

MEETING REPORT

Insights into modern disease from our distant evolutionary past

Bryndis Yngvadottir*,¹

¹*The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*

An EMBO workshop entitled ‘Human Evolution and Disease’ was held recently (6–9 December 2006, Hyderabad, India) where 141 scientists from many disciplines came together to discuss recent studies of human variation, origins and dispersal, natural selection and disease susceptibility. The meeting tackled the subject of human evolution and disease from the different perspectives of archaeology, linguistics, genetics and genomics based on both new and publicly available data sets. In this report, we highlight the latest fashion crazes in the discipline, in particular, the use of large public data sets and new methods to analyse modern human variation and the links between human evolution and disease susceptibility.

European Journal of Human Genetics (2007) 15, 603–606. doi:10.1038/sj.ejhg.5201811; published online 14 March 2007

Keywords: conference; human evolution; disease; human variation; archaeology; linguistics

‘Human Evolution and Disease’ was the title of a multi-disciplinary EMBO workshop organized by Kumarasamy Thangaraj (CCMB, India) and Chris Tyler-Smith (The Wellcome Trust Sanger Institute, UK) held in The Centre for Cellular and Molecular Biology (CCMB) Hyderabad, India 6–9 December 2006, where 141 scientists from 11 countries came together to discuss recent advances in these two linked fields. A full list of abstracts can be obtained at <http://www.ccmb.res.in/events/pastevents/embo2006/meetingreport>.

Studies of human origins have commonly tried to combine evidence from archaeology, linguistics and genetics. Genetics, motivated by the aim of understanding complex disorders, is now producing large and often publicly available data sets, such as the genome sequences from human,¹ chimpanzee² and Neanderthal³ (to name a few) as well as data on normal variation from the HapMap⁴ and copy number variation (CNV)⁵ consortia. Although the HapMap samples are without doubt the most widely

used sources, people are increasingly analysing larger samples, both from the HGDP-CEPH worldwide panel of diverse populations⁶ and their own distinctive local populations. For example, Gyaneshwer Chaubey (University of Tartu and Estonian Biocentre, Estonia) cited 10 000 Indian samples in one study. It is easy to see how such powerful resources could benefit evolutionary studies, but this meeting also illustrated the importance of an evolutionary perspective for understanding disease. The origins and migrations of our species, as well as past natural selection, influence complex disease susceptibility, and will provide increasingly important inputs into medical genetics.

Conflicting ideas about modern human origins and dispersals

The origin and dispersals of anatomically modern humans (AMH) has long been a subject of hot debate in genetics, archaeology and palaeontology. Current evidence generally supports a recent single origin in East Africa and subsequent spread throughout the world. However, the agreement among scientists seems to stop there, as the timing and routes of these dispersals are still debated.

Paul Mellars (University of Cambridge, UK) addressed two key questions. Why, if AMH originated in Africa 150–

*Correspondence: B Yngvadottir, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, UK.

Tel: +44 0 1223 834244; Fax: +44 0 1223 494919;

E-mail: by1@sanger.ac.uk

Received 1 February 2007; accepted 2 February 2007; published online 14 March 2007

200 KYA (thousand years ago) did they only disperse out of Africa 50–60 KYA? To answer this, he pointed to the archaeological signature revealing major technological, economic and social developments in southern Africa roughly 60–80 KYA, which could have been crucial driving forces for AMH to expand their range successfully throughout the world. And was there one major exodus, or two? He suggested that archaeological evidence for distinct ‘southern route’ and ‘northern route’ stone tool traditions was likely to reflect just the materials available and ‘cultural drift’, so archaeology does not require more than one expansion.

The prehistory of Southeast Asia has stimulated a wide range of speculations, for example, concerning the unique Andaman Islanders and the origin and migration route of ancient Austronesians into Polynesia. These topics provide a flavour of the continuing debates about dispersals.

Lalji Singh (CCMB, India) gave evidence from mtDNA sequences that two ancient maternal lineages, M31 and M32, in the Onge and Great Andamanese do not match any other populations in the world, indicating that the Andaman Islanders have survived in complete genetic isolation from other South and Southeast Asian populations, as the migration of AMH out of Africa.⁷ However, although the Great Andamanese and Onge share these genetic lineages, their languages seem to be very distinct from one another. Anvita Abbi (Jawaharlal Nehru University, India) proposed that their languages belong to different language families: Proto Great Andamanese and Proto Ang, which probably arose in the Indian subcontinent and gave rise to Great Andamanese and Ong/Jarawa, respectively, over the last several thousand years. As George van Driem (Leiden, The Netherlands) reminded us, although linguistics, archaeology and genetics have often been used in combination to provide support for human origins, the evidence from the three disciplines does not always tell the same story. Languages, in particular, may be better seen as leaves that have fallen to the ground and lost ancient information about their relationships, rather than as branches on an informative evolutionary tree.

