Genomic Med. (2008) 2:253–271
DOI 10.1007/s11568-009-9096-3

ABSTRACTS

Genome variation, diversity and evolution

108: Updating phylogeny of mitochondrial DNA macrohaplogroup M in India

[1]**Chandrasekar Adimoolam**, [2]V. R. Rao

[1]DNA Polymorphism of the Contemporary Indian Populations Group, Anthropological Survey of India, S.R.C., Manav Bhavan, Bogadi, Mysore-570026, India, [2]DNA Polymorphism of the Contemporary Indian Populations Group, Anthropological Survey of India, 27, Jawaharlal Nehru Road, Kolkata-16, India

To elucidate the 'southern route' hypothesis of modern human migration, construct maternal phylogeny and prehistoric dispersals of modern human in Indian sub continent, a diverse subset of 642 mitochondrial DNA (mtDNA) lineages those belongs macrohaplogroup M was chosen for complete genome sequencing from the collection of 2,887 control-region sequences sampled from 26 tribal populations of India. On the basis of complete mtDNA sequencing, we have identified 13 new haplogroups M51 to M63; redefined/solidified and fully characterized haplogroups M2, M3, M4, M5, M6, M8-C'Z, M9, M10, M11, M12-G, D4, D5, M18, M30, M33, M35, M37, M38, M39, M40, M41, M43 and M49 which were previously described by control and/or coding-region polymorphisms. Our findings indicate that the greatest haplogroup diversity and in situ origin of Indian mtDNA lineages. The current Indian mtDNA gene pool was shaped by the initial settlers and was galvanized by minor events of gene flow from the east and west to the restricted zones. Northeast Indian mtDNA pool harbors all East Asian lineages associated with the Late Upper Paleolithic and/or early Neolithic dispersals. East Asian genes in northeast India established through admixture rather than replacement.

109: Evolutionary dynamics of the human Y chromosome

Doris Bachtrog

University of California San Diego, 9500 Gilman Drive, 92093 La Jolla, CA, USA

Y chromosomes, whose absence or presence determine the sex in humans, carry few functional genes and an abundance of repetitive junk DNA. Sex chromosomes arose from ordinary autosomes but the Y chromosome degenerates due to its lack of genetic recombination. For example, the human Y chromosome originated about 150–200 million years ago and most of its originally >1,000 genes have since been lost. This has led several researchers to propose that men will go extinct because of Y chromosome degeneration. By assuming that the human Y chromosome degenerates at a constant rate, some authors predict that the human Y chromosome will lose its last gene within the next 125,000–10 million years, leading to the extinction of males. However, the assumption of a constant rate of degeneration is not based on an evolutionary model. Both positive and negative selection models have been proposed to account for the loss of functional genes on a non-recombining Y chromosome, and I develop dynamic evolutionary models to predict the rate of Y degeneration over evolutionary time. I show that the rate of degeneration is not constant over time, but instead declines steadily due to the steady decrease in the number of active genes on a degenerating Y chromosome. Interestingly, strikingly different evolutionary dynamics are predicted for positive and negative selection models; negative selection is the dominant process in the very initial stages of Y chromosome evolution, while positive selection becomes most important at intermediate stages of degeneration. Using these models, together with estimated mutation and selection parameters for humans, I show that the human Y will likely retain most of its genes after another 300 million years of evolution. This contradicts recent predictions of the forthcoming loss of all functional genes on the human Y within 10 MY.

110: Polymorphisms in ACTN3 and KCNJ11 genes and athletic performance in South Indian athletes

Anandan Balakrishnan, Shridevi Venkataramani, Sambantham Shanmugan, Paramasivam Arumugam, Chaithanya Chitteti, Jayaraman Gopalswamy

Department of Genetics, Dr. ALM PG IBMS, University of Madras, Taramani Campus, Chennai 600 113, India

Genetics plays a major role in determining performance and success of an individual in a sport. Elite athletic performance, a complex trait is said to largely depend on complex interactions of multiple genetic and environmental factors with inter-individual and inter-population differences. The aim of this study was to analyze ACTN3 (R577X) and KCNJ11 (E23 K) polymorphisms in athletes and controls. Athletes were divided into different groups according to their sporting category (short distance athletes and long distance athletes) and performance levels

(below average, average and outstanding). This study involved 503 track and field athletes and 591 controls. Genotyping was performed by PCR-RFLP. The frequency of the R allele of ACTN3 gene was significantly higher in athletes compared to controls. The frequency of RR genotype was higher in male athletes (19%) compared to the control males ($\sim$14%). The RR genotype and R allele were significantly associated (OR = 1.3908, $P = 0.006$) with male short distance athletes. Also, the R allele was associated with the outstanding performance of male athletes (OR = 1.8710, $P = 0.0002$). The KCNJ11 genotype frequencies of control and athlete groups differed significantly in males ($P = 0.023$) but not in females ($P = 0.46$). The frequency of E allele of the KCNJ11 was greater in male athletes (0.60) than in male controls (0.548). The EE genotype is over-represented and the E allele was significantly associated with outstanding short distance athletes. We therefore conclude that the ACTN3 gene R577X polymorphism and KCNJ11 gene E23K polymorphism to be associated with athletic performance.

111: Allele frequency and haplotype diversity in three hyaluronan metabolic genes in 24 Indian populations

[1]**Kaustuv Basu**, [2,3]Arijit Mukhopadhyay, [1]Ilora Ghosh, [4]The Indian Genome Variation Consortium, [1]Kasturi Datta

[1]Biochemistry Laboratory, School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India, [2]Genomics and Molecular Medicine, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, [3]G.N. Ramachandran Centre for Genome Informatics, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, [4]Nodal Laboratory, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India

Hyaluronan is an important space-filling component of the extracellular matrix and it is involved in different cellular processes. Several functional studies indicate that the genes encoding Hyaluronan Binding Protein 1 (C1QBP), Hyaluronan Synthase 2 (HAS2), and Hyaluronidase 3 (HYAL3) are responsible for various diseases like atherosclerosis, breast cancer and probably many more. Our present study is the first attempt to characterize the SNPs of these genes to analyze the allele frequency distribution in the general population. As a part of Indian Genome Variation (IGV) project, we were interested to characterize the variations in the unaffected individuals to test their efficacy as markers and to analyze signature of biological selection at any of these loci. From these three genes we selected a total of 11 SNPs with a minimum minor allele frequency of 10% in Caucasians from the HapMap database. We genotyped them in 548 normal individuals, from 24 different endogamous Indian populations, using MALDI-TOF based Sequenom platform. Overall genetic differentiation among all populations (AMOVA = 3.5%; FST = 0.035; $P < 0.001$) was comparable with the differentiation observed when grouped independently by language, geography and ethnicity (FST = 0.037–0.04; $P < 0.001$). AMOVA analysis showed negligible differentiation among Dravidian (DR) speaking caste (LP) and tribes (IP) (-1.34%; FST = 0.01) with respect to Indo-European (I.E) speaking LP and IP (3.30%; FST = 0.04). But among populations within DR speaking LP and IP, variation was higher (2.98%) than among populations within I.E speaking LP and IP (1.47%). On the basis of the genetic distance (DA) and principal component analysis we clustered all the populations in major 3 groups which are ethnically and linguistically distinct. Haplotype diversity analysis was performed to identify major haplotypes in each population which can serve as the basal data to identify at-risk haplotype for future disease-associated studies related to hyaluronan metabolism. Interestingly, for *HYAL3*, one haplotype, major (24–72.7%) in 22 out of 24 populations, was found in significantly lower frequency in an I.E-LP from north India (17.4%) and a tribal population from western India (9.1%). More detailed analysis of these genes and their potential relation with various diseases will be discussed in the meeting.

