more or less close proximity between the corpora on the sole basis of the presence/absence of a patronym in various places without taking into account the variability of patronymic frequencies, the latter data being potentially unreliable as far as the medieval period is concerned.

The exhaustive reading of the household census of the bailiwick of Dijon for the years 1376, 1424, 1470 and 1610 makes it possible to construct a corpus
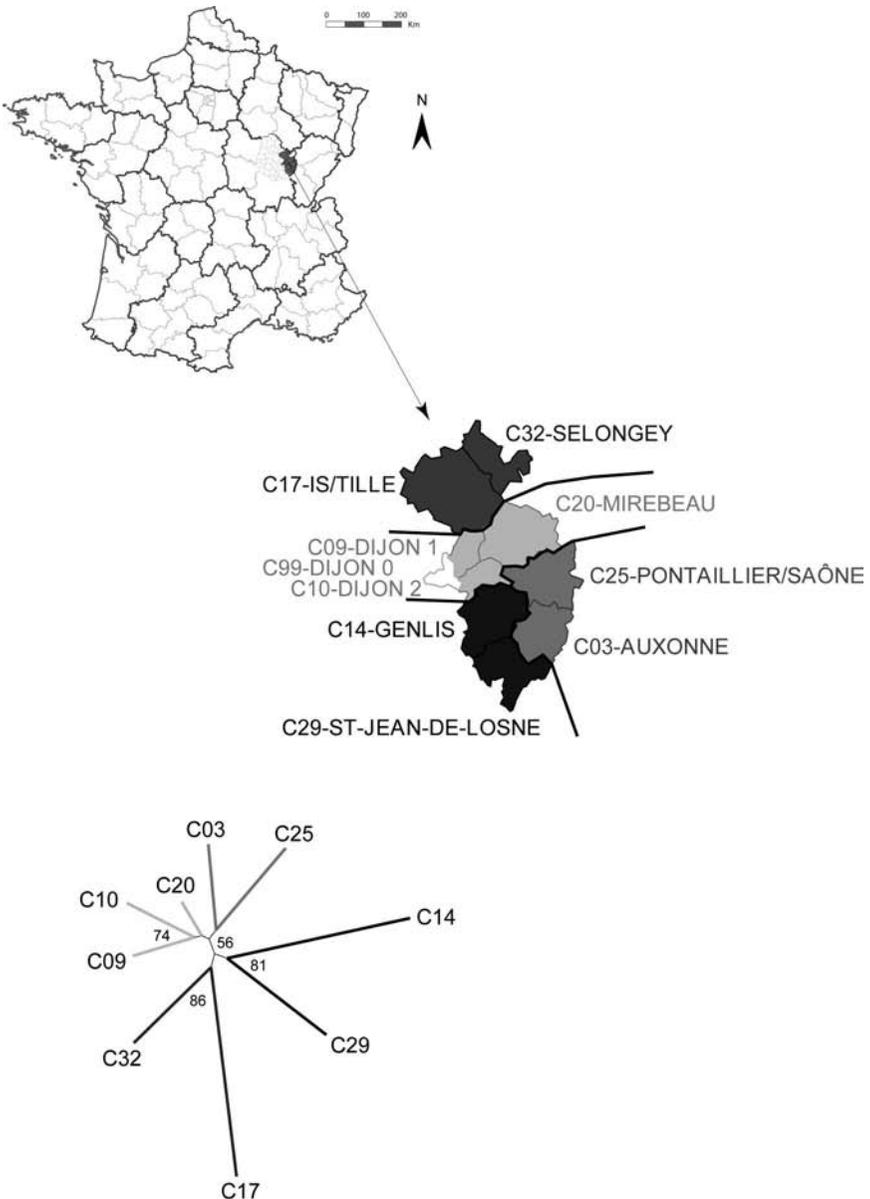
Figure 15. Division of cantons based upon the presence/absence of (sur)names.

of more than 35,500 occurrences distributed over 288 continuously documented localities grouped together by canton (on the basis of present-day administrative divisions). The anthroponymic structure of the populations thus observed

highlights four groups within each of which the patronymic proximity suggests more intense exchanges. The relationships between the cantons can be represented in the form of a tree constructed by neighbor-joining (Saitou and Nei 1987) with bootstrap values (Felsenstein 1985) (Figure 15). The comparison with 20th-century data, taken from the *Registre français des noms patronymiques* [French register of patronymic names] for the period 1891–1940, reveals an astonishing stability: the present-day anthroponymic structure was already in place, with few differences, in the Middle Ages. This result needs to be further refined, but it does seem to suggest that the most recent migrations have not, at this scale of analysis, had a destructuring effect on micro-regional patronymic corpora, and hence that the privileged axes of population interchanges have not undergone any fundamental changes.

The (re)constitution of patronymic corpora for past periods is a difficult exercise, but the problems inherent in historical documents are not insurmountable. It is surprising to discover, as far as the regions which it has been possible to investigate are concerned, that many of the points that seem to characterize contemporary corpora (diversity of corpora, a high degree of local specificity for most patronyms, renewal of the overall corpus, yet stability of the most frequent names in the results, etc.) already seem to be in place in 14th- and 15th-century France.

The data for 1376–1610 make it possible to identify, from a surname perspective, four groups: (1) Selongey and Is-sur-Tille, which correspond to the enclaved, afforested land of "La Montagne"; (2) Mirebeau and the cantons lying to the east of Dijon on the Côte and near the capital; both (3) and (4) the low-lying land on the plain of the Saône, divided by the Tille and its marshes, which were later drained and were long an almost impassable barrier and thereby a de facto limit on people's movements: Pontarlier and Auxonne are on the left bank (to the east) of the river; and Genlis and Saint-Jean-de-Losne are on the right bank (to the west).

Reconstructing Past Genetic Structures in Recently Transformed Populations: Surnames and Y-Chromosomes in the Upper Savio Valley (Central Apennines, Italy) [Alessio Boattini, Antonella Useli, Davide Pettener]

Many of the preceding contributors (Bloothooft et al., Brunet et al., Chareille, Coates and Hanks, Dräger) focused on the efficacy of surnames in tracing movements of people as well as in reconstructing historical changes in migration patterns and/or similarity/dissimilarity coefficients between populations. These features make surnames an interesting tool for human-population genetics inferences per se.

Recently, in the context of molecular anthropology studies focused on the variability of the Y-chromosome—with which surnames share a patrilineal ancestry (King and Jobling 2009)—the study of surnames found a new field of application. Most frequently, surnames have been advocated to design more

careful sampling strategies (Boattini et al. 2010a; Manni et al. 2005). Surnames have been used to increase the "archaeogenetic" power of genetic studies through the analysis of historical records and pedigrees (Bowden et al., 2008; Boattini et al., 2011). In this way, researchers were able to infer "past" genetic structures of populations by selecting those individuals who carry surnames that were proved to be present in a certain area at the time of surname introduction. In particular, Manni et al. (2005) introduced a "general" surname method, based on Self-Organizing Maps (SOMs), that provides an efficient identification of groups of surnames that share a geographic origin and migration history. The method was first tested in the case study of the Netherlands (Manni et al. 2005; Manni et al. 2008), then successfully replicated in microgeographic contexts (Boattini et al. 2010a, 2010b; Rodriguez Diaz and Blanco-Villegas 2010).

Here we apply the SOMs methodology in order to unravel the genetic structure of a population that was subjected to radical transformations during the last century. The Upper Savio Valley—a mountain population located in Italian Central Apennines—experienced a series of demographic phenomena that were common to a great part of Italian mountain communities: major depopulation and migrations toward the most important urban centers. In this study, we will compare surname clusters identified by SOMs with Y-chromosome variability in the Upper Savio Valley. Our main purposes are: (1) to test the power of the SOMs method to discover "real" (biologically significant) clusters, and, if this condition is met, (2) to search for historical changes in surname structure of the population and (3) to identify remnants of historic genetic structures within the investigated area.

The Data and Methods Surname analysis is based on 10,202 records from conscription lists for the years 1828–2005, corresponding to individuals born between 1808 and 1987. Following historic/geographic criteria, the Upper Savio Valley was subdivided into five areas (A, B, C, D, E), of which A and B correspond to the main urban centers of the valley—where the great part of the population is currently settled—while C, D and E are very rural areas, that nowadays are largely deserted (Figure 16).

Surname distributions were analyzed with SOMs. The SOMs method is a clustering technique through neural networks based on "competitive learning," an adaptive process in which the cells ("neurons") simulating a neural network ("map") gradually become sensitive to different input categories (Kohonen 1984). The main idea is that different neurons specialize to represent different types of input vectors; in doing so they interact with the neighboring neurons by means of a "neighborhood function." This procedure will result in the differen-tiation of the whole map-space: (a) identical vectors will be mapped at the same neuron, (b) slightly different ones at close neurons, while (c) very different vectors will be mapped at far neurons. The shape (rectangular or square) and size (number of cells) of the SOMs are defined by the user. The size of the map determines the maximum number of different clusters; therefore, larger maps will
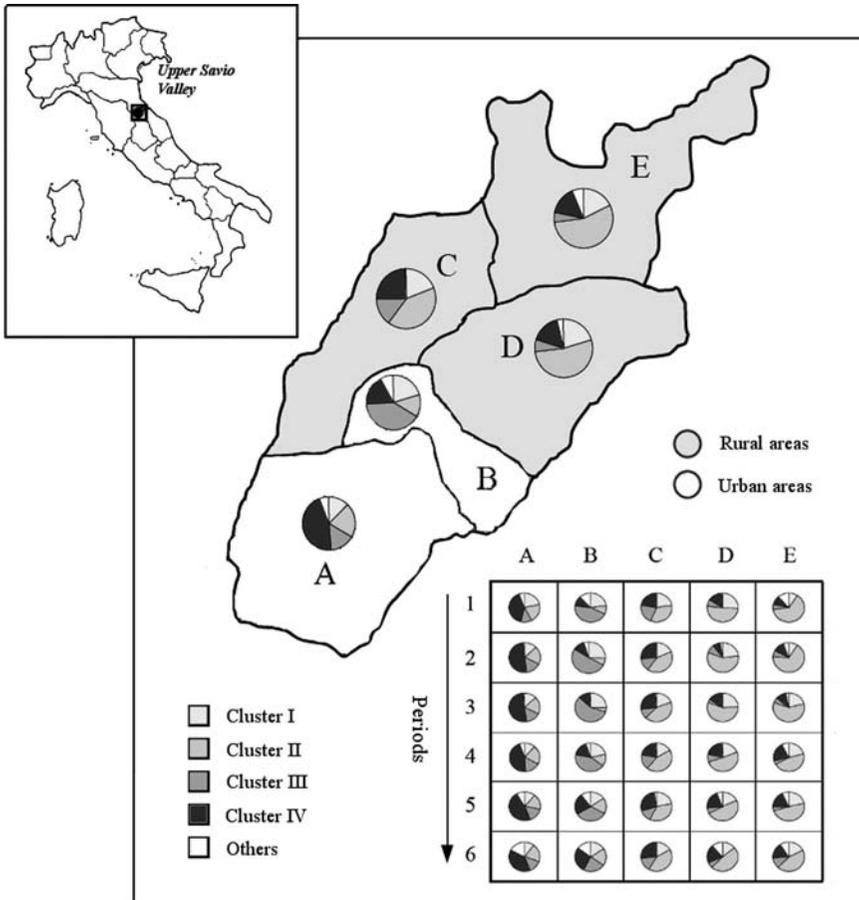
Figure 16. Geographic location and frequencies of the main surname clusters from SOMs with their temporal changes (right, below) in the Upper Savio Valley.

classify items (surnames, in this study) more accurately than smaller ones. Nevertheless, it may happen that some cells remain empty, while others collect many items. Manni et al. (2005) demonstrated that the SOMs method can be considered a "blind" automated approach to identify the geographic origin of surnames.

For the study of Y-chromosome variability, we collected peripheral blood samples from 59 individuals who were selected on the basis of (a) pertinence of their surname to one of the main SOMs clusters (see below), and (b) ascertained patrilineal residence in the Upper Savio Valley for the last three generations. For each sample, 31 binary polymorphisms (YAP, M213, M9, 92R7, M173, SRY1532, P25, TAT, M22, M70, 12f2, M170, M62, M172, M26, M201, M34, M81, M78, M35, M96, M123, M167, M17, M153, M18, M37, M126, M73, M65, M160) and 12 short tandem repeats [STRs] (DYS391, DYS389I, DYS439, DYS393, DYS390, DYS385a/b, DYS438, DYS437, DYS19, DYS392, DYS389II) were typed.

Hg	Frequencies	
	Rural areas	*Urban areas*
E1b1b1a-M78	0.07	0.00
E1b1b1b-M81	0.03	0.03
G-M201	0.03	0.23
I*(xI2a2)-M170	0.00	0.07
J*(xJ2)-12f2	0.03	0.00
J2-M172	0.21	0.20
R1*(xR1a,R1b1)-M173	0.03	0.00
R1a1-M17	0.03	0.00
R1b1-P25	0.55	0.43
T-M70	0.00	0.03
Sample size	29	30

HG R1b1 - P25
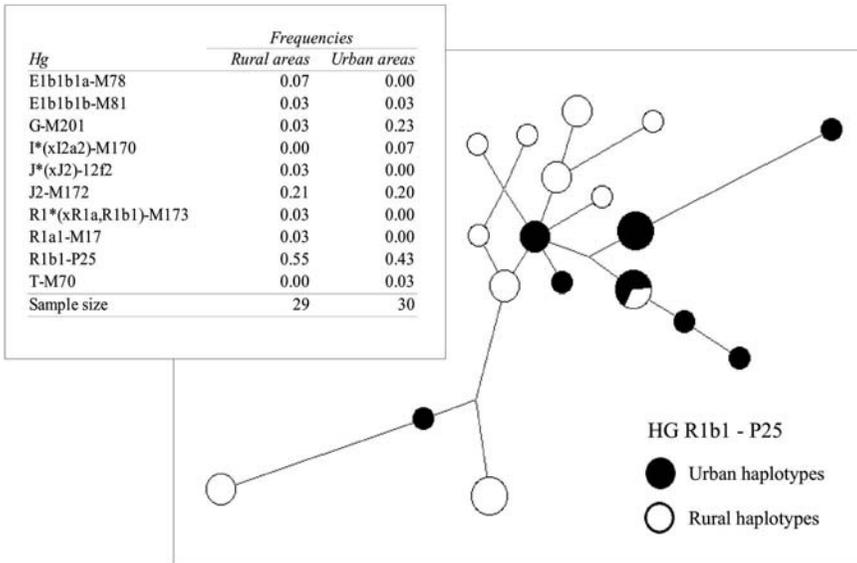
● Urban haplotypes

○ Rural haplotypes

Figure 17. Haplogroup frequencies and network of the R1b1–P25 haplogroup in the rural and urban subpopulations.

Results and Discussion. The geographic distribution of surnames was analyzed using SOMs. This revealed four main surname clusters: clusters I (33 items) and II (99 items) are mainly represented in areas C, D and E; thus, these groups of surnames may be considered as indigenous to rural areas, while clusters III (72 items) and IV (125 items) are mostly found in areas A and B; thus, the corresponding surnames very likely had their origin in the urban centers of the Upper Savio Valley (Figure 17). For some of these, we were able to confirm their inferred place of origin based on 16th-century surname information for two Upper Savio Valley parishes from previous research (Boattini & Pettener 2005). As a second step, we explored diachronic changes in SOMs cluster frequencies by subdividing our data according to six 30-year intervals (referring to the year of birth: 1808–1837, 1838–1867, 1868–1897, 1898–1927, 1928–1957, 1958–1987).

All the considered areas show a temporal increase in the degree of within-area surname diversity (Figure 16), particularly for the two more recent periods. These results were confirmed by continuous descending Fst patterns for the Upper Savio Valley for the whole historic interval considered (results not shown) and suggest that our population was characterized by considerable internal mobility (in particular toward the urban areas).

These results suggest strongly that social-cultural factors gave rise to a reproductive barrier between inhabitants of the chief towns and those of the surrounding areas, despite their sharing the very same environment. Neverthe-less, historical changes in SOMs cluster frequencies and Fst show a shift toward

a higher degree of surname homogeneity between areas, meaning that the reproductive barrier has been disappearing, especially during the last two periods (i.e., the second half of the 20th century). Unfortunately, our study was not able to discriminate between monophyletic and polyphyletic surnames, as was the case for Manni et al. (2005), but this was expected given the microgeographic setting of this research; regarding this last point, analogous results were obtained for the Alpine isolate Val di Scalve (Boattini et al. 2010a).

The next step of our research was to verify if SOMs results were confirmed by Y-chromosome analyses. The 59 total samples were divided into two groups corresponding to: 29 individuals whose surnames are included in clusters I and II (rural), and 30 individuals whose surnames are included in clusters III and IV (urban). While haplogroup frequencies between the two subpopulations were not significantly different (with the exception of haplogroup G, that was found almost exclusively in the urban sub-population) (Figure 17), Fst calculations based on STR haplotypes revealed a slight but significant differentiation (Fst = 0.022, p = 0.02). This means that these differences lay mainly within haplogroups, as is clearly demonstrated by a network representation of haplogroup R1b1-P25 (Figure 2), the most widespread in the Upper Savio Valley, to which corresponds Fst = 0.074, p = 0.02. "Urban" haplotypes mostly cluster in the same branch of the network, while "rural" ones form different branches (stemming from the same "urban" haplotype). Summing up, it seems very likely that the two subpopulations evolved from the same ancestral population, a process that—for historical reasons—probably had its origins during the late Middle Ages.

In conclusion, we can affirm that surname results, as obtained with the SOMs, are confirmed and enhanced by Y-chromosome data. Furthermore, the combined use of cultural markers (surnames) and molecular markers (Y-chromosomes), enabled us to bring to light a "fossil" reproductive barrier between two different groups of individuals—urban and rural ones—within the same population and environment. The demographic changes that intervened during the studied period and in particular in the second half of the 20th century (increased population mobility, depopulation of the rural areas), caused that barrier to disappear. At a more general level, this study underlines the contribution that surname analysis can bring to molecular anthropology studies and in particular to those aimed at the reconstruction of genetic histories of populations.

Cross-Disciplinary Perspectives in the Spatial Analysis of Names [Pablo Mateos, Paul Longley, and James Cheshire]

Looking at the recent past, surnames have provided a reliable indicator of population structure (Lasker 1985), migration flows (Darlu and Ruffié 1992; Piazza et al. 1987), intermarriage (Bugelski 1961), endogamy and genetic

inheritance (Jobling 2001). Moreover, focusing on disentangling contemporary migration and population diversity, surnames have also proved very useful to classify populations by ethno-cultural origin in health registers (Lakha et al. 2011; Lauderdale and Kestebaum 2000), electoral candidates (Dancygier 2010; Wood et al. 2011), inventors (Natham 2010), and even in social communities on the web such as MySpace and Facebook (Chang et al. 2010) or Wikipedia (Ambekar et al. 2009). Some of these studies (Bloothooft, this article) also use forenames as well as surnames frequencies, since the former are also intergenerationally assigned, following cultural norms, social networks, and persistent group practices (Alford 1988). Taken together both forenames and surnames can provide even more refined understandings of population structure and relatedness (Mateos et al. 2011).

Moving from the age of migration to the age of DNA research, let's illustrate how such cross-disciplinary understandings of people's names can help to disentangle some of the overlapping and recent episodes that conform the origins of our common past at much finer geographical scales. We will do so through a few examples also presented at the Paris workshop by a team of geographers at University College London (UCL) (Longley, Mateos and Cheshire).

In a recent study on the genetic structure of Great Britain (Winney et al. 2012) UCL geographers provided a set of methodologies that were key in enhancing the sampling of thousands of DNA donors as well as the statistical analysis of population substructures in genotyped information. The contributions by geographers were twofold. First, they geocoded (assigning a pair of geographic coordinates) the places of birth of 3,865 subjects and that of their parents and grandparents across Great Britain (circa 27,000 of places of birth). This then facilitated the task of calculating mean distances between grandparental places of birth, determining distance and population size thresholds to filter out persons with non-local or "urban" grandparents contributing to their DNA makeup. Of the sample, 75% turned to have a mean distance between grandparental places of birth of 37.3 km, while 70% of grandparental places of birth were classified as "rural." Second, they established whether each subject's surname actually originated close to the area where his/her grandparents were born. This obviously mainly worked for monophyletic surnames, or those with no more than two or three distinct areas of origin, but the task of determining such "area of origin" was eased through the use of a historical population register, as opposed to a contemporary one. The 1881 census with the names of 29-million respondents was geocoded at a small area level for which relative-surname frequency distributions were computed (through a collaboration between geographers and historians). This allowed dividing the sample between DNA donors with "local" versus "non-local" surnames to look for further genetic patterns. Given the relatively old ages of the donors, the fact that only grandparental places of births were used, and that surname areas of origin were taken from 1881 frequency distributions, the study substantially improved the knowledge about a

sample taken in the 2000s by "bringing it back" to a population closer to the late 19th century, regardless of contemporary locations of DNA donors. For detailed results about substructures found in Great Britain and some of the historic explanations, see Winney et al. (2012). Tyler-Smith and Xue (2012) have praised this detailed approach to the geographic, historical and onomastic analysis of genetic data and highlighted its value in improving disease association studies. As they put it, this "microcosmic survey of genetic variation in a set of small islands off the western coast of the Eurasian continent is revealing the level of differentiation that builds up over millennia via events well documented by archaeology and history, so these alternative data sets can be compared to address questions about the initial peopling of the area, and its subsequent reshaping by internal and external forces" (Tyler-Smith and Xue 2012: 130).

Other more advanced techniques to determine a surname's "core region" (or regions) of origin have been also developed by the UCL Geography team using a probability approach to create "surname surfaces" with which to identify spatial concentrations of surnames and compare them over time (Cheshire and Longley 2012). This is done through Kernel Density Estimation (KDE), conforming a backdrop against which a DNA-sampling strategy could be designed. Furthermore, geographical-analysis methods can also complement the existing range of approaches to classify continuous space into discrete cultural regions identifying barriers to population inter-action through surnames, frequencies and densities. In this vein two recent studies published by this team were also presented at this workshop using a range of clustering, areal classification, and spatial-analysis techniques (Cheshire et al. 2011; Longley et al. 2011).

Finally, a different approach to surname-origin classification, also pre-sented by the UCL team at the Paris workshop, draws from recent advances in physics to cluster large and complex networks (Mateos et al. 2011). Naming networks were constructed linking surnames through the forenames they share in 17 countries at the individual-person level drawn from the aforementioned UCL Worldnames database (a network comprising 118-million people, 4.6-million unique surnames and 1.5-million unique forenames, linked through 46.3 million unique forename-surname pairs). Algorithms to search for community structure in very large networks were used to identify clusters of cultural ethnic and linguistic origin. Clusters of surnames were automatically classified and identi-fied by their cultural origin by checking them against surname dictionaries to validate the methodology (with results varying between 0.71 to 1.0 in terms of sensitivity and specificity values). Such approach permits the automatic classi-fication of large numbers of surnames into clusters of cultural commonality that can be then further analyzed for linguistic, historic, or geographical patterns. This methodology automatically identified clusters of these surnames originated outside Europe, which, given that most of the data was drawn from Europe, shows its value to disentangle recent migration episodes as well as population structure in the non-Western world.
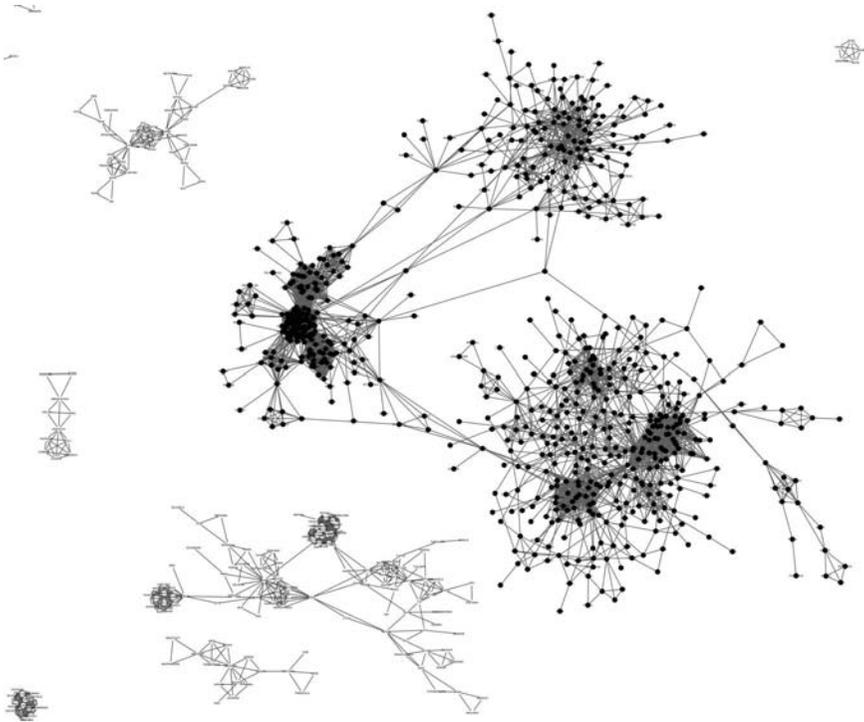
Figure 18. A naming network of Auckland, New Zealand (adapted from Mateos et al. 2011).

The network shows a selection of surnames in the city of Auckland (NZ) using data from the 2006 Electoral Register. Each node represents a unique surname and the links are common forenames shared between a surname pair. The network's topological structure reveals clear clustering in Auckland's naming practices, reflecting closely-knit social networks and ethno-cultural customs that prevent cross-cultural (fore-)naming. For example the bottom third of the figure contains a subnetwork of names that are Tongan, Samoan and other Pacific Islanders. The full version of this network including the actual surnames can be visualized in an online version available at http://www.onomap.org/naming-networks/fig2.aspx: this Figure can be navigated with full panning and zooming capabilities for flexible exploration. For more details on how the network was built see Mateos et al. 2011.

In a final example drawn from the same paper (Mateos et al. 2011), the population of a single city, Auckland, New Zealand was also clustered using a naming networks approach (see Figure 18 for a visualization of such network). It identified clear clusters of Pacific populations that retain intramarriage practices in contemporary New Zealand, in particular Tongan, Samoan, and to a lesser extent Maori surnames as well as other Pacific Islanders. The fact that this was done automatically is particularly striking, especially when there is a lack of

knowledge of the Pacific name corpora, their frequencies and origins in the literature. The surname-classification methods used in this example, combined with those proposed in Winney et al. (2012) could be used to design a much more efficient DNA-sampling strategy in Auckland without the expense of conducting research in each of the Pacific Islands, magnified by the fact that there are probably more Samoans and Tongans in Auckland than on those islands. As such, surnames can be used to enhance sampling strategies in urban areas for populations that may have perhaps diluted from the areas of origin as a result of mass migration. This application has already been proposed in the population-genetics literature for rural-urban migration in Western Europe (Manni et al., 2005). The aforementioned methods brought in by geographers to conventional population-genetics research, should be complemented by those presented in this article by historians, linguists and anthropologists (e.g., see those led by Darlu, Chareille, Coates, and Dräger). Such cross-disciplinary research approaches will be invaluable to improve methods to classify the ethno-linguistic origin of surnames, pinpoint them to historical areas of origin and trace subsequent waves of migration and specific-population dynamics over space. We hope to have made a small contribution to the wealth of cross-disciplinary approaches presented in this paper, and instilled some curiosity amongs researchers in population genetics in enhancing their research findings through the incorporation of methods of spatial analysis of names.

Final Remarks

In this article, we have provided an overview of some ongoing research addressing surnames in Western Europe, Canada and, for a very limited aspect, New Zealand. While similar studies are conducted in other regions and continents (see Colantonio et al. 2003), we think that the examples presented are representative of the kind of research questions that family names allow; questions that often go beyond the simplistic use of surnames as a proxy to Y-chromosome diversity. As this paper mainly addresses a readership of anthropologists and population geneticists, let us start our discussion from this latter traditional use of surnames. Continuous technological improvements have made possible the analysis of very large portions of our DNA. Full genome sequencing will soon become an easy and widespread technique allowing very deep inference about regional and microregional genetic differences, that can be explained by demographic factors that, in turn, often rely on historical and cultural processes. Family names of patrilineal descent have proved to mirror a single locus on the Y-chromosome (King and Jobling, 2009). However, they have a temporal depth that is quite limited (between 4 and ±30 generations) when compared to the scale of demographic processes inferred with molecular markers. *In this context, why should anthropologists take into consideration surname information that, albeit easier to collect than DNA data, is sometimes tricky to interpret, as it is suggested in this summary paper?*

The easiest answer is that surnames allow a retrospective look at human variation. They enable comparisons between recent and ancient surname corpora because historical documents often report surname information over several successive generations, and they have a degree of polymorphism that (for the moment) is larger than the one inferred from DNA markers.

Isn't this approach similar to the scientific interest in ancient DNA technology, which is now being applied to entire populations? Once extant human diversity has been satisfactorily described (this is not too far away), one of the major questions will be to explain when and how it arose. Direct evidence may come from ancient DNA studies but they will remain geographically uneven as the molecule inevitably degrades in unfavorable climates. This is why Boattini and coworkers (this paper) most appropriately use the expression of *Archaeo-genetic Power* to define the interest of surnames in anthropology and biodemography as they allow genetic samplings that are quite representative of past populations. Today, in an age of global migrations (Castles and Miller 2009), surnames have indeed the potential to allow an intermediate level of access to the recent past and to small geographical scales that are difficult to address otherwise. In most societies, the interest of surname research is ultimately related to their hereditary nature, but also to their role as markers of group identity (Alford, 1998), making them useful to classify populations according to ancestral proximity. Studies in this area are all based on one simple assumption: the distribution of people's names over space and time is far from random, even in today's highly mobile societies. Surnames have already proved very useful to depict migration phenomena in different periods making possible the identification of past genetic isolates and population structures that have been modified or have disappeared nowadays. This is where the potential of surnames in population studies goes well beyond the tradition analogy with the non-recombinant part of the Y-chromosome.

Future research will probably be directed to (a) determine the most probable geographical, temporal and cultural origin of surnames; (b) distinguish polyphyletic from monophyletic surnames; (c) identify common surname linages in variations of spellings; (d) establish finely detailed surname frequency distribution across space and time, (e) delineate areas of surname origin and barriers to cultural and population interaction, and (f) combine the above advances to tease out the different population episodes that have been overlaid across space and over time. None of such challenging tasks will arise without a close collaboration of different disciplines, as most of this paper's contributions clearly show.

Literature Cited

Alford, R. 1988. *Naming and Identity: A Cross-Cultural Study of Personal Naming Practices.* New Haven, CT: Hraf Press.

Angers, D., and P. Chareille. 2010. Patronymes et migrations en Normandie de la fin du XIVe à la fin du XVe siècle: premiers résultats. In *Anthroponymie et Migrations dans la Chrétienté Médiévale*, M. Bourin and P. Martinez Sopena, eds. Madrid, Spain: Casa de Velázquez (Colección de la Casa de Velázquez 116), 275–316.

Ambekar, A., C. Ward, J. Mohammed, S. Male, and S. Skiena. 2009. Name-ethnicity classification from open sources. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; June 28–July 1; Paris, France, pp. 49–58. Available: http://delivery.acm.org/10.1145/1560000/1557032/p49-ambekar.pdf?key1 = 1557032&key2 = 1502083521&- coll = GUIDE&dl = GUIDE&CFID = 53350992&CFTOKEN=96858509 Accessed 2010 Dec 18.

Archer, S. 2003. *The British 19th-Century Surname Atlas.* CD-ROM. Dartford: Archer Software. [New edition (2011).]

Beck, P., and P. Chareille. 1997. Espaces migratoires et aire d'influence de la ville de Dijon à la fin du XIVe siècle. *Cahiers de Recherches Médiévales (XIIIe-XVe s.)* 3:17–32.

Beck, P., and P. Chareille. 1998. Sédentarité et mobilité à Dijon à la fin du XIVe siècle. In *La ville au Moyen Âge*, Vol. 2: *Sociétés et pouvoirs dans la ville*, N. Coulet and O. Guyotjeannin, eds. Paris, France: Éditions du CTHS, 95–104.

Black, G. F. 1946. *The Surnames of Scotland: Their Origin, Meaning and History.* New York: New York Public Library. [Reprinted Edinburgh: Birlinn (1993).]

Bloothooft et al. (this article)

Bloothooft, G. 2011. Linguistics and geography, the surname case. In *Sound and Sounds, Studies Presented to M. E. H. (Bert) Schouten*, W. Zonneveld, H. Quené, and W. Heeren, eds.Utrecht, UiL-OTS, 9–20.

Bloothooft, G., and D. Onland. 2011. Socioeconomic determinants of first names. *Names* 59:1:25–41.

Bloothooft, G., and L. Groot. 2008. Name clustering on the basis of parental preferences. *Names* 56:3:111–163.

Boattini et al. (this article)

Boattini, A., C. Griso, and D. Pettener. 2010b. Are ethnic minorities synonymous for genetic isolates? Comparing Walser and Romance populations in the Upper Lys Valley (Western Alps). *J. Anthropol. Sci.* 89:161–173.

Boattini, A., D. Luiselli, M. Sazzini et al. 2011. Linking Italy and the Balkans. A Y-chromosome perspective form the Arbereshe of Calabria. *Ann. Hum. Biol.* 38:59–68.

Boattini, A., M. E. Pedrosi, D. Luiselli et al. 2010a. Dissecting a human isolate: Novel sampling criteria for analysis of the genetic structure of the Val di Scalve (Italian Pre-Alps). *Ann. Hum. Biol.* 37:604–609.

Boattini, A., and D. Pettener. 2005. Tra crinali e confini: mobilità matrimoniale e barriere riproduttive in Romagna Toscana (Bagno di Romagna, 1572−1930). In *Il Matrimonio in Situazioni Estreme: Isole e Isolati Demografici,* M. Breschi and A. Fornasin, eds. Udine: Forum, 127–142.

Bourin, M., and P. Martínez Sopena, eds. 2010. *Anthroponymie et Déplacements Dans la Chrétienté Médiévale.* Madrid, Spain: Casa de Velázquez (Collection de la Casa de Velázquez, 116).

Bowden, G. R., P. Balaresque, T. E. King et al. 2008. Excavating past population structures by surname-based sampling: The genetic legacy of the Vikings in northwest England. *Mol. Biol. Evol.* 25:301–309.

Bugelski, B. R. 1961. Assimilation through intermarriage. *Social Forces.* 40:148.

Castles, S., and M. Miller. 2009. *The Age of Migration.* Palgrave Macmillan: London.

Cavalli Sforza, L. L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press.

Cavalli-Sforza, L. L., A. Moroni, and G. Zei. 2004. *Consanguinity, Inbreeding and Genetic Dfift in Italy*. Princeton, NJ: Princeton University Press, 90–148.

Chang, J., I. Rosenn, and L. Backstrom et al. 2010. ePluribus: Ethnicity on social networks. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media 23–26 May; Washington*. Association for the Advancement of Artificial Intelligence (AAAI), pp. 18–25. Available: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1534/1828. Accessed 2011 Feb 03.

Chareille, P., and P. Darlu. 2010. Anthroponymie et migration: Quelques outils d'analyse et leur application à l'étude des déplacements dans les domaines de Saint-Germain-des-Près au IXe siècle. In *Anthroponymie et Migrations Dans la Chrétienté Médiévale,* M. Bourin and P. Martinez Sopena, eds. Madrid, Spain: Casa de Velázquez (Collection de la Casa de Velázquez 116), 41–73.

Chen, K., and L. L. Cavalli-Sforza. 1983. Surnames in Taiwan: Interpretations based on geography and history. *Hum. Biol.* 55:367–374.

Cheshire, J. A., P. Mateos, and P.A. Longley. 2011. Delineating Europe's cultural regions: Population structure and surname clustering. *Hum. Biol.* 83(5):573–598.

Cheshire, J. A., and P. A. Longley. 2012. Identifying spatial concentrations of surnames. *International Journal of Geographical Information Science* 26 (2):309−325.

Ciecœlikowa, A., ed. 2007−. *Antroponimia Polski od XVI do XVIII Wieku* [Polish anthroponymy from the XVI to the XVIIIth century], vol. 1 (A−G); vol. 2 (H−Mą). Kraków: Lexis (2007 and 2009, respectively). [Vol. 3 is at press and vol. 4 in preparation.]

Cottle, B. 1967. *The Penguin Dictionary of Surnames.* Harmondsworth: Penguin. [3rd, fully revised, ed. by John Titford (2009).]

Dammel, A., and M. Schmuck. 2008. Der deutsche familiennamenatlas (DFA): Relevanz computergestützter familiennamengeographie für die dialektgeographie. In *Sprachgeographie Digital: Die Neue Generation der Sprachatlanten*, E., S., and W. König, eds. Hildesheim et al.: Olms. 73–104, 254–260.

Dancygier, R. M. 2010. *Immigration and Conflict in Europe.* New York: Cambridge University Press.

Darlu, P., and A. Degioanni. 2007. Localisation de l'origine géographique de migrants par la méthode patronymique: exemple de quelques villes de France au début du XXème siècle. *Espace géographique* 3:251–265.

Darlu, P., and J. Ruffié. 1992. L'immigration dans les départements français étudiée par la méthode des patronymes. *Population* 3:719–734.

Darlu, P., G. Brunet, and D. Barbero. 2011. Spatial and temporal analyses of surname distributions to estimate mobility and changes in historical demography: The example of Savoy (France) from the XVIIIth to XXth century. In *Navigating Time and Space in Population Studies,* M. P. Gutmann, G. D. Deane, E. R. Merchant, and K. M. Sylvester, eds. Series. International Studies in Population, vol. 9, Springer, 1st Edition, 2011, XII.

Darlu, P., A. Degioanni, and J. Ruffié. 1997. Quelques statistiques sur la distribution des patronymes en France. *Population* 3:607–634.

de Bhulbh, S. (*c.* 1997) *Sloinnte na h-Éireann = Irish surnames.* Faing, Co. Luimnigh: Comhar-Chumann Íde Naofa. [2nd edn. titled *Sloinnte uile Éireann = All Ireland surnames* (2002).]

De Felice, E. 1978. *Dizionario Dei Cognomi Italiani.* Milano, Italy: Mondadori.

Degioanni, A., and P. Darlu. 2001. A Bayesian approach to infer geographical origins of migrants through surnames. *Ann. Hum. Biol.* 28:537–545.

de Woulfe, P. 1906. *Sloinnte Gaedheal is Gall = Irish names and surnames.* Dublin: M. H. Gill [2nd edn (1922/3).]

Dräger, K., and M. Schmuck. 2009. The German Surname Atlas Project —Computer-based surname geography. In *Names in Multi-lingual, Multi-cultural and Multi-ethnic Contact. Proceedings of the 23rd International Congress of Onomastic Sciences,* W. Ahrens, S. Embleton, and A. Lapierre, eds. August 17−22, 2008, York University, Toronto, Canada. [CD-Rom].

Emery, R. 1952. The use of the surname in the study of medieval economic history. *Medievalia et Humanistica* 7:43–50.

Emery, R. 1955. A further note on medieval surnames. *Medievalia et Humanistica* 9:104–106.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.

Grünert, H. 1958. *Die Altenburgischen Personennamen. Ein Beitrag zur mitteldeutschen Namenforschung.* Tübingen: Niemeyer.

Hanks, P., and F. Hodges, eds. 1988. *A Dictionary of Surnames.* Oxford: Oxford University Press.

Hanks, P., ed. 1990. *Dictionary of American Family Names*, 3 vols. Oxford: Oxford University Press.

Hanks, P., P. McClure, and R. Coates (forthcoming 2012). Family names of the United Kingdom: A new research project in British anthroponomastics. *Proceedings of the 24th International Congress of Onomastic Sciences, Barcelona, Spain, 5−9 September 2011.*

Hellfritzsch, V. 2007, *Personennamen Südwestsachsens. Die Personennamen der Städte Zwickau und Chemnitz bis zum Jahre 1500 und ihre Sprachgeschichtliche Bedeutung.* Leipzig: Leipziger Universitätsverlag.

Hey, D. G. 2000. *Family Names and Family History.* London: Hambledon and London.

INSEE (1985) *Registre français des noms patronymiques.*

Jobling, M. A. 2001. In the name of the father: Surnames and genetics. *Trends in Genetics* 17:353–357.

Karlin, S., and J. McGregor. 1967. The number of mutant forms maintained in a population. *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics, and Probability* 4:415–438.

Kedar, B. 1973. Toponymic surnames as evidence of origin: Some medieval views. *Viator* 4:123–129.

King, T. E., and M. A. Jobling. 2009. What's in a name? Y-chromosome and the genetic genealogy revolution. *Trends Genet.* 25:351–360.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43:59–69.

Kohonen, T. 1984. *Self-organization and Associative Memory.* Berlin: Springer.

Kunze, K. 2004. *Dtv-Atlas Namenkunde: Vor- und Familiennamen in Deutschen Sprachgebiet.* 4th edn. München: dtv.

Kunze, K., and D. Nübling, eds. 2009. Deutscher Familiennamenatlas. Band 1: Graphematik/ Phonologie der Familiennamen I: Vokalismus. Von Christian Bochenek, Kathrin Dräger. Berlin, New York: de Gruyter.

Kunze, K., and D. Nübling, eds. 2011. Deutscher Familiennamenatlas. Band 2: Graphematik/Phonologie der Familiennamen II: Konsonantismus. Von Antje Dammel, Kathrin Dräger, Rita Heuser, Mirjam Schmuck. Berlin, New York: de Gruyter.

Kunze, K., and D. Nübling, eds. 2012. Deutscher Familiennamenatlas. Band 3: Morphologie der Familiennamen. Von Fabian Fahlbusch, Rita Heuser, Jessica Nowak, Mirjam Schmuck. Berlin, New York: de Gruyter.

Lakha, F., D. Gorman, and P. Mateos.2011. Name analysis to classify populations by ethnicity in public health: Validation of Onomap in Scotland. *Public Health* 125 (10): 688−696.

Lauderdale, D. S., and B. Kestenbaum. 2000. Asian American ethnic identification by surname. *Population Research and Policy Review* 19(3):283−300.

Longley, P. A., J. A. Cheshire, and P. Mateos.2011. Creating a regional geography of Britain through the spatial analysis of surnames. *Geoforum* 42(4): 506–516.

MacLysaght, E.1985. *The Surnames of Ireland*, 6th edn. Blackrock Co. Dublin: Irish Academic Press.

Mandemakers, K. 2000. The Netherlands. Historical sample of the Netherlands. In *Handbook of International Historical Microdata for Population Research* (Minnesota Population Center Minneapolis 2000), P. Kelly Hall, R. McCaa, and G. Thorvaldsen, eds., 149–177.

Manni, F., W. Heeringa, B. Toupance et al. 2008. Do surname differences mirror dialect variation? *Hum. Biol.* 81:41–64.

Manni, F., B. Toupance, A. Sabbagh et al. 2005. New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. *Am. J. Phys. Anthropol.* 126(2):214–228.

Manni, F., B. Toupance, A. Sabbagh et al. 2005. A new method for surname studies of ancient patrilinear population structures, and possible application to improvement of Y-chromosome sampling. *Am. J. Phys. Anthropol.* 126:214–228.

Mateos, P. (forthcoming) *Ethnicity, Geography and Populations: Tracing Diversity and Migration Through People's Names.* Springer: Heidelberg.

Mateos, P. 2007. A review of name-based ethnicity classification methods and their potential in population studies, *Population Space and Place* 13(4):243–263.

Mateos, P., P. A. Longley, and D. O'Sullivan 2011. Ethnicity and population structure in personal naming networks. *PloS ONE* 6 (9): e22943.

Mateos, P., R. Webber, and P. Longley. 2007. The Cultural, Ethnic, and Linguistic Classification of Populations and Neighborhoods Using Personal Names, CASA working paper 116, UCL, London http://discovery.ucl.ac.uk/3472/

McKinley, R. A. 1975. *The Surnames of Norfolk & Suffolk.* Oxford: Leopard's Head Press.

McKinley, R. A. 1977. *The Surnames of Oxfordshire.* Oxford: Leopard's Head Press.

McKinley, R. A. 1981. *The Surnames of Lancashire.* Oxford: Leopard's Head Press.

McKinley, R. A. 1988. *The Surnames of Sussex.* Oxford: Leopard's Head Press.

McKinley, R. A. 1990. *A History of British surnames.* Harlow: Longman.

McKinley, R. A., G. Redmonds et al. 1973–1998. A Series of County-Based Surname Studies. Oxford: Leopard's Head.

Morgan, T. J., and P. Morgan. 1985. *Welsh Surnames.* Cardiff: University of Wales Press.

Nathan, M. 2011. The Economics of Super-diversity: Findings from British Cities, 2001–2006. SERC Discussion Paper Series, SERCDP0068, London: London School of Economics.

Nei, M. 1973. Genetic distance between populations. *American Naturalist* 106:283–292.

Neumann, I. 1970. *Die bäuerlichen Familiennamen des Landkreises Oschatz.* Berlin: Akademie-Verlag.

Neumann, I. 1981. *Die Familiennamen der Stadtbewohner in den Kreisen Oschatz, Riesa und Großenhain bis 1600.* Berlin: Akademie-Verlag.

Nicholls, K., ed. 1994. *The Irish Fiants of the Tudor Sovereign during the Reigns of Henry VIII, Edward VI, Philip and Mary and Elizabeth I*, 4 vols. Dublin: Burke.

Nübling, D., and K. Kunze. 2005. Familiennamenforschung morgen: Der deutsche Familiennamenatlas (DFA). In *Namenforschung Morgen: Ideen, Perspektiven, Visionen*, A. Brendler and S. Brendler, eds. Hamburg: Baar, 141–151.

Nübling, D., and K. Kunze. 2006. New perspectives on Müller, Meyer, Schmidt: Computer-based surname geography and the German Surname Atlas project. *Studia Anthroponymica Scandinavica* 24: 53–85.

Piazza, A., S. Rendine, G. Zei et al. 1987. Migration rates of human populations from surname distribution. *Nature* 329:714–716.

Postles, D. 1995. *The Surnames of Devon.* Oxford: Leopard's Head Press.

Postles, D. 1998. *The Surnames of Leicestershire & Rutland.* Oxford: Leopard's Head Press.

Reaney, P. H. 1958, 1976. *A Dictionary of British Surnames.* London: Routledge and Kegan Paul. [Third edn. by Reaney and Richard M. Wilson, *A Dictionary of English Surnames* (1991). R&W.]

Redmonds, G. 1973. *The Surnames of Yorkshire, West Riding.* Oxford: Leopard's Head Press.

Redmonds, G. 2002. *Surnames and Genealogy: A New Approach.* Bury: Federation of Family History Societies.

Rodriguez Diaz, R., and M. J. Blanco Villegas. 2010. Genetic structure of a rural region in Spain: Distribution of surnames and gene flow. *Hum. Biol.* 82:301–314.

Rohlfs, G. 1997. *Studio e Ricerche su Lingue e Dialetti di Italia.* Fireze, Italia: Sansoni Editore.

Saitou, N., and M. Nei. 1987. The Neighbor-Joining Method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4(4):406–425.

Schmuck, M. 2009. Personennamen als quelle der grammatikalisierung. Der *ing*-Diminutiv in Mecklenburg-Vorpommern. *Beiträge zur Namenforschung (N.F.)* 44: 35–65.

Tooth, E. 2000. *The Distinctive Ssurnames of North Staffordshire*, 2 vols. Leek: Churnet Valley Books.

Tyler-Smith, C., and Y. Xue. 2012. A British approach to sampling. European journal of human genetics. *Europ. J. Hum. Genet.* 20(2):129–130.

Walther, H. 1993. *Zur Namenkunde und Siedlungsgeschichte Sachsens und Thüringens. Ausgewählte Beiträge 1953−1991.* Leipzig: Reprint-Verlag.

Wijsman, E., G. Zei, A. Moroni et al. 1984. Surnames in Sardinia. II. Computation of migration matrices from surname distribution in different periods. *Ann. Hum. Genet.* 48:65–78.

Winney, B., A. Boumertit, T. Day et al. 2012. People of the British Isles: Preliminary analysis of genotypes and surnames in a UK-control population. *Europ. J. Hum. Genet.* 20(2):203–210.

Wood, J., D. Badawood, J. Dykes et al. 2011. BallotMaps: Detecting name bias in alphabetically ordered ballot papers. *IEEE Transactions on Visualization and Computer Graphics* 17 (12):2384−2391. doi: 10.1109/TVCG. 2011.174.

Yasuda, N., L. L. Cavalli-Sforza, M. Skolnick et al. 1974. The evolution of surnames: An analysis of their distribution and extinction. *Theoretical Population Biology* 5:123–142.

Zei, G., G. Barbujani, A. Lisa et al. 1993. Barrier to gene flow estimated by surname distribution in Italy. *Ann. Hum. Genet.* 57:123–140.

Zei, G., R. Guglielmino et al. 1983. Surname in Sardinia. I. Fit of frequency distributions for neutral alleles and genetic population structure. *Annals Hum. Genet.* 47:329–352.

Zei, G., A. Lisa, O. Fiorani et al. 2003. From surnames to history of Y-chromosome: The Sardinian population as a paradigm. *Europ. J. Hum. Genet.* 11(10):802–807.