1992M

Insights from low-coverage whole Y chromosome sequencing of 1,244 individuals. *Y. Xue for The 1000 Genomes Project 1000Y Group.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The 1000 Genomes Project Phase 3 has sequenced 1,244 males belonging to 26 populations from Africa, South and East Asia, Europe and the Americas. In addition to these low coverage (~4-6x) sequences, the project also includes over two hundred males sequenced to high coverage by Complete Genomics and several sets of SNP genotype data used for validation. The group generated a union set of 80,895 Y-SNP, 2,830 Y-MNP and 6,076 short indel calls by combining results from seven different callers. Approximately 6,000 Y-STRs and ~1,000 large structural variants were also called, using two callers for each class. Validation suggests that the Y-SNP, Y-STR and large deletion and duplication calls have very high quality, but that the Y-MNPs and indels do not. Using 59,666 stable high confidence Y-SNPs, we have constructed a phylogenetic tree, to which the more complex classes of variant can be added. The tree recapitulates and extends the established phylogeny. It confirms a very rapid Paleolithic expansion (in number) of Y lineages post-dating the movement out of Africa, and Neolithic or later expansions of independent Y lineages in Africa, Europe, East Asia and South Asia. We observe different patterns in different continental regions, suggesting that this male expansion was extremely rapid in Europe, rapid in Africa, and less rapid in South and East Asia. These data thus provide powerful new insights into male evolutionary history and promise further insights into Y-chromosomal mutation and selection processes.

1993S

Exploring the Y-Chromosome Variation of Modern Panamanians. *A. Achilli[1], V. Battaglia[2], V. Grugni[2], U.A. Perego[1,3], H. Lancioni[1], M. Tribaldos[4], A. Olivieri[2], I. Cardinali[1], E. Rizzi[5], A. Raveane[2], M.R. Capodiferro[1], S.R. Woodward[3,6], J.M. Pascale[4], R. Cooke[7], N. Myres[3,6], A. Torroni[2], J. Motta[4], O. Semino[2].* 1) Dept. of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy; 2) Dept. of Biology and Biotechnology, University of Pavia, Pavia, Italy; 3) Sorenson Molecular Genealogy Foundation, Salt Lake City, Utah, USA; 4) Gorgas Memorial Institute for Health Studies, Panama City, Panama; 5) Institute of Biomedical Technologie, National Research Council (ITB-CNR), Segrate, Milano, Italy; 6) Ancestry, Provo, Utah, USA; 7) Smithsonian Tropical Research Institute, Panama City, Panama.

The Isthmus of Panama - the narrow neck of land connecting the northern and southern American landmasses - was a forced corridor for the Paleo-Indian expansion that originated from Beringia ~15-17,000 years ago. Archaeological findings suggest that some descendants of the earliest migrants remained on the isthmus, while accounts from early European explorers witnessed the presence of two main indigenous groups (the Cueva and the Coclé) in pre-Columbian times - populations that have since disappeared due to disease, warfare, and enslavement following the Spanish conquest. Today's indigenous groups total about 5.3% of the Panamanian population, and are mainly represented by the Ngöbé, Buglé, Kuna, Emberá, and Wounan tribes, which traditionally appear to have settled in Panama from surrounding regions after the autochthonous natives were decimated. However, there is no evidence that the ancestral indigenous gene pool was completely replaced. If this was the case, the populations of modern Panama should have retained at least a fraction of the native pre-Columbian gene pool, possibly at a variable extent, given the differential degree of geographical and genetic isolation of the different Panamanian communities during the past five centuries. A recent study of the mtDNA history of the modern Panamanian population (Perego et al., 2012), based on a sample of 1565 individuals with Native American maternal ancestry, concludes that (1) the first settlement of Panama occurred quite rapidly after the initial colonization of the American continent, (2) based on complete sequence analyses, the founder ages of the most common lineages point to an ancient expansion supporting the antiquity of the Pacific coastal route, 3) the mitochondrial gene pool exemplifies the link between pre- Columbian and modern Panamanian populations (in fact, 83% of modern Panamanians clusters into native pan-American lineages). It appears that the Spanish conquistadores and additional more recent European demographic influences did not contribute significantly to today's genetic composition of Panama, at least with regard to the maternal side. In this study, we have now tested the same scenario from the paternal side by employing the analysis of the Y-chromosome variation in modern Panamanians.