One of the more famous debates within archaeology, linguistics and genetics is about the origin and settlement history of the Polynesians, where different people end up telling different stories according to the type of evidence used. A model favoured by many linguists and archaeologists is the ‘express train’ model.⁸ This uses linguistic evidence placing Taiwan as the origin of the Austronesian language family and archaeological evidence of Lapita pottery found in Polynesia also with an assumed origin in Taiwan. The model predicts a recent and rapid expansion of Austronesian-speaking farmers from Taiwan who sailed off to Polynesia via Near Oceania 3.6–6 KYA without mixing with indigenous populations on their way. However, the ‘entangled-bank’ model⁹ proposes a long (perhaps 40 KYA) history of cultural and genetic transmission

between indigenous populations in Near Oceania during the settlement of Polynesia. Furthermore, a modified version of the express train is the ‘slow boat’ model,¹⁰ which assumes that the dispersal was not a rapid one but rather a slow migration through Near Oceania allowing cultural and genetic admixture with indigenous populations on the way to Polynesia. The origin of Polynesians is still being discussed as two talks focused on resolving these models with genetic data.

Another interpretation of the slow boat model would place the origins of Polynesians in eastern Indonesia rather than Taiwan.¹¹ In accordance with this, Martin Richards (University of Leeds, UK) introduced mtDNA results from populations in Malaysia and Indonesia indicating a substantial indigenous population stratum throughout Southeast Asia extending back to the earliest modern human settlement, at least 50 KYA. Richards therefore proposed that Austronesian speakers arose from indigenous populations within Island Southeast Asia through dispersals from west to east rather than from north (Taiwan) to south. He furthermore pointed out that the linguistic tree could be interpreted not only with an origin in Taiwan but alternatively with origins in the Philippines or Borneo.

In order to test the previously mentioned models and infer the origins of Polynesians, Manfred Kayser (Erasmus University, The Netherlands) presented results comparing mtDNA, Y chromosome and autosomal data from Western and Central Polynesian Islands with potential source populations from East and Southeast Asia as well as island and mainland New Guinea. The results were best fitted with a slow boat from Asia through Melanesia with a stop in New Guinea. The Y chromosome and mtDNA data showed a gradual west to east decrease in diversity indicating the direction of settlement being west to east. There was also evidence of sex-biased admixture with Melanesians: Polynesian men seem to have ‘mingled’ more with the Melanesians than did the women, most likely as a result of matrilocality. Pre-Polynesians are therefore suggested to have experienced a genetic but not linguistic admixture with a maternal transfer of Austronesian languages.

Defining boundaries: geography, language and genes

In recent years, the idea of genetic boundaries has become increasingly debated. Do they exist, and if so, what creates a genetic boundary? Is it language, geography, culture or a combination of these?

From a worldwide perspective, Guido Barbujani (University of Ferrara, Italy) pointed out the difficulties of associating genetic differences with sharp boundaries. Barbujani introduced a reanalysis of 377 microsatellites in the CEPH human diversity panel⁶ with a new statistical method that detected zones of increased genomic change.

Barbujani's group identified nine population clusters from the same data used previously to identify six clusters.¹² Although it is possible to cluster genotypes according to geography, language or other criteria, the clusters found depend on the assumptions of the model used. To date, there is no robust overall genetic subdivision of human-kind.¹³

The Himalayan region, however, provides a rare example of a location that does contain a well-defined geographic, linguistic and genetic boundary. Chris Tyler-Smith presented a study of autosomal, mtDNA and Y-chromosomal markers revealing that, despite the presence of the highest mountain range on earth, genetic variation actually correlated better with language than with geography. In contrast, Gyaneshwer Chaubey presented a pattern of a genetic variant, the R7 mtDNA haplogroup, which fitted better with geography than language in the Indian subcontinent.

