

Mark A. JOBLING (UK)
Molecular Anthropology in the Genomic Era

University of Leicester, UK

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, suspect that the diversity of these markers is strongly influenced by natural selection, and researchers interested in investigating the human past have since then sought neutral marker phenotypes and adaptive influences as a nuisance.

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 Y chromosome. The 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 can 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 structure.

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. While the resolution limit has been reached now that its 16.5kb can be sequenced readily in many individuals, this is not so with the Y - only a small number of the available STRs is normally analysed, and a resequencing of megabases of this chromosome is now possible with careful application of new technologies. This reveals 100s of new SNPs per chromosome analysed, posing challenges in unifying datasets and standardising methodology and nomenclature. Recent sequence analyses of Y chromosomes separated by only a few generations identifies lineage-specific markers (2009), perhaps promising the phylogenetic resolution needed to distinguish between different migration events very close in time. Members of the general public, through their obsessive genetic genealogy, are contributing useful scientific insights.

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 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 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 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 variation (Hammer et al., 2008).

Using genetics to test hypotheses based on historical, archaeological or linguistic evidence often uses a cherry-picking approach to the other disciplines that lacks objectivity. Although 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 East African patterns of modern genetic diversity (Handley et al., 2007), underscoring the importance of early events when populations were small.

By contrast, there are researchers who regard phenotypes and selection as the important stuff, and population structure and history as the distracting nuisance. Unfortunately, although phenotypes of humans are of particularly acute interest, our species is not a model organism. The kinds of controlled experiments we might carry out on mice are impossible, so we must 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 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, pig hair thickness and tooth morphology (Kimura et al., 2009), and the fascinating (no, seriously) trait of earwax type (Yoshiura et al., 2006). Other traits promise to be less tractable, with 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, but each contributes only a tiny amount (a few millimetres) of the variance of the trait. Tellingly, Francis Galton's Victorian back-of-an-envelope approach to height prediction 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 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 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 we think, the trait is complex and multigenic, then it will be more difficult to understand. We may hypothesise a common origin of pygmy groups to explain the common phenotype, but this would make pinpoint the specific locus or loci responsible for the phenotype amongst the loci shared simply through recent common origin. Then again, the detection of phenotypically important loci in populations will be difficult because of small sample sizes, and grant applications (often damned by reviewers as 'fishing expeditions') will tend to founder upon the unforgiving reefs of calculations.

The role of natural selection in the development of short stature is mysterious, and it is difficult to regard selective explanations based on the ease of moving about in forests as anything but ludicrous 'Just So' stories. Darwin would probably advocate sexual selection here, but proving him right is not going to be easy. Even when we can see clear selective advantages to particular adaptations, the problem of drift is a bugbear 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 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 in one place and may even have a different physiological and genetic basis. It seems likely that admixture-based approaches will be useful here.

In the sunlit distance, glimpsed through a glass darkly, lies the brave and bright new world of whole genome sequences (www.1000genomes.org), unswayed by ascertainment bias and variants. 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 mine and use such sequences, and think what questions they would like to address with them. Surely, the more sequence, the better? If we knew the sequences of all the genomes (we'd 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 like recording and classifying of the environments and the phenotypes (Samuels et al., 2009), rather than the genotypes, will then become crucial, and the anthropologists (and the ethicists) the world.

Jorge ROCHA (Portugal) (1) (2)
The peopling of Africa

(1) IPATIMUP-Institute of Pathology and Molecular Immunology of the University of Porto.
(2) Department of Zoology and Anthropology, Faculty of Sciences, University of Porto.

[Figure not reported here]

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 recent 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 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 rain 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 to data analysis, 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 reached an unprecedented level of precision. The aim of this lecture is to provide an overview of the genetic evidence on African population history that became available with these recent advances.

A significant part of our present understanding of African genetic variation is based on the study of mitochondrial DNA (mtDNA) and the non-recombining portion of the Y chromosome (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, of recombination, haplotype phylogenies can be easily reconstructed, providing a temporal framework for mutation accumulation, which can be related to the geographic distribution of 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 oldest basal NRY-haplogroup A-M91 suggests an ancestral link of the southern African Khoes-Sans click-speaking groups to East Africa. The relatively old NRY B2b-M112 haplogroup is common ancestry of Khoes-Sans and Pygmy hunter-gatherer groups. A lineage within the younger E3b-M35* paragroup suggests that pastoralism might have been introduced to southern Africa 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 speaking agriculturists. Among the mtDNA haplotypes, the basal L0d clade is almost exclusive to the southern African Khoes-Sans but is also found in the click-speaking Sandwe from Tar confirming the ancient link of the Khoes-Sans to Eastern Africa. The younger haplogroup L1c, which probably originated in central Africa, is crucial to assess the ancestral relationship between Pygmy hunter-gatherers and their neighboring Bantu-speaking farmers. However, an important limitation of studies based on the NRY or mtDNA markers is that they amount to the study 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 Sarah Tishkoff's landmark study on 2432 individuals from 11 using a panel of 1327 polymorphic markers (Tishkoff et al., 2009).

In brief, the study as shown that most African genetic variation can be sorted into 14 ancestral population clusters and that most populations exhibited high levels of mixed ancestry, reflecting 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 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 I family (paralleled by the distribution of NRY haplogroup E3a-M2); iii) another group comprising Chaditic and Nilo-Saharan-speaking populations from Nigeria, Cameroon, Chad and south (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-speaking populations from Sudan, Kenya and Tanzania; and v) a group with noncontiguous geographic distribution consisting of Pygmy and southern African Khoes-Sans population evidence for shared ancestry among hunter-gatherers (consistent with the distribution of NRY haplogroup B2b, but not with mtDNA, since haplogroup L1c seems to preferentially link West Pygmies to neighboring Bantu agriculturists). In spite of the major advance provided by this study, it is important to note that regions like the Sahel, the Atlantic West Africa, Namibia, 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 on African populations.

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 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 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 population sizes (Bisoli 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 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 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 culture 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, intermingling among the peoples of southern Angola, which has become one of my own research interests, has generated intriguingly discordant combinations of ethnicity, language and lifestyle that we discussed in the lecture to illustrate the usefulness of local patterns in understanding major tendencies (Coelho et al., 2009).

A final aspect of the recent advances in understanding genetic diversity within Africa is related with data analysis. Datasets based on multiple, independently evolving genetic systems are particularly well suited to simulation-based inferential frameworks that are aimed to distinguish between alternative models of population history and to estimate key microevolutionary

under a given model. Recent applications of rejection algorithms and Approximate Bayesian Computation to infer the branching history of Pygmy and agricultural populations provide examples of the usefulness of new computational methods to address population history in Africa (Patin et al., 2009; Verdu et al., 2009). With the rapid accumulation of multilocus gen 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 with

Jeroen PIJPE (The Netherlands)*

Skewed male population substructure among an agriculturalist Ghanaian tribe

Socio-economic and cultural factors might play an important role in explaining differences in human population genetic structure. To explain patterns in population substructure, studies analyzed genetic differences among widely dispersed populations, and did not consider differences among clans within the same tribe and/or village. We conducted a detailed tribal sex geographic study to investigate the influence of socio-economic and anthropological factors on population genetic structure. We analyzed the DNA of 205 males from the Bimoba tribe in a single village of Farfar in the Upper East Region of Ghana. These males belong to 6 different clans and were living in 93 different compounds scattered over an area of approximately 4 km². We found a striking, skewed male population substructure due to an almost complete lack of male mediated gene flow among clans, as reported by 15 Y-chromosome Short Tandem R (STRs) and a series of biallelic Single Nucleotide Polymorphisms (SNPs) defining the Y-chromosome haplogroup lineage E1b1a. We found a markedly skewed male population substructure almost complete lack of male gene flow among clans of the Bimoba tribe within one single village. Males were classified in Y-haplogroups E1b1a*, E1b1a7a* and E1b1a8*. In contrast, mtDNA HVR-1 sequence, and from 15 autosomal STRs indicate a virtually random female mediated gene flow among clans.

On the micro-geographic scale of a single village, population genetic structure among a traditional agricultural people is deeply influenced by the social structures. The Y-chromosome is highly skewed by clan(-group) membership, whereas female mediated gene flow is not bound by such social structures. This pattern can be explained by the patrilocal and patrilineal of such societies, and by past migration events. The Bimoba offer a valuable insight into the cultural processes that have shaped genetic variation in humans.

*With:

Tom van der Hulle, Hans J. Meij, Kristiaan van der Gaag, P. Eline Slagboom, Rudi G.J. Westendorp, Peter de Knijff

Sergio TOFANELLI (Italy)*

Malagasy admixture: the tale of a recent encounter between deep-rooted lineages and beyond

We fit the history of Malagasy admixture in a highly resolved phylo-genetic framework by typing a large set of uni-parentally transmitted markers in unrelated individuals from inland ethnic groups. The uniqueness of Malagasy was confirmed to be due to a recent encounter between gene pools (Insular Southeast Asian and sub-Saharan African) that have been shaped 60,000 years of independent evolution. The distribution of the two ancestral components was ethnic and sex biased, with the Asian ancestry appearing more conserved in the female than male gene pool and in inland than in coastal groups. Thanks to forward simulations and the use of a novel and more accurate measure of genetic distance (DHS), the focus about the Malagasy lineages was enlarged in space and pushed back in time. Complex underlying demographies after the admixture event could make the search of univocal ancestries inconclusive link between Malagasy and Bornean (Maanyan) vocabulary misleading. The pattern of diffusion was compatible with a primary admixture of proto-Malay people with Bantu speaking western-like pool of haplotypes, followed by a secondary flow of Southeastern Bantu speakers unpaired for gender and geography. Some groups appear suitable cases for admixture models aimed at detecting disease-associated variants that differ markedly in frequency between the two parental populations

*With:

Stefania Bertoncini (Università di Pisa, Pisa, Italy), Loredana Castri (Università di Bologna, Bologna, Italy), Donata Luiselli (Università di Bologna, Bologna, Italy), Francesc Calafell (Univ Pompeu Fabra, Barcelona, Spain), Giuseppe Donati (Oxford Brookes University, Oxford, UK), Giorgio Paoli (Università di Pisa, Pisa, Italy)

Fulvio CRUCIANI (Italy)*

Human Y-chromosome haplogroup R1b1a (R-V88): A paternal genetic record of early-mid Holocene trans-Saharan connections

Human Y chromosomes belonging to haplogroup R-P25 are quite rare in Africa, being found mainly in Asia and Europe. However, a group of P25 Y chromosomes that are not defined by presence of a downstream derived marker (the paralogous R-P25*) are found concentrated in the central-western part of the African continent, where they can be detected at frequency 95%. Phylogenetic evidence and coalescence time estimates suggest that R-P25* chromosomes (or their phylogenetic ancestor) may have been carried to Africa by an Asia-to-Africa back-migration in prehistoric times. Here we describe six new mutations that define the relationships among the African R-P25* Y chromosomes and between these African chromosomes and previous P25 Eurasian sub-lineages. The incorporation of these new mutations into a phylogeny of the R-P25 haplogroup led to the identification of a new clade (R1b1a or R-V88) encompassing African R-P25*, about half of the few European/west Asian R-P25*, and the R-M18 chromosomes. A world-wide phylogeographic analysis of the R-P25 haplogroup provided strong support for an Asia-to-Africa back-migration hypothesis. The analysis of the distribution of the R-V88 haplogroup in more than 1,800 males from 69 African populations, revealed a striking genetic cline between the Chadic-speaking peoples from the central Sahel and several other Afroasiatic speaking groups from North Africa. The R-V88 coalescence time was estimated at 9,200-5,600 years ago, early-mid Holocene. We suggest that R-V88 is a paternal genetic record of the proposed mid-Holocene migration of proto-Chadic Afroasiatic speakers through the Central Sahara into the Sahel Basin.

* With:

Beniamino Trombetta (1), Daniele Sellitto (2), Andrea Massaia (1), Giovanni Destro-Bisol (3), Elizabeth Watson (4) Eliane Beraud Colomb (5), Jean-Michel Dugoujon (6), Pedro Moral (7) Scozzari (1) (1) Dipartimento di Genetica e Biologia Molecolare, Sapienza Università di Roma, Rome 00185, Italy; (2) Istituto di Biologia e Patologia Molecolari, Consiglio Nazionale delle Ricerche, Rome, Italy; (3) Dipartimento di Biologia Animale e dell'Uomo, Sapienza Università di Roma, Rome 00185, Italy; (4) The Swedish Museum of Natural History, Stockholm, Sweden; (5) Laboratoire d'Immunologie, Hôpital de la Sainte-Marguerite, Marseille, France; (6) Laboratoire d'Anthropologie, FRE 2960, Centre National de la Recherche Scientifique (CNRS) Université Paul Sabatier Toulouse, France; (7) Department of Biologia Animal, Universitat de Barcelona, Barcelona, Spain.

A. RANCIARO (USA) (1)*

The Genetic Basis of Lactase Persistence in Africa

In most individuals, the ability to digest lactose, the sugar present in milk, declines rapidly after weaning because of decreasing levels of the enzyme lactase (encoded by the LCT gene) in the intestine. However, there are individuals who maintain the ability to digest milk into adulthood due to a genetic adaptation in populations that have a history of pastoralism. In order to identify variants associated with the lactase persistence (LP) trait and to study the evolutionary history of LP in Africa, we resequenced 1.7 kb of intron 9 and 3.3 kb of intron 13 of the MCM6 gene (associated with LP in Europeans) upstream of LCT, and 2.0 kb of the promoter region of the LCT gene.

A total of 973 individuals representing 77 different groups from Africa (n=68), Asia (n=3), Middle Eastern (n=3) and Europe (n=3) were used in this study.

We analyzed genotype/phenotype associations in 410 individuals for which we measured lactase activity (Lactose Tolerant Test) and identified three variants significantly associated with LP in Africans (G/C-14010, T/G-13915, and C/G-13907). We also identify a strong signature of recent positive selection in several East African pastoralist groups. Levels of nucleotide diversity tests of neutrality were performed and the negative trend in Tajima's D test is consistent with positive directional selection. Simulations were performed to rule out the possibility of demographic effect. Microsatellite haplotype analysis was also used to reconstruct the origin and spread of the LP associated variants in Africa. Our results indicate that mutations associated with LP independently in African populations. Additionally, we find evidence for an East African origin for the spread of pastoralism into South Africa.

*With:

A. Ranciaro, J. Hirbo, F. Reed, M. Campbell, H. Muntaser, O. Sabah, G. Destro-Bisol, Alain Froment, Maritha J. Kotze, Thomas B. Nyambo, S. A. Tishkoff, 10 (1) Dept Genetics, University of Pennsylvania, Philadelphia, PA; (2) Dept. of Biology, University of Maryland, College Park, MD; (3) Dept. of Evolutionary Ecology, Max Planck Institute for Biological Anthropology, Leipzig, Germany; (4) Inst. of Endemic Diseases, University of Khartoum, Sudan; (5) KEMRI, Nairobi, Kenya; (6) Dipt. Biologia Animale e dell'Uomo, Università "La Sapienza", Roma UMR 208, IRD-MNH, Musée de l'Homme, Paris; (7) Dept of Pathology, Faculty of Health Sciences, University of Stellenbosch, Tygerberg, South Africa; (8) Dept. of Biochemistry, Muhimbili University of Health and Allied Sciences (MUHAS), Dar es Salaam, Tanzania; (9) Department of Genetic and Biology, University of Pennsylvania, Philadelphia, PA.

Sara PIACENTINI (Italy)*

GSTM1 and GSTT1 gene polymorphisms in European and African populations

Glutathione S-transferases (GSTs) are a superfamily of multifunctional proteins with fundamental roles in cellular detoxification. GSTs have been grouped into numerous classes; many polymorphic and their polymorphism may contribute to promoting individual differences in their response to xenobiotics. This study is focused on GSTT1 and GSTM1 gene polymorphisms of particular interest as anthropogenetic markers in studies on intra- and inter-population variability. The frequencies of null genotypes obtained by PCR multiplex in two population samples of European-origin and two population samples of African-origin were compared with those from other populations of different origin reported in the literature with the aim of contributing to the geography of these markers. On the basis of the results obtained from the statistical analyses, it is possible to distinguish three main clusters represented by population samples of African or European origin respectively. Other populations are randomly distributed outside these three main clusters, perhaps because of the low sample sizes.

Data from PCR allele-specific technique highlighted the possible presence of a SNP (rs11550605) in the third exon of the GSTT1 gene that leads to a change of the aminoacidic residue at position 104 (T104P). In the present study this substitution was not found and the studied population does not exhibit this mutation. The sequencing of a 480 bp fragment of the GSTT1 gene (from the fourth intron and a short part of the fourth exon) in a sample of the Italian population was performed to assess the presence of SNPs in the GSTT1 gene; the sequences were then compared with the reference one. No nucleotide substitution was highlighted, so confirming the data in the literature indicating low frequencies of mutations and heterozygosity.

* With:

Renato Polimanti, Maria Fuciarelli. Department of Biology, University of Rome "Tor Vergata"

Sarah MARKS (UK)*

Molecular characterization of low recombination genomic regions for bio-anthropological studies: The peopling of Southern Africa

It is now generally accepted that modern humans originated in Africa. In recent years there have been studies into the genetic variation of current human populations in order to understand population movement around the world (Underhill et al 2001, Li et al 2008). Southern African genetic history, however, has so far been understudied, despite possibilities that this region may have played a significant role in the origins of modern humans (McBrearty and Stringer 2007). For this study, a total of 10 genomic regions have been selected, all on different chromosomes characterised by reduced recombination rate (<0.1cM/Mb, with many being <0.001cM/Mb). Each region is unlinked to a coding region and contains a minimum of 2 Short Tandem Repeat (STR) markers. These STRs have been combined in three STR multiplex reactions. Analysis of these regions in samples from Southern Africa will be used to address questions of interest about population history.

history in this area: 1) What (if any) admixture is present between Bantu-speaking populations and hunter-gatherer and pastoralist Khoisan speakers; 2) Whether gene flow occurred between populations that were already present in the area before the arrival of the Bantu agriculturalists; 3) How the genetic diversity of Southern African populations compares with that of other populations, and thus how it relates to the emergence of modern humans.

* With:
Cristian Capelli

Valeria MONTANO (1)*

A genetic perspective on the spread of Bantu communities

Every anthropologist working on sub-Saharan African populations is supposed to have faced the Bantu expansion's big deal. Bantu is a language family that is currently spoken in a wide Saharan Africa, longitudinally extended from Cameroon to Kenya until the extreme South, excluding a part of South Africa, Botswana, Namibia and all Madagascar. The most accepted hypothesis about Bantus' origin was put forward by Joseph Greenberg, who located the first communities in the Benue River Valley, across South-Eastern Nigeria and West Cameroon (Greenberg, 1963). Starting from linguistic and historical hypotheses, Molecular Anthropologists have analyzed the genetic structure of Bantu communities and proposed that the present day distribution of lineages of Y-chromosome and mitochondrial DNA (mtDNA) could be a result of the expansion of Bantu speaking peoples. The present work focuses on the Bantu populations that are still in genetic and cultural continuity with the first ancient communities, between Nigeria and Cameroon and other Bantu communities of Cameroon, Gabon and Congo. All these populations are supposed to have been involved in the Western stream Bantu migration (Vansina 1984; Belezza et al., 2005). Our aim is to gain insights into the population dynamics underlying the expansion of Bantu languages through the analysis of the classical uniparental genetic systems (Y-chromosome and mtDNA). Seventeen populations have been analyzed for 21 SNPs and 17 Y-chromosome and for the hypervariable region I of the mtDNA. The results show different signals of structuration for the two genetic systems, opening an avenue to test hypotheses about the spread of Bantu languages.

