

Molecular Anthropology in the Genomic Era

Giovanni Destro-Bisol^{1,2}, Mark A. Jobling³, Jorge Rocha⁴, John Novembre⁵, Martin B. Richards⁶, Connie Mulligan⁷, Chiara Batini^{2*} & Franz Manni⁸

1) *Istituto Italiano di Antropologia, Roma, Italy* - e-mail: isita@isita-org.com

2) *Department of Animal and Human Biology, University of Rome "La Sapienza"*

3) *Department of Genetics, University of Leicester, UK* - e-mail: maj4@leicester.ac.uk

4) *IPATIMUP - Institute of Pathology and Molecular Immunology of the University of Porto; Department of Biology, Faculty of Sciences, University of Porto* - e-mail: jrocha@ipatimup.pt

5) *Department of Ecology and Evolutionary Biology; Interdepartmental Program in Bioinformatics; University of California, Los Angeles, California, USA* - e-mail: jnovembre@ucla.edu

6) *Institute of Integrative & Comparative Biology, Faculty of Biological Sciences, University of Leeds, UK*
e-mail: m.b.richards@leeds.ac.uk

7) *Department of Anthropology, University of Florida, USA* - e-mail: cmulligan@ufl.edu

8) *CNRS UMR 7206 - USM 104 National Museum of Natural History - Musée de l'Homme, Paris, France*
e-mail: manni@mnhn.fr

9) *Department of Genetics, University of Leicester, UK (current address)* - e-mail: cb334@leicester.ac.uk

Summary – *Molecular Anthropology is a relatively young field of research. In fact, less than 50 years have passed since the symposium "Classification and Human Evolution" (1962, Burg Wartenstein, Austria), where the term was formally introduced by Emil Zuckerkandl. In this time, Molecular Anthropology has developed both methodologically and theoretically and extended its applications, so covering key aspects of human evolution such as the reconstruction of the history of human populations and peopling processes, the characterization of DNA in extinct humans and the role of adaptive processes in shaping the genetic diversity of our species. In the current scientific panorama, molecular anthropologists have to face a double challenge. As members of the anthropological community, we are strongly committed to the integration of biological findings and other lines of evidence (e.g. linguistic and archaeological), while keeping in line with methodological innovations which are moving the approach from the genetic to the genomic level. In this framework, the meeting "DNA Polymorphisms in Human Populations: Molecular Anthropology in the Genomic Era" (Rome, December 3-5, 2009) offered an opportunity for discussion among scholars from different disciplines, while paying attention to the impact of recent methodological innovations. Here we present an overview of the meeting and discuss perspectives and prospects of Molecular Anthropology in the genomic era.*

Keywords – *Genomics, Human populations, Genetic variation, Interdisciplinary approaches.*

The dawn of Molecular Anthropology can, at least formally, be traced back to the 1962 symposium "Classification and Human Evolution"

at Burg Wartenstein in Austria. In that context, the American biologist of Austrian origin Emil Zuckerkandl first introduced the term "Molecular

Anthropology” to designate the study of primate phylogeny and human evolution through the genetic information decoded by proteins and polynucleotides (Sommer, 2008).

In the fifty years since that symposium, Molecular Anthropology has not only moved its focus onto the molecule which encodes the genetic information, deoxyribonucleic acid (DNA), but it has also extended its application well beyond the aspects initially implied by Zuckerkandl, so becoming one of the most promising and rapidly growing sub-fields of Anthropology. In fact, insights into several important issues have been obtained using a molecular approach, leading to a substantial advance in our knowledge of various key aspects of human evolution. These include, among others, the reconstruction of the history of human populations and peopling processes, the characterization of DNA in extinct humans and ancient populations and the role of adaptive processes in shaping the genetic diversity of our species (Jobling *et al.*, 2004).

The pioneering study on mitochondrial variation in worldwide populations by Rebecca Cann and coworkers in the late eighties is one of the most celebrated applications of Molecular Anthropology, due to its important implications for the understanding of the origin and diffusion of anatomically modern *Homo sapiens* (Cann *et al.*, 1987). Their findings were claimed to be a substantial argument in favour of the recent African origin (RAO) of our species and led to the spread of the popular concept of “mitochondrial Eve”. The initial results have been subsequently challenged by further studies which have extended and improved sampling, increased genetic information and incorporated demographic aspects (e.g. Vigilant *et al.*, 1991; Templeton, 1992; Relethford, 1998). Interestingly, it was soon understood that the genetic evidence, although powerful, needs to be considered jointly with paleontological and archaeological evidence in order to achieve a more comprehensive view on the emergence of our species and evaluate the relevant hypotheses more carefully. This was thoroughly and elegantly pursued by Chris Stringer and Peter Andrews in their seminal paper “Genetic and fossil evidence for

the origin of modern humans” (1988). Such contribution is also worth noting for the systematic comparison between theoretical expectations and findings of the RAO and multiregional models on modern human evolution, providing an alternative to most of the previous papers based on descriptive and circumstantial approaches.

The Human Genome Diversity Project (HGDP) may be regarded to as another turning point for Molecular Anthropology. Promoted by Luigi Luca Cavalli-Sforza and others in the early 1990s, HGDP aimed to explore human differences and history by looking at genomes from numerous indigenous populations across the globe, involving anthropologists, geneticists, medical doctors, linguists, and other scholars (Cavalli-Sforza, 2005). This project was designed to offer an opportunity for systematic research providing a shared set of DNA samples to laboratories working on human genetic variation, which was obtained through the use of immortalized lymphoblastoid cells collected from populations of particular anthropological interest (Cann *et al.*, 2002). Unfortunately, HGDP also raised important controversies, mostly of ethical nature (Ikilic & Paul, 2009), which slowed down the initiative. Nonetheless, HGDP played a key role in the change of perspective of Molecular Anthropology from genetics to genomics, coherently with its mission to explore the mutual benefits between groups involved in the initiatives of the Human Genome Organization (HUGO) and the laboratories working on human diversity. The most recent large-scale projects aimed at analysing human variation include HapMap and 1000Genomes (International HapMap Consortium, 2003; Via *et al.*, 2010). Either through the analysis of common variants, in the former, or through the discovery of rare variants within different genomes, in the latter, both these initiatives are moving the focus of human diversity studies to the genomic level.

Those mentioned above can be considered as paradigmatic examples of the double challenge that molecular anthropologists have to face even in the current scientific panorama. In fact, as members of a discipline, Anthropology,

which is strongly committed to the integration of different forms of knowledge, we need to foster the debate with researchers belonging to sister disciplines (e.g. linguists, archaeologists and primatologists). At the same time, the increasing demand for exhaustive analyses of human genome variation requires constant methodological and theoretical updates, which constitute a drive towards increasing specialization. While the promotion of interdisciplinary debate in Molecular Anthropology has already been at the centre of various initiatives, among which the series of conferences organized by Colin Renfrew are especially worth mentioning (Renfrew & Boyle, 2000; Bellwood & Renfrew, 2002; Forster & Renfrew, 2006), the continuous development of genomic approaches to human diversity is opening unprecedented opportunities and raising further issues which render a renewed focus necessary.

Coherently with this background, the meeting “DNA Polymorphisms in Human Populations: Molecular Anthropology in the Genomic Era”, organized in Rome (December 3-5, 2009) by the Istituto Italiano di Antropologia and the National Museum of Natural History of Paris, offered an opportunity to foster dialogue among researchers from different disciplines, while paying attention to the impact of recent innovations in theory and practice of molecular studies on human evolution. The first three sessions provided an updated view of the genetic variability continent-by-continent and highlighted the issues that still require investigation. Topics under discussion included both results and inferences obtained through “traditional” approaches (e.g. data from unilinear markers) as well as new and next-generation DNA sequencing methods. The closing session was dedicated to the dialogue about theoretical and practical aspects of interdisciplinary interactions in the Genomic era, putting molecular anthropologists face-to-face with researchers from Paleoanthropology, Archaeology, Linguistics and Medicine (see the JASs forum “Molecular Anthropology in the Genomic era: interdisciplinary perspectives” in this JASs issue). All the abstracts of oral and

poster presentations are available at <http://www.isita-org.com/MolAnthroGenomics/2009.htm>.

Here we summarize the contents of the invited lectures from the Congress and comment on some issues raised during the meeting. This report does not only aim to provide JASs readers with an overview of the Congress, but could also represent a useful reference for future initiatives designed to evaluate the state of the art and discuss perspectives and prospects of Molecular Anthropology in the genomic era.