Human variation

The analysis of CNV in the human genome is becoming increasingly important as its role in evolution, genetic diversity and disease is seen as comparable with that of SNP variation. Richard Redon (The Wellcome Trust Sanger Institute, UK) introduced results from the first CNV map of the human genome where 1477 CNV regions covered ~12% of the genome in the four HapMap populations.⁵ CNV is extensive, genome-wide, complex and likely to have major functional impact. The importance of CNV in human evolution is, however, poorly understood and Yali Xue (The Wellcome Trust Sanger Institute, UK) described a search for recent positive selection in the CNVs identified by Redon *et al.*⁵ For example, the gene *UGT2B17* revealed a striking pattern of global differentiation and unusual nucleotide diversity at the breakpoint. Indeed, the investigation of the mysterious world of CNVs has only just begun and is likely to yield some interesting findings in the next few years.

As more genomic data become publicly available, we see the advantages of combining data sets. By using the recently published sequence from the Neanderthal genome³ and the HapMap data,⁴ Sridhar Kudaravalli (University of Chicago, USA) could begin to study the population genetics of humans and Neanderthals. He estimated the split time between the two species at ~400 KYA and found no evidence for admixture between humans and Neanderthals.

With recent advances in sequencing technologies and DNA extraction, the use of ancient DNA in the study of human origins is becoming ever more popular. However, there is still no reliable method to authenticate the endogenous template from a highly degraded and possibly contaminated sample that has been amplified by PCR. To

address this issue, Agnar Helgason (deCODE Genetics, Iceland) proposed a novel method using the information obtained from post-mortem damage observed among cloned sequences and applying an evolutionary model to determine the 'real' endogenous template.

Searching for selection and disease candidate genes

In a workshop on human evolution and disease, the subject of natural selection provided a central theme. Although new beneficial traits are expected to trigger positive selection, disease-causing mutations would likely be washed away by negative selection. When AMH migrated out of Africa ~50 KYA, they had to adapt to new environments, nutritional sources, parasites and diseases, which may well have caused selection to leave a detectable signature in their genome. However, as there is no single test for selection that applies to all circumstances and all types of data, and because chance or demographic events can mimic the effects of selection, the choice of which test to use can be a tricky one. Furthermore, deciding whether the statistic used deviates from neutral expectations is also an important issue. Although many people use empirical comparisons to achieve this, Agnar Helgason promoted the use of coalescent simulations, which took into account known demographic parameters and were conditioned on allele frequencies and the recombination map.¹⁴ Furthermore, both Rasmus Nielsen (University of Copenhagen, Denmark) and Chris Ponting (University of Oxford, UK) emphasized the utility of the allele frequency spectrum to reveal selection, because positively selected genes tend to show an excess of high frequency alleles. Nielsen further went on to argue that genes with both an excess of low frequency alleles and a high ratio of fixed differences between species to polymorphic differences within species are more likely to be disease-associated than other genes. As the search for selection tends to be focused on coding regions because their function is understood, Ponting reminded us that the majority of regions estimated to be affected by purifying selection (negative selection) are non-coding and that 1/7 disease causing mutations are in fact found outside the coding region.

Although the identification of positive selection is still a goal of many scientists, Mark Stoneking (Max Planck Institute for Evolutionary Anthropology, Germany) pointed out that there are rather few well-supported examples of genetic variation being altered because of a phenotypic effect influenced by natural selection. He gave a sobering example of *TRPV6*, which regulates calcium uptake. The gene showed a signature of positive selection dating to about 11 KYA in all non-African populations and was suspected of involvement in dairying. However, functional analyses found no significant differences

between the ancestral and derived alleles in their calcium channel activity. Stoneking noted that identifying the genetic basis even of phenotypes that differ greatly among populations can be problematic.

Jaume Bertranpetit (University Pompeu Fabra, Spain) presented a study aimed at identifying selection mediated by pathogens. As infectious diseases have strong geographical structure, he expected to find similarly high geographical structure in variants of genes related to host-pathogen interaction, but found no excess in this type of gene. Footprints of pathogen-driven selective forces were neither seen in a global picture of genetic diversity nor in the whole pathway.

Sudhir Kumar (Centre for Evolutionary Functional Genomics, USA) emphasized the importance of evolutionary analyses for understanding patterns of human disease mutations. He showed how comparison of amino-acid substitutions across different species can help distinguish disease-associated mutations (DAMs) from neutral amino-acid variations. DAMs tend to be found in positions conserved across species, and a human variant in a position that is variable between species is unlikely to cause disease. However, 10% of DAMs do occur in positions that vary between species, emphasizing the importance of species-specific effects.