*With:

Marcari V.1, Anayale O.3, Comas D.2, Destro-Bisol G.1.
1Dipartimento di Biologia Animale e dell'Uomo, Università di Roma "La Sapienza", Rome, Italy
Istituto Italiano di Antropologia, Rome, Italy
2Unitat de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat "Pompeu Fabra", Barcelona, Spain
3Department of Zoology, Ibadan University, Ibadan, Nigeria

John NOVEMBRE (USA)

Maps and migrations: Insights to the genetic structure of Europe from single nucleotide polymorphism data and principal components analysis

Department of Ecology and Evolutionary Biology; Interdepartmental Program in Bioinformatics; University of California - Los Angeles

Due to ease of accessibility, patterns of genetic variation in samples of European individuals have been some of the most carefully characterized throughout the world and arguably across species. Despite the intensive attention given to these samples, basic questions still remain unanswered regarding what the dominant patterns are and what ancestral events explain them. The challenge has been due to how statistical methods have been applied to emerging data sets. In particular, non-model based methods, especially principal components analysis (PCA), played an important role in how genetic data from these samples have been interpreted.

While PCA was first pioneered in the 1970s to summarize patterns in allele frequencies among samples, a novel form of individual-based PCA has recently become popular in human population genetics (e.g. Price et al. 2006). This resurgence is mainly due to the fact that when doing genome-wide association mapping for disease susceptibility loci, PCA coordinates can be used to control for population stratification. Further, 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 populations.

To understand its behavior more concretely, several recent theoretical studies have helped make clear how PCA behaves in different settings. Key results are that: (1) Under models of differentiated populations, PCA will identify easily definable clusters. Notably, detecting such clusters behaves like a phase transition whereby if the number of markers and number of individuals increased above a certain threshold, clusters suddenly become detectable (Patterson et al. 2006). (2) Under models of continuous, spatial population structure, the PCA coordinates for (or even populations) take the form of gradients (or more complex sinusoidal functions) over geographic space, even if the total population is at a demographic equilibrium. These results are interesting from the stand-point of interdisciplinary studies as this behavior of PCA has been essentially understood in some sub-disciplines of science (e.g. meteorology, image analysis, time, but their relevance was only recently noted within the population genetics community (Novembre and Stephens 2008). (3) The expected PCA coordinates for each individual in a sample derived from average pair-wise coalescent times among individuals in the sample; doing so reveals how PCA is dependent on relative sample-sizes and connects PCA to coalescent theory (2009).

These theoretical insights help greatly with the interpretation of PCA results from many recent large-scale single nucleotide polymorphism (SNP) studies. For example, in a recent collaboration between GlaxoSmithKline and academic scientists, several thousand European individuals were sampled and genotyped using the Affymetrix 500K SNP genotyping platform (the POPRE Nelson et al. 2008, Novembre et al. 2008). The results show a striking correspondence between genetics and geography even at fine spatial scales. Studies by other groups at both the finer spatial scales (e.g. within Finland and Iceland, e.g. Lao et al. 2008, Sabatti et al. 2009) also support this connection between genetics and geography (although in some cases the relative sample sizes or the presence of outlier populations distorts the basic pattern).

One 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 suggest about the peopling of Europe? To address this question we merged SNP data from putative isolates sampled as part of the Human Genome Diversity Project (e.g. French Basque Sardinians, Orcadians, and the Adygei, Cann et al. 2002) with SNP data from the POPRES European samples. We also merge in novel SNP data from the Sorbs, a previously uncharacterized speaking putative isolate from Eastern Germany. Our analysis of the Sorbs has been part of an interdisciplinary collaboration with medical geneticists from the University of Leipzig (Ton and Stumvoll) as well as a historian from UCLA (Patrick Geary).

A second major unanswered question regards how to interpret PCA. While (Novembre and Stephens 2008) show gradients can arise in PCA even under general conditions where spatial autocorrelation exists in data (e.g. equilibrium stepping-stone models), if an expansion has recently occurred, does the direction of the PC1 gradient indicate its direction? Surprisingly, the direction of the gradient in PC1, under many expansion parameter settings, does not align with the expansion wave (François et al. 2009). To explain this phenomenon, we must consider the "a phenomenon that takes place during the expansion of a population due to serial founder effects and the spatial patterns that are left behind by "surfing" alleles.

In sum, PCA is subject to a variety of behaviors that are sometimes easily misunderstood. Nonetheless, PCA can serve as a flexible exploratory tool for visualizing major patterns of population structure in a sample and for quality control (e.g. identifying outliers and batch effects in genotyping assays). Ultimately though, methods that are tailored to detect specific demographic events (e.g. the decay of diversity with distance from an origin or patterns of allele surfing) will be the most powerful way forward in illuminating the peopling of Europe.

Alessio BOATTINI (Italy)*

The Genographic Project in Italy: Y-chromosome preliminary results and perspectives

The Genographic Project is a five year genetic anthropology study (concluding in 2011) aimed to explore the migratory history of the human species by analysing around 100,000 DNA samples from 10,000 worldwide populations. Our research group is currently collaborating with the Genographic Center for Western and Central Europe (principal investigators: David Comas, Jaume Bertranpetit, Begona Martínez-Cruz) in order to sample and unravel the genetic variability of the Italian population(s).

A preliminary biodemographic analysis, based on around 80,000 surnames from more than 15 millions Italian individuals, was performed in order to design an accurate sampling strategy resulting sampling map served as a template for the actual DNA sampling campaign, to which collaborated actively local Blood Transfusion Centers. 28 sampling points were selected as samples collected (around 40 individuals per sample). Informed consent and pedigree information up to the third generation were obtained from each participant. The study includes only individuals whose four grandparents were born in the same sampling area.

At present Y-chromosome typing is in progress: around 680 individuals were typed for 80 SNPs and 19 STRs.

Preliminary results show that 34 haplogroups are represented in our sample, only five of them exceeding the 5% of the total. R1b1b2 lineages are the most frequent, recurring in around 40% of the samples. Other most common haplogroups are G2a, I2a2, E1b1b1a and J2a.

Besides Y-chromosome typing completion, mtDNA variability will be investigated by sequencing the HVS-I and analysing 22 biallelic markers.

These data will allow to explore the Italian genetic history and to shed light on the most important historical events related to the peopling of the country. In particular, lineage-specific investigations will serve as a powerful tool to unravel the complicated regional patterns characterising Italian history.

* With:

A. Boattini, D. Yang Yao, A. Useli, B. Martínez-Cruz, G. Ciani, D. Comas, J. Bertranpetit, D. Luiselli, D. Pettener.

Carla CALO' (Italy)*

Analysis of Y-chromosome polymorphisms in the linguistic isolate of Carloforte (Sardinia)

Carloforte is the only village located on the small island of San Pietro, off the southwestern coast of Sardinia (Italy). San Pietro was first populated in 1738 by emigrants coming from Tabarka (Tunisia) and originating from Pegli (Liguria, Italy).

For about 10 generations, those Genoese migrants had very little contact with the mainland populations of both Tunisia and Sardinia, maintaining a separate cultural as well as genetic identity. The cultural aspect is evident in the Pegli dialect, which is still spoken today, making the Carloforte population a linguistic isolate (Vona et al., 1996). Earlier studies based on matrimonial classical genetic markers and incidence of a specific disease provided evidence that Carloforte is a genetic isolate as well (Vona et al., 1996; Heath et al., 2001). Carloforte is characterized by a remarkably high endogamy rate (75.42%) and high percentage of consanguineous marriages (6.62%, $\alpha = 1.63 \times 10^{-3}$). Interestingly, in Carloforte the highest inbreeding values (Fit) were observed after 1850, due to a positive shift towards consanguineous marriages.

In this paper we present further data on the genetic structure of the Carloforte population by reporting on the distribution of 17 Y-STRs and Y-haplogroups.

Individuals from Carloforte selected for the present study (N=43) were proven descendants of the village founders. Moreover, the participants were chosen for not having ancestors in the region at least up to the grandparental generation. For comparison we selected a sample from Sulcis Iglesiente (southern-western of Sardinia), the nearest region to Carloforte.

Results on Y-chromosome (STRs and haplogroups) confirmed the genetic peculiarity of Carloforte, that turned out to be genetically differentiated from Sulcis Iglesiente. It is worth noting that Carloforte population shares common haplotypes with the Peninsular Italian population and not with Sulcis Iglesiente. Y-chromosome analysis confirmed the cultural and genetic isolation of Carloforte.

*With:

Vona G., Ghiani M.E., Scudiero C.M., Mameli A., Robledo R., Corrias L.

R. LELLI (Italy)**The peopling of Southern Italy: A maternal view*

Since prehistoric times Southern Italy has been a cultural crossroads of the Mediterranean basin. Genetic data on the peoples of Basilicata and Calabria are scarce and, particularly, no mtDNA variability has been published. In this study 415 individuals from Southern Italy was analysed for mtDNA in order to provide their classification into haplogroups. Median-joining network analysis was applied to observe relationship between the major lineages of the Southern Italians. Mitochondrial DNA haplotypes of populations from Apulia, Basilicata, Calabria, Campania and Sicily are compared, using multivariate analysis, with those of other Italian and Mediterranean populations, so as to investigate their genetic relationships. The haplogroup distribution in the Southern Italian samples falls within the typical pattern of mtDNA variability of Western Eurasia. The comparison with other Mediterranean countries substantial homogeneity of the area, which is probably related to the historic contact through the Mediterranean Sea. The bulk of the data demonstrated that Southern Italy shows the typical mtDNA pattern of Mediterranean basin variability, even though it is likely that Southern Italy was less affected of the LGM, which reduced genetic diversity in Europe.

* With:

OTTONI, C., MARTINEZ-LABARGA, C., RICKARDS, O.

Valentina COIA (Italy) (1,2)**Italian oriental Alps: mtDNA variation in geographically and linguistically isolated populations*

As a result of ancient and complex peopling processes and the presence of physical barriers, the alpine area provides unique opportunities for anthropological and genetic studies of ge and linguistic isolation. In the framework of the projects "Biodiversity and history of the populations from Trentino (BIOSTRE)" and "Isolating the isolates. Case study 2: Eastern Alpine (PRIN projects 2007-2009)", we investigated the genetic structure of twelve populations from the eastern alpine area, including four linguistically isolated groups. Our database includes individuals relative to Italian (Val di Sole, Val di Non, Val di Fiemme, Valle di Primiero, Valle dell'Adige, Valle Giudicarie and Valle del Fersina from Trentino and Val Cadore from Veneto) [(Altipiano di Luserna from Trentino, Sauris (Province of Udine), Sappada (Province of Belluno)] and Ladin speaking communities (Val di Fassa from Trentino). In this contribution, we report on variation of the mitochondrial DNA (sequencing of the hypervariable region 1 and typing of 17 SNPs) and compare the results obtained with literature neighbouring European populations. Our first results show a substantial differentiation between the linguistically isolated populations, more evident for the German-speaking communities, and other populations in terms of inter-genetic diversity and genetic signatures of demographic history. At the same time, we observed a worth noting heterogeneity among the linguistically isolated populations, even of common linguistic background (e.g. among Ladin groups from the Dolomites or between Sappada and Sauris communities).

*With:

Cinzia Battaglia¹, Vera Damiani¹, Fabrizio Rufo¹, Federica Crivellaro⁴, Patrizia Parisi⁵, Federica Trombetta⁵, Ilaria Boschi⁵, Laura Baldassarri⁵, Cristian Capelli³, Stefano Grimaldi², An Pedrotti² and G. Destro-Biso^{1,6}¹ Dipartimento di Biologia Animale e dell'Uomo, Università "La Sapienza" di Roma, Italia² Dipartimento di Filosofia, Storia e Beni Culturali, Università degli Studi di Trento, Italia³ Department of Zoology, University of Oxford, OX1 3PS, UK, Oxford⁴ Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK⁵ Istituto di Medicina Legale e delle Assicurazioni, Università Cattolica di Roma, Italia⁶ Istituto Italiano di Antropologia, Roma, Italia

The research is supported by the "Provincia Autonoma di Trento" (Post-doc PAT to V.C.) the M.I.U.R. (to G.D-B.) and the Istituto Italiano di Antropologia (to G.D-B.).

Maria LOPEZ-VALENZUELA (Spain)**Analysis of Y-chromosome variation in Gypsies*

Gypsies are one of the most interesting ethnic groups in Europe due to their semi-nomadic life style and their uncertain origin. Despite the lack of a written history, it has been suggested that they originated in the Indian subcontinent and they arrived to Europe in recent historical times. During their diasporas, Gypsy groups split as they migrate across Europe, although the exact endogamic costumes and isolation is not known. In the present study, Gypsies from several locations across Europe are compared to their neighbouring populations. The analysis of a set of chromosome STRs (Short Tandem Repeats) shows that Gypsies exhibit a different haplotype composition and reduced genetic diversity compared to the European groups, suggesting a recent bottleneck during the colonization of Europe. However, the Gypsy groups show a certain degree of genetic flow from European groups, our results show that the extent of this flow is different from country to country. An haplogroup prediction based on the Y-STR profiles reveals the high frequency of haplogroup H in the Gypsy groups. This haplogroup is shared by many Gypsy groups and is restricted to the Indian subcontinent. Our results are in accordance with previous studies that suggest that Gypsies originated in India in recent times from a small number of founders.

*With:

Begoña Martínez-Cruz and David Comas.

Roberto RODRIGUEZ-DIAZ (Spain)**Distribution of Surnames and Genetic Flow in a Rural Spanish Region: Genetic Structure*

The study of isolated populations with a subsistence economy is of special interest regarding biodemography. The reason is that the conditions to which they have been subjected are those present in great part of the history of our species, thus the conclusions drawn can be widely extrapolated. In this work two techniques have been employed (SOM and Monmonier application) in this field, to study the genetic structure of a small rural Spanish region: Fuentes Carrionas. It has its origins in the information contained in the matrimonial contracts and reconstructions between 1880 and 1979. The coefficients of relationships (Hedrick) have been calculated among the populations (by isonymy and progenitor-descendant matrix) and their relationship with geographical distances (Mantel) has been studied. Later, using Monmonier's algorithm the genetic barriers have been examined. Finally, applying self-organised maps, distribution of surnames has been studied.

The analysis demonstrates that isonymy and genetic flow offer similar results ($p < 0.01$) and that geographical distance is significantly related to both ($p < 0.01$), thus it seems to be the factor of isolation. Nonetheless, the genetic barriers show a region divided in two.

The distribution of surnames presents an identical division, 29.56% appear almost exclusively in the north-eastern half and 21.91% in the south-west. Both techniques yield coherent and complementary results. Their combined use allows a very detailed study, from which other results arise, which, although geographical distance has been the most determining factor, of different nature exist (orographic and socio-economic) which have marked the genetic structure of Fuentes Carrionas.

*With:

María José Blanco-Villegas (University of Salamanca, Spain)

Ashot HARUTYUNYAN**When genetics and history clash: The origins of Hamshen Armenians*

The Hamshenis are an isolated geographic group of Armenians with a strong ethnic identity who, until the early decades of the twentieth century, inhabited the Pontus area on the south of the Black Sea. Scholars hold different views concerning their origins. Broadly there are three alternative regions of origin suggested: (a) eastern Armenia (an area roughly covered by the state of Armenia), (b) western Armenia (an area that is now part of Turkey), and (c) Central Asia; with 'Armenia', in this context, meaning the historical country as it existed in antiquity. To establish whether data from the non-recombining portion of the Y chromosome would support one or another of the suggestions, so far as paternal descent is concerned, we screened 12 males of Hamsheni descent for 12 biallelic and 6 microsatellite Y chromosome markers and compared them with previously published data from populations representative of the three regions. DNA samples were collected in 82 residents of two villages (Novomichailovskiy and Tenginka) on the Black Sea coastal area of Russia (Krasnodar region) who refer to themselves as Hamshenis. Most of the ancestors of the extant group moved to these villages in 1915 from their settlements in the region of Samsun (westward from original area Hamshen). We found differences between the Hamshenis and other groups and support for western Armenia as the most likely region of origin. Specifically, the Hamshenis share their 'modal haplotype', one that encountered at the highest frequency (15.9%), only with Western Armenians (2.8%), while this haplotype is completely absent in other groups. The haplotype distribution and paternal genetic distances suggest a high degree of genetic isolation for the Hamshenis consistent with their retention of a distinct dialect of Armenian.

*With:

Mark Thomas, Ashot Margaryan, Neil Bradman and Levon Yepiskoposyan.

Renato POLIMANTI (Italy)**GST polymorphism in the Italian population: Anthropogenetic marker or marker of susceptibility?*

Glutathione S-Transferases (GSTs) (EC 2.5.1.18) are a supergene family of enzymes with roles in the cellular detoxification of a wide range of exogenous and endogenous compounds. Seven GST classes encoding cytosolic enzymes have been described (Alpha, Mu, Pi, Sigma, Theta, Zeta and Omega). Examples of allelic variation have been identified in several of these. The distribution of polymorphisms related to the cytosolic GSTs has been described in different populations and there is a growing literature showing associations between GST genotype and disease outcome. In the present study some cytosolic GST polymorphisms (GSTA1*-69C/T, GSTM1*0, GSTP1*I105V, GSTO1*A140D, GSTO1*E155del, GSTO1*E208K, GSTO2*N142D and GSTI1 investigated in an Italian population sample. GSTO1*E155del and GSTO1*E208K alleles were detected using the Confronting Two-Pair Primers (CTPP) analysis and allele specific PCR re while the analyses of other genetic polymorphisms were performed by PCR-RFLP method. The aim of this study is to clarify the geography of these genetic markers and the relationship between GST gene polymorphism, ethnicity and the prevalence of certain diseases. The results obtained were compared to those found for other populations in previous studies. The comparison of different situations: For some polymorphisms the allele frequencies proved to have different patterns among African, Asian and European populations, while for other GST gene polymorphisms the allele frequencies were not significantly different among the populations considered. Analyzing the mortality and the morbidity of diseases linked to GST gene polymorphisms we tried to evaluate the possible selection effect of diseases on GST genetic variability. In conclusion the final outcome of this research should lead to a better understanding of interactions between genetic variability and disease susceptibility.

* With:

Sara Piacentini and Maria Fuciarelli. Department of Biology, University of Rome "Tor Vergata".

Andrea NOVELLETTO (Italy)*Diet-driven dynamics of NAT2 variants in dispersed human populations*

Genetic variation at NAT2 has been long recognized as the cause of differential ability to metabolize a wide variety of drugs of therapeutic use. We explored the pattern of genetic variation in human populations that significantly extend the geographic range and resolution of previous surveys, to test the hypothesis that different dietary regimens and lifestyles may explain in population differences in NAT2 variation.

The entire coding region was resequenced in 98 subjects and six polymorphic positions were genotyped in 150 additional subjects. A single previously undescribed variant was found (3'12Y>H).

Our results can be summarized and interpreted as follows: 1) the NAT2 coding region is poorly differentiated in the population samples examined; 2) the major determinant of inter-population diversity are the phenotypic proportions; 3) population dispersals were not accompanied by a concomitant accumulation of molecular diversity; 4) the data fit the distribution obtained in populations already attaining polymorphic frequencies at the time of exit out of Africa.

Conversely, haplotype frequencies significantly differ across groups of populations with different subsistence styles. The pool of fast haplotypes show a strong decreasing trend in the order of gatherers/pastoralists/agriculturalist.

Based on previous biochemical evidence, we suggest the diminished dietary availability of folates resulting from the nutritional shift, as the possible cause of the fitness increase associated with haplotypes carrying mutations that reduce enzymatic activity.

We then propose that the present NAT2 diversity in human populations is the result of three distinct processes: i) presence of variation for slow-causing sites in widely dispersed populations (possibly as neutral variation) before major shifts to pastoralism and/or agriculture; ii) independent emergence of selective advantage for multiple slow-causing mutations in population from H-G to pastoralism/agriculture; iii) further introgression of slow-causing variants into populations anchored to H-G by later gene flow.

*With:

Francesca Luca, Giuseppina Bubba, Massimo Basile, Radim Brdicka, Emmanuel Michalodimitrakis, Olga Rickards, Galina Vershuby, Lluís Quintana-Murci, Andrey I. Kozlov, Andrea Nov

Elena GIGLI (Italy)**An improved PCR method for endogenous DNA retrieval in contaminated Neanderthal samples based on the use of blocking primers*

Neanderthal skeletal remains are usually contaminated with modern human DNA derived from handling and washing of the specimens during excavation. Despite the fact that the distinct haplotypes allow the design of specific primer pairs, for instance in most of the mitochondrial DNA (mtDNA) hypervariable region 1 (HVR1), the human contaminants can often outnumber endogenous DNA, thus preventing a successful retrieval of Neanderthal sequences. We have developed a novel PCR method, based on the use of blocking primers that preferentially bind human contaminant DNA and block their amplification, and greatly improve the efficiency of Neanderthal DNA retrieval. We tested the method in four El Sidron Neanderthal samples (two two bone fragments) with different contamination levels and taphonomic conditions, and we have been able to significantly increase the Neanderthal yield from figures around 25.23% (5- to 90.18% (75.3-100%).

*With:

Morten Rasmussen, Sergi Civit, Antonio Rosas, Marco de la Rasilla, Javier Fortea, M. Thomas P. Gilbert, Eske Willerslev, Carles Lalueza-Fox.

A. ZAULI (Italy)**HaPlone: A user-friendly web-based application for the management of molecular anthropology data*

The "BiBi - Biodiversity and Bioinformatics" project (University of Bologna Strategic Projects) was aimed to collect and organize molecular anthropology samples and data. One of the goals was the design, the implementation, the test and maintenance of a data base useful for theoretical studies".

To that end we developed HaPlone, a web-based application built on top of the state-of-the-art Plone (<http://plone.org>) content management system. The application allows to store, in and retrieve data through the familiar interface of a standard web browser.

Data are stored in the "Subject" data structure, containing both personal and molecular data, that can be inserted, inspected and edited using a user-friendly interface. For each subject application calculates on-the-fly the haplogroup, based on its tested UEPs, as well as the most recent common ancestor for each sexual lineage, based on the stored population.

The system also takes care of checking data consistency and flags the user for potential errors, such as inconsistent or conflicting UEPs or out-of-range STRs within a given subject. Population subsets can be easily selected (by location, haplogroup, sex, MRCA) using simple query forms, whose reports also provide basic statistics and charts on the selected sets. Full selected subject data can be readily exported to CSV files for processing by other applications such as spreadsheets or statistical packages.

By leveraging on Plone access-control features, the application can handle selective access to stored data, allowing fine-grained control on what it can be accessed by an anonymous or authenticated user, so it can be used both for internal information sharing and data dissemination simultaneously.

Currently the prototype handles only Y-chromosome molecular data (UEPs and STRs), but work is planned to extend data handling to mtDNA data too.

*With:

A. Boattini, A. Eusebi, M. Amico, I. Rossi, D. Luiselli, R. Casadio, D. Pettener.

Avshalom ZOOSMANN-DISKIN (Israel)*The origin of Eastern European Jews revealed by autosomal and sex chromosomal polymorphisms*

Objective: This study aims to establish the likely origin of Eastern European Jews.

Methods: This is done by genetic distance analysis of autosomal markers and haplotypes on the X and Y chromosomes.

Results: According to the autosomal polymorphisms the investigated Jewish populations do not share a common origin, and Eastern European Jews are closer to Italians in particular than Europeans in general than to the other Jewish populations. The similarity of Eastern European Jews to Italians and Europeans is also supported by the X chromosomal haplotypes. In contrast to the Y-chromosomal haplotypes Eastern European Jews are closest to the non-Jewish populations of the Eastern Mediterranean. The autosomal genetic distance matrix has a correlation (0.789) with geography, whereas the X-chromosomal and Y-chromosomal matrices have only a moderate correlation (0.375 and 0.425 respectively).

Conclusions: The close genetic resemblance to Italians accords with the historical presumption that Ashkenazi Jews started their migrations across Europe in Italy and with historical evidence that conversion to Judaism was common in ancient Rome. The reasons for the discrepancy between the results based on the autosomes and the X chromosome on the one hand and the Y chromosome on the other are discussed.

Luca PAGANI (UK)**Characterization, through re-sequencing, of genetic variants associated with high altitude adaptation in North Caucasian ethnic groups*

We are searching for signals of positive selection at candidate genes for high-altitude adaptation in North Caucasian highlanders using Illumina indexed re-sequencing. A total of 55 unresequenced individuals from three ethnic groups settled in ancient villages located over 2,000 meters above sea level were selected as the study population. Caucasian lowlanders (Adygei, n = 20) and 1 chimpanzee were used as controls. Archeological evidence suggests a long history (>10,000 years) of living at high altitudes, making the Daghestani populations suitable for adaptation to hypoxic stress, but possibly also experiencing an unusual demographic history. In order to disentangle selective and demographic effects, fifteen candidate genes involved in metabolism (HIF1 α ; PHD1; PHD2; PHD3; VHL; EPO; EPOR; VEGF; EDN1; NOS3; ACE; α -globin) were re-sequenced together with 27 putatively neutral control regions chosen from the Hominid Project and the ENCODE3 Project.

The regions of interest were amplified by long-PCR, checked by gel electrophoresis and those belonging to the same individual pooled and purified using the QIAquick PCR Purification Kit. The pooled sample was indexed by adding an eight-nucleotide tag according to a protocol developed at the WTSI, and sequenced using the Illumina GAII platform. The resulting reads were then aligned to the reference sequence. False positive and false negative SNP call rates were measured and an optimal set of SNP calls established.

Among more than 1000 novel SNPs, we found non-synonymous variants within the HIF1 α , ACE, EPOR and NOS3 genes that could be considered as candidates for hypoxia adaptation. SNPs in the neutral regions are being used to investigate the demographic history of the populations and the candidate targets of positive selection are being scanned for signals of positive selection.

*With:

Qasim Ayub (The Wellcome Trust Sanger Institute); Daniel MacArthur (The Wellcome Trust Sanger Institute); Yali Xue (The Wellcome Trust Sanger Institute); Iwanka Kozarewa (The Wellcome Trust Sanger Institute); Daniel Turner (The Wellcome Trust Sanger Institute); Sergio Tofanelli (Università di Pisa, Pisa, Italy); Kazima Bulayeva (Vavilov Institute, Moscow, Russia); Kenneth K. Kidd (University, Connecticut, USA); Giorgio Paoli (Università di Pisa,); Chris Tyler-Smith (The Wellcome Trust Sanger Institute).

Turi KING (UK)**Genome-wide analysis of coancestry among men sharing British surnames*

Men who share uncommon British surnames frequently share high-resolution Y chromosome haplotypes, providing unambiguous evidence of common paternal ancestry within the last 1000 years. We are using whole-genome analysis to examine the degree of autosomal coancestry among apparently unrelated men who share both a surname and a Y haplotype. Such groups are because they represent easily ascertained cohorts midway between the pedigree and the population, and could have utility in genetic epidemiological studies.

We are currently analyzing 80 men bearing six surnames with associated spelling variants, using the Affymetrix SNP 6.0 chip to type 906,600 SNPs, and the homozygosity haplotype method to search for shared autosomal segments.

*With:

Mark Jobling.

Daniel FALUSH (UK)**A new statistical method to infer population admixture events using genetic variation data*

We present a novel statistical method that uses densely-spaced Single-Nucleotide-Polymorphism (SNP) data to identify the major admixture events occurring throughout a population's history. The model has several advantages over leading available analytical approaches in this area, such as principal-components-analysis and STRUCTURE. In particular it can simultaneously (i) test

advantage of the information inherent in patterns of linkage disequilibrium, i.e. non-random associations amongst neighbouring SNPs along a chromosome, (ii) efficiently analyse hundreds of individuals at hundreds of thousands of SNPs genome-wide, and (iii) allow for relatively straight-forward interpretation and direct inference of key historical parameters, such as the times of major admixture events. Using simulated data matched to currently available human datasets, we show that our model can identify and accurately date admixture events that occurred between 7 and 150 generations ago. As our technique exploits the rich information in genetic data to infer details of a population's admixture history, it marks a powerful complementary anthropological research and can help to resolve a number of existing controversies. We present results from applications of our model genome-wide 650K SNP data for individuals from wide populations of the Human Genome Diversity Panel (Science 319, 1100-1104). The analysis identifies several important admixture events, some of which are historically well established identification of recent European genetic influx into the Maya Native American population), others that can be placed into a clear historical context (e.g. an East Asian genetic influx into Central and South Asian populations dated precisely to the era of the Mongol empire), and some that are to our knowledge novel (e.g. admixture in the Cambodian population between Central/South Asian source and an East Asian source dated to around the period of the Cambodian Empire). Plus bonus, unveiling of project X, also involving Daniel Lawson.

*With:
Garrett Hellenthal and Simon Myers.

Marco SAZZINI (Italy)*

Is Molecular Anthropology ready for the Next-Generation Sequencing Technologies revolution? A whole transcriptome sequencing case study

In recent years, a new generation of non-Sanger-based sequencing technologies has succeeded in sequencing DNA in a massively parallel fashion, enabling a huge reduction in the per-sequencing cost. This has brought genomics back from large genome centres into laboratories of small academic consortia, making the attainment of that comprehensive perspective also in the field of Molecular Anthropology research. The potential to explore the full spectrum of the human genome variability in an extremely more detailed way respect to the study variants launched by the HapMap project is thus imminent and will presumably establish new baselines for human evolutionary and complex diseases studies.

Nevertheless, a whole-genome sequencing approach still results prohibitively expensive due to the high sequence coverage required, as well as the huge amount of data rapidly produced. Next-generation sequencing technologies (NGSTs) has turned out to be an outstanding analytical challenge, making bioinformatics expertise and facilities essential. Several questions regard to the speed and ease of NGSTs assimilation into the mainstream of Molecular Anthropology.

In the attempt to deal with such questions, we describe results from a Whole-Transcriptome Shotgun Sequencing (RNA-Seq) case study in which 30 million 36 bp cDNA reads were generated from an individual sequenced by means of the Illumina technology. The mapping of reads to the human genome reference sequence led to the identification of more than 2,000 single nucleotide substitutions, as well as to the achievement of exhaustive alternative splicing and gene expression profiles. A comprehensive qualitative and quantitative picture of a human transcriptome drawn, demonstrating that NGSTs actually provide new promising opportunities to deepen the knowledge of human genome variation by simultaneously assaying a wide spectrum of genomic features in a time and cost-efficient way.

* With:
Paolo Garagnani¹, Alessio Boattini¹, Ilaria Iacobucci², Alberto Ferrarini³, Enrico Giacomelli³, Luciano Xumerle⁴, Giovanni Malerba⁴, Massimo DelleDonne³, Giovanni Martinelli², Donata

Roscoe STANYON (Italy)*

Evolutionary Molecular Cytogenetics provides a Pictorial Legacy of Human Origins and an Explicative Foundation for Contemporary Genomics

The evolution of the human genome is an integral part of our understanding of human origins. Chromosome painting in over 50 species of primates has allowed us to trace the origin of the genome and reconstruct the karyotype of long extinct ancestors. Recently, in situ hybridization of about 900 BAC (Bacterial Artificial Chromosomes) in an array of primate species allowed the evolution of marker order within each human chromosome. Classically, centromere position was considered highly conserved but the BAC hybridizations revealed that centromeres shift their position forming Evolutionary New Centromeres (ENCs). On the evolutionary line between macaques and humans there are 14 ENCs. An evolutionary perspective can provide underlying explicative grounds for contemporary genomic phenomena. Knowledge of ENCs provides an explanation for the clustering of human clinical neocentromeres. Clinical neocent cluster at "hotspots" that frequently are sites of deactivated centromeres or harbor ENCs in various primate species. Chromosome 14 and 15 in the ancestral primate genome were a single chromosome. This chromosome was fissioned in the ancestor of hominoids. The original centromere was deactivated and two new centromeres formed, one for chromosome 14 and one for chromosome 15. Clinical neocentromere cluster at the domain of the inactivated centromere at 15q25. The cluster of clinical neocentromeres at 3q26 is the locus where a ENC formed in primates. We recently reported on a clinical neocentromere at chr6:26,407-26,491 kb, precisely where our ancestor had a centromere which was deactivated in the human line after divergence from lesser apes. The centromere jumped back to its original position 17 million years ago. We can hypothesize that clinical neocentromeres and evolutionary neocentromere are two faces of the same coin, an example of Dobzansky's dictum that "nothing in biology makes sense except in the light of evolution."

* With:
Francesca Bigoni, Department of Evolutionary Biology, Laboratory of Anthropology, University of Florence.

Vincenza COLONNA (Italy)*

Detection of genetic structure in isolated populations: effects of consanguinity, divergence time and effective population size

Small human populations tend to diverge genetically from source populations because of several factors (e.g., geographical, social, religious, linguistic) resulting in reproductive isolation with small effective population sizes (N_e) at the founding event, genetic drift can rapidly cause genetic differentiation from the source population. Further, subdivision may increase and significantly contribute to reduce the effective population size. Thus, in the absence of migration, isolates can rapidly diverge from source populations, even when they separated recently. Genetic clustering could be observed even in closely-related populations.

With this study we investigate the influence of the study design on the extent of clustering in two cases:

In the first case, we considered the presence of familial groups in the sample. We used genetic data from two isolated villages with a common origin, presenting a high degree of structure which extensive genealogical data are available. We analyzed structuring after removing pairs of relatives (to variable degrees of relatedness) in samples from the two villages. We observed measures of population structuring decreased with the removal of familial groups, demonstrating that, indeed, observed genetic structuring is a consequence of consanguinity. Further, estimated the numbers of markers and sample sizes required to observe this effect.

In the second case, we considered the effect of divergence time (t) and N_e , assuming constant population size, in a more general case of an isolate separating from a source population. We considered variable lapses of time within a maximum of 50 generations, and we expected to observe decreasing clustering with increasing N_e and decreasing t . This expectation was confirmed by our simulated data. We provide quantitative estimates of this effect, as a function of N_e , t , and of the numbers of markers and individuals considered.

*With:
§RR Ferrucci, #M Ciullo, *G Barbuiani
§Dipartimento di Biologia ed Evoluzione, Università di Ferrara, Ferrara, Italy
#Istituto di Genetica e Biofisica "A. Buzzati-Traverso", CNR, Napoli, Italy

Francisc CALAFELL (Spain)*

Recombination-based human population genomics

Most inferences in human population genetics are based on the non-recombining mtDNA and NRY, and a hurdle usually cited in relation to autosomal and X-chromosome data is the lack of recombination. We have turned this argument around by studying human population diversity using recombination events as genetic markers. To infer past recombination events, we used a software called IRIS, which uses the patterns of adjacent SNPs created due to linkage disequilibrium by means of a combinatoric as well as a statistical algorithm based on pattern-switching. A preliminary run of the model, about 7Mb in the X chromosome were studied in the males of the 11 populations of the HapMap3 database, and 5166 recombination events were located. The terms of the position and of the haplotypes carrying the signal of the past event. Presence/absence of a particular recombination event was coded for each chromosome studied, and a string was termed a "recombinotype". We confirmed that our analysis correlated with recombination rates inferred through methods based on linkage disequilibrium and on sperm typing. We analyzed recombination events and recombinotypes with the same toolkit available for SNPs and haplotypes. Individual ancestry and population substructure were detectable with higher resolution using recombination events as markers rather than when using traditional allele frequencies. In addition, recombination analysis revealed an exclusive component within the African population that could correspond to the trace of ancestral hunter-gatherer African populations. The use of recombination events as genetic markers opens the door not only for human population genetics but for a deeper understanding on how recombination shapes genomes.

*With:
Marta Melé, Asif Javed, Laxmi Parida, Jaume Bertranpetit.

Martin RICHARDS (UK)

Archaeogenetics and the peopling of Asia

Abstract to be announced

Laure SEGUREL (France)*

Looking for genetic adaptations to diet from a comparative study of herders and agriculturalists in Central Asia

During the vast majority of their past, humans have been hunter-gatherers, with a diet poor in carbohydrates and a variable availability of food. This dietary pattern could have led to selective pressure for insulin resistance, a phenotype saving the precious glucose. Nowadays, in industrialized conditions (a high quantity and density of food with less physical activity) adaptations might have become an important genetic burden, leading for example to type II diabetes or other "civilization diseases". However, since the Neolithic transition, nearly 10,000 years ago, while hunter-gatherers and herders would still need to select for insulin resistance, farmers could have seen this selective constraint released, thanks to high levels of carbohydrates in their diet. According to these hypotheses, i.e. the thrifty genotype (Neel, 1962) and the carnivore connection (Colagiuri & Miller, 2002), past genetic adaptations to lifestyle are therefore responsible for important health disparities between ethnic groups. To test these hypotheses, we have collected phenotypic and genetic data in Central Asia, for Tajiks and Kyrgyz, known to respect the term farmer and herder populations. We have found that herders have nearly twice more risk to be insulin resistant than farmers, which is consistent with the previous evolutionary hypothesis. Furthermore, tests of neutrality on 11 candidate genes, known to be associated with type II diabetes, have revealed signals of balancing and local selection on some genes, which could have been involved in past adaptations to diet. However, for these genes, the causative mutation has been found in higher frequency in farmers. Further analyses based on haplotypic data will be used to understand how strongly and when these selection events have occurred.

*With:

Frederick DELFIN*

Y-chromosome genetic diversity of Filipino Negrito and non-Negrito groups

The Philippines are considered to be a strategic crossroad for human migrations in the Asia-Pacific region, and the origins and diversity of Filipino groups have been popularly explained as the result of several migratory incursions of populations from neighboring geographic regions. Of particular interest are Filipino Negritos, who with their characteristic short stature, kinked skin and traditional hunter-gatherer mode of subsistence have been considered to be descended from the earliest migration of modern humans to the Philippines, which may also be the migration to the Asia-Pacific region. As such a historical distinction between Negrito and non-Negrito Filipino groups has been perpetuated. Despite considerable anthropological interest groups, there is a paucity of genetic data on Filipino groups. We surveyed Y-chromosome diversity in 16 Filipino language groups (including six Negrito groups) and found extensive genetic diversity within, and heterogeneity among, both Negrito and non-Negrito groups. We find no Y-chromosome genetic support for the dichotomy between Negrito and non-Negrito groups. Filipino appear to have diverse genetic affinities with different populations in the Asia-Pacific region. Intriguingly, we find genetic links between some Negrito groups and indigenous Australians support the view that Negrito groups are descended from an early migration of modern humans to the Asia-Pacific region.

* With:

Jazelyn M. Salvador, Gayvelline C. Calacal, Henry B. Perdigon, Kristina A. Tabbada, Lilian P. Villamor, Saturnina C. Halos, Ellen Gunnarsdóttir, Sean Myles, David A. Hughes, Shuhua Xu, Lao, Manfred Kayser, Matthew E. Hurler, Mark Stoneking and Maria Corazon A. De Ungria.

Ellen Droefn GUNNARSDOTTIR (Germany)*

High-throughput sequencing of complete mtDNA genomes in three groups from the Philippines

The Philippines is vastly rich in cultural and ethnic diversity; of 85 million inhabitants there are over 100 ethno-linguistic groups spread over 7000 islands. All languages spoken in the P today belong to the Austronesian language family and it is believed that the majority of the inhabitants are descended from Austronesian farmers who migrated from Taiwan around 4000 years ago. But old human fossils dating as far back as 47,000 BP indicate that the Philippines were settled much earlier than the Austronesian expansion. It has been proposed that "Ne" in the Philippines have a distinct genetic origin because of their physical appearance (short stature, dark skin color, frizzy hair). To support this hypothesis the linguistic diversity of these groups accounts for a quarter of all the linguistic diversity in the Philippines, even though they only make up ~0.03% of the total population. However, little is known about genetic diversity in the Philippines; Surigaonon, Manobo and Mamanwa (a Negrito group). The data support the hypothesis that the Negritos represent an early, separate migration to the Philippines, as a unique haplogroup, containing previously unreported mutations, that branches off at the root of macrohaplogroup N. This study demonstrates the advantages of high-throughput sequencing of complete mtDNA genomes, both by giving unbiased estimates of genetic diversity and by refining the mitochondrial phylogenetic tree.

*With:

Mark Stoneking.

Irina PUGACH (Germany)*

A genetic record of Australian aborigines based on large-scale genotyping data

Australia holds some of the earliest archaeological evidence for the expansion of modern humans out of Africa, with initial occupation of Sahul (the Australia-New Guinea landmass) 40,000 years ago. Australia and New Guinea were separated by rising waters only during the end of last glaciation 8,000 years ago, which, if this landmass was settled by one population to around 40,000 years of shared history for the Australians and New Guineans. Studies of mtDNA and Y-chromosome genetic variation reveal little or no association between these two and the nature of this dissimilarity is still being debated. We are currently working with large-scale genotyping data from Australian aborigines, produced using the Affymetrix SNP Array platform. We have carried out principal component analysis on more than 750,000 autosomal SNPs, and our results suggest that the ancient association between Australia and New Guinea indeed exist, but that these populations must have separated very early in the history of Sahul. The extent of isolation of Australia following initial colonization is also a matter of debate. It has been suggested that gene flow to Australia from the Indian subcontinent occurred at the time of the introduction of dingo, and the appearance of microliths, during the Holocene. Maximum-likelihood based software frappe, we were able to detect a signal which reflects migration from India to Australia in times before European contact. Strikingly, we also detect indicative of ancient shared ancestry between Indian populations and Australia. It is possible, that this signal reflects the first human dispersal from Africa, through India to Oceania. We are looking for haplotypes with significantly longer than expected ranges of linkage disequilibrium (LD) to identify genomic regions bearing signatures of local positive selection.

* With:

Rostislav Matveyev, Kun Tang, David Lopez Herraes, Marc Bauchet, Peter Nurnberg, Manfred Kayser and Mark Stoneking.

Mannis VAN HOVEN (The Netherlands)*

Unexpected island effects at an extreme: human genetic diversity in Nias

The Indonesian island of Nias is located ~110 km west of North Sumatra. Its ~600,000 inhabitants speak a unique (Austronesian) language and are considered a separate ethnic group in Southeast Asia. To investigate the genetic affinities of Nias islanders, we analyzed paternally inherited Y chromosome (NRY) and maternally inherited mitochondrial (mt)DNA markers in a representative sample of >400 individuals from across the island. Surprisingly, basically only two NRY haplogroups were observed, one predominantly in the north and the other only in the south. Nineteen mtDNA haplogroups were observed, one of them being present at a frequency of 40% and all others at frequencies below 10%. Both Y-chromosome and mtDNA diversity and mtDNA hypervariable segment 1 (HVS1) diversity were found to be highly reduced in Nias as compared to other regional populations. Y-STR diversity was even lower than most Polynesian islands where, unlike Nias, reduced diversity is expected due to their remote geographic location. These observations suggest an unexpected and previously undetected bottleneck in Nias' population history and show that Nias forms an exception to the general pattern of high genetic diversity in Island Southeast Asian populations.

* With:

Marja van Schoor, Johannes Hämmerle, Lea Brown, Ingo Kennerknecht, Manfred Kayser.

Connie MULLIGAN (USA)*

A genetic perspective on peopling of the Americas

* Connie J. Mulligan, Department of Anthropology, University of Florida, Gainesville, FL 32605, USA

The colonization of the Americas represents the most recent major human occupation of an uninhabited land mass on the planet. Therefore, we may be able to ask increasingly specific 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. Certain aspects of the colonization that are agreed upon by the scientific community, i.e. a single migration originated from an East Asian source and crossed over the Bering land bridge entering North America (summarized in Fig. 1 and Kitchen et al. 2008). This process created a strong population bottleneck such that modern Native Americans show significant reduced 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 on 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 (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 the Americas, most likely around the Lake Baikal region, is widely accepted based on mtDNA and Y chromosome data. The idea of an early European migration to the Americas prior to Columbus' voyage in the 1490s was once proposed based on presumed Caucasian features of the famous Kennewick Man discovered in the state of Washington, but support for this idea has largely disappeared based on comparative skeletal analyses. The number of migrations was initially under debate, but 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 r (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 an unexposed Bering land bridge after the last glacial maximum (LGM) ~18,000-23,000 years ago.

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 tip of South America ~14,500 years ago (Dillehay 2008). Simple simulation studies show that a rapid expansion is necessary to maintain frequencies of the major mitochondrial haplogroups in 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 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 distinct waves of migration into the Americas. The first date is generally based on the initial diversification of New World-specific haplogroups. For example, mtDNA divergence dates of ~30,000-40,000 years ago (Bonatto and Salzano 1997), reflecting the initial diversification of New World genetic variation as the populations diverged from ancestral Asians but entry to 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 different datasets but use different mitochondrial genome substitution rates, i.e. 'fast' substitution rates (e.g. ~1.7 x 10⁻⁸ substitutions/site/year) support a post-LGM entry and 'slow' substitution rates (~1.26 x 10⁻⁸ substitutions/site/year) support a pre-LGM entry. Endicott and Ho (2008) recommend that substitution rate estimates should be based on an 'internal calibration' of the 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 several 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 entry to the Americas (first mentioned by Bonatto and Salzano 1997). Mulligan et al. (2008) estimated that ~7000-15,000 years were required to generate the New World-specific variation. It further proposed that the migrating population occupied Beringia during this period of isolation. Paleoeological data from ancient eastern Beringia are indicative of productive, dry grass 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

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) 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 nuclear and mitochondrial, 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 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 post-LGM. 2) A better understanding of the period prior to entry to the Americas is also worthy of study, i.e. Was Beringia the occupied land mass? How long was the occupation? What 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 events that occurred after the initial entry to the Americas. Studies that look for correlation between genetics and linguistics have a checkered history in terms of providing general insights, 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 toward 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 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 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, if 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. Coupled 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.

Rebecca JUST (USA)*

Characterization of a Native American mtDNA haplogroup C lineage

A new mtDNA haplogroup C founder lineage ("C4c") was recently identified in two Native Americans from Colombia [1]. The aim of the present study was to generate additional entire mtDNA sequences to further characterize this clade.

An American mtDNA control region population database was searched for potential C4 lineages (i.e. any haplogroup C sequence not attributable to C1 or C5). Entire mtDNA sequences were generated for a subset of the samples identified and from five private donors (n=21). A C4 phylogenetic tree incorporating both the newly generated and published entire mtDNA genomes was constructed and considered in comparison to recently published phylogenies [1-3].

We propose a revised definition of C4 in which previous Asian clades C4a and C4b are re-designated C4a1 and C4a2, and which now also includes former branch C7 [2,4]. Fourteen of the generated sequences and the previously published Colombian genome cluster together and comprise a Native American clade we term C4a3. Coalescent times estimated for the C4a3 are in agreement with previously published estimates for the divergence of Native American and Asian lineages [4]. The inclusion of two new Native American sequences within the C4a3 may indicate additional haplogroup C4 American founders. These data refine the poorly characterized Native American C4a3 founder lineage and modify the haplogroup C phylogeny.

[1] Tamm et al. Beringian standstill and spread of Native American founders. *PLoS One* 2007; 9:e829.

[2] Volodko et al. Mitochondrial genome diversity in Arctic Siberians, with particular reference to the evolutionary history of Beringia and the Pleistocene Peopling of the Americas. *Am J Hum Genet*; 82:1084-1100.

[3] van Oven and Kayser. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):e386-E394. <http://www.phyloree.org>.

[4] Soares et al. Correcting for purifying selection: an improved mitochondrial molecular clock. *Am J Hum Genet* 2009; 84:740-759.

*With:

Josefina M.B. Motti, Erin M. Gorden, Jodi A. Irwin, Jessica L. Saunier, Melissa K. Scheible, Michael D. Coble, Claudio M. Bravi.

Jodi IRWIN (USA)*

Geographic substructure in the mitochondrial DNA distribution of U.S. "Hispanic" populations

Self-described United States "Hispanics" represent individuals of Mexican, Central and South American, Cuban, Native American, Puerto Rican and even Spanish-speaking European origin. Demographic and historical data, the contribution from these various source populations to the US "Hispanic" population is highly dependent on geography within the US. Indeed, these contributions have been clearly detected in the gene pools of assorted regional populations via autosomal markers (Bertoni et al. 2003). While the available mitochondrial DNA data hint at discontinuity (Allard et al. 2003), the mtDNA picture is far from complete.

We recently analyzed 853 samples representing five regional populations from New York, California, Texas, Florida and Puerto Rico, and confirmed that dramatic differences exist among lineage compositions of regional "Hispanic" populations. While 39% of New York "Hispanics" exhibited African mtDNA haplotypes, only 5% and 1% of "Hispanics" from southern California and Florida, respectively, reflected African-derived lineages. Likewise, the proportion of Native American haplotypes differed dramatically between regional populations. 44% of the New York sample exhibited Native American haplotypes, as opposed to 70% and 78% of the California and Texas samples, respectively.

In an effort to more finely characterize the pattern of mtDNA variation across US "Hispanics", we have now extended our analyses to include data from over 2800 samples that represent geographic coverage of the United States (including Hawaii and Alaska). The results of these analyses provide the most detailed picture yet of the complex mtDNA landscape of US "Hispanics".

*With:

Thomas Parsons, Jessica Saunier, Melissa Scheible, Kimberly Sturk, Toni Diegoli, Michael Coble.

Mark WHITTEN (1) (Germany)

Investigating potential ascertainment bias in sample selection using complete mitochondrial DNA genome sequences of Siberian populations

Previous research on mtDNA HVR1 sequences from Siberian populations (Sakha (Yakut), Tuvan, Even, Evenk, and Yukaghir) has uncovered a high percentage of sequence type sharing making it difficult to detect putative admixture. Sequencing complete mtDNA genomes allows for more fine-scaled analyses to be performed which should provide a better understanding of the history of these populations.

Traditionally, complete mtDNA sequences have been generated using Sanger sequencing methods. However, presumably because of the increased costs and time involved with sequencing samples in a collection, the majority of the complete mtDNA genomes deposited in GenBank either come from studies focusing on specific haplogroups of interest or from studies where samples with unique mtDNA HVR1 sequences were chosen for complete mtDNA sequencing. An underlying assumption of the latter selection process is that if HVR1 sequences are identical in individuals then their complete mtDNA genomes should also be identical. However, if this assumption is incorrect, an ascertainment bias is potentially introduced.

To investigate this possible bias, we sequenced the complete mtDNA genomes of nearly 400 Siberian samples using a novel protocol that combines the preparation of indexed libraries with hybridization enrichment of mtDNA for sequencing on the Illumina Genome Analyzer II. This is a rapid, cost-effective method for sequencing complete mtDNA genomes to high (>50-fold).

We examined whether selecting samples for complete mtDNA sequencing based on HVR1 identity excludes potentially informative polymorphic sites and found that only around one-third of individuals with identical HVR1 sequences also have identical complete mtDNA genomes. Currently, we are determining the extent to which this biased sampling affects population prehistory. The higher resolution achieved with complete sequences is expected to shed light on the degree of admixture between groups whose languages show signs of contact.

*With:

MINGKUN LI (2), CESARE DE FILIPPO (1), BRIGITTE PAKENDORF (1)

1. Junior Scientists Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, 2. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology.

Lutz ROEWER (Germany)*

Local evolution in the Amazon basin - studies of Y chromosome markers

The tribal population under study, the Waorani, live in the Amazonian region in the east of Ecuador. The Waorani have a forest-dwelling hunter-gathering lifestyle, specific cultural practices and distinctive physical traits, e.g. a low mean stature. While Waorani groups with 6-10 families traditionally practised a nomadic lifestyle moving regularly from one camp to another, the culture is becoming sedentary. The aim of our study was twofold. First we study whether traditional family and clan system partially survived or broke down under the influence of acculturation and admixture with neighbouring ethnic groups and Western lifestyle. We compared two completely different communities, one with a size of ~80 inhabitants living according to their traditional school village with ~400 people of mixed ethnic ancestry which have abandoned their culture and family structures to a large extent. Altogether swabs of 154 individuals of these two groups were collected with group informed consent and analysed for known Y chromosome markers. Secondly we addressed the question of demographic history of the Waorani. For this purpose we compared the polymorphisms found in this isolated population with indigenous populations of the Amazon basin and other parts of South America as collected in the YHRD 3.0 repository.

*With:

Geppert M, Willuweit S, Zweynert S, Baeta M, Nunez C, Martínez-Jarreta B, Vacas-Cruz O, González-Solorzano J, González-Andrade F

David COMAS (Spain)

The Genographic Project: insights into Western/Central European variation

Institut de Biologia Evolutiva (UPF-CSIC), CEXS-UPF-PRBB, Barcelona

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: 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 major colonisations at a continental level. However, some migration routes, especially those within continents, are far from being completely understood. In order to shed light on the recent human migrations, National Geographic and IBM, with the participation of the Waitt Family Foundation, have launched the Genographic Project. This international project aims to provide data in order to reconstruct human movements through the analysis of uniparental genomes (mitochondrial DNA and Y chromosome). Several groups, one in each continent, are collected and performing the proper analyses. 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 resolution of the migrations within Europe and neighbouring geographic areas.

Guido BARBUJANI (Italy)*

Inference of demographic processes from comparisons of ancient and modern DNAs

* Dept. Biology and Evolution, University of Ferrara

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 inference 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 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 do 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 sample 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, (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 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 *n* 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 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. If 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 in a coalescent simulation, with parameters sampled from appropriately broad distributions of priors. An arbitrary number (threshold) of simulation experiments showing the shortest 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 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 differ 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-scale 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 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 for 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 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 Etruscan origins are unclear, with ancient historians suggesting either that they immigrated from Anatolia, or alternatively that they represent an autochthonous population (Barker and 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, models 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 Etruscan difference is unlikely to result from systematic errors in the ancient DNA sequences (Mateiu and 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 models of the genealogical relationships among samples (Figure 1). We found no evidence of genealogical continuity for two Tuscan communities, Murlo and Volterra, for which Model 2 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 models 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 basis 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' 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 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 population mosaic of mtDNAs, and cannot be regarded as globally descended from the people who inhabited the same regions in preclassical times.

Claudio FRANCESCHI (Italy)*

Genetics and anthropology of aging, age-related diseases and Chagas's disease: the experience of the Bologna group

We will present data regarding our experience on the role of genetics in human aging and longevity (Europe) and in Chagas's disease (Argentina) stressing the importance of an integrated/multidisciplinary approach where genetics and anthropology play a major role.

AGING AND LONGEVITY: 1. The AKEA study of exceptional longevity in Sardinia. In order to identify the centenarians all over the island and to collect data (health status, genealogical and biological samples (blood)) a complex logistical organization was set up. We will review some of the major data and the necessity of a more comprehensive approach for the interpretation of results. 2. The GEHA (Genetics of Healthy Aging) in Europe (2004-2010). Within the framework of this project 2500 old sibpairs (both members of the family being older than 90 years) and the same number of younger controls matched for sex and ethnicity in 11 European countries. The genome wide linkage analysis on the 90+ sibpairs has been completed and a GWAS is in progress. We will focus on the logistical and methodological problems related to the complexity of the populations involved. Finally we will present recent genetic data we have obtained from the study of the genetics of Chagas's disease in the Chaco region of Argentina. We will stress the critical importance for such study of a strict collaboration between anthropologists, medical and population geneticists. Indeed, this study has been possible owing to the optimal relationship with the Wichí population resulting from previous careful field study performed by the anthropologists. This collaboration has made possible to address in the best correct way difficult problems related to the informed consent and the obtainment of the approval of the Ethical Committees in Bologna and in Argentina. Data on population admixture (mtDNA and Y chromosome markers) of the two communities (Wichí and "criollos") living in the Chaco region will be presented.

*With:

Claudio Franceschi, Federica Sevini, Francesco Lescai, Donata Luiselli, Davide Pettener, Cristina Dasso, and Zeldina Franceschi

Mark COLLARD (Canada)

Anthropology and Archaeology

"Given the recent convergence of evolutionary psychology and human behavioral ecology-sociobiology, one might expect that the next generation of researchers will rapidly untangle all the mysteries of human behavior and cognition. Unfortunately, I do not think that this will happen quickly. The main reason is that no branch of the evolutionary social sciences has an adequate understanding of human culture."

Kim Hill in *The Evolution of Mind: Fundamental Questions and Controversies*, ed. by S.W. Gangestad and J.A. Simpson (Guilford Press, 2007), p. 351.

Kim Hill's assessment of the state of "evolutionary culture studies" in the foregoing quotation is overly pessimistic. For many years it was certainly the case that attempts to develop an approach to human culture were not only few and far between, but also largely theoretical. However, over the last 10 years the situation has changed, and there is now a reasonably solid body of empirical work in which cultural data are analyzed within the framework of evolutionary theory. This development should be of particular interest to molecular anthropologists because it has been driven in large part by the use of techniques developed by molecular biologists and other evolutionary biologists. In this paper, I will review two of the main "threads" within this literature. One is the application of population genetic modeling to cultural data. Most of the studies that form this thread have employed the neutral model, but recently researchers have also been developing selection-based population genetic models. The other thread I will discuss is the application of the cladistic method of phylogenetic reconstruction to cultural data. The studies in which this method has been adopted have addressed a range of issues from the processes that give rise to population-level cultural diversity to the colonization of the New World to the evolution of ancient systems. In the final part of my talk I will outline some of the problems that will have to be overcome before the evolutionary analysis of culture becomes mainstream and suggest ways molecular anthropologists can help.

Quentin ATKINSON (UK)*

The prospects for tracing deep language ancestry

* *Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 64 Banbury Road, Oxford OX2 6PN, UNITED KINGDOM*

"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 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 languages. Proposed language super-families - such as Amerind in the Americas and Nostratic and Eurasiatic in Eurasia - or global language classifications like those controversially link 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 ancestor 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 due to chance (Ringe, 1998). However, recent advances in the available data and methods (Dunn et al., 2005; Pagel, 2000; Pagel et al., 2007; Reesink et al., in press) have 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 homologies or 'cognates' identified on the basis of form and meaning correspondences across languages. One has 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 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 on average roughly 20% of cognates are lost every 1000 years. As shown in Figure 1 (blue line), such a rate implies that a pair of languages that diverged just 4,500 years ago (separate 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 signals 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 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 items 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 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 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 (red 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 (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. 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 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. 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

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 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. 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 useful 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. (2008), 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 to 10,000 years. More recently, Reesink, Singer and Dunn (In press), have used structural data to classify the languages of the ancient super-continent Sahul into recognized major groups 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 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.

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 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 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 Structures (Haspelmath et al. 2005)) but more work is required on these lines if we are to fully capitalise on the linguistic legacy of our cultural past.