1994M

Population specific patterns of novel haplotype groups at the *PAH* locus. *G. Povysil[1], S. Wieser[2], S. Hochreiter[1], J. Zschocke[2].* 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria; 2) Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria.

The phenylalanine hydroxylase (*PAH*) gene is of particular interest for population genetic studies because the distribution patterns of well-defined phenylketonuria (PKU [MIM 261600]) mutations can be linked with distinct SNP haplotypes for the assessment of ancient migration. Through family segregation analysis and molecular haplotyping with long-range PCR in PKU patients (232 *PAH* mutant alleles) and controls (157 *PAH* normal alleles) from various European countries we identified five major haplotypes in the distal 15 kb region of the *PAH* gene. Haplotypes differ by 3-16 specific SNPs each and have been quite stable over the last millennia. The 29 common European PKU mutations can be linked to specific haplotypes with little evidence of recombination in the *PAH* gene. The results were compared to available sequencing data of Africans, East Asians, Europeans, and Admixed Americans from the 1000 Genomes Project. Additional data from chimpanzee, orangutan, and macaque, as well as high coverage sequences of Neandertal and Denisova, were used in conjunction with these data to establish a possible evolutionary tree of haplotype emergence. There are five major distal *PAH* haplotypes that can be found in all continental populations, but at different relative frequencies. For Europeans and Asians they make up more than 98% of all *PAH* alleles. While Europeans have comparably high frequencies for all of them, the most common haplotype in Asians amounts to almost 78% of *PAH* alleles. In contrast, Africans have many very rare haplotypes that can only be found in Africans, or Africans and Admixed Americans. The ancestral haplotype that matches the sequences of chimpanzee, orangutan, and macaque, has only been found in Africans and one Admixed American individual. The haplotype that matches the Neandertal and Denisova sequence can be found in Africans, Admixed Americans and one Asian individual. Interestingly, additional variations on the Asian haplotype are more similar to the Neandertal than to other present day individuals with this haplotype. The combination of disease mutations and common gene variants with molecular haplotyping and available genetic data from different countries allows a unique insight in the genetic history of human populations.

1995S

High-coverage sequencing of diverse Y chromosomes via in-solution capture. *A.F. Adams[1], G.D. Poznik[1], R.E.W. Berl[2], N.A. Hammond[1], F.L. Mendez[1], O.E. Cornejo[2], M. Napel[1], M. Schertler[1], P.A. Underhill[1], M.W. Feldman[1], C.D. Bustamante[1].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) School of Biological Sciences, Washington State University, Pullman, WA.

As the longest stretch of non-recombining DNA in the human genome, the Y chromosome provides unique insight into the demographic and evolutionary history of modern humans. However, most large-scale studies of human Y chromosome diversity have been confined to a small number of STRs and SNVs. We conduct high-throughput sequencing of Y chromosomes to study male-line demography and evolution.

Having previously determined which regions of the Y chromosome are amenable to short-read sequencing, we designed a set of NimbleGen biotinylated DNA probes to target these 10.4 Mb. We then performed in-solution, hybridization-based capture, using 105 individuals from the 1000 Genomes Project to optimize and validate our capture protocol. Because our approach yields higher coverage than that obtained by the 1000 Genomes Project, we are able to increase resolution at the tips of the phylogeny—the portion most informative of recent demography, such as admixture within the Americas. Following protocol optimization, we performed Y-chromosome capture on 34 saliva-extracted samples from three southern Ethiopian minority ethnic groups. These data provide an unprecedented look at the population structure of southern Ethiopia.

Our capture protocol can increase coverage of the Y-chromosome target regions by two orders of magnitude, as compared to whole-genome shotgun sequencing. We achieve >30× coverage of the targeted region for each of 24 samples multiplexed in a single HiSeq lane. We report some of the first high-coverage sequencing of the hgA and hgA0 lineages and uncover deep novel substructure within the poorly characterized E2 haplogroup. We are now working to apply this capture method to a large panel of individuals from a wide range of populations in order to gain a greater understanding of human Y chromosomal diversity.