Taking the focus down to a specific disease, Inês Barroso (The Wellcome Trust Sanger Institute, UK) noted the strong genetic component of type II diabetes. The explanation for why a largely genetic disease is so common may be found in the 'thrifty gene' theory proposed by James Neel.¹⁵ Thrifty genes are thought to have been selected in ancient times to enable fat storage to protect people against starvation during times of famine, but with the readily available calories accompanying Western lifestyles carriers became prone to obesity and diabetes. Association studies are now beginning to identify susceptibility genes. Agnar Helgason reported an evolutionary investigation of one such gene *TCF7L2*, whose association has been replicated 10 times.¹⁶ Interestingly, he found evidence for positive selection ~4–11 KYA, corresponding to the onset of agriculture, but not acting on the risk variant (HapB). The recently selected variant (HapA) was protective against diabetes but surprisingly was associated with an increased, rather than decreased, body/mass index. The reality seems to be more complex than the neat thrifty gene theory.

Around 30 million people suffer from heart disease and yet the aetiology often remains unknown. Two hundred mutations have been identified in 20 different genes but *MYBPC3* is thought to be involved in ~45% of the total cardiomyopathy. To shed light on this, Kumarasamy Thangaraj (CCMB, India) described a study of ~6000 Indian samples revealing a 25 bp deletion in *MYBPC3* thought to induce hypertrophy in both homozygous and heterozygous state, but with a remarkable 4% or so

prevalence. The deletion seemed to have a single origin and to have drifted to this high frequency, perhaps behaving in an evolutionarily neutral way because of its late onset.

The workshop often contrasted the results from different disciplines. Although they all investigate a common history, archaeologists may be able to provide the best insights into the timing of events, whereas linguists and geneticists can infer the structure of ancient phylogenies by looking at the modern variety, be it in language or DNA. Evolution has strongly influenced contemporary disease susceptibility and now a number of large data sets and methods provide a common ground to investigate both areas. This work has only just begun and perhaps soon meetings that combine human evolution and disease will be the norm.

Acknowledgements

I thank Chris Tyler-Smith for his helpful comments and The Wellcome Trust for the funding enabling me to attend the workshop. The main sponsors of the workshop were EMBO and CCMB. Additionally, the hosts from CCMB did a great job taking care of the visitors.

References

- 1 The International Human Genome Mapping Consortium: A physical map of the human genome. *Nature* 2001; **409**: 934–941.
- 2 The Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005; **437**: 69–87.
- 3 Noonan JP, Coop G, Kudaravalli S *et al*: Sequencing and analysis of Neanderthal genomic DNA. *Science* 2006; **314**: 1113–1118.
- 4 The International HapMap Consortium: A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.
- 5 Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
- 6 Cann HM, de Toma C, Cazes L *et al*: A human genome diversity cell line panel. *Science* 2002; **296**: 261–262.
- 7 Thangaraj K, Chaubey G, Kivisild T *et al*: Reconstructing the origin of Andaman Islanders. *Science* 2005; **308**: 996.
- 8 Diamond JM: Express Train to Polynesia. *Nature* 1988; **336**: 307–308.
- 9 Terrell JE, Hunt TL, Gosden C: The dimensions of social life in the Pacific - Human diversity and the myth of the primitive isolate. *Curr Anthropol* 1997; **38**: 155–195.
- 10 Kayser M, Brauer S, Weiss G *et al*: Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 2000; **10**: 1237–1246.
- 11 Richards M, Oppenheimer S, Sykes B: mtDNA suggests Polynesian origins in Eastern Indonesia. *Am J Hum Genet* 1998; **63**: 1234–1236.
- 12 Rosenberg NA, Pritchard JK, Weber JL *et al*: Genetic structure of human populations. *Science* 2002; **298**: 2381–2385.
- 13 Barbujani G, Belle EM: Genomic boundaries between human populations. *Hum Hered* 2006; **61**: 15–21.
- 14 Sabeti PC, Walsh E, Schaffner SF *et al*: The case for selection at CCR5-Delta32. *PLoS Biol* 2005; **3**: e378.
- 15 Neel JV: Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am J Hum Genet* 1962; **14**: 353–362.
- 16 Grant SF, Thorleifsson G, Reynisdottir I *et al*: Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38**: 320–323.