112: Frequency of C3435T single nucleotide ABCB1 genetic polymorphism in native and Mestizo Mexican populations

[1,3]**Joaquín Becerra-Contreras**, [1]Fernando Rivas, [1,4]Sandra G. Orozco-Flores, [1,3]Edagar R. Ochoa-Martínez, [1,2]Juan Manuel Oliva-Ortiz, [1,2]Luz Berenice López-Hernández, [1]Marisela Casas-Castañeda, [1,2]Lucila Sandoval-Ramírez

[1]División de Genética, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Jalisco, Mexico, [2]Doctorado en Genética Humana, CUCS. Universidad de Guadalajara, Guadalajara, Jalisco, Mexico, [3]Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco, Mexico, [4]Centro Universitario de Ciencias Exactas e Ingeniería, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico

Background: P-glycoprotein is a membrane protein encoded by the ABCB1 gene, which demonstrates functional polymorphism. It is present in endothelial cells of the blood, brain barrier, intestine, liver, kidney, etc. and is an important factor regulating the bioavailability of many therapeutics. The ABCB1 single nucleotide polymorphism 3435C > T is associated with higher P-glycoprotein expression. Objectives: To determine allelic and genotype frequencies of C3435T single nucleotide ABCB1 genetic polymorphism in Native and Mestizo Mexican populations. Methods: 415 DNA of unrelated individuals was analyzed; 48 Huicholes (North of Jalisco and east of Nayarit), 48 Purepecha (Lake region of Michoacan state), 60 Tarahumaras (South-east of Chihuahua) and 215 Mestizos (Oaxaca-Guerrero coast, Chihuahua and Veracruz states). Results: It was found a high percentage of heterozygous in Native and Mestizo populations. The distribution of Huichol population was homozygous for the wild-type allele (CC) was 27.08, 50% were compound heterozygous with a mutant T-allele (TC), and 22.91% were homozygous for the mutant allele (TT). In Purepecha population the frequencies were: CC (25%), TC (45.83%) and TT (29.16%), in Tarahumara population: CC (16.66%), TC (63.33%) and TT (20.0%) and in Mestizos: Oaxaca-Guerrero coast: CC: (27.38%), TC (46.42%) and TT (26.19%), Chihuahua state population: CC (22.22%), TC (44.44%) and TT (33.33%) and Veracruz state population: CC (22.09%), TC (43.02) and TT (34.88%). Discussion: Allelic and genotype frequencies of 3435C > T changed meaningfully in populations with different origins, with a prevalence of the allele C in certain parts of the world (>70% in African populations), on the contrary in Caucasian and Asian populations, where this allele shows frequencies from 30 to 50%. Conclusions: We found a difference in the distribution of frequencies in the wild-type allele between a Germany population (reference population) and the Native populations (Tarahumara and Purepecha). On the other hand the percentage of heterozygous is high in the other populations (Mestizo and Huichol).

113: The growth hormone receptor gene and its association with mandibular height—a PCR based research study

Aditi Bhardwaj

Sharad Pawar Dental College and Hospital, Sawangi (Meghe), Wardha, Maharashtra, India

Multiple genes and environmental influences are involved in the growth of the mandible, it is difficult to determine the relationship between the mandibular shape and a particular gene. It is realized that the Growth hormone/Growth Hormone Receptor (GHR) play an important role in the skeletal growth and development as the GHR is present on the mandibular condyle. Malocclusion is not a disease but

a variation of occlusion and is an expression of multifactorial traits. In this paper quantitative trait analysis would be used to investigate relation between mandibular shape and Growth Hormone Receptor gene (GHR) by determining single nucleotide polymorphism (SNP) and select SNP as appropriate markers and determining the relationship between GHR, SNP and cephalometric linear measurements.

114: Structure of human genomic regions by functionality: an assessment of selective effects

Nidhan Kumar Biswas, Somosree Sarkar, Badal Dey, Partha P. Majumder

Indian Statistical Institute, Human Genetics Unit, 203 B.T. Road, Kolkata-108, India

Introduction: The structure of the human genome is expected to be strongly influenced by natural selection. Selective effects are most likely on functional, disease-associated genomic regions. Hypothesis: Genomic regions with decreasing functionality exhibit decreasing effects of natural selection. Methodology: To test this hypothesis, we selected four genomic regions of decreasing functionality: (a) Atherosclerosis-associated: Selectin L (*SELL*); (b) Housekeeping: Beta-Actin (*ACTB*) and (c) two gene-desert regions on chromosomes 2 (GDES1: 11 kb) and 6 (GDES2: 12 kb). We assayed variation in these regions by double-pass resequencing ($n = 72$) and genotyping ($n = 360$) among 432 individuals chosen from 8 ethnic groups representing the socio-cultural and geographical diversity of India. The nature and extent of variation and the selective effects were estimated by analysis of site-frequency spectra, tests of neutrality and network analysis for these (Indian) and global (HapMap) data. Results: (1) Site-frequency spectra compared with neutral expectations: for (a) *SELL*, significantly ($P < 10^{-9}$) greater intermediate frequency variants; (b) *ACTB*, no significant ($P > 0.05$) deviation from neutrality; and (c) Marginally (GDES1: $P = 0.03$) or no (GDES2: $P > 0.05$) deviation from neutrality. (2) Neutrality tests: Tajima D and Fu & Li D* and F* statistics are all non-significant ($P > 0.1$) for all genomic regions. (3) Median-joining haplotype networks: (a) *SELL*, high-frequency modal haplotypes are separated by long branches; multiple star-like phylogenies radiating from larger nodes; (b) *ACTB*, GDES1 and GDES2, no striking features of haplotypes networks were observed. (4) Comparison with HapMap data (no comparable data for *ACTB*): (a) *SELL*, large differences in haplotypes frequencies between India and HapMap populations; many novel haplotypes in India; (b) GDES1 and GDES2, nearly all haplotypes are shared with HapMap populations, with frequencies in India similar to CEU. Conclusion: Genomic variation and haplotype structure in *SELL* (a gene associated with atherosclerosis) are primarily determined by natural selection, while those for the housekeeping gene *ACTB* and the gene-desert regions conform to neutral expectations. These findings are supportive of our hypothesis.

115: The genetic origin of Austro-Asiatic speakers

Gyaneshwer Chaubey

Estonian Biocentre, Tartu University, Riia 23, Tartu, Estonia

The origin of Austro-Asiatic (AA) speakers dispersed in South-East (SE) and South Asia is enigmatic. Previous studies proposed the association of NRY haplogroup O2a (M95) with the origin and diffusion of this language group. The high frequency of Y chromosomal haplogroup O2a (M95) in AA populations suggests their homeland to SE Asia, while the prevalence of South Asian specific mtDNA haplogroups corroborates local matrilineal origin of AA speakers in the Indian subcontinent, not in SE Asia. To test the different scenarios emerged by archaeological, archeaobotanical, linguistic and genetic studies, we analyzed 17 Y-STR microsatellite markers in M95 derived SE, East and South Asian samples. Besides this, we studied the association of 1540T/C polymorphism of the autosomal EDAR and hair thickness globally. Significantly higher Y-STR diversity and calculated from it O2a coalescent time in SE Asian than in Indian AA populations suggest the origin of AA speakers in SE Asia and further migration to India. Furthermore, the presence of 1540C polymorphism in EDAR gene (reported to be most frequent in East Asian populations) in AA and Tibeto-Burman populations advocate a SE and East Asian genetic input into Indian AA populations. Our results confirm a shared patrilineal genepool of the Indian AA speaking populations and those inhabiting SE Asia.

116: Red cell enzyme polymorphism in Gond, Korku and Keer Tribes inhabiting Pachmari biosphere of Madhya Pradesh

[1]**Ruchira Chaudhary**, [2]S. M. S. Chahal, [1]Bharti Deepak

[1]Department of Zoology, Govt. MVM College Dept Barkatullah Univ., Jehangirabad, Bhopal, MP, India, [2]Department of Huamn Biology, Punjabi University, Patiala, Punjab, India

The present study deals with the genetic analysis of Gond, Korku and Keer tribes of Madhya Pradesh. The blood samples were collected at random from a total of 200, 154 and 131 apparently healthy and not closely related individuals of either sex of the Gond of Hoshangabad, Korku of Betul and Keer of Sehore districts of the state, respectively. The samples were analyzed for phenotypes of A1A2BO and Rh (D) blood groups by standard tube method and for Red Cell Enzymes by electrophoresis. Haemolysates were prepared using freezing and thawing method and stored at −20°C. Prepared haemolysates were used for isoenzyme typing by biochemical technique of electrophoresis and specific staining protocols. The present study shows high incidence of A1 allele and low of B allele in Korku, and presence of A2 allele in few cases of both Gond and Korku, as well as low d frequency in serological markers. In Biochemical trait, presence of less common phenotype in ACP i.e. ACP*C is recorded in Gond tribe. The allele frequencies of ADA, PGM1, ACP1 and AK1 indicate closeness between Gonds of Hoshangabad, Bhil and Korku of Pachmari (Hoshangabad) but difference from Keer. All these tribes depicts great genetic variation in reference of isozymes polymorphism. The Chi square (χ^2) test for goodness of fit revealed no significant deviation between observed and expected number in any marker, except ESD, suggesting that the tribe is in genetic equilibrium. The present study will help to characterize genetically the Gond, Korku and Keer tribes of Madhya Pradesh.

117: HapMap based whole genome scan for positive selection signals and the construction of database SNP@Evolution

Libin Deng, Feng Cheng, Wei Chen, Changqing Zeng

Beijing Institute of Genomics, the Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, China

Taking the advantage of high-throughput genomic surveys, we performed a genome-wide scan for selection signals across the human genome. Using HapMap phase II dataset, we first applied boxplot to identify low HET outlier windows based on empirical distributions of genotypes. From 2,081 clusters containing contiguous low HET windows, we detected 1,461 regions among HapMap populations as the candidates subjecting to strong selection (Pboot < 0.01). To reveal the

possible functional changes caused by these selective events, single locus FST analysis was applied to identify outliers showing significant population differentiation in genic regions. Based on the empirical distribution of FST, we identified 1,853 candidate genes subjected to geographically restricted positive selection. Furthermore, candidate regions were supported by population differentiation test. Total of 803 out of 1,853 genes with high FST values were found in candidate regions which span only 6% of whole genome size, indicating a much higher frequency of their distributions than that in the genomic background (χ^2 test, $P < 0.01$). Furthermore, although most events were exclusive to a single geographic group, a significant fraction of selection (about 35%) was experienced by non-African populations, suggesting geographically restrictive selection sweep might play an important role in population differentiation during recent expansion of modern humans. Based on our results which provide new clues to understand recent selective sweeps and population differentiation, we further built SNP@Evolution as an integrative and hiberarchy database focusing on selection. The graphic user interface was constructed with Generic Model Organism Database toolkit to develop Gbrowser to illustrate various datasets. To capture multiple signals of positive selection across the genome, empirical P values of HET, FST, and iHS (integrated Haplotype Score) of most annotated genes were computed and integrated to demonstrate outliers as shown in data query interface. Available at http://bighapmap.big.ac.cn, we expect SNP@Evolution to become a valuable database to study nature selection and population differentiation.

118: Involvement of functional polymorphisms in the promoter regions of matrix metalloproteinase-2, -3 and -9 in the susceptibility of gastric cancer

[1]**Sanjib Dey**, [2]Debjit Saha, [1]Snehasikta Swarnakar

[1]Indian Institute of Chemical Biology, 4, Raja Subodh Chandra Mullik Road, Jadavpur, Kolkata-700032, India, [2]Ruby General Hospital, Kasba Golpark, Eastern Metropoliton Bypass, Kolkata 700107, India

Gastric cancer (GC) is one of the most common threats to healthcare in many Asian countries. Increased expression of the matrix metalloproteinases (MMPs), a family of zinc-dependent endopeptidases that degrade a range of extracellular matrix (ECM) proteins are associated to GC. MMPs regulate various cell behaviors including tumor development and metastasis and, hence are believed to be target genes for genetic predisposition studies. Genetic variations in several MMP promoters influence the transcription and expression of MMPs which result in pathophysiological conditions. Expression of MMP-2, -3 and -9 has been closely related to lymph node metastasis, vascular invasion and cancer prognosis. We have examined the serum samples for MMP-2, -9 and -3 activities of 50 GC patients of different clinical stages and of 36 controls using gelatin and casein zymography. Our result shows significant upregulation in MMP-9 activity while dual regulation in MMP-2 and very little changes in MMP-3 activity in serum in GC patients compare to that of control. The expression of MMP-9 and -2 was parallel to their activities as judged by Western blotting. In addition, we have prepared blood genomic DNA from GC patients and control subjects to identify any correlation between their SNPs in MMPs promoter and risk of GC. Several reports described that MMP2-1306C/T, MMP3 -1171-5A/6A and MMP9-1562C/T SNPs have been associated with initiation, invasion, and metastasis to various cancers. However, a very few reports exist on relationship between those SNPs and risk of GC. We conducted a hospital based case-control study in Eastern Indian population to assess the effects of these MMP promoter SNPs on susceptibility and clinical staging in GC patients using DNA sequencing, PCR restriction fragment length polymorphism(RFLP) and Taqman probe assays. A statistically valid association of those SNPs with the risk, occurrence and progression of GC will be tested in a large sample population in future.

119: The prevalence of variant alleles of functionally significant polymorphisms in the methylenetetrahydrofolate reductase, factor V, and prothrombin genes in the Sri Lankan population

[1]**V. H. W. Dissanayake**, [1]L. Y. Weerasekara, [1]G. C. Gammulla, [1]S. L. D. Jayaweera, [1]C. Kariyawasam, [1]B. Perera, [1]L. Y. S. Sandamal, [2]F. Broughton Pipkin, [3]L. Morgan, [1]R. W. Jayasekara

[1]Human Genetics Unit, Faculty of Medicine, University of Colombo, Kynsey Road, Colombo 00800, Sri Lanka, [2]Division of Obstetrics, School of Human Development, University of Nottingham, D Floor, East Block, Queen's Medical Centre, Nottingham NG7 2UH, UK, [3]Division of Clinical Chemistry, School of Molecular Medical Sciences, University of Nottingham, A Floor, West Block, Queen's Medical Centre, Nottingham NG7 2UH, UK

Introduction: Polymorphisms in the methylenetetrahydrofolate reductase (MTHFR), Factor V (F5), and Prothrombin (F2) genes are known to be associated with thrombophilia. We studied the prevalence of the variant alleles of the MTHFR 677C > T and 1298A > C; F5 1691G > A (Leiden) and 4070A > G; and F2 20210G > A polymorphisms in the Sri Lankan population. Methods: DNA samples of 240 (50% male) volunteers belonging to the Sinhalese (80), Sri Lankan Tamil (80) and Moor (80) ethnic groups recruited in Colombo, Sri Lanka were genotyped at these polymorphic sites using PCR/RFLP methods. Results: The genotype frequencies were as follows: Sinhalese—MTHFR 677C > T: CC 78%,CT 20%,TT 3%; 1298A > C: AA 67%; AC 33%; CC 0%; F5 1691G > A: GG 96%, GA 4%, AA 0%; 4070A > G : AA 88%, AG 13%,GG 0%; F2 20210G > A: GG 100%, GA 0%, AA 0%. Tamil—MTHFR 677C > T: CC 83%,CT 17%,TT 0%; 1298A > C: AA 75%; AC 23%; CC 2%; F5 1691G > A: GG 94%, GA 6%, AA 0%; 4070A > G : AA 90%, AG 10%,GG 0%; F2 20210G > A: GG 100%, GA 0%, AA 0%. Moor—MTHFR 677C > T: CC 83%,CT 17%,TT 0%; 1298A > C: AA 69%; AC 31%; CC 0%; F5 1691G > A: GG 96%, GA 4%, AA 0%; 4070A > G : AA 85%, AG 14%,GG 1%; F2 20210G > A: GG 100%, GA 0%, AA 0%.
Conclusion: All the polymorphisms studied except F2 20210G > A were polymorphic. This information would be useful when designing genetic studies and offering clinical genetic testing in the Sri Lankan population.
Acknowledgement: This work was funded by the National Science Foundation, Sri Lanka and the IRQUE Project Research Grants, University of Colombo, Sri Lanka.

120: The genetic legacy of Indian muslims

[1]**M. Eaaswarkhanth**, [1]I. Haque, [2]Z. Ravesh, [3]F. A. Khan, [4]T. Kivisild, [5]C. T. Smith, [6]L. Singh, [6]K. Thangaraj

[1]National DNA Analysis Center, Central Forensic Science Laboratory, 30, Gorachand Road, Park Circus, Kolkata-700014, India, [2]Department of Genetics, Osmania University, Hyderabad, India, [3]State Forensic Science Laboratory, Lucknow, India, [4]Leverhulme Center for Human Evolutionary Studies, University of Cambridge, Cambridge, UK, [5]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, CB10 1SA, UK, [6]Centre for Cellular and Molecular Biology, Hyderabad-500007, India

The extent of genetic contribution besides the spread of Islam from Arabian and Middle Eastern countries to the Indian sub-continent has been scarcely studied. In the direction of addressing this concern, a total of 476 individuals belonging to six Muslim groups from different geographical regions of India were analyzed for variations in mitochondrial, Y-chromosome, and an autosomal DNA markers. All the samples were highly diverged under the branches of major mtDNA haplogroups—M, N and R. Interestingly, we found L0a2 haplogroup in three individuals of Dawoodi Bohra Muslims (Tamil Nadu). Analyzing 431 Y-chromosomes, we found diverged Y haplogroups, of which E3b1a and J*(xJ2) were informative. The contribution of the maternal L0a2 (0.59%) and paternal E3b1a (0.63%) and J*(xJ2) (5.05%) haplogroups can be attributed to their close affinities with Middle East. Both the maternal and paternal admixture contributions from the neighboring non-Muslim Indian populations were found to be maximum and minimum from the Arabian and Iranian populations. On sequencing *LCT/MCM6* gene in 476 individuals, C13910T mutation was observed with a frequency of about 0.091% in Indian Muslim populations. Our extensive analysis revealed the existence of genetic signatures of Middle East in some of the contemporary Indian Muslims.

121: Yin-Yang haplotype structure at the ACE Locus in Indian populations

[1]**Shabana Farheen**, [1]B. Mukhopadhyay, [2]C. S. Chakrabarti, [1]Partha P. Majumder

[1]Human Genetics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India, [2]Department of Zoology, Burdwan University, West Bengal, India

Angiotensin I-converting enzyme (ACE) is one of the most intensely studied candidate genes in the context of human cardiovascular and psychiatric diseases. In spite of the strong indications that the impact of the ACE gene on these diseases is through the joint—not individual—effect of SNPs in this gene, no work has been done to study the haplotype structure and diversity of this gene in India. We have investigated the allele and haplotype profiles of eleven polymorphisms in the regulatory (rs4277405, rs4459609, rs1800764, rs4292, rs4291), exonic (rs4309, rs4331, rs4343) and intronic (C5144T, A5967G and ACE *Alu* insertion/deletion) regions covering 17.4 kb of the ACE gene. These polymorphisms were assayed by PCR-RFLP in 12 Indian populations from diverse socio-linguistic groups inhabiting different geographical regions. All the SNPs studied were highly polymorphic ($0.18 \leq P \leq 0.57$) except for C5144T which was monomorphic in all the Indian populations. The frequency of the alleles at the SNP loci was significantly different in two populations, Kadar and Kammanaidu. Overall the frequencies of the SNPs were similar to those of the Caucasians. A total of 33 haplotypes were identified of which two haplotypes with a cumulative frequency of more than 80% were identified in the Indian populations. Haplotype TATATCTG**I**A accounted for 40–70% of the Indian haplotypes, while frequency of CCCTCTCA**D**G varied from 18 to 40%. It may be noted that these two most frequent haplotypes have complementary nucleotides at all sites; Yin-Yang haplotypes. Population-specific haplotypes occurred at extremely low frequencies. The phylogenetic relationships reconstructed using haplotype frequencies revealed that the Indian populations were distinct from the African populations and showed a closer affinity with the Caucasians. A median joining network of haplotypes constructed using data from the HapMap and the Indian populations showed that the haplotypes found in the Yoruban, Japanese and Chinese populations clustered separately from the haplotypes found in India, which clustered with the Caucasian (CEU) haplotypes. [This work has been funded by the Department of Science and Technology WOS-A grant to SF].

122: Haplotype diversity in disease candidate genes analyzed in the Indian genome variation project

[1]**Sudheer Giddaluru**, [1]Prashant K. Singh, [2]Sangeeta Khanna, [2]Amit K. Mandal, [1,2]Arijit Mukhopadhyay, [1,2]Mitali Mukerji, [3]The Indian Genome Variation Consortium, [1,2]Samir K. Brahmachari

[1]Genomics and Molecular Medicine, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, [2]G.N. Ramachandran Centre for Genome Informatics, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, [3]Nodal Laboratory, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India

Differences in haplotype frequencies across genetically distinct populations add a source of heterogeneity in disease association studies. This restricts us in selecting population specific tagging SNPs whenever the haplotype frequencies vary across populations. Initial analysis of a diverse set of 21 genes of Indian Genome Variation Data phase-I revealed high mean haplotype diversity values for majority of genes with two to five major haplotypes shared across populations despite the presence of highly differentiating SNPs. Extending the analysis to phase-II, haplotype frequencies and the diversity were estimated among 24 endogamous populations that belong to four linguistic and different geographical regions of India: A total of 536 genes that are potential candidate genes for different complex diseases like diabetes, schizophrenia, bipolar disorder, hypertension, asthma, etc. were included after applying the selection criteria of a minimum of three SNPs/gene with each having a MAF more than 5%, and follow Hardy–Weinberg equilibrium in minimum 22 populations. Preliminary analysis of Phase-II data in 93 genes revealed mean haplotype diversity ranging from 0.4 to 0.93. Further when searched for significant over- or under-representation of genes using Gene Ontology (GO) terms, it is observed that set of genes involved in similar biological process have similar diversity across populations. For example, it is observed that genes belonging to synaptic transmission and regulation of nucleotide biosynthesis showed high haplotype diversity (0.7–0.8), and genes belonging to protein kinase cascade and cell development pathways showed low haplotype diversity (0.4–0.5). Extensive analysis of Phase II will give further insight into the pathways and biological processes that are more tolerant to genetic variation and thus provide important guidelines for disease association studies.

123: Genomic medicine and developing countries: creating a room of their own

[1]**Billie-Jo Hardy**, [1,2]Béatrice Séguin, [1,3]Peter A. Singer, [1,3]Abdallah S. Daar

[1]McLaughlin-Rotman Centre for Global Health, Program on Life Sciences, Ethics and Policy, University Health Network and University of Toronto, MaRS Centre, South Tower, Suite 406, 101 College Street, Toronto, Ontario, M5G 1L7, Canada, [2]Leslie Dan Faculty of Pharmacy, University of Toronto, 144 College Street, Toronto, Ontario, M5S 3M2, Canada, [3]McLaughlin Centre for Molecular Medicine, MaRS Centre, TMDT, 101 College Street, 14th Floor, Suite 701, Toronto, Ontario, M5G 1L7, Canada

In 2002, the World Health Organization recommended, in its report entitled, 'Genomics and World Health', that all nations, including developing nations, harness the advantages of emerging genomic sciences and associated technologies to ensure global health equity. In light of this, we examined the opportunities and implications to global health of the emerging trend to apply knowledge of human genomic

variation towards the understanding of disease susceptibility and drug response. Over a period of 3 years, we performed four case studies of institutes which have either initiated or proposed large scale genotyping initiatives in Mexico, India, Thailand and South Africa. Using well established empirical case study methodology we studied the National Institute for Genomic Medicine in Mexico, the Indian Genome Variation Consortium in India, the Thai SNP Discovery Project/Thai Center for Excellence in Life Sciences Pharmacogenomic Project in Thailand, and the Africa Genome Education Institute/Division of Human Genetics, University of Cape Town in South Africa. We will discuss the results according the following five themes: political will, individual leadership, genomic sovereignty, knowledge-based economy and local health benefits. We will also draw comparisons and make recommendations across these initiatives, highlighting lessons learnt and how they may apply to other developing countries more generally.

124: The Mexican genome diversity project: analysis of genetic structure in Mestizo and Amerindian populations of Mexico

A. Hidalgo-Miranda, L. Uribe-Figueroa, I. Silva-Zolezzi, J. C. Fernandez-Lopez, J. Estrada-Gil, G. Ortiz-Ramos, R. Mojica-Espinosa, J. Cruz-Colin, A. Contreras, D. Velazquez, H. Hernandez-Lemus, S. March-Mifsut, G. Jimenez-Sanchez

National Institute of Genomic Medicine, Mexico, Mexico D.F., Mexico

Most of Mexico's population is represented by Mestizos resulting from admixture of Amerindian, Spaniards and, in a lesser extent, African populations. This admixture process has led to particular genomic ancestry structure. To optimize the use of the human genome information to improve healthcare in the Mexican populations, we are systematically evaluating the genomic variability of the Mexican Mestizo population. We genotyped 300 self-defined Mestizos from six regions [Guanajuato (GUA); Guerrero (GUE); Sonora (SON), Veracruz (VER); Yucatan (YUC) and Zacatecas (ZAC)] using the Affymetrix 100 K SNP array set. Heterozygosity (HET) values, Fst and principal component analysis were determined. HET was higher in Mexican subpopulations compared to the Asian HapMap samples, SON showing the highest (0.287) and GUE the lowest (0.274). Fst values from five of six pairwise comparisons between Mestizos (SON-GUE: 0.019, SON-VER: 0.013; SON-YUC: 0.012; SON-GUA: 0.011; SON-ZAC: 0.006) where higher that those of the Japanese-Chinese samples of the HapMap (JPT-CHB: 0.007). PCA separated HapMap samples in three clusters. Mestizos showed a wide distribution between the CEU and other source of variation not present in HapMap data. To demonstrate that the Amerindian population represents the third ancestral component in the Mestizo population, 33 Zapotec (ZAP) Amerindian samples were included in the analyses. Fst values showed that GUE and VER are the most similar to the Amerindian sample (ZAP-GUE: 0.032; ZAP-VER: 0.038) and SON the most different (ZAP-SON: 0.082). Fst values for the SON-ZAP comparison was higher than all comparisons between Mestizos and CEU or Asian samples. As expected, the highest Fst value was obtained in the African-Amerindian comparison (ZAP-YRI: 0.238). PCA analysis showed that the ZAP clustered between the CEU and the Asian population. Mestizos from GUE and VER were closely located to the ZAP cluster, while samples from SON were located at the other end of the distribution, closer to the CEU. Our results indicate significant heterogeneity between populations within Mexico, as well as between Mexicans and the HapMap populations. We are including additional Mexican Amerindians and increasing the SNP density to better understand the admixture process of the Mexican Mestizo population, and develop more suitable tools to analyze the genetic bases of complex diseases in the Mexican populations.

125: Genomic analysis of mitochondrial DNA in the Indigenous groups of Malaysia

[1]**Lih-Chun Hong**, [2]Mun-Yik Fong, [3]Juli Edo, [1]Maude Elvira Phipps

[1]University of Malaya, Department of Molecular Medicine, Faculty of Medicine, 50603 Kuala Lumpur, Malaysia, [2]University of Malaya, Department of Parasitology, Faculty of Medicine, 50603 Kuala Lumpur, Malaysia, [3]University of Malaya, Center for Malaysian Pribumi Studies, Institute of Research Management and Consultancy, 50603 Kuala Lumpur, Malaysia

Polymorphisms in the control and coding regions of mitochondrial DNA (mtDNA) were analyzed in four Indigenous groups ($N = 188$). A total of 64 different mtDNA haplotypes were elucidated. Of the 64 sequences observed, 44% were unique and 56% were shared by more than one individual. We report seven novel polymorphisms, with three in the HVS-1 and four in HVS-2 segments that have hitherto not been reported. These polymorphisms have been submitted to MITOMAP database. Most mtDNA variants belonged to three major haplogroups M, N and R. The majority of the haplogroups were found to be exclusive to the Indigenous groups in Malaysia. The most frequent were haplogroups including M21a and R21, which both reflect very old lineages, deeply rooted in South East Asian populations. This was evident in about 80% of the Negritos (Jehai and Kensiu) that were genotyped. Haplogroups N21 (24%) and N22 (8%) were predominantly found in Temuan whilst haplogroup N9a6a was predominant in Bidayuh (40%). Haplogroups indicative of South and Central Asia distributions were observed in variable frequencies. The Temuan comprise the most diverse group followed by Jehai and Bidayuh. The least diverse were the Kensiu. In our study cohort, the lower levels of genetic diversity in these four groups in contrast to majority of the Asian population (0.798–0.872) are indicative of genetic drift and may reflect bottleneck effects due to relatively long periods of isolation. The 9 bp deletion in the MT-CO2/MT-TK (tRNALys) intergenic region of mtDNA (n.p. 8271–8281) was detected in 10.1% of the participants investigated. This deletion was identified in a Kensiu individual, five Bidayuh and the majority of Temuan (28%), but not detected in Jehai. Our results support the current view that Malaysia is home to some of the oldest groups to have migrated out of Africa.

126: Analysis of genome-wide copy number variations in diverse Indian populations

[1]**Pankaj Jha**, [1]Pramod Gautam, [2]Vinod Scaria, [1,2]Samir K. Brahmachari, [3]The Indian Genome Variation Consortium, [1]Mitali Mukerji

[1]Genomics and Molecular Medicine, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, India, [2]G.N. Ramachandran Knowledge Centre for Genome Informatics, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, India, [3]Nodal Laboratory, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, India

There has been a considerable excitement on the potential significance of copy-number variations (CNVs) in the human genome due to their involvement in a number of diseases as well as effects on global

gene expression. A large number of common CNVs are enriched in certain biological processes and are reported to be selected in response to diet and environment. In contrast to SNPs, the frequency of CNVs differs considerably in related populations. Thus depending on the presence/absence or difference in frequency of CNVs, the etiology and prevalence of a disease could differ between populations. In this study, we report the extent of structural variations in 511 samples from 26 reference populations representing the genetic spectrum of India and sharing different affinities with global populations. Additionally, a random set of 52 samples from these populations were used as reference. The analysis of CNVs was carried out using the Affymetrix (50 K) array. We generated intensity files using the GCOS software followed by copy number detection using Affymetrix Genotype Console (ver. 2.1). Using in-house developed programs, the CNV regions were mapped to genomic co-ordinates on the hg18 build of the human genome. Validation of a representative set through RT-PCR and Illumina Infinium (370 duo) chips have been done. In order to characterize these CNV regions we first looked for their overlap with reported CNV regions and observed them to be present in one third of the total samples. As previously reported, we also observed correlation of CNV regions with Alu repeats and segmental duplications. However, neither the length nor the GC content of the chromosomes correlated with CNVs. Among these CNV regions we observed a highly significant enrichment of genes involved in cell proliferation and differentiation, signal transduction, organ development, response to environment, stress, regulation of cellular and metabolic processes, nucleosome positioning etc. in both duplicated and deleted set. Thus CNVs could modulate core biological processes and have system wide effects. However, we observed that CNVs were not always shared in populations which are genetically related by SNPs. Thus an extensive genome wide CNV map on the reference set of Indian populations would be enormously valuable for identification of novel structural variants associated with common and complex diseases.

127: Patterns of high-resolution recombination rates in human genes

[1]**Mamoru Kato**, [1]Fuyuki Miya, [2]Yonehiro Kanemura, [1]Toshihiro Tanaka, [1,3]Yusuke Nakamura, [1]Tatsuhiko Tsunoda

[1]SNP Research Center, RIKEN, 1-7-22 Suehiro, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan, [2]Institute for Clinical Research, Osaka National Hospital, 2-1-14 Hoenzaka, Chuo-ku, Osaka City, Osaka 540-0006, Japan, [3]Human Genome Center, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

The HapMap project has recently found considerable variation in high-resolution recombination rates along the chromosomes of the human genome. Although the associations between this variation and genomic sequence features, such as genic regions, provide information on haplotype diversity and natural selection in these regions, the associations are not well understood. Here, we performed microarray experiments to identify genes specifically expressed in human tissues and investigated patterns of recombination rates within tissue-specific genes. We also examined the patterns within genes classified by the Gene Ontology. Surprisingly, genes specifically expressed in the frontal lobe, which is a brain region involved in human cognitive abilities, had low recombination rates, whereas genes specifically expressed in the cerebellum, which is a brain region with primitive functions shared by all vertebrate species, had high rates. We also found that genes for intra-cellular vital activities, including metabolism and transcription, had low recombination rates, whereas genes involved in inter-cellular information transmission mediated by channels and transporters had high rates. These findings suggest that natural selection forms the recombination rate tendencies according to the physiological functions exerted in the genes. For example, the low recombination rates in frontal lobe-specific genes may indicate that a few haplotypes have been rapidly widespread across the population because higher cognitive abilities are advantageous. Frontal lobe-specific genes with extremely low recombination rates may be candidates for genes related to cognitive abilities that human species have recently obtained.

128: Y-chromosome diversity in Southern-Ural Bashkirs

[1]**Elza Khusnutdinova**, [1]Artem Lobov, [1]Bayazit Yunusbaev, [2]Rinat Yusupov, [1]Marina Bermisheva, [1]Ildus Kutuev, [1]Rita Khusainova, [3]Richard Villems

[1]Institute of Biochemistry and Genetics, Pr. Oktyabrya 71, Ufa, Russia, [2]Institute of Language and Literature, Pr. Oktyabrya 71, Ufa, Russia, [3]Estonian Biocenter, Riia str. 23, Tartu, Estonia

Bashkirs are Turkic-speaking ethnic group, occupying forest-steppe and steppe landscapes of southern Ural region of Russia. Medieval Age Turkic nomads were vehicle of Bashkir ethnogenesis, which largely involved indigenous tribes of Southern Urals, namely Finno-Ugric and Indo-Iranian tribes. Here we present phylogenetic analysis of Y-chromosome lineages in a sample of 587 Bashkirs drawn from different parts of the southern Ural region and neighboring areas: Abzelilovskiy ($N = 152$), Sterlibashevskiy ($N = 54$), Baimakskiy ($N = 95$), and Burzyanskiy ($N = 82$) districts of Bashkortostan republic, Orenburg ($N = 79$), Perm ($N = 72$), Samara and Saratov ($N = 51$) Oblasts of Russia. Obtained samples of Y-chromosomes were analyzed using 24 biallelic markers of the Y chromosome non-recombining region. A total of 17 haplogroups were identified among which R1b1b2-M269, R1a-SRY 1532, and N1c-Tat lineages were predominant. N1c-Tat haplogroup was present in all studied subpopulations with low to high frequencies. Most probable source population that contributed N1c-Tat lineage into Bashkirs is Finno-Ugric tribes. Y-chromosome lineages specific to Northeast Asian populations (C3c-M48) or East Asian populations in general (O-M175) were absent or found with very low overall frequency (less than 10%). Overall prevalence of typical West Eurasian (R1a-SRY 1532 and R1b1b2-M269) and North Eurasian (N1c-Tat) lineages could be interpreted in to ways: (1) Turkic-speaking newcomers were indistinguishable from populations of Northcentral Eurasia with respect to Y-chromosome composition; (2) Genetic input associated with their arrival was limited and Y-chromosome diversity in Bashkirs largely represent genetic heritage of Pre-Turkic settlement process.

129: Simple sequence repeats as beneficial mutators in the evolution of *Yersinia pestis*

Pankaj Kumar, H. A. Nagarajaram

Laboratory of Computational Biology, Center for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad, India

Yersinia pestis is a notorious pathogen well known for causing major pandemics in the past and, in the recent times it has been dreaded for its potential use in bio-terrorism. Here we report an in-depth study of *Yersinia Pestis* and *Yersinia pseudotuberculosis* (an ancestor of *Yersinia pestis*) genomes for the distribution, abundance and polymorphism of perfect Simple Sequence Repeats (SSRs)—the

sequences generally known for reversible hyper mutations in the form of insertions and deletions of their repeating motifs and also for imparting novel functions to the genomes. The genomes are abundant with perfect SSRs, approximately one SSR tract at every 4 bp of the genome. Vast majority of the perfect SSRs are short (less than 8 bp). Long perfect SSRs are rarely found suggesting strong selection against expansion of perfect SSRs. Cross-genome comparisons reveal 391 polymorphic perfect SSR tracts which form the evidences of inter-species and intra-species SSR polymorphisms. As compared to *Yersinia pseudotuberculosis* some of the ORFs in *Yersinia pestis* have suffered premature terminations due to SSR mutations indicating the role of SSRs as one of the factors causing gene loss which is believed to be the major determinant in the gain of virulence in *Yersinia pestis*. Our studies further reveal interesting association of some of the perfect SSRs mutations in pathogenicity, adaptation and evolution of *Yersinia pestis*.

130: Genetic portrait of modern Belarusians: mitochondrial DNA and Y-chromosome perspective

[1]**Alena Kushniarevich**, [1]Larysa Sivitskaya, [1]Nina Danilenko, [2]Richard Villems, [1]Oleg Davydenko

[1]Institute of Genetics and Cytology, Academicheskaya Str 27, Belarus, [2]Estonian Biocenter, Riia Str 23, Estonia

To explore the genetic structure of modern Belarusians and to trace their paternal and maternal history the high resolution analysis of non-recombining portion of Y-chromosome (NRY) and mitochondrial DNA (mtDNA) polymorphism was performed. In total 576 NRYs and 292 mtDNAs were analyzed. Samples were selected from DNA-bank of native Belarusians on conditions that they: represent all six ethno geographical regions of present-day Belarus, descend from unrelated healthy individuals and allow the analysis of both sex-associated genetic loci. Analysis of NRY diversity was performed by hierarchical typing of 26 informative biallelic markers according to established phylogeny. Phylogenetic state of mtDNAs was determined by sequencing of HVSI region (16020–16400 nps) following RFLP analysis of coding region diagnostic sites. Revealed composition of paternal and maternal gene pools is typical for East Europeans. Distinctive features of mtDNA pool of Belarusians are: predominance of limited number of haplogroups, high diversity of maternal genetic lineages, and minor (3%) presence of Asiatic type of mtDNA (D-lineages). About 88% of mtDNA diversity is determined by R-descendents: family HV which includes haplogroup H (37%), HV-group (4%), V (6%), U-branch (23%) and cluster JT (17%). Paternal gene pool of Belarusians is characterized in its turn by domination of only three genetic lineages—haplogroups R1a (51%), I1b (16%) and Finno-Ugric haplogroup N3 (9%) covering thus about 76% of total Y-chromosome diversity; the rest falls to R1b, I1a, I1c and E, J lineages. Components of Y-chromosomal and mtDNA gene pools of modern Belarusians reflect multiple gene flows separated in time, dominating source of which was southern or south-eastern European region. Analysis of distribution of mtDNA and Y-chromosomal components among six subpopulations has revealed that genetically Belarusians are low structured population. A separation of subpopulations along the south-north line, which is demonstrated particularly in distribution of Y chromosomal lineages R1b, I1a and I1b, N3 and G-chromosomes, has been noted; east-west gradient is insignificant. In sense of mtDNA diversity Belarusians show relatively high level of homogeneity. Belarusians demonstrate the highest genetic affinity with two other Eastern Slavs—Russians and Ukrainians in context of their genetic history; presence of Finno-Ugric N3 chromosomes points to common genetic roots of Balts and East Slavs.

131: A genome-wide scan of balancing selection in the human genome

G. Laval, L. B. Barreiro, E. Patin, L. Quintana-Murci

Institut Pasteur, 25, rue Dr Roux, Paris, France

Inferences concerning the action of natural selection in the human genome have increased our understanding of the evolutionary forces that affect the human genome, have augmented our knowledge of gene function and promise to increase our understanding of the genetic basis of disease. The recent availability of genome-scale genotyping data has led to the identification of regions of the human genome that seem to have been targeted by natural selection. These studies have focused on the action of local positive selection and negative selection, while little is known about the action of balancing selection—selection that increases variability within a population. We report here on a genome-wide scan for signals of balancing selection, at the worldwide level and at the continental level. One of the signatures of balancing selection is higher within-population heterozygosity (e.g. minor allele frequency close to 0.5) and lower population genetic distances (i.e. FST) with respect to neutral expectations—the average FST computed on the human genome. To unmask the extent to which balancing selection have influenced human genome variability, we developed a statistical framework that was applied to three genome-wide SNP datasets: the HapMap Phase II ($\sim$3 million SNPs) Frazer et al. (2007), Perlegen ($\sim$1 million SNPs) Hinds et al. (2005) and the Human Genome Diversity Project -CEPH ($\sim$600,000 SNPs) Li et al. (2008). By applying a number of highly conservative filters, we provide a balancing selection map of the human genome. In addition, our analyses show that several biological functions are significantly enriched in candidate genes under balancing selection. Overall, our study suggests that the effects of balancing selection in the human genome are more important than previously thought.

Frazer et al. (2007) Nature 449:851–861
Hinds et al. (2005) Science 307:1072–1079
Li et al. (2008) Science 319:1100–1104

132: Evolutionary constraints on hub and non-hub proteins in human protein interaction network: insight from protein connectivity and intrinsic disorder

Baisali Manna, Bratati Kahali, Tanusree Bhattacharya, T. C. Ghosh

Bose Institute, P1/12 CIT Road Scheme, VII M, Kankurgachi, Kolkata-700054, India

It has been claimed that proteins with more interacting partners (hubs) are structurally more disordered and have a slow evolutionary rate. Here, in this paper we analyzed the evolutionary rate and structural disorderness of human hub and non-hub proteins present/absent in protein complexes. We observed that both non-hub and hub proteins present in protein complexes, are characterized by high structural disorderness. There exists no significant difference in average evolutionary rate of complex-forming hub and non-hub proteins while we have found a significant difference in the average evolutionary rate between hub and non-hub proteins which are not present in protein complexes. We concluded that higher disorderness in complex forming non-hub proteins facilitates higher number of interactions with a large number of protein subunits. High interaction among protein subunits of complex forming non-hub proteins imposes a selective constraint on their evolutionary rate.

133: High resolution phylogeographic study of mtDNA macrohaplogroup M in South Asia

[1]**Mait Metspalu**, [1]Gyaneshwer Chaubey, [1]Erik Prank, [1]Anu Solnik, [1]Monika Karmin, [1]Ene Metspalu, [2]Doron Behar, [3]Juan J. Sanchez, [4]Phillip Endicott, [5]Sadagopal Shanmugalakshmi, [5]Karuppiah Balakrishnan, [6]Ramasamy Pitchappan, [7]Toomas Kivisild, [1]Richard Villems

[1]Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, Riia 23, Tartu, Estonia, [2]Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa, Israel, [3]National Institute of Toxicology and Forensic Science, Canary Islands Delegation, Campus de Ciencias de la Salud, 38320 La Laguna, Tenerife, Spain, [4]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK, [5]School of Biotechnology, Bharathidasan University, Trichirappalli, Tiruchirappalli 620 024, India, [6]Department of Immunology, School of Biological Sciences, Madurai Kamaraj University, Madurai, 625 021, India, [7]Leverhulme Centre of Human Evolutionary Studies, The Henry Wellcome Building, University of Cambridge, Fitzwilliam Street, Cambridge, CB2 1QH, UK

An absolute majority of the extant South Asian (SA) maternal genetic lineages have been shown to be autochthonous and largely shared by caste and tribal communities as well as between speakers of different language groups. Hence, most of the extant SA populace share their maternal roots in the population that colonised SA in Late Pleistocene. Within the boundaries of this general picture there are many questions that still demand attention. Here we introduce 52 new complete mtDNA sequences that significantly refine the phylogeny of haplogroup (hg) M. The new SA hg M phylogeny contains 166 mtDNA types in 25 (10 new) distinct lineages that arise from the pan Eurasian founder haplotype M* and further eight (2 new) that stem from the founder M4'30 which stands one mutational step away from M*. The number of SA maternal lineages that coalesce to the most recent common ancestor (MRCA) in the three pan Eurasian founder haplotypes is higher than in other regions. This implies different demographic history with probably less severe population bottlenecks and higher effective population size during Late Pleistocene. The average synonymous mutational distance (rho = 5.3) to the root of the SA hg M phylogeny yields a signal for expansion at around 35K years ago (KYA). Interestingly this is about 10K YA younger than that of the hg M lineages in East Asia and Oceania. It has been argued that in SA secondary population expansion(s) might have deflated the overall age estimate. However, the distribution of unweighted contributions of the individual basal SA hg M lineages to the overall age estimate does not deviate significantly from Poisson distribution. On the other hand, the distribution of age estimates for all nested hg M clades is bimodal. However, the peaks in the distribution occur at rho one and three. Thus, there is some evidence for discrete expansion waves in the demographic history of SA which, however, do not coincide with the proposed event at around 30K YA. We further developed a Snapshot genotyping panel to investigate the phylogeography of the main new branches of the SA hg M phylogeny. We were able to simultaneously genotype for 26 polymorphisms in altogether 2,300 samples from different geographic regions of India. 6 hgs out of the identified 30 had significantly different frequencies in northern and southern India. Detailed phylogeographic analyses permitted evaluating different scenarios relating genes with languages or geography in several case studies.

134: Darwin's dream: multi-parametric consesus phylogenetic tree in 3-D

[1]**Sohan P. Modak**, [2]M. Milner Kumar

[1]Institute of Genomics and Integrative Biology, Mall Road, Delhi 110007, India, [2]Zoology Department, Karnatak University Dharwad, Dharwad, India

Darwin's Dream (DD) is a powerful modular software tool that constructs, visualizes and animates multiparametric phylogenetic trees in a 3D space to study molecular evolution, and evolutionary hierarchy. In conventional method, the branching pattern in uniparametric trees for different genes or proteins differs among the same set of species. DD retrieves sequences for Genes/RNA/Polypeptide) from online biological databases, performs Multiple Sequence Alignment (MSA) by CLUSTALW, T-COFFEE, Mafft or Probcons and constructs multiparametric phylogenetic trees. Estimated evolutionary distances for three traits among all species pairs are plotted in 3D with respect to a reference species at 0,0 with the rest placed in the 3D space. All pairs Euclidian distances are estimated and subjected to Multidimensional scaling (MDS) to get unique 3D representations of UPGMA, NJ, ME or MP trees. We transpose the topologies of 2D trees using Euclidian distances over the positions of species in 3D obtained by MDS by our All pairs shortest distance algorithm to find the branch points. Mean Euclidian distances are bootstrapped to estimate Confidence Limits (CL) of phylogenetic trees. DD, written in Python, visualizes high-resolution 3D consensus trees and our Fidelity algorithm compares species clades at various nodes against benchmark taxonomic tree to validate phylogenetic relationships. Other dependencies include TREE-PUZZLE and PHYLIP.

From MSA of aa sequences of three mitochondrial polypeptides, we have built consensus tree for 79 species. Here, Urochordate locate between Mollusca and Arthropoda and primitive to Cephalochordates and Hemichordates that are sister taxa of echinoderms. From the Cyclostomes stem, mammals branch out separately from other vertebrates. The former begins with Marsupials and branches into one leading to primates and the other to remaining mammals. The phylogenetic trees based on aa sequences of 15 tRNA synthetases from 129 Archae and Prokaryotes exhibit higher fidelity index than for 16 s rRNA. DD trees visualize consensus phylogeny in 3D that compares well to concatenated sequence trees and can compare 'n' number of sequences and species. DD detects cases of convergent evolution and long-branch attraction.

135: Genetic variation and haplotype structures of innate immunity genes in Eastern India

[1]**Souvik Mukherjee**, [1]B. Bairagya, [1]P. Bhattacharya, [2]S. K. Bhattacharya, [1]B. Dey, [1]U. Dey, [1]T. Ghosh, [1]S. Maiti, [1]K. Mishra, [1]S. Mukherjee, [3]K. Narayanasamy, [1]S. Poddar, [1]N. Sarkar Roy, [1]P. Sengupta, [3]S. Sharma, [2]D. Sur, [1]D. Sutradhar, [4]D. K. Wagener, [1]Partha P. Majumder

[1]TCG-ISI Centre for Population Genomics, Kolkata, India, [2]National Institute of Cholera and Enteric Diseases, Kolkata, India, [3]The Centre for Genomic Applications, New Delhi, India, [4]Research Triangle Institute International, Rockville, USA

Introduction: Being evolutionarily ancient, innate immune system may have been highly optimized by natural selection. If true, diversity in the innate immunity genes is expected to be low. It has recently been emphasized that the innate immune system may be plastic and

continuing to evolve due to natural selection imposed by pathogen diversity.
Aim: To investigate the nature and extent of genetic variation in deciphering the role of natural selection in innate immunity genes.
Methodology: The study participants belonged to two different communities—Muslim and Hindu—residing in a slum area of Kolkata, India, exposed to a high load of microbial pathogens. Genomic variation was assayed in 171 individuals by resequencing ~75 kb of DNA comprising 12 innate immunity genes—Cathelicidin Antimicrobial Peptide (*CAMP*), Defensins (*DEFA4,DEFA5,DEFA6* and *DEFB1*), Mannose Binding Lectin (*MBL2*) and Toll-like Receptors(*TLR1, TLR2,TLR4,TLR5,TLR6* and *TLR9*). Haplotypes were determined for each gene from the genotype data and haplotype networks estimated.
Results: (a) Almost half of the 548 DNA variants discovered was novel. (b) The SNP density in these innate immunity genes (>3 SNPs/kb) exceeds the highest density observed for any autosomal chromosome in the human genome, indicative of plasticity. (c) Gene diversities were low to moderate (from 1.73% for *CAMP* to 21.8% for *TLR9*). Haplotype diversities were high for most genes (from 62% for *TLR9* to 88% for *DEFA6*). (d) For many of the genes, the ratio of SNPs to haplotypes is >1, indicating that there is repression of recombination or, more likely, that many recombinant haplotypes are selected against. (e) The genetic positioning of the study populations among the HapMap populations based on data of these genes substantially differed from data of other genes. (f) While the major allele at 76% of variable sites in our data coincided with the ancestral (chimpanzee) allele, at the remaining 24% of sites the major allele is not the ancestral allele. This observation is indicative of the action of natural selection.
Conclusions: This is the first extensive population-based study of natural variation in a large number of innate immunity genes. The high genomic variation and the distinct haplotype structure of innate immunity genes observed among individuals have possibly resulted from the impact of co-evolution with pathogens. We are now formally testing the hypothesis of natural selection acting on these genes.

136: Application of Indian genome variation data as guidelines for future disease association studies from population sub-structure perspective

[1,2]**Arijit Mukhopadhyay**, [1]Ikhlak Ahmed, [2]Samira Bahl, [1]Amit Chaurasia, [2]Prashant Singh, [1]Meenakshi Anurag, [1]Debasis Dash, [3]Partha P. Majumder, [1,2]S. K. Brahmachari, [4]I. G. V. Consortium, [1,2]Mitali Mukerji

[1]G.N. Ramachandran Centre for Genome Informatics, Institute of Genomics and Integrative Biology, Mall Road, Delhi-110007, India, [2]Genomics and Molecular Medicine, Institute of Genomics and Integrative Biology, Mall Road, Delhi-110007, India, [3]Human Genetics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata-700108, India, [4]Nodal Laboratory, Institute of Genomics and Integrative Biology, Mall Road, Delhi-110007, India

The recently published results of the Indian Genome Variation (IGV) study indicate that 55 diverse populations can be grouped into four major population clusters. Cluster-1 contains populations mainly from Austro-Asiatic (AA) linguistic background, cluster-2 is mostly Tibeto-Burman (TB), cluster-3 is mainly Indo-European (I.E) large populations and cluster-4 is predominated by Dravidian (DR) populations. This implies that for a case-control study, depending on the selection of cases and the genes under study, the data generated by the IGV project can be directly used potentially obviating the need of collecting a new cohort of controls. The focus of this study is to test the hypothesis by estimating the level of population sub-structure present in these clusters using the Genomic Control (GC) method to measure the chi-square inflation factor (λ). We have identified 52 SNPs from intergenic regions each having heterozygosity of more than 0.4 in each of the 55 populations under study and selected them as neutral markers. The freely available software GC.r was modified in-house to automate multiple iterations. All the samples from different populations under one cluster were pooled together. The GC was performed within and between clusters for 1,000 iterations. The average value of λ within cluster was found to be ~1.0 ($\lambda = 1$ for no stratification), while that of between clusters ranged from 2 to 5. This implies that the identified population clusters are genetically homogeneous. Currently, more in-depth analysis is in progress with larger number of neutral markers to achieve a better level of significance. Also, we have tested the effect of pooling controls from the same cluster and from different clusters in case-control studies in our own research on schizophrenia. A particular haplotype of SYNGR1 was found to be associated with patients when compared with the ethnically matched controls from southern part of India. Replacing the controls with that of IGVdb we found that only when the cluster of IE large populations or a pool of I.E and DR large populations are used as controls, the significant association was reproduced (*P*-value 0.03 and 0.002, respectively). Using control from any other cluster either showed no association or a reverse association with controls was observed. Thus our data demonstrates that taking control data from clusters identified in IGV project, corresponding to the ethnicity of the case cohort can reliably detect disease association.

137: Phylogeography of selected tribal populations of Orissa, India

Amrita Nandan, Vishwas Sharma, Varun Kumar Sharma, Kumaraswamy Thangaraj, Lalji Singh

Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India

Having 62 different tribal populations, Orissa considered as the third state to have highest frequency of tribal populations. Most of these tribes are untouched by the modern human exploitation due to inaccessible hilly landscape. In the present study, we have analyzed the mitochondrial and Y-chromosomal markers of 108 individuals belonging to five tribal populations (Savara, Bhumia, Gadaba, Dhurva and Bonda). The phylogenetic tree constructed on the basis of complete mitochondrial sequences illustrate various ever define subgroups of R8 with unique set of mutations. Further, comparative analysis of these sequences with the genetic database shows the absence of these subgroups (R8a1, R8a2, R8a3a, R8a3b, R8b1, R8b2a, and R8b2b); reflects in situ origin of these subgroups. The preliminary analysis of Y-chromosome shows the presence of F*, H, O, J, L and R haplogroup with the highest frequency of haplogroup H. Both the mtDNA and Y chromosome results suggest that the tribal populations of Orissa are very ancient. Comparative data from various ethnic populations would unravel the origin of these populations. Detailed information would be made available at the time of presentation.

138: Single nucleotide polymorphisms of ADH and ALDH2 gene among the Bondo-Highlanders of Orissa

[1]**Jayanta Kumar Nayak**, [1]B. N. Sarkar, [2]P. K. Das, [1]V. R. Rao

[1]Anthropological Survey of India, 27, Jawaharlal Nehru Road, Kolkata-700 016, India, [2]Department of Anthropology, Utkal University, Vani vihar, Bhubaneswar, Orissa-750 004, India

The studied Austro-Asiatic population (the Bondo-Highlanders) is a unique population with a typical life style of intaking alcoholic

products having alcohol content varying from 4 to 29%, starting from 6 months of life. But no phenotypic expression of alcoholism is found. Present work is carried out on the Bondos with the hypothesis that they may be genetically protected. Variants of different class I ADH and ALDH2 genes are responsible for providing protection to the form alcoholism. Three SNPs; ADH1B Arg47His (Exon 3), ADH1B Arg369Cys (Exon 9) and ADH1C Val349Ile (Exon 8), are functionally validated in terms of phenotype–genotype correlations and are in specific linkage disequilibrium (LD) with the non-coding SNPs. ALDH2 degrades acetaldehyde metabolized from ethanol and its encoding gene ALDH2 has a functional polymorphism; ALDH2 Glu487Lys associated with low enzyme activity. Since Glu487Lys of this locus is fixed for the functional subunit in all non-East Asian populations, this polymorphism was examined along with ALDH2 for other exonic and intronic loci to identify informative markers for studying the role of this gene in this studied population. For the collection of informations on alcoholic dependence, Subjects were interviewed with the modified 24-item Michigan Alcoholism Screening Test (MAST). Genomic DNA were extracted from 110 unrelated adult Bondos with written consent approved by the ethical guidelines of the Anthropological Survey of India. Direct sequencing of PCR products by ABI 3730 Automated DNA analyzer identified all SNPs. Statistical analysis were done by appropriate softwares. The MAST scores below three is considered absolutely normal. As per MAST scores the observed frequency in normals was 74.5% (104) and in risk group the observed frequency was 5.5%(6). The studied population was conspicuous by the complete absence of the African specific allele ADH1B*369 Cys, the frequency of East Asian specific ADH1B*47 His is 0.054 is also negligible. The important functional variant in the studied population was ADH1C*349Ile (ODDs ratio 1.314 and 95%CI 0.7–2.464) and ALDH2 Glu487Lys (ODDs ratio 0.906 and 95%CI 0.247–3.27). ADH and ALDH2 study provides a baseline for future research into the role of the ADH and ALDH2 locus in alcoholism. The ALDH2 Glu487Lys site was found in high frequency to high metabolic allele and therefore highly unlike to contribute to phenotypes tied to alcohol metabolism.

139: Population structure inference and genetic matching in European samples using genome-wide marker sets

[1]**Michael Nothnagel**, [1]Timothy T. Lu, [2]Oscar Lao Grueco, [1]Olaf Junge, [1]Sandra Freitag-Wolf, [1]Amke Caliebe, [2]Manfred Kayser, [1]Michael Krawczak

[1]Christian Albrechts University, Institute of Medical Informatics and Statistics, Brunswiker Str. 10, 24105 Kiel, Germany, [2]Erasmus University Medical Center, Department of Forensic Molecular Biology, Dr. Molewaterplein 50, 3015 GE, Rotterdam, Netherlands

Genetic association studies can be confounded by population stratification. Genetic matching, such as between cases and controls, by use of large numbers of genetic markers can prevent systematic ancestry differences. Here, we compare the genetic structure of European populations using more than 2,400 samples with genome-wide single nucleotide polymorphism (SNP) data from 23 different regions. Furthermore, we investigate if a small number of ancestry-sensitive markers (ASM) are sufficient to allow a genetic matching in European sample sets with the same accuracy as the complete, genome-wide marker set. Our results indicate that, besides a small number of highly informative markers, the great majority of markers contribute only little, but independent information for matching and a large number of markers are required for reliable matching within Europe.

140: Establishment of a national repository for human genetic research: an initiative

P. B. S. V. Padmanabham, V. R. Rao

Anthropological Survey of India, 27 Jawaharlal Nehru Road, Kolkata-700016, India

An important outcome of the Anthropological Survey of India's National Conference 'Human Origins, Genome and People of India' was a panel discussion, 'DNA banking of named Indian Populations'. The issues that were discussed centered on societal challenges that have arisen with advancement of science and technology in the post genomic era and its application to Human Genetics and Molecular Medicine. Till date, the existing DNA banks in the country are being managed by individual institutions with an approval from the institutional ethical committee. They are not guided by any uniform norms with regard to collection, storage and access to genetic material in their possession. And publications in several peer reviewed national and international journals on Indian populations do not acknowledge the source/s of DNA. While recognizing the existence of guidelines from ICMR and DBT, the panelists recommended for an inter-ministerial committee to examine them, in view of the national interests.

An inter-ministerial committee, involving the ministries of Culture, Science and Technology, Health, Defence and Law met and discussed about the (a) importance of preserving genetic heritage of Indian populations, in view of their un-paralleled and unique diversity (b) potential of these biological samples for bio-medical and bio-technological applications, (c) inherent commercial applications there of, (d) importance of conserving historical and ancient skeletal material in terms of bio-molecular approaches and (e) sensitivity to local/tribal beliefs and customs while collecting blood samples from the same group again and again by different researchers etc. The inter-ministerial committee recommended the Government to set up a National Advisory Committee for establishing a National Repository of Human Genetic Resources and Data. The constitution of the Committee and the terms of reference were gazetted on December, 10, 2005. The National Advisory Committee, after due deliberations in a series of meetings, submitted its report recommending for an establishment of a permanent National Repository for Human Genetic Resources in the country.

The present paper discusses development of bio-banks in south-east Asian countries vis-à-vis the initiative on developments related to the National Repository for Human Genetic Resources in India, for the benefit of common man.

141: Relative nucleotide substitution rates from allele frequency spectrum of human gene loci

L. Park

Natural Science Research Institute, Yonsei University, 134 Shinchon-Dong, Seodaemun-Ku, Seoul, South Korea

The mutation rate is one of the most important parameters in genetics. Due to the insufficiency of genetic data in the past, estimates typically relied upon phylogenetic methodologies, which usually involve several assumptions. Though there have been attempts at direct estimates, most prior efforts paid little attention to identify the substitution rate of each nucleotide (A, C, G, or T) for another. To estimate the nucleotide substitution rate of one base for another, the basic theory of the recurrent mutation model was applied. Using large-scale, high-quality, public re-sequencing data, population genetic parameters ($4\ N\mu$ and $4\ Nv$) for each pair of base substitutions were obtained, and the relative mutation rates were studied. For all six

possible pairs of bases, allele frequency spectrums were examined to obtain population genetic parameters. The estimates for all twelve parameters were between 0.2 and 0.6, and were used to identify the relative mutation rate of each base for another. The relative mutation rates exhibited good agreement with the matched substitution rates of bases in the base pairs. Transition mutation rates were noted to be a bit higher than transversions, but the number of polymorphic sites for transition was about twice that for transversions and four times for CpG sites. It was noteworthy that the T-to-C transition mutation rate was higher than the C-to-T transition mutation rate, contrary to the prior expectations and reports. This current result might be reasonable, because it can explain the higher number of low frequency C-to-T transitions from the distribution of T and C transitions. With insertion and deletion mutations, the deletion to insertion rate was higher than the reverse. Using a simple recurrent mutation model and real re-sequenced data, relative mutation rates were obtained for each base substitution. The higher T-to-C mutation rate than C-to-T, especially in CpGs, implies a possible maintenance system of CpG islands in gene loci. Population differences also were identified between African and European descendents, probably due to past population history.

142: Fine-scale recombination and linkage disequilibrium at CYP2C and CYP2D cytochrome P450 gene subfamily regions in European populations and implications for association studies of complex pharmacogenetic traits

[1]**V. N. Pimenoff**, [2]G. Lavall, [3]D. Comas, [1]J. U. Palo, [4]I. Gut, [5]H. Cann, [2]L. Excoffier, [1]A. Sajantila

[1]Department of Forensic Medicine, University of Helsinki, P.O. Box 40, Helsinki, Finