The prospects for tracing deep language ancestry

Quentin D. Atkinson

Institute of Cognitive & Evolutionary Anthropology, University of Oxford, 64 Banbury Road, Oxford, OX2 6PN, U.K.
Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zeland
e-mail: q.atkinson@auckland.ac.nz

“If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, had to be included, such an arrangement would, I think, be the only possible one.”

The Origin of Species, Charles Darwin (1859)

Whilst there is now broad agreement that our genetic ancestry can be traced back to a late Pleistocene origin in Africa, there is no such consensus about the roots of the world’s 6000 or so languages. Proposed language super-families – such as Amerind in the Americas and Nostratic and Eurasian in Eurasia – or global language classifications like those controversially linked to the human genetic tree (Cavalli-Sforza et al., 1988), are viewed with scepticism by most linguists. Words are thought to evolve too rapidly to allow reliable identification of common ancestry beyond a limit of ~8ky BP (Ringe, 1998) and when apparent ‘long-range’ relationships are identified, proponents have been unable to provide statistical verification that any resemblances are beyond what would be expected by chance (Ringe, 1998). However, recent advances in the available data and methods (Dunn et al., 2005; Pagel, 2000; Pagel et al., 2007; Reesnik, Singer & Dunn, 2009) suggest the established ~8ky limit may need to be re-evaluated (Gray, 2005), potentially greatly extending the time depth over which language ancestry is informative about human prehistory.

Most claims for long-range language relationships rest on putative lexical homologues or ‘cognates’ identified on the basis of form and meaning correspondences across languages. One reason many have found this evidence hard to swallow is that the rate of replacement of cognates through time appears to be too rapid and too unpredictable to leave any reliable signal after just a few thousand years. For example, Morris Swadesh’s (Swadesh, 1952) early attempts to derive a single lexical retention rate found that even among a set of 200 relatively stable basic vocabulary terms, on average roughly 20% of cognates are lost every 1000 years. As shown in Figure 1 (bold line), such a rate implies that a pair of languages that diverged just 4,500 years ago (separated by 9,000 years of change) is expected to share only five cognates from an initial 200 in the Swadesh list. After 7,000 years, this number drops below one. Under this scenario, proposals for language classifications stretching back to the early Neolithic and beyond seem completely untenable – the number of cognates at such time depths will be too few to allow genuine historical signal to be distinguished from chance resemblances.

However, not all words are created equal - some evolve more slowly than others. Pagel (2000) has shown that a model of lexical evolution that allows rates of change to differ across meanings fits the observed distribution of lexical divergence in Indo-European better than Swadesh’s constant rate model. More recent work has revealed that the rate at which different
Swadesh list meanings evolve is correlated across language families (Pagel & Meade, 2006) and that the frequency with which a meaning is used in everyday speech, together with its part of speech, can explain almost 50% of the variation in rates of lexical replacement (Pagel et al., 2007). Thus, commonly used pronouns (such as I, you and we) and numerals (one, two, four and five) evolve roughly 100 times slower than the rarer, more rapidly evolving Swadesh adjectives and verbs (such as dirty, or to throw) (Pagel et al., 2007). This predictable variation in rates of lexical replacement dramatically increases the feasibility of reconstructing deep language ancestry.

Figure 1 (grey line) shows the expected number of surviving cognates shared between language pairs for a given separation time based on the empirically derived rate distribution from Pagel et al., (2007). Whilst under a constant rate model it would take only 4,500 years to reduce the cognate pool from 200 to five, allowing for rate variation extends this threshold beyond 20,000 years. Even languages that separated 50kya, perhaps contemporaneous with the African exodus, are expected to share at least two cognates. Of course, even if cognates exist at such time depths, there remains the problem of identifying them and demonstrating that any similarities are beyond what would be expected by chance, but the predictability of rates across meanings may help here too. Based on information about word frequency, part of speech or rates of change within language families, one can predict not just how many cognates should be shared between a pair of languages given some time of separation, but which meanings are more likely to produce cognate forms. Finding cognate forms for two or three meanings from a possible 200 may not constitute convincing evidence for a relationship, but if those meanings are also a priori expected to be the most stable, then a case for common ancestry can be made.

As well as words, structural features of language, such as the set of phonemes a language uses, its gender system or favoured word order, can also provide information about language ancestry. Although we currently lack rate estimates for structural data of the kind mentioned above, some structural features are claimed to be highly stable (Nichols, 1992) and so may prove decisive in identifying long-range language relationships. Indeed, some of the most promising recent research testing deep ancestry hypotheses makes use of structural language features. Dunn et al., (2005), for example, were able to use structural data together with phylogenetic inference techniques from evolutionary biology to identify historical signal in the Papuan languages likely to date back over 10,000 years. More recently, Reesnik et al. (2009), have used structural data to classify the languages of the ancient supercontinent Sahul into recognized major groups, some of which are likely to be just as old or perhaps much older. These findings are among the first to demonstrate language relationships beyond the traditionally held ~8ky limit. As in the case of the lexical data, if a set of highly stable structural features can be identified, it should be possible to push this time horizon back substantially further.

![Fig. 1 – the expected number of Swadesh 200 meaning list cognates surviving to the present plotted against language divergence time for Swadesh’s (1952, 1955) constant rate model (bold line) and for a model incorporating the empirical distribution of rates derived from Indo-European (grey line; Pagel, Atkinson and Meade, 2007). For comparison, a dashed line is drawn at five surviving cognates. The colour version of this figure is available at the JASs website.](image-url)
From our origins in Africa, the story of human evolution is largely one of cultural change. Language genealogies track cultures in a way that genes cannot (Friedlaender et al., 2009) and so are crucial to our understanding of human prehistory. The findings discussed here suggest that we should in principle be able to trace language ancestry back beyond the Neolithic, perhaps even as far as our expansion from Africa. Comparative analysis and hypothesis testing on a global scale will require high-quality and easily accessible lexical and structural language databases covering a large fraction of the world’s languages. Some important steps are now being taken in this direction (e.g., the World Atlas of Language Stuctures (Haspelmath et al., 2005)) but more work is needed along these lines if we are to fully capitalise on the linguistic legacy of our cultural past.

References


Inference of demographic processes from comparisons of ancient and modern DNAs

Guido Barbujani

Dept. Biology and Evolution, University of Ferrara, Ferrara, Italy
e-mail: g.barbujani@unife.it

Our ability to infer past demographic changes has substantially improved with the development of methods for the reliable typing of DNA from ancient specimens. However, the inferential process remains complicated, because ancient samples are small and the genetic information they yield is generally limited to one marker, mtDNA. Therefore, whenever dealing with ancient DNA evidence, besides asking what is the demographic model best accounting for the observed patterns in the data, one has also to consider whether there is enough statistical power in the data to discriminate among alternative models. To address the main question, one basically compares scenarios of genetic continuity between ancient and modern samples with scenarios in which the samples belong to different branches of the genealogical tree.

Computer simulation of explicit demographic models is an effective means to test hypotheses on the relationships between ancient and modern samples. Serial coalescent approaches, in particular (Anderson et al., 2005), allow one to generate genealogies from the present back to the common ancestor, in which individuals are added at various moments in time, representing modern and ancient samples. By attributing a DNA sequence to the common ancestor of the whole genealogy, and by randomly distributing mutations on the genealogical tree, one thus generates many simulated datasets. The sequences themselves are arbitrary (in fact, strings of 0s and 1s), but their differences are not, as they reflect the consequences of the genealogical and of the mutational processes. Therefore, one can estimate from them summary statistics, describing how genetic variation would be if the model is true.

Algorithms of Approximate Bayesian Computations (ABC: Beaumont et al., 2002) allow comparisons among models, as well as the estimation of the relevant demographic parameters. In short, genetic diversity in the data is summarized by a number of observed summary statistics. Millions of realizations of the demographic process assumed under each model are generated by Serial coalescent simulation, with parameters sampled from appropriately broad distributions of priors. An arbitrary number (threshold) of simulation experiments showing the shortest Euclidean distance between observed and simulated summary statistics are then retained, and the model parameters are estimated from them. By counting how often each specific model generated data falling within the best-fitting simulation replicates, one estimates a global posterior probability for each model. Algorithms exist for testing whether the parameters estimated under each model depart significantly from the observed statistics, and whether there is enough power in the data to discriminate among models.

Two applications of this method to ancient DNA data from populations of pre-classical Italy, are giving rather different descriptions of the evolution of genetic diversity through a time-bracket of some 2,500 years. In Sardinia, two modern populations separated in space by just 120 km, Ogliastra and Gallura, showed
very different relationships with a sample of 23 individuals from Bronze-age burials. A direct genealogical continuity between Bronze-age Sardinians and the current people of Ogliastra (a genetic isolate), but not Gallura, showed a much higher probability than any alternative scenarios, regardless of the method chosen for comparing models (Table 1). Also, there was evidence that genetic diversity in Gallura evolved largely independently, owing in part to gene flow from mainland Italy (Ghirotto et al., 2009).

In Tuscany, we are currently investigating the demographic scenarios accounting for the observed relationships among modern and ancient (Etruscan) inhabitants of the area. The Etruscans’ biological origins are unclear, with ancient historians suggesting either that they immigrated from Anatolia, or alternatively that they represent an autochthonous population (Barker & Rasmussen, 1998); equally obscure are their genealogical relationships with current inhabitants of Tuscany. We had available a set of 20 Etruscan sequences (Vernesi et al., 2004). In general, moderns Tuscan sampled in the areas of highest density of Etruscan sites show some mtDNA resemblance with people of the Eastern Mediterranean shore (Achilli et al., 2007), but not with the Etruscans, and that difference is unlikely to result from systematic errors in the ancient DNA sequences (Mateiu & Rannala, 2008).

In a preliminary ABC analysis of several modern and ancient samples, the latter comprising Etruscans and Middle-age people from Tuscany (Guimaraes et al., 2009), we compared three basic models of the genealogical relationships among samples (Fig. 1). We found no evidence of genealogical continuity for two Tuscan communities, Murlo and Volterra, for which Model 2 was clearly supported by data. On the contrary, Model 1 received strong statistical support when we compared with the ancient samples a third Tuscan area, Casentino. In addition, we could fit model 1 also to the mtDNA sequences from a population of the Western coast of Anatolia, where Herodotus placed the putative origin of the Etruscans. To make sure that those findings had a biological meaning, we also compared the Etruscans with other modern Italian samples, finding again no evidence of genealogical continuity. The apparent common ancestry does not clearly imply that modern Western Anatolians and Casentino people are both descended from the Etruscans, but rather that they share common ancestors who did not differ much from the Etruscans. Herodotus proposed an origin of the Etruscan culture in a migration episode...
from Anatolia less than 3000 years ago. To test whether genetic data give any support to this interpretation, we are currently estimating by IM methods the likely time of separation of the two modern samples.

In general, the comparisons of ancient and modern DNA suggest that genetic traces of the ancient inhabitants of a region can be found among the modern people, but modern populations are a mosaic of mtDNAs, and cannot be regarded as globally descended from the people who inhabited the same regions in preclassical times.

References


---

*Fig. 1 – Schematic representation of three models of demographic relationships among Etruscan (E), Medieval Tuscan (M) and Contemporary (C) populations. In Models 2 and 3, an ancestral population that underwent a split is designated by an a.*
Integrating anthropological genetics with cultural anthropology and archaeology: new opportunities

Mark Collard

Laboratory of Human Evolutionary Studies, Department of Archaeology, Simon Fraser University, Canada
e-mail: mark.collard@shaw.ca

The process of integrating anthropological genetics with cultural anthropology and archaeology has been frustratingly slow. Recently, however, there have been three developments in cultural anthropology and archaeology that have the potential to accelerate the process.

The first is the recognition that analytical techniques from evolutionary biology can be fruitfully applied to cultural data. The techniques that have been used most frequently to date are the neutral model and cladistics. Among the cultural datasets that have been analyzed with the neutral model are decorated pottery from Neolithic Germany (Shennan & Wilkinson, 2001; Bentley & Shennan, 2003) and pre-contact America (Neiman, 1995; Lipo et al., 1997; Kohler et al., 2004; Eerkens & Lipo, 2005), and baby names and dog breeds from 20th century America (Bentley et al., 2004; Herzog et al., 2004). Some cultural applications of cladistics have focused on the evolution of particular types of artifact, such as projectile points (O’Brien et al., 2001) and psalteries (Temkin & Eldredge, 2004). Others have sought to clarify the processes that generate population-level cultural diversity (e.g. Collard & Shennan, 2000; Tehrani & Collard, 2002; Jordan & Shennan, 2003; Collard et al., 2006; Tehrani & Collard, 2009). Still others have used cladistics to investigate colonization processes in prehistory (e.g. Buchanan & Collard, 2007; Coward et al., 2007). The recognition that evolutionary biological techniques like the neutral model and cladistics can be applied to cultural data has at least two potential benefits for the integration of genetic anthropology with cultural anthropology and archaeology. The most obvious is that it opens up the possibility of carrying out analyses in which genetic and cultural data are formally analyzed together as opposed to one dataset being formally analyzed and the results interpreted in the light of previous work on the other dataset, which is the norm at the moment. Less obviously, but no less importantly, the increasing familiarity of cultural anthropologists and archaeologists with the terminology, concepts and epistemology of evolutionary biology should increase the effectiveness of interactions between genetic anthropologists, cultural anthropologists and archaeologists.

The second development that has the potential to accelerate the process of integrating anthropological genetics with cultural anthropology and archaeology is the introduction of a method of estimating changes in prehistoric demography from radiocarbon dates (Erlandson et al., 2001; Gkiasta et al., 2003; Gamble et al., 2005; Shennan & Edinborough, 2007; Buchanan et al., 2008; Collard et al., 2008; Niekus, 2009). The most recent version of the method proceeds by collating a large sample of radiocarbon dates for the location and time range of interest. The dataset is then reduced in such a way that each site-phase in the sample is only represented by a single date. Subsequently, the one-date-per-site-phase dates are calibrated and their summed probability distribution calculated. The major peaks and troughs in the summed probability distribution are taken to reflect fluctuations in population size. The rationale of this approach
is that, because the number of occupations in a given time period can be expected to relate monotonically to population size, changes in summed probability distributions of calibrated 14C dates derived from different occupations serve as a proxy for changes in population size. The method has been used to investigate a number of important issues, including the transition to farming in Europe (e.g. Gkiasta et al., 2003; Shennan & Edinborough, 2007) and the claim that the Clovis Palaeoindians of North America were decimated by an extraterrestrial impact at 12,900 calBP (e.g. Buchanan et al., 2008; Collard et al., 2008). The introduction of radiocarbon-based demographic modeling has the potential to accelerate the process of integrating genetic anthropology with cultural anthropology and archaeology not only because palaeodemography is one of the main shared interests of genetic anthropologists and archaeologists but also because it can be easily combined with genetic estimates of past population size to generate novel insights about prehistory.

The third development that has the potential to accelerate the process of integrating anthropological genetics with cultural anthropology and archaeology is the growth of interest among cultural anthropologists and archaeologists in a body of theoretical work known as ‘gene-culture co-evolutionary theory’ or ‘dual inheritance theory’ (Durham 1979, 1991; Pulliam & Dunford 1980; Cavalli-Sforza & Feldman 1981; Boyd & Richerson 1985). In dual inheritance theory, genes and culture are viewed as two distinct but interacting systems of information transmission. They both involve the transmission of phenotype-influencing information but operate via different mechanisms. The genetic system is based on reproduction, while the cultural one involves social learning. With this difference in mind, dual inheritance theorists hold that genetic evolution and cultural evolution are similar in that they are both based on the process that Darwin referred to as descent with modification, but they also accept that the nature of social learning is such that cultural evolution is influenced by forces that have no obvious equivalents in genetic evolution. Most notably, individuals can choose to copy practices from nonkin, and they are also able to modify or discard practices in the light of experience. The significance of these processes is that cultural evolution cannot be assumed to be always in step with genetic evolution. Sometimes it will be, but frequently it will not. Our ability to learn from nonkin means that cultural patterns will often not coincide with genetic patterns. Likewise, our ability to learn from other individuals and to pass on those behaviors to yet other individuals throughout our lives means that cultural evolution will often be faster than genetic evolution. Dual inheritance theory even allows for the possibility that the transmission of some cultural traits might be maladaptive from a genetic point of view. The reason the growth of interest in dual inheritance theory has the potential to accelerate the process of integrating genetic anthropology with cultural anthropology and archaeology is that dual inheritance theory makes it possible for genes and culture to be analyzed within the same theoretical framework rather than, as has been the case in the past, being dealt with in different and often incompatible paradigms.

References


Buchanan B., Collard M. & Edinborough K. 2008. Paleoindian demography and the


The spread of *Homo sapiens* out of Africa and the subsequent continent colonisations and migrations have been reconstructed through the analyses of data reported by several disciplines, such as paleoanthropology, archaeology, linguistics and genetics. The joint effort of these disciplines has allowed having a broad knowledge about the tempo and mode of the origin of our species and the major colonisations at a continental level (Jobling *et al*., 2004). However, some local migration routes, especially those within continents, are far from being completely understood.

In order to shed light on the reconstruction of human migrations through the analysis of genetic markers, National Geographic and IBM, with the participation of the Waitt Family Foundation, have launched the Genographic Project (http://genographic.nationalgeographic.com). This international project aims to provide genetic data in order to reconstruct human movements through the analysis of uniparental genomes (mitochondrial DNA and Y chromosome). The project has three main subprojects: the public participation, the legacy fund, and research project.

The public part of the project aims to encourage general public to know more about their own ancestors. Each individual interested in knowing his/her deep ancestry along a single line of direct descent (maternal or paternal) can acquire a kit that is sent to his/her home. In about eight weeks, the results of the maternal or paternal markers will be deposited anonymously in the webpage of the Genographic Project and the individual can access to the data with his/her code. The sequence of the first hypervariable region of the mitochondrial DNA or some STRs and SNPs of the Y chromosome are provided to the participants in order to classify their lineage into the human phylogenetic tree and the information of the migration history of this particular branch is also available for the participant. The funds collected in this part of the public participation support the rest of the project. The participants can also introduce their data to an anonymous general database that can be used for global analyses. In this way, a global analysis of more than 78,000 mitochondrial results of the database has been published (Behar *et al*., 2007).

The Legacy Fund, a subproject of Genographic, through the extension of grants, aims to empower indigenous and traditional peoples by supporting locally-led efforts that can also raise global awareness about the cultural loss indigenous and traditional communities face.

Finally, the third part of the Genographic Project is the research project. This part aims to unravel the migration history of the human species at a population level. Several groups of scientists, one in each continent, are collecting samples representative of the geographical regions in order to have a general coverage of the human variation. The goal of the project is to collect a total of 100,000 samples to unravel major and minor migrations of the human species.

Within Western/Central Europe, we are proceeding with the sample collection and the first analyses of the results. These analyses will allow us to provide a finer resolution of...
the migrations within Europe and neighbouring geographic areas. More than 5,000 samples have already been collected and we are proceeding with the analyses of the Y-chromosome and mitochondrial DNA. The main questions to answer in Europe are the ones that have been posed for several years and not totally answered by genetic markers. To what extent the current European gene pool has been influenced by both Palaeolithic and Neolithic migrations? What was the role of the Last Glacial Maximum refugees and the subsequent recolonization of the continent? Was the Neolithic migration homogeneous across Europe? Can we correlate differences in archaeological Neolithic sites (linear pottery culture in Central Europe and impressed ware pottery in the Mediterranean) with different migratory waves? What was the influence of more recent migrations in the European genetic landscape? In order to address these questions we will use massive sample sizes, high resolution markers in the uniparental genomes, as well as standardized methods of analysis. We have been collaborating with other regional centres within the Genographic Project in order to address local genetic issues. In this way, we have demonstrated the genetic impact of the Crusaders in the Near East using Y-chromosome markers (Zalloua et al., 2008), or the genetic impact of Phoenicians in the Mediterranean (Zalloua et al., 2008).

With the Genographic Project, we aim to shed light to the migration history of our species and test some hypotheses posed by different disciplines. We hope to use the present genetic knowledge to provide in the future the answers about our remote past.

References


Complex traits are difficult to disentangle, as the phenotypes result from the interaction of different basic components where genetics interacts with environment, epigenetics and stochasticity (Cevenini et al., 2008). Studies in humans represent a particular challenge, owing to the pervasive and profound influence of cultural habits and specific evolutionary history on the different human populations. We will discuss the importance of adopting an integrated and multidisciplinary approach in genetic studies on two complex traits, such as aging and Chagas disease, as the most appropriate and correct methodology of data collection and analysis.

**Longevity**

Longevity is a very complex trait (Leroi et al., 2005; Kirkwood et al., 2005), and the phenotype of long-living people is not easy to define (Passarino et al., 2007). The genetics of longevity has peculiar and unexpected characteristics such as allele timing (Bonafè et al., 2004; Invidia et al., 2009) increased homozygosity at several loci (Bonafè et al., 2001a; Cardelli et al., 2008) gender prevalence (Passarino et al., 2002) and specificity (Bonafè et al., 2001b), strong dependence on the demographic structure (Yashin et al., 1999; Passarino et al., 2002; Cardelli et al., 2008) and epidemiological specificities of the different populations, likely resulting from its post-reproductive occurrence (De Benedictis & Franceschi, 2006), which in turn suggests a minor or negligible role of natural selection. Thus, the genetics of longevity can be highly population-specific, requires a large number of subjects (Lescai et al., 2009a) and different results in different populations are not unexpected and cannot be taken, at first glance, as a simple lack of reproducibility and validation failure (Cellini et al., 2005). The data so far collected indicate that a variety of genes are involved (Franceschi et al., 2005; Franceschi et al., 2007a; Salvioli et al., 2006; Capri et al., 2008) but it interesting to note that those involved in
major pathways and functions such as IGF1/insulin pathway (Bonafè et al., 2003), inflammation and immune responses (Carrieri et al., 2004; Franceschi et al., 2005), oxidative stress and lipid metabolism (Salvioli et al., 2005; Lescai et al., 2009b) and sirtuins (Bellizzi et al., 2005) are well represented. Thus the genetics of longevity requires *ad hoc* study design which are not easy to perform in different populations (Franceschi et al., 2007; De Rango et al., 2008).

We will report our experience regarding two studies which have peculiar and unique characteristics, *i.e.*:

1. **The AKEA study of exceptional longevity in Sardinia** (Deiana et al., 1999). In order to identify the centenarians all over the island and to collect data (health status, genealogical data) and biological samples (blood) a complex logistical organization was set up. We will review some of the major data we have collected regarding the prevalence of centenarians and their sex ratio and geographical clustering, as well as their phenotype and genotype (Passarino et al., 2001; Deiana et al., 2002; Carru et al., 2003; Lio et al., 2003; Poulain et al., 2004; Pes et al., 2004; Caselli et al., 2006; Polidori et al., 2007, Scola et al., 2008).

2. **The GEHA (GEnetics of Healthy Aging) Project** (2004-2010) in Europe (Franceschi et al., 2007b). The aim of the project is to identify genes and gene variants involved in human longevity and healthy aging, with particular attention to the cross-talk between the nuclear and the mitochondrial genomes. To this aim, within the framework of this European Union (EU) project, 2500 old sibpairs (both sibs being 90+) and the same number of sex- and ethnicity-matched younger controls have been recruited in 11 European countries. The genome wide linkage analysis on the 90+ sibpairs has been completed and a GWAS study in the oldest members of the 90+ sibpairs is in progress. Moreover, mtDNA haplogroups and sub-haplogroups have been analysed in all the recruited people, in order to confirm in a large sample of subjects derived from different geographical areas previous results suggesting a correlation between human longevity and mtDNA variants, both germ-line inherited and somatically acquired (De Benedictis et al, 1999; Zhang et al., 2003; Rose et al., 2007), which appears to be population-specific (Dato et al., 2004). Indeed, such studies on mtDNA variants at the population level may be biased by a variety of methodological issues (Santoro et al., 2006; Raule et al., 2007; Salvioli et al., 2008).

In both studies, even if at a different scale, major logistical, ethical and genetical problems related to the complexity of the populations involved, emerged, stressing the necessity of a more comprehensive and integrative approach, capable of including demographers and historical demographers, anthropologists and local historians, besides geneticists, population geneticists and biogerontologists, for the best and more correct interpretation of the genetic results.

**Chagas disease**

Chagas disease affects several millions individuals in South America, where the *Trypanosoma cruzi* is endemic and causes higher mortality than any other parasite. Its main vector is the haematophagous *Triatoma infestans*. Clinical manifestations range from asymptomatic infections to potentially fatal cardiopathy. Rural populations are the most affected, due to housing conditions and underlying poverty and low education. Recently, cases of infected subjects have been reported in USA and Europe as a result of the massive migration from South America, arising problems for blood donors, organ transplantations and new births. The aim of our study was to test the hypothesis of a different genetic susceptibility to the *Trypanosoma cruzi* infection and its clinical outcomes, in two populations, such as the Wichi (Amerindian natives) and the Criollos (a complex admixture of people of European and native origin), living in the same geographic region.
of Argentinian Gran Chaco (Mission Nueva Pompeya) where the infection is endemic (about 60-70% in both populations), but having different cultural habits and lifestyles as well as different historical-genetic-demographic characteristics. A critical issue in such field studies is represented by the cultural and anthropological mismatch between the native populations and the researchers coming from outside. Indeed, the anthropologists working since several years with the Wichí and Criollos in Mission Nueva Pompeya were fully aware that the conceptualization of diseases and particularly that regarding Chagas disease was substantially different from the one adopted by the geneticists and medical doctors. To this regard, we would like to stress the critical importance of the strict collaboration between the anthropologists, the medical doctors and the population geneticists which allowed us to overcome the non trivial and usually neglected difficulties (cultural and logistical) related to the approval of the Ethical Committees in Bologna and in Argentina, the signature of the informed consent and the collection of biological samples. Indeed, this study has been possible owing to the full involvement of the native populations and the reciprocal respect between the Wichí and Criollos and the anthropologists, resulting from their previous careful field studies. It is worth mentioning that the anthropologists devoted a particular attention to the preparation of a bilingual (Wichí and Spanish) booklets aiming to explain to the two communities the characteristics of the Chagas disease as well the implications of their participation to the study, taking into account their cultural characteristics.

Within this scenario, a preliminary step was to assess the genetic admixture of the populations under study, taking into account cultural/anthropological and socio/economical variables. To this aim a total of 600 individuals from the two populations was analysed. With two sets of uniparental markers (mtDNA and Y chromosome) in order to estimate the degree of admixture and parental contributions in both populations (Yang Yao et al., 2009). We characterised 17 STRs loci and 16 SNPs for the Y chromosome and the mtDNA haplogroups by sequencing the entire Control Region (D-loop) followed by RFLP analysis of the coding region. The Y chromosome SNPs definition shows a typical, but very high, prevalence of Amerindian haplogroups in Wichí, while in Criollos almost the same percentage represents European (admix analyses highlight the main role of Spain and Italy) and Amerindian contribution. The mtDNA haplogroup analysis revealed an Amerindian origin for both populations, with the presence of European haplogroups only in <2% of Criollos. These results shed lights on the extreme admixture of the analysed populations and may help in explaining the distribution of Chagas morbidity in the Gran Chaco region. In addition, we are assessing the patterns of seropositivity in specific familiar groups in relation to their different ethnic origin (Wichí and Criollos), and are studying polymorphisms of candidate autosomal genes, to unravel the genetic basis of Chagas susceptibility.

Acknowledgements

The GEHA study is an integrated project funded by EU, FP6; the investigations on Chagas’s disease have been funded by the Consorzio Interuniversitario Italiano per l’Argentina (CUIA), projects 2006 and 2007.

References


Bonafé M., Cardelli M., Marchegiani F., Cavallone L., Giovagnetti S., Olivier F., Lisa R., Pieri C. & Franceschi C. 2001a. Increase


JASs forum: Interdisciplinary views on Molecular Anthropology in the Genomic Era