Molecular anthropology in the genomic era, an overview¹

Molecular Anthropology: past and present

The first attempts to understand the history of human population movement, demographic change and admixture through genetics used protein markers, such as blood groups and HLA (Cavalli-Sforza *et al.*, 1994). We now suspect that the diversity of these markers is strongly influenced by natural selection, and researchers interested in investigating human history have since sought neutral markers, regarding phenotypes and adaptive influences as a disturbance. Prominent amongst these markers have been the non-recombining region of the Y chromosome and mitochondrial (mt)DNA, despite ongoing concerns about regional selection on the latter (Balloux *et al.*, 2009), and most major questions and many populations have now been addressed to some degree using small numbers of informative sites on these loci. Their uniparental modes of inheritance continue to illuminate sex-biased processes, and the coinheritance of Y haplotypes with patrilineal surnames allows the exploitation of these cultural labels in the investigation of past population structures (King *et al.*, 2009). Issues of ascertainment bias of markers here are fading with the use of multiple Y-STRs and increasing numbers of Y-SNPs, and with increased resolution of mtDNA analysis. The entire mtDNA (approximately 16.5kb) can be now readily sequenced in

¹ lecture presented by Mark A. Jobling

many individuals, whereas for the Y-chromosome only a small number of the available STRs are routinely analysed, and the resequencing of megabases of this chromosome is now possible with the careful application of new technologies. This reveals hundreds of new SNPs per chromosome analysed, posing challenges for unifying datasets and standardizing methodology and nomenclature. Recent sequence analyses of Y chromosomes separated by only a few generations have identified lineage-specific markers (Xue *et al.*, 2009), which represents an important step towards reaching the phylogenetic resolution needed to distinguish between different migration events which are very close in time. Members of the general public, through their obsessions with genetic genealogy, are helping to provide useful scientific data. Genome-wide SNP typing is now affordable and offers interesting insights into the geographical patterning of common autosomal variation (Novembre *et al.*, 2008). It suffers from the Eurocentric ascertainment bias of common SNPs, and a similar bias in the population distribution of available genome-wide association study data (Need & Goldstein, 2009). Because of the tag-SNP-based designs of marker sets, it also lacks much of the potential temporal resolution provided by the evolutionary relationships among haplotypes. Conventional resequencing of multiple specific X-chromosomal and autosomal segments, and the typing of markers in low-recombination regions, can provide some of this resolution, and has thrown light on the history of sex-specific behaviours (Hammer *et al.*, 2008).

Using genetics to test hypotheses based on historical, archaeological or linguistic evidence often uses a 'cherry-picking' approach when considering the other disciplines, which lacks objectivity. Although most of the tractable questions seem likely to be those linked to relatively recent events, one of the most impressive findings of recent years has been the remarkable explanatory power of simple distance from East Africa for patterns of modern genetic diversity (Ramachandran *et al.*, 2005), underscoring the importance of early events when populations were small.

From phenotype to genotype (and back)

By contrast, there are researchers who regard phenotypes and selection as the important issues, and population structure and history as the distraction. Unfortunately, although the phenotypes of humans are of particularly interest, our species is not a model organism. The kinds of controlled experiments we might carry out on mice are impossible (Terwilliger & Lee, 2007), so we must make do with the 'experiments of nature' represented by anthropologically interesting populations, while at the same time trying to account for the complex influence of a complex environment that includes the epitome of defining human complex phenotypes, culture. Some anthropologically interesting phenotypes are yielding to the power of genetic and genomic analysis, including resistance or susceptibility to some pathogens, dietary adaptation, pigmentation, hair thickness and tooth morphology (Kimura *et al.*, 2009). Other traits promise to be less tractable, with the tractability depending on the often unknown underlying genetic architecture. Stature is a good example - in outbred populations in the developed world, dozens of loci have been identified in huge samples, but each contributes only a tiny amount (a few millimetres) to the variance of the trait. Tellingly, Francis Galton's Victorian back-of-an-envelope approach to height prediction greatly outperforms the technological might of twenty-first century genomics (Aulchenko *et al.*, 2009). Here, the common-disease-common-variant hypothesis seems to be losing the battle to hypothetical copy-number variants, rare mutations, gene-gene interactions and epigenetics (Manolio *et al.*, 2009). Short stature among pygmy populations is a well-known example of an anthropologically interesting phenotype, but its elucidation falls foul of the problem of unknown genetic architecture, both within and between populations. If one or a few loci explain it, and if candidate loci translate from Europe to the rest of the world, then simple approaches may bear fruit. But if, as seems likely, the trait is complex and multi-genic, then it will more difficult to understand. We may hypothesise a common origin of pygmy

groups to explain the common phenotype (Patin *et al.*, 2009), but this would make it difficult to pinpoint the specific locus or loci responsible for the phenotype amongst the loci shared simply through recent common origin. Moreover, the detection of phenotypically important loci within populations will be difficult because of small sample sizes, and grant applications (often damned by reviewers as ‘fishing expeditions’) will tend to face the insoluble problem of power calculations. The role of natural selection in the development of short stature is mysterious, and certainly more complex than simple ‘Just So’ stories based on the ease of moving about in forests (Migliano *et al.*, 2007). Even when we can see clear selective advantages in particular adaptation, the problem of drift represents one of the major difficulties of studies of poorly understood phenotypes. We can use genome-wide approaches to seek segments of DNA showing frequency elevations in populations living, for example, at high altitude, but how do we distinguish between adaptation and drift as explanations for frequency differences? And can we identify suitable control populations, in which drift has not also been a problem? If we want to support findings by ‘replication’ in other high-altitude populations, we face the problem that the adaptation may have arisen independently, and may even have a different physiological and genetic basis. It seems likely that admixture-based approaches will be useful here. In the distance, however, lies the brave and bright new world of whole genome sequences (www.1000genomes.org; Via *et al.*, 2010), uncompromised by ascertainment bias and rich with rare variants – recent investigations of African genome sequences are already starting to show how much diversity will be revealed (Schuster *et al.*, 2010). Although the new methods are still too expensive to be applied to most anthropologically interesting samples, this is likely to change soon, and molecular anthropologists should learn how to mine and use such sequences, and think what questions they would like to address with them. Surely, the more sequences, the better? If we knew the sequences of all the genomes of everyone, we

would be able to learn everything that could be learned about the relationships among individuals and populations, the processes of mutation, and the influence of selection. It seems likely that the quality of recording and classification of the environments and the phenotypes (Samuels *et al.*, 2009), rather than the genotypes, will then become the crucial factor, and the anthropologists (and the ethicists) will inherit the world.

The peopling of Africa²

Despite Africa’s central role in human evolution, African populations have been less well characterized than other groups in most studies addressing human genetic variation. Until recently, inferences about human population history typically relied on few African populations that were assumed to be representative of the whole continental diversity. While this limitation did not challenge the validity of general conclusions about the origins and global distribution of human genetic variability, insufficient sampling has certainly hampered our perception of how human diversity was shaped within Africa. With the highest time depth of human history and over 2000 ethnolinguistic groups dwelling in landscapes that range from the driest deserts to the most humid forests, Africa could hardly be understood without a more comprehensive population sampling.

In the last decade, improvements in sampling coverage, together with the increasing availability of highly informative genetic markers and the use of new approaches regarding data analysis, had a tremendous impact in the assessment of Africa’s genetic variation. Although the amount and quality of genetic data is still far from being fully satisfactory, the current genetic portrait of Africa has reached an unprecedented level of precision.

Unilinear markers

A significant part of our present understanding of African genetic variation is based on the study of mitochondrial DNA (mtDNA) and the non-

² lecture presented by Jorge Rocha

recombining portion of the Y chromosome (NRY) (Cruciani *et al.*, 2002; Salas *et al.*, 2002). Because of their uniparental patterns of inheritance and lower effective population size, mtDNA and NRY haplotypes provide complementary information about female- and male-specific aspects of genetic variation and are especially sensitive to the effects of drift. MtDNA and NRY markers tend to be highly geographically structured and, due to lack of recombination, haplotype phylogenies can be easily reconstructed, providing a temporal framework for mutation accumulation, which can be related to the geographic distribution of different lineages. Several NRY and mtDNA haplogroups are particularly informative because their origins appear to be geographically and temporally distinct from each other. For example, the distribution of the oldest basal NRY-haplogroup A-M91 suggests an ancestral link of the southern African Khoe-San click-speaking groups to East Africa. The relatively old NRY B2b-M112 haplogroup points to the common ancestry of Khoe-San and Pygmy hunter-gatherer groups. A lineage within the younger E3b-M35* paragroup suggests that pastoralism might have been introduced to southern African from East Africa prior to Bantu migrations. The relatively young E3a-M2 haplogroup is widespread in Niger-Kordofonian-speaking populations and provides a marker for the expansion of Bantu-speaking agriculturists. Among the mtDNA haplotypes, the basal L0d clade is almost exclusive to the southern African Khoe-San but is also found in the click-speaking Sandwe from Tanzania confirming the ancient link of the Khoe-San to Eastern Africa. The younger haplogroup L1c, which probably originated in central Africa, is crucial to assess the ancestral relationship between western Pygmy hunter-gatherers and their neighboring Bantu-speaking farmers.

A multilocus approach

An important limitation of studies based on the NRY and mtDNA markers is that they amount to the characterization of only two genetic systems, which, due to the stochasticity of evolutionary processes, are insufficiently robust to generate meaningful estimates of

relevant population history parameters. Multilocus approaches designed to overcome this difficulty have received a remarkable boost with the recent publication of Tishkoff's landmark study on 2,432 individuals from 113 populations using a panel of 1,327 polymorphic markers (Tishkoff *et al.*, 2009). In brief, the study showed that most African genetic variation can be sorted into 14 ancestral population clusters and that most populations exhibited high levels of mixed ancestry, consistent with historical migrations across the continent. Consideration of geographic data along with clustering analysis distinguished five major groups of clusters, including (Fig. 1): i) a contiguous northern fringe encompassing Berber, Cushitic and Semitic Afroasiatic speakers from Saharan and East Africa; ii) a widespread group corresponding to the distribution of the Niger-Kordofonian language family (paralleled by the distribution of NRY haplogroup E3a-M2); iii) another group comprising Chadic and Nilo-Saharan-speaking populations from Nigeria, Cameroon, Chad and southern Sudan (some of which share a lineage within NRY haplogroup R that may have been introduced into Africa by a back migration originating in Asia; Cruciani *et al.*, 2002); iv) a group with Nilo-Saharan and Cushitic-speaking populations from Sudan, Kenya and Tanzania; and v) a group of noncontiguous geographic distribution consisting of Pygmy and southern Africa Khoe-San populations, providing evidence for shared ancestry among hunter-gatherers (consistent with the distribution of NRY haplogroup B2b, but not with mtDNA, since the three main hunter-gatherer groups are characterized by very distinct haplogroups.) In spite of the major advances provided by this study, it is important to note that regions like the Sahel, the Atlantic West Africa, Namibia, Angola and the central corridor comprising the DR of Congo, Central Zimbabwe and the Zambia, remain sparsely sampled. On the other hand, to make full use of the framework provided by Tishkoff's investigation, it is crucial to generate increasingly comparable datasets. This could be achieved by defining a minimum subset of highly informative markers to be used in future works concerning other African populations.

Prospects for future studies

To disentangle the spatial-temporal processes that gave rise to the emergent portrait of African genetic diversity, it will be important to address both deep-time and more fine-scale questions, combining continent-wide studies with more detailed pictures provided by regional or local case studies. Moreover, an interesting approach to interpret the basic properties of the observed genetic variation is to focus on discordance among different sets of genetic data, or between genetic data and non-genetic aspects of human variation. For example, the discrepancy between the patterns of genetic variation in NRY and mtDNA has provided important insights about the influence of sociocultural factors in shaping differences in male and female migration rates and effective sizes (Destro-Bisol *et al.*, 2004). Discordance between levels and patterns of genetic variation in nuclear and uniparental markers may be useful to reduce the number of population history models that are compatible with the data. On the other hand, differences between geographic patterns at putatively selected loci and neutral loci may be used to evaluate the strength of selection and to analyze the influence of demographic processes in spreading selected variants (Coop *et al.*, 2009). Finally, dissociation of common trends in the relationships between genetics, linguistics and lifestyles provide unique opportunities to analyze the impact of admixture between different populations and to analyze how major shifts in genetic and cultural patterns occur. For example, interactions among the peoples of southern Angola has generated intriguingly discordant combinations of ethnicity, language and lifestyle (Coelho *et al.*, 2009).

A final aspect of the recent advances in understanding genetic diversity within Africa is related to data analysis. Datasets based on multiple, independently evolving genetic systems are particularly well suited to simulation-based inferential frameworks which aim to distinguish between alternative models of population history and to estimate key microevolutionary parameters under a given model. Recent applications of rejection algorithms and Approximate Bayesian Computation to infer the branching history of

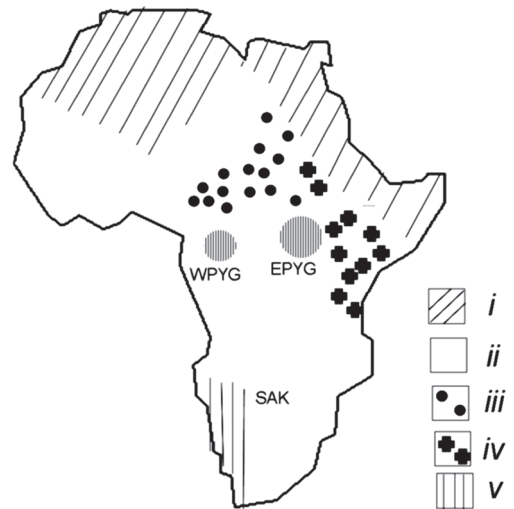


Fig. 1 - Geographic location of major groups of ancestral clusters within Africa. Patterns are identified by the number used in the text to describe the population composition of each group of clusters. WPYG=western Pygmies; EPYG=Eastern Pygmies; SAK= southern African Khoes-Sans. For details see text (modified from Tishkoff *et al.*, 2009).

Pygmy and agricultural populations provide excellent examples of the usefulness of new computational methods in addressing population history in Africa (Patin *et al.*, 2009; Verdu *et al.*, 2009). With the rapid accumulation of multilocus genotype data and the significant increase in sampling density, it is expected that similar inferential frameworks will be successfully extended to explicit geographical modeling of human dispersals within Africa.

Maps and migrations: insights to the genetic structure of Europe from SNP data and PC analysis³

The genetic variation of European individuals has been one of the most carefully characterized throughout the world and arguably across

³ lecture presented by John Novembre

any species. Major pre-historic events invoked to explain European genetic variation include the initial colonization of Europe, contractions and expansions from glacial refugia due to Pleistocene climate change, demographic expansions associated with the Neolithic innovation of agriculture in the Near East, and more recent population movements, such as those associated with the early medieval period (Barbujani & Goldstein, 2004). Resolving which events had a predominant effect on European genetic variation has been a long-standing goal of anthropological genetics, and is also relevant to the design and interpretation of genome-wide association studies, population genetics tests of recent positive selection, and personal ancestry testing. Despite the intensive attention, basic questions still remain unanswered regarding what the dominant patterns are in European genetic variation and what ancestral events explain them.

Recent progress has been made due to the advent of high throughput SNP genotyping technologies, which have made it possible to examine patterns of genetic variation in European samples at an unprecedented scale. The application of SNP genotyping technology to European populations has been facilitated by a growing recognition of the importance of population genetic variation for mapping the variants underlying heritable disease traits and pharmacogenomic traits. For example, several thousand European individuals were sampled and genotyped using the Affymetrix 500K SNP genotyping platform as part of a collaboration between GlaxoSmithKline and academic scientists (the POPRES project, Nelson *et al.*, 2008; Novembre *et al.*, 2008; Auton *et al.*, 2009; data available via dbGAP).

To analyze patterns of variation in such a large set of polymorphic loci, researchers have been turning towards multivariate statistical methods, chiefly principal components analysis (PCA). While PCA was first pioneered in the 1970s to summarize patterns in sample allele frequencies (e.g. Menozzi *et al.*, 1978), a novel form of individual-based PCA has recently become popular for analyzing SNP data (e.g. Price *et al.*, 2006). This resurgence of PCA is mainly due to

the fact that when doing genome-wide association mapping for disease susceptibility loci, PC coordinates can be used as covariates to control for population stratification. Furthermore, individual-based PCA has been argued to be attractive because it does not presume pre-defined groups, nor does it assume a discrete set of ancestral populations.

An individual-based PCA plot of the European POPRES individuals shows a striking resemblance to geographic maps of Europe (Novembre *et al.*, 2008; see Fig. 2). These results stand in contrast to alternative possibilities, such as clustering of European populations by language family (e.g. Romance, Slavic, Germanic languages). Notably, Hungarian individuals in the sample cluster with their geographic neighbours, a result which one might have found surprising given they are local linguistic outliers because they speak a non-Indo-European language. PCA analyses by other groups at both similar (e.g. Lao *et al.*, 2008; Heath *et al.*, 2008) and finer spatial scales (e.g. within Finland and Iceland, e.g. Lao *et al.*, 2008; Sabatti *et al.*, 2009) also evidence plots that resemble the geographic arrangement of populations (although in some cases the influence of relative sample sizes and/or the presence of outlier populations distorts the basic pattern).

Why these PCA plots resemble geographic maps at all is an interesting question. Insight can be gained mathematically by considering cases in which sampling is roughly uniform across space, and the pattern of observed covariance amongst individuals decays with geographic distance (an isolation-by-distance pattern). In these settings, PC coordinates will typically be a function of the geographic position of each individual, and PC1 and PC2 will form perpendicular gradients over geographic space (Novembre & Stephens, 2008). This behaviour of PCA has been understood in essence by some sub-disciplines of science (e.g. meteorology, image analysis) for some time, but their relevance was only recently noted within the population genetics community (Novembre & Stephens, 2008). Importantly, the map-producing behaviour of PCA is based on observed

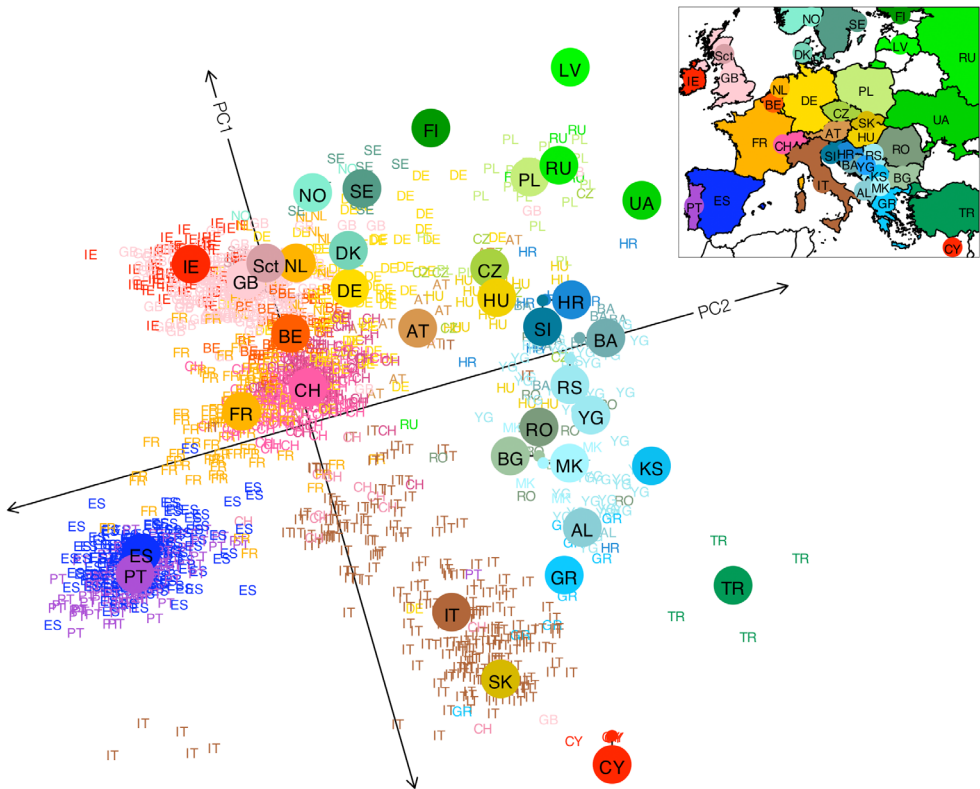


Fig. 2 – A principal component representation of genetic data from 1,387 Europeans (reprinted from Novembre *et al.*, 2008). List of abbreviations: AL, Albania; AT, Austria; BA, Bosnia-Herzegovina; BE, Belgium; BG, Bulgaria; CH, Switzerland; CY, Cyprus; CZ, Czech Republic; DE, Germany; DK, Denmark; ES, Spain; FI, Finland; FR, France; GB, United Kingdom; GR, Greece; HR, Croatia; HU, Hungary; IE, Ireland; IT, Italy; KS, Kosovo; LV, Latvia; MK, Macedonia; NO, Norway; NL, Netherlands; PL, Poland; PT, Portugal; RO, Romania; RS, Serbia and Montenegro; RU, Russia, Sct, Scotland; SE, Sweden; SI, Slovenia; SK, Slovakia; TR, Turkey; UA, Ukraine; YG, Yugoslavia. See Novembre *et al.* (2008) for further details.

patterns of spatial covariation in the data. Because various processes or events may give rise to the same pattern in which covariances decay with distance, it is still unclear which processes/events gives rise to the observed PCs in European populations. It certainly at some level must involve geographically restricted mating, but how much of the spatial covariance is due to on-going geographically restricted mating versus more ancestral population movements is unclear.

Another major question that remains from this initial round of SNP studies is: how do putative European population isolates fit into the

broader context of European genetic diversity and what does it suggest about the peopling of Europe? An exciting arena of future research is to use large panels of SNPs to understand the fine-scale relationships of population isolates to their geographic neighbours. Recent results from SNP studies in the Basque question whether the Basque are as isolated as previously supposed (Laayouni *et al.*, 2010; Garagnani *et al.*, 2009). On-going research is investigating the genetic origins of the Sorbs, a previously uncharacterized, Slavic-speaking putative isolate from Eastern Germany (Veeramah *et al.*, in preparation).

As studies of SNP diversity move forward, one important caution is that while PCA is a powerful tool for visualizing fine-scale population structure, PCA can be dependent on relative sample sizes (Novembre & Stephens 2009; McVean, 2009). As a result, the exact direction of PC1 can vary from study to study (contrast for example Novembre *et al.*, 2008 to Heath *et al.*, 2008). The expected PCA coordinates for each individual in a sample can be derived from average pair-wise coalescent times among individuals in the sample (McVean, 2009), and doing so helps explain observations that PCA is dependent on relative sample-sizes (Novembre & Stephens, 2009). In turn, we expect methods which are tailored to detect specific demographic signatures (e.g. the decay of diversity with distance from a putative origin) to be a powerful way forward in illuminating the peopling of Europe.

Archaeogenetics and the peopling of Asia⁴

Global patterns of human genetic diversity suggest that modern human variation is broadly (albeit shallowly) structured at continental level, with South Asia and East Asia (and probably also Southeast Asia) forming genetic clusters or domains distinct both from each other and from (Native) America, Australasia, west Eurasia and sub-Saharan Africa. This has been shown by analysing multiple autosomal microsatellites using the STRUCTURE software (Rosenberg *et al.*, 2006). However, evidence is accumulating, especially from the non-recombining marker systems, mitochondrial DNA (mtDNA) and the non-recombining part of the Y chromosome (NRY), that this is the result of sequential colonisation and expansion from very small founder groups who dispersed from an East African homeland within the last 70,000 years (ky) or so (Macaulay *et al.*, 2005; Metspalu *et al.*, 2006; Richards *et al.*, 2006).

Recent archaeological and fossil evidence suggests that anatomically modern humans were

settled in Southeast Asia by at least 50 kya, implying that South Asia was already inhabited by this time, although unequivocal evidence from the Subcontinent is more recent. Genetic estimates are much less precise, but a recent new calibration of the mtDNA mutation rate, which employs the entire variation in the mtDNA genome for maximum precision and makes allowance for the action of purifying selection, therefore also maximising accuracy, provides at least one molecular clock that can be employed for phylogeographic reconstructions (Soares *et al.*, 2009). This suggests that modern humans first settled in Asia 60–70 kya – somewhat earlier than the earliest widely accepted archaeological evidence, but matching some less widely accepted evidence from Australia and perhaps also China.

It was initially assumed that Eurasia had been settled by modern humans *via* northeast Africa and the Levant, ~50 kya, and Y-chromosome evidence has been used to argue for a Central Asian “heartland” from which much of the Old World was settled (Wells *et al.*, 2001). However, the aforementioned dating evidence from Australia suggested an earlier dispersal from the Horn of Africa across the Red Sea and along the tropical southern Asian coastline. This was supported by the extremely high number of basal mtDNA haplogroup R and M lineages in India (Sun *et al.*, 2006), and by similarities between industries associated with modern humans in South Africa ~60 kya and South Asia at least 35 kya (Mellars, 2006).

Analysis of complete mtDNA genomes sequences from so-called “relict” populations in South Asia, Southeast Asia and Australasia have been used to address this question. Modern non-African populations throughout the world, with the exception of populations or regions with a recent African ancestry, harbour mtDNAs from just three major founder clades, M, N and (nested closely within N) R, all of which belong to the L3 clade, which is of sub-Saharan African origin ~70 kya. Aboriginal populations in South Asia, Southeast Asia and Australasia display mtDNA profiles that include basal lineages belonging to all three of the mtDNA founder clades, indicating that even the most ancient populations on the

⁴ lecture presented by Martin B. Richards

southern coast of Asia were part of the same, single dispersal out of Africa (Macaulay *et al.*, 2005).

This pattern, and the molecular-clock timing of the dispersal to at least 60 kya, suggest that the primary expansion was along the southern coastal route, with the Asian continental heartland (including Southwest Asia, and ultimately Europe) taking place subsequently along various corridors as climatic conditions allowed, most likely after 50 kya. These dates seem to exclude the possibility, suggested on archaeological grounds as well as on earlier genetic analyses, that the dispersal into South and Southeast Asia took place before the volcanic eruption of Toba in Sumatra ~74 kya, which is therefore unlikely to have had an impact on Asian populations. Moreover, the dispersal seems to have been extremely rapid, within the space of a few thousand years, since it led to the divergence of the distinct domains of basal mtDNA lineages in each region, rather than a pattern of nesting (such as occurred in the settlement of the Americas from East Asia and the Remote Pacific from Southeast Asia/Near Oceania).

There is relatively little differentiation between ethnic and language groups within South Asia, which is similar to other parts of Eurasia. The Indian Subcontinent has long been seen as having been deeply affected by migrations from the north, and the non-recombining markers and autosomal SNP analysis indeed suggest genetic gradients, but these have arisen from a variety of distinct prehistoric dispersals, with little or no impact attributable to the putative Aryan migrations that are thought to have led to the establishment of the caste system. There are mtDNAs in India that originated in Southwest Asia but they probably arrived not long from the time of first settlement, and only a tiny minority that appear to have arrived during historical times. The demic impact of the Southwest Asian Neolithic appears to have been similarly minor for most of the Subcontinent, despite some claims to the contrary (Chaubey *et al.*, 2006).

Southeast Asia was settled by the southern coastal route by ~55 kya according to the mtDNA clock, when much of Island Southeast

Asia formed part of the mainland as the Sunda continent. Dental patterns, as well as genetic diversity, suggest that East Asia was initially settled from the south, although there is a suggestion in Y-chromosome patterns of an early offshoot from the southern route east of the Himalayas into the region of the Tibetan plateau, sometimes referred to as the “mammoth steppe”. The northeast Asian coast was reached at least 30 kya; some mtDNA and Y-chromosome lineages in Japan appear to trace to this time. Genetic and fossil data indicate discontinuities in the prehistory of East Asia; there are suggestions of subsequent re-dispersals from north to south, which may be in part due to Neolithic expansions, but seem likely to also reflect the expansion of Han Chinese people within the last 1,500 years or so. The impact of the Last Glacial Maximum is also likely to have been severe in continental East Asia, whereas refugial areas existed within Southeast Asia. Sea-level rises beginning ~19 kya had their maximal impact, however, in Southeast Asia; the Sunda continent was inundated leading to wide scale dispersals of lineages across what is now Island Southeast Asia which may have had a much greater demographic impact than the subsequent Holocene spread of the Neolithic across Southeast Asia and into the Pacific islands (Soares *et al.*, 2008).

A Genetic Perspective on Peopling of the Americas⁵

The colonization of the Americas represents the most recent major human occupation of an uninhabited land mass on the planet. The recency of this event suggests that it may have left a substantial signature in the genome. Therefore, we may be able to ask increasingly specific questions and provide more detailed information about this process than for other older and more complicated processes such as the initial migration of anatomically modern humans out of Africa. There are certain aspects of the colonization that are agreed upon by the scientific community, i.e.

⁵ lecture presented by Connie Mulligan

a single migration originated from an East Asian source and crossed over the Bering land bridge before entering North America (summarized in Fig. 2 and Kitchen *et al.*, 2008). This process created a strong population bottleneck such that modern Native Americans show significant reductions in genetic variation relative to other global populations and, furthermore, genetic variation throughout the Americas shows evidence of substantial genetic drift. Less consensus has been reached for other parameters of the colonization process such as the timing of the migration (both leaving Asia and entering the Americas), size of the founding population, nature of the migration from Asia (continuous movement versus several short-range migrations), and migration route(s) taken within the Americas.

Consensus on peopling of the Americas

An East Asian source population for all indigenous Native Americans, most likely around the Lake Baikal region, is widely accepted based on mtDNA and Y chromosome data. The alternative idea of an early European migration to the Americas prior to Columbus' voyage in the 1490s to account for some Native American genetic diversity was once proposed based on presumed Caucasoid features of the famous 'Kennewick Man' discovered in the state of Washington; support for this idea has largely disappeared based on comparative skeletal analyses. The number of migrations to the Americas was initially under debate, but has converged on a single migration based on a wealth of data including mitochondrial DNA (mtDNA), Y chromosome markers, short nuclear DNA sequences, and autosomal microsatellite markers (Mulligan *et al.*, 2004; Wang *et al.*, 2007; Fagundes *et al.*, 2008) and most recently, X chromosome sequence and nuclear single nucleotide polymorphism (SNP) data (Bourgeois *et al.*, 2009; Gutenkunst *et al.*, 2009). Furthermore, most geneticists believe there was virtually no ancient gene flow between Asia and the Americas after the initial migration, likely reflecting inundation of the exposed Bering land bridge after the last glacial maximum (LGM) ~18-23 kya.

Once humans entered the Americas, it appears that their movement may have been very rapid based on archaeological evidence of human occupation at Monte Verde at the southern extent of South America ~14.5 kya (Dillehay, 2008). Simple simulation studies show that a rapid expansion is necessary to maintain frequencies of the major mitochondrial haplogroups into the southern reaches of the Americas (Fix, 2004). Empirical and simulation data suggest that genetic drift has played a significant role in determining patterns of Native American genetic diversity as evidenced by greater differentiation and population structure throughout the Americas relative to other continents, reflecting the rapid dispersal, small population size, and genetic isolation of Native American groups. Native American genetic diversity also shows evidence of substantial admixture, particularly through the incursion of European Y chromosomes (Wang *et al.*, 2007).

Debated points on peopling of the Americas

Of the issues still under active debate, the timing of the migration is a critical point. First, it must be established that there are at least two relevant dates, the migration out of Asia and the entry into the Americas. The first date is generally based on the initial diversification of New World-specific haplogroups. For example, mtDNA data support a date of ~30-40 kya (Bonatto & Salzano, 1997), reflecting the initial diversification of New World genetic variation as the populations diverged from ancestral Asians but prior to their entry into the New World. The timing of entry to the Americas is more debated and dates generally fall into periods that are pre- and post-LGM. Different dates are frequently based on similar mtDNA datasets but use different mitochondrial genome substitution rates, i.e. 'fast' substitution rates (e.g. $\sim 1.7 \times 10^{-8}$ substitutions/site/year) support a post-LGM entry and 'slow' substitution rates (e.g. $\sim 1.26 \times 10^{-8}$ substitutions/site/year) support a pre-LGM entry. Endicott & Ho (2008) recommend that substitution rate estimates should be based on an 'internal calibration' of the underlying phylogeny used in the rate estimation; their estimates of the mitochondrial coding genome

substitution rate generally support younger dates, i.e. post-LGM entry.

The tempo of the migration has recently received widespread attention, e.g. Tamm *et al.* 2007. This issue can be viewed as an investigation of the movement of people (was it a continuous movement or a series of short-range migrations?) or a focus on when (and where) did the genetic variation that is specific to and ubiquitous throughout the New World occur? There are mitochondrial variants that define New World-specific haplogroups, e.g. C1b, C1d, X2a (Tamm *et al.*, 2007) prompting researchers to propose a period of population isolation prior to expansion into the Americas (first mentioned by Bonatto & Salzano in 1997). Mulligan *et al.* (2008) estimated that ~7000-15,000 years were required to generate the New World-specific variation. It has been further proposed that the migrating population occupied Beringia during this period of isolation. Paleocological data from ancient eastern Beringia are indicative of productive, dry grassland suggesting that Beringia was able to sustain at least small populations of humans and other large mammals. The lack of archaeological data for human occupation of Beringia most likely reflects the fact that the proposed occupation sites are now inundated.

The size of the founding population has also been the subject of considerable study. New estimates based on mtDNA coding genomes and short nuclear sequences support an effective population size of ~1,000-2,000 individuals (Fagundes *et al.*, 2007; Mulligan *et al.*, 2008). Once the population entered the Americas, there is considerable interest in determining the exact route(s) taken by the migrants. The distribution of two specific mtDNA haplogroups was used to support both coastal and inland routes (Perego *et al.*, 2009), but simulation and empirical studies of whole mitochondrial genomes and hundreds of autosomal microsatellite markers strongly support coastal routes over inland routes (Fix, 2004; Wang *et al.*, 2007; Fagundes *et al.*, 2008).

Future research

There are multiple aspects of the peopling of the Americas that are still subject to debate and,

thus, warrant attention. **1)** Better estimates of substitution rates, both mitochondrial and nuclear, are necessary to provide robust support for age estimates of key events within the colonization process. This is particularly true for estimates of entry to the Americas since a pre-LGM entry implies that the migrant population overcame severe climatic and geologic, i.e. North American ice sheets, obstacles to survive that would not have been present if their entry postdated the LGM. **2)** A better understanding of the period prior to entry into the Americas is also worthy of study, i.e. Was Beringia the occupied land mass? How long was the occupation? What proportion of the population actually entered the Americas? **3)** Continued investigation of patterns of genetic variation within the Americas is necessary in order to better understand the various regional colonization events that occurred after the initial entry into the Americas. Studies that look for correlation between genetics and linguistics have a checkered history in terms of providing general insights; most likely, correlation between linguistics and genetics will reflect unique regional histories and not general trends or processes during the course of colonization. **4)** There is a move towards more simulation of data and modeling of alternative evolutionary scenarios in addition to continued collection of empirical data. The simulation and modeling approaches have the advantage of statistically determining the goodness of fit between empirical data and alternative scenarios. For example, the support for a coastal and inland route within the Americas was supported by the differential distribution of two distinctive mitochondrial haplogroups (Perego *et al.*, 2009); it would be informative to know how often such a distribution occurs by random chance and, thus, if the actual distribution is sufficiently unique to require explanation via separate migration routes within the Americas. **5)** A broad perspective on the colonization process is also valuable. Comparison with other colonization processes, i.e. migration out of Africa, provides a complementary perspective and allows general inferences on the colonization process to be formulated.

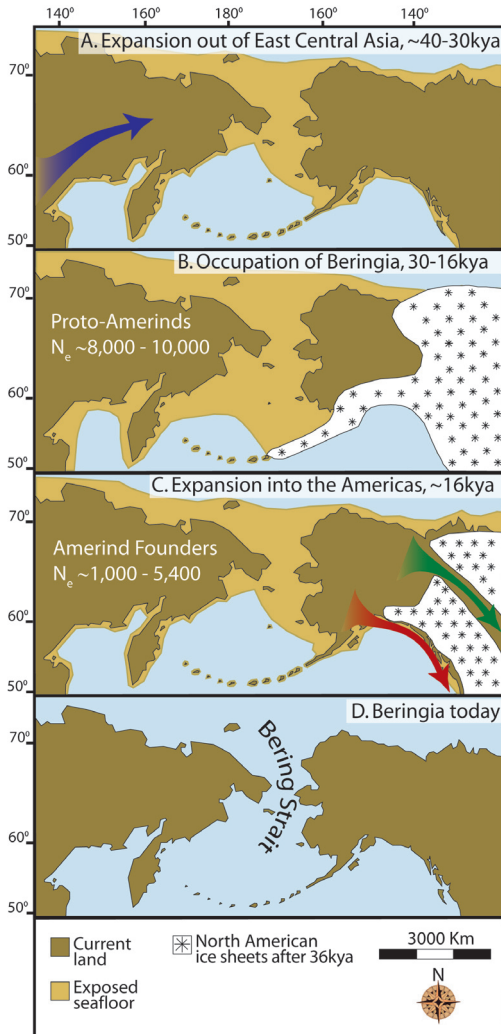


Fig. 3 - Maps depicting a three-step colonization model for the peopling of the Americas. (A) Divergence, then gradual population expansion of the Amerind ancestors from an East Central Asian gene pool (blue arrow). (B) Proto-Amerind occupation of Beringia with little to no population growth for ~15,000 years. (C) Rapid colonization of the New World by a founder group migrating southward through the ice free, inland corridor between the eastern Laurentide and western Cordilleran Ice Sheets (green arrow) and/or along the Pacific coast (red arrow). The lowest frame depicts Beringia as it is today. Modified from Kitchen et al. (2008).

Perspectives and prospects for Molecular Anthropology in the Genomic Era

The congress “DNA Polymorphisms in Human Populations: Molecular Anthropology in the Genomic Era” offered an important opportunity to scholars and students to discuss some topical aspects of research in human evolution. This initiative provided us with a picture of the various ways in which the genetic structure of human populations can be explored, showing the versatility of researchers in using different sampling schemes, exploring variation at diverse geographic scales, looking at genes which are neutral or amenable to selection and focusing on whole genomes or specific lineages.

A general impression we obtained from most of the presentations given in the course of the meeting is that modelling and comparison of evolutionary scenarios by data simulations are finally becoming a widespread alternative to the descriptive reports and ad-hoc explanations which have represented the standard for population studies up to a few years ago. In fact, many contributions have compared evolutionary histories using new computational methods which are being developed to take a growing number of variables into account. This substantial change of perspective seems to demonstrate the consciousness, acquired by Molecular Anthropologists, of the importance of moving towards hypothesis testing approaches, and follows the path set out by Stringer and Andrews (1988), with the further advantage of using quantitative methodologies. It may also stimulate further advancements, since the availability of multilocus data and dense sampling will, hopefully, make it possible to test spatial models of human migrations more carefully, which is another key issue in the reconstruction of the prehistory of our species.

Thanks to the genomic approach, some important results have been already achieved and further developments are to be expected. These include the recent calibration of the mtDNA mutation rate which takes into account the effect of purifying selection and makes phylogeographic reconstructions more reliable (Soares *et al.*, 2009).

Furthermore, it has been pointed out that the set up of broad panels of genetic informative loci which have proved useful for the investigation of large geographic areas inhabited by genetically heterogeneous populations, such as that studied by Tishkoff *et al.* (2009), provides a framework for optimizing cost/benefit ratios in future studies which aim to fill sampling gaps and gain a more complete picture of genetic diversity and population history.

Regarding future prospects, it has been envisioned that the power of genomic approaches will also help Anthropology overcome some of its inherent limitations. This could materialize if the study of entire genomes opens up new avenues for the identification of genetic determinants underlying complex phenotypes of special interest for human evolutionary biology, such as stature and high altitude adaptation.

While the advent of genomics is already revolutionizing research in Molecular Anthropology and promises to continue to do so in the near future, some interdisciplinary lines of anthropological research have maintained all their relevance or seem to be destined to attract even more interest. The lectures summarized here draw attention to the importance of studies of well defined populations to help clarify issues of general interest, such as the relations between cultural and biological changes or the assessment of hypotheses on routes of major migratory events in human prehistory. This is the case of human groups with unusual combinations of genetic, linguistic and lifestyle features in Africa, relict populations in Asia and Australasia and European isolates (see also Destro-Bisol *et al.*, 2008 in this Journal). It is also worth noting that new research avenues opened up by genomics revitalize interest in environmental aspects, viewed either as variables which act as co-determinants of phenotypic variation in complex traits, or paleo-ecological changes which could have had a deep impact on past human mass migrations.

In conclusion, our congress showed that the combination between interdisciplinary approaches and methodological and theoretical innovations has become an essential aspect for studies of human evolution at molecular level.

Even more importantly, we have learned that making this integration more complete and fruitful will be crucial in achieving new targets and will extend applications to other anthropological questions. We hope that “DNA Polymorphisms in Human Populations: Molecular Anthropology in the Genomic Era” will make a significant contribution in this direction.

Acknowledgements

The congress “DNA Polymorphisms in Human Populations: Molecular Anthropology in the Genomic Era” was supported by The University of Rome “La Sapienza”, Ministero della Ricerca e dell’Istruzione, Ministero per i Beni e le Attività Culturali, Ateneo Federato della Scienza e della Tecnica and by the journal “Human Biology” (Wayne State University Press, Detroit MI, USA). GDB acknowledges funds from the MIUR (PRIN project 2007-2009 n. 2007TYXE3X) and the Università di Roma “La Sapienza” (prot. C26A09EA9C/2009). MAJ is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant no. 087576), and gratefully acknowledges support from the Wellcome Trust. JR was partially financed by the research grants from Fundação para a Ciência e a Tecnologia (FCT): PPCDT/BIA-BDE/56654/2004 and PTDC/BIA-BDE/68999/2006. We express our gratitude to the members of the scientific committee: Cristian Capelli, Oscar Lao Grueso, Andres Moreno-Estrada and Paul Verdu. Finally, we warmly thank Paolo Anagnostou, Cinzia Battaglia, Valentina Coia, Vera Damiani, Francesco Montinaro, Veronica Marcari and Nancy Wise for their help during the Congress.

References

- Aulchenko Y.S., Struchalin M.V., Belonogova N.M., Axenovich T.I., Weedon M.N., Hofman A., Uitterlinden A.G., Kayser M., Oostra B.A., van Duijn C.M., Janssens A.C. & Borodin P.M. 2009. Predicting human height by Victorian and genomic methods. *Eur. J. Hum. Genet.*, 17: 1070-1075.

- Auton A., Bryc K., Boyko A.R., Lohmueller K.E., Novembre J., Reynolds A., Indap A., Wright M.H., Degenhardt J.D., Gutenkunst R.N., King K.S., Nelson M.R. & Bustamante C.D. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.*, 19:795-803.
- Balloux F., Handley L.J., Jombart T., Liu H. & Manica A. 2009. Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc. R. Soc. Lond. B. Sci.*, 276: 3447-3455.
- Barbujani G. & Goldstein D.B. 2004. Africans and Asians abroad: genetic diversity in Europe. *Annu. Rev. Genomics Hum. Genet.*, 5:119-150.
- Bellwood P. & Renfrew C. (eds) 2002. *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, University of Cambridge, UK.
- Bonato S.L. & Salzano F. M. 1997. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am. J. Hum. Genet.*, 61:1413-1423.
- Bourgeois S., Yotova V., Wang S., Bourtoumieu S., Moreau C., Michalski R., Moisan J.P., Hill K., Hurtado A.M., Ruiz-Linares A. & Labuda D. 2009. X-chromosome lineages and the settlement of the Americas. *Am. J. Phys. Anthropol.*, 140:417-428.
- Cann H.M., de Toma C., Cazes L., Legrand M.F., Morel V., Piouffre L., Bodmer J., Bodmer W.F., Bonne-Tamir B., Cambon-Thomsen A., Chen Z., Chu J., Carcassi C., Contu L., Du R., Excoffier L., Ferrara G.B., Friedlaender J.S., Groot H., Gurwitz D., Jenkins T., Herrera R.J., Huang X., Kidd J., Kidd K.K., Langaney A., Lin A.A., Mehdi S.Q., Parham P., Piazza A., Pistillo M.P., Qian Y., Shu Q., Xu J., Zhu S., Weber J.L., Greely H.T., Feldman M.W., Thomas G., Dausset J. & Cavalli-Sforza L.L. 2002. A human genome diversity cell line panel. *Science*, 296: 261-262.
- Cann R.L., Stoneking M. & Wilson A.C. 1987. Mitochondrial DNA and human evolution. *Nature*, 325: 31-36.
- Cavalli-Sforza L.L. 2005. The Human Genome Diversity Project: past, present and future. *Nature Rev. Genet.*, 6: 333-340.
- Cavalli-Sforza L.L., Menozzi P. & Piazza A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, New Jersey.
- Chaubey G., Metspalu M., Kivisild T. & Villems R. 2006. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays*, 29: 91-100.
- Coelho M., Sequeira F., Luiselli D., Belez S. & Rocha J. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol.*, 21: 9-80.
- Coop G., Pickrell J.K., Novembre J., Kudaravalli S., Li J., Absher D., Myers R.M., Cavalli-Sforza L.L., Feldman M.W. & Pritchard J.K. 2009. The role of geography in human adaptation. *Plos Genet.*, 5:e1000500.
- Cruciani F., Santolamazza P., Shen P., Macaulay V., Moral P., Olckers A., Modiano D., Holmes S., Destro-Bisol G., Coia V., Wallace D.C., Oefner P.J., Torroni A., Cavalli-Sforza L.L., Scozzari R. & Underhill P.A. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.*, 70: 1197-1214.
- Destro-Bisol G., Donati F., Coia V., Boschi I., Verginelli F., Caglià A., Tofanelli S., Spedini G. & Capelli C. 2004. Variation of female and male lineages in Sub-Saharan populations: the importance of sociocultural factors. *Mol. Biol. Evol.*, 21: 1673-1682.
- Destro-Bisol G., Anagnostou P., Batini C., Battaglia C., Bertoni S., Bottini A., Caciagli L., Calò C.M., Capelli C., Capocasa M., Castrì L., Ciani G., Coia V., Corrias L., Crivellaro F., Ghiani M.E., Luiselli D., Mela C., Melis A., Montano V., Paoli G., Sanna E., Rufo F., Sazzini M., Taglioli L., Tofanelli S., Useli A., Vona G. & Pettener D. 2008. Italian isolates today: geographic and linguistic factors shaping human biodiversity. *J. Anthropol. Sci.*, 86: 179-188.
- Dillehay T.D. 2008. Probing deeper into first American studies. *Proc. Natl. Acad. Sci. U.S.A.*, 106: 971-978.

- Endicott P. & Ho S. 2008. A Bayesian evaluation of human mitochondrial substitution rates. *Am. J. Hum. Genet.*, 82: 895-902.
- Fagundes N.J.R., Ray N., Beaumont M., Neuenschwander S., Salzano F.M., Bonatto S.L. & Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. U.S.A.*, 104:17614-17619.
- Fagundes N.J.R., Kanitz R., Eckert R., Valls A.C.S., Bogo M.R., Salzano F.M., Smith D.G., Silva W.A. Jr, Zago M.A., Ribeiro-dos-Santos A.K., Santos S.E., Petzl-Erler M.L. & Bonatto S.L. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.*, 82: 583-592.
- Fix A.G. 2004. Rapid deployment of the five founding Amerind mtNDA haplogroups via coastal and riverine colonization. *Am. J. Phys. Anthropol.*, 128:430-436.
- Forster P. & Renfrew C. (eds) 2006. *Phylogenetic Methods and the Prehistory of Languages*. McDonald Institute for Archaeological Research, University of Cambridge, UK.
- Francois O., Currat M., Ray N., Han E., Excoffier L. & Novembre, J. 2009. Principal component analysis under population genetic models of range expansion and admixture (under review).
- Garagnani P., Laayouni H., González-Neira A., Sikora M., Luiselli D., Bertranpetit J. & Calafell F. Isolated populations as treasure troves in genetic epidemiology: the case of the Basques. *Eur. J. Hum. Genet.*, 17:1490-1494.
- Gutenkunst R.N., Hernandez R.D., Williamson S.H. & Bustamante C.D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *Plos Genet.*, 5:e1000695.
- Hammer M.F., Mendez F.L., Cox M.P., Woerner A.E. & Wall J.D. 2008. Sex-biased evolutionary forces shape genomic patterns of human diversity. *Plos Genet.*, 4: e1000202.
- Handley L.J., Manica A., Goudet J. & Balloux F. 2007. Going the distance: human population genetics in a clinal world. *Trends Genet.*, 23: 432-439.
- Heath S.C., Gut I.G., Brennan P., McKay J.D., Bencko V., Fabianova E., Foretova L., Georges M., Janout V., Kabesch M., Krokan H.E., Elvestad M.B., Lissowska J., Mates D., Rudnai P., Skorpén F., Schreiber S., Soria J.M., Syvänen A., Meneton P., Herçberg S., Galan P., Szeszenia-Dabrowska N., Zaridze D., Génin E., Cardon L.R. & Lathrop M. 2008. Investigation of the fine structure of European populations with applications to disease association studies. *Eur. J. Hum. Genet.*, 16: 1413-1429.
- Ilkic I. & Paul. N.W. 2009. Ethical aspects of genome diversity research: genome research into cultural diversity or cultural diversity in genome research? *Med. Health Care Philos.*, 12:25-34.
- International HapMap Consortium 2003. The International HapMap Project. *Nature*, 426:789-796.
- Jobling M., Hurles M & Tyler Smith C. 2004. *Human Evolutionary Genetics: Origins, Peoples and Disease*. Garland Science Publishing, New York.
- Kimura R., Yamaguchi T., Takeda M., Kondo O., Toma T., Haneji K., Hanihara T., Matsukusa H., Kawamura S., Maki K., Osawa M., Ishida H. & Oota H. 2009. A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.*, 85: 528-535.
- King T.E. & Jobling M.A. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.*, 25: 351-360.
- Kitchen D., Miyamoto M.M. & Mulligan C.J. 2008. A three-stage colonization model for the peopling of the Americas, *Plos One*, 2: e1596.
- Laayouni H, Calafell F, Bertranpetit J. 2010. A genome-wide survey does not show the genetic distinctiveness of Basques. *Hum Genet.*, 127:455-458.
- Lao O., Lu T.T., Nothnagel M., Junge O., Freitag-Wolf S., Caliebe A., Balasckova M., Bertranpetit J., Bindoff L.A., Comas D., Holmlund G., Kouvatsi A., Macek M., Mollet I., Parson W., Palo J., Ploski R., Sajantila A., Tagliabracci A., Gether U., Werge T., Rivadeneira F., Hofman A., Uitterlinden A.G., Gieger C., Wichmann H.E., Rütger A., Schreiber S., Becker C., Nürnberg P,

- Nelson M.R., Krawczak M. & Kayser M. 2008. Correlation between genetic and geographic structure in Europe. *Curr. Biol.*, 18: 1241-1248.
- Macaulay V., Hill C., Achilli A., Rengo C., Clarke D., Meehan W., Blackburn J., Semino O., Scozzari R., Cruciani F., Taha A., Shaari N.K., Raja J.M., Ismail P., Zainuddin Z.A., Goodwin W., Bulbeck D., Bandelt H.-J., Oppenheimer S., Torroni A. & Richards M. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, 308: 1034-1036.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorff L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R., Chakravarti A., Cho J.H., Guttmacher A.E., Kong A., Kruglyak L., Mardis E., Rotimi C.N., Slatkin M., Valle D., Whittemore A.S., Boehnke M., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F., McCarroll S.A. & Visscher P.M. 2009. Finding the missing heritability of complex diseases. *Nature*, 461: 747-753.
- McVean G. 2009. A genealogical interpretation of principal components analysis. *Plos Genet.*, 5: e1000686.
- Mellars P. 2006. Going East: New genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science*, 313: 796-800.
- Menozzi P, Piazza A, Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in Europeans. *Science*, 201:786-792.
- Metspalu M., Kivisild T., Bandelt H.J., Richards M. & Villems R. 2006. The pioneer settlement of modern humans in Asia. In H.J. Bandelt, V. Macaulay & M. Richards (eds): *Human mitochondrial DNA and the evolution of Homo sapiens*, pp 181-199. Springer, Berlin.
- Migliano A.B., Vinicius L. & Lahr M.M. 2007. Life history trade-offs explain the evolution of human pygmies. *Proc. Natl. Acad. Sci. U.S.A.*, 104: 20216-20219.
- Mulligan C.J., Hunley K., Cole S., Long J.C. 2004. Population genetics, history, and health patterns in native Americans. *Annu. Rev. Genom. Hum. Genet.*, 5: 295-315.
- Mulligan C.J., Kitchen A. & Miyamoto M.M. 2008. Updated three-stage model for the peopling of the Americas. *Plos One*, 3: e3199.
- Need A.C. & Goldstein D.B. 2009. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.*, 25: 489-494.
- Nelson M.R., Bryc K., King K.S., Indap A., Boyko A.R., Novembre J., Briley L.P., Maruyama Y., Waterworth D.M., Waeber G., Vollenweider P., Oksenberg J.R., Hauser S.L., Stirnadel H.A., Kooner J.S., Chambers J.C., Jones B., Mooser V., Bustamante C.D., Roses A.D., Burns D.K., Ehm M.G. & Lai E.H. 2008. The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.*, 83: 347-358.
- Novembre J., Johnson T., Bryc K., Kutalik Z., Boyko A.R., Auton A., Indap A., King K.S., Bergmann S., Nelson M.R., Stephens M. & Bustamante C.D. 2008. Genes mirror geography within Europe. *Nature*, 456: 98-101.
- Novembre J. & Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nature Genet.*, 40: 646-649.
- Patin E., Laval G., Barreiro L.B., Salas A., Semino O., Santachiara-Benerecetti S., Kidd K.K., Kidd J.R., Van der Veen L., Hombert J.M., Gessain A., Froment A., Bahuchet S., Heyer E. & Quintana-Murci L. 2009. Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *Plos Genet.*, 4:e1000448.
- Perego U.A., Achilli A., Angerhofer N., Accetturo M., Pala M., Olivieri A., Kashani B.H., Ritchie K.H., Scozzari R., Kong Q.P., Myres N.M., Salas A., Semino O., Bandelt H.J., Woodward S.R. & Torroni A. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.*, 19:1-8.
- Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.*, 38: 904-909.
- Ramachandran S., Deshpande O., Roseman C.C., Rosenberg N.A., Feldman M.W. & Cavalli-Sforza L.L. 2005. Support from the relationship of genetic and geographic distance in human

- populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U.S.A.*, 102: 15942-15947.
- Relethford J.H. 1998. Genetics of modern human origins and diversity. *Annu. Rev. Anthropol.*, 27: 1-23.
- Renfrew C. & Boyle K. (eds) 2000. *Archaeogenetics: DNA and the Population Prehistory of Europe*. McDonald Institute for Archaeological Research, University of Cambridge, UK.
- Richards M., Bandelt H.J., Kivisild T. & Oppenheimer S. 2006. A model for the dispersal of modern humans out of Africa. In H.J. Bandelt, V. Macaulay & M. Richards (eds): *Human mitochondrial DNA and the evolution of Homo sapiens*, pp. 225-265. Springer, Berlin.
- Rosenberg N.A., Mahajan S., Gonzalez-Quevedo C., Blum M.G., Nino-Rosales L., Nini V., Das P., Hegde M., Molinari L., Zapata G., Weber J.L., Belmont J.W. & Patel P.I. 2006. Low levels of genetic divergence across geographically and linguistically diverse populations from India. *Plos Genet.*, 2: e215.
- Sabatti C., Service S.K., Hartikainen A.L., Pouta A., Ripatti S., Brodsky J., Jones C.G., Zaitlen N.A., Varilo T., Kaakinen M., Sovio U., Ruokonen A., Laitinen J., Jakkula E., Coin L., Hoggart C., Collins A., Turunen H., Gabriel S., Elliot P., McCarthy M.I., Daly M.J., Järvelin M.R., Freimer N.B. & Peltonen L. 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genet.*, 41: 35-46.
- Salas A., Richards M., De la Fe T., Lareu M.V., Sobrino B., Sanchez-Diz P., Macaulay V. & Carracedo A. 2002. The making of the African mtDNA landscape. *Am. J. Hum. Genet.*, 71: 1082-1111.
- Samuels D.C., Burn D.J. & Chinnery P.F. 2009. Detecting new neurodegenerative disease genes: does phenotype accuracy limit the horizon? *Trends Genet.*, 25: 486-488.
- Schuster S.C., Miller W., Ratan A., Tomsho L.P., Giardine B., Kasson L.R., Harris R.S., Petersen D.C., Zhao F., Qi J., Alkan C., Kidd J.M., Sun Y., Drautz D.I., Bouffard P., Muzny D.M., Reid J.G., Nazareth L.V., Wang Q., Burhans R., Riemer C., Wittekindt N.E., Moorjani P., Tindall E.A., Danko C.G., Teo W.S., Buboltz A.M., Zhang Z., Ma Q., Oosthuysen A., Steenkamp A.W., Oostuisen H., Venter P., Gajewski J., Zhang Y., Pugh B.F., Makova K.D., Nekrutenko A., Mardis E.R., Patterson N., Pringle T.H., Chiaromonte F., Mullikin J.C., Eichler E.E., Hardison R.C., Gibbs R.A., Harkins T.T. & Hayes V.M. 2010. Complete Khoisan and Bantu genomes from southern Africa. *Nature*, 463: 943-947
- Soares P., Ermini L., Thomson N., Mormina M., Rito T., Röhl A., Salas A., Oppenheimer S., Macaulay V. & Richards M.B. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.*, 84: 740-759.
- Soares P., Trejaut J.A., Loo J.-H., Hill C., Mormina M., Lee C.L., Chen Y.M., Hudjashov G., Forster P., Macaulay V., Bulbeck D., Oppenheimer S., Lin M. & Richards M.B. 2008. Climate change and post-glacial human dispersals in Southeast Asia. *Mol Biol. Evol.*, 25: 1209-1218.
- Sommer M. 2008. History in the gene: Negotiations Between Molecular and organismal anthropology. *J. Hist. Biol.*, 41:473-528
- Stringer C. & Andrews P. 1998. Genetic and fossil evidence for the origin of modern humans. *Science*, 239: 1263 - 1268.
- Sun C., Kong Q.P., Palanichamy M.G., Agrawal S., Bandelt H.J., Yao Y.G., Khan F, Zhu C.L., K. Chaudhuri T. & Zhang Y.P. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.*, 23: 683-690.
- Tamm E., Kivisild T., Reidla M., Metspalu M., Smith D.G., Mulligan C.J., Bravi C.M., Rickards O., Martinez-Labarga C., Khusnutdinova E.K., Fedorova S.A., Golubenko M.V., Stepanov V.A., Gubina M.A., Zhadanov S.I., Ossipova L.P., Damba L., Voevoda M.I., Dipierri J.E., Villemers R. & Malhi R.S. 2007. Beringian standstill and spread of Native American founders. *Plos One*, 2: e829.
- Templeton A.R. 1992. Human origins and analysis of mitochondrial DNA sequences. *Science*, 255:737.

- Terwilliger J.D. & Lee J.H. 2007. Natural experiments in human gene mapping: the intersection of anthropological genetics and genetic epidemiology. In Crawford M.H. (ed): *Anthropological Genetics: theory, methods and applications*, pp 38-76. Cambridge University Press, Cambridge.
- Tishkoff S.A., Reed F.A., Friedlaender F.R., Ehret C., Ranciaro A., Froment A., Hirbo J.B., Awomoyi A.A., Bodo J.M., Doumbo O., Ibrahim M., Juma A.T., Kotze M.J., Lema G., Moore J.H., Mortensen H., Nyambo T.B., Omar S.A., Powell K., Pretorius G.S., Smith M.W., Thera M.A., Wambebe C., Weber J.L. & Williams S.M. 2009. The genetic structure and history of Africans and African Americans. *Science*, 22: 1035-1044.
- Verdu P., Austerlitz F., Estoup A., Vitalis R., Georges M., Théry S., Froment A., Le Bomin S., Gessain A., Hombert J.M., Van der Veen L., Quintana-Murci L., Bahuchet S. & Heyer E. 2009. Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Curr. Biol.*, 19: 312-318.
- Via M., Gignoux C. & Burchard E.G. 2010. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med.*, 2:3.
- Vigilant L., Stoneking M., Harpending H., Hawkes K. & Wilson A.C. 1991. African populations and the evolution of human mitochondrial DNA. *Science*, 253: 1503-1507.
- Wang S., Lewis C.M., Jakobsson M., Ramachandran S., Ray N., Bedoya G., Rojas W., Parra M.V., Molina J.A., Gallo C., Mazzotti G., Poletti G., Hill K., Hurtado A.M., Labuda D., Klitz W., Barrantes R., Bortolini M.C., Salzano F.M., Petzl-Erler M.L., Tsuneto L.T., Llop E., Rothhammer F., Excoffier L., Feldman M.W., Rosenberg N.A. & Ruiz-Linares A. 2007. Genetic variation and population structure in Native Americans. *Plos Genet.*, 3:e185.
- Wells R.S., Yuldasheva N., Ruzibakiev R., Underhill P.A., Evseeva I., Blue-Smith J., Jin L., Su B., Pitchappan R., Shanmugalakshmi S., Balakrishnan K., Read M., Pearson N.M., Zerja L.T., Webster M.T., Zholoshvili I., Jamarjashvili E., Gambarov S., Nikbin B., Dostiev A., Aknazarov O., Zalloua P., Tsoy I., Kitaev M., Mirrakhimov M., Chariev A. & Bodmer W.F. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. U.S.A.*, 98: 10244-10249.
- Xue Y., Wang Q., Long Q., Ng B.L., Swerdlow H., Burton J., Skuce C., Taylor R., Abdellah Z., Zhao Y, Asan, MacArthur D.G., Quail M.A., Carter N.P., Yang H. & Tyler-Smith C. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.*, 19: 1453-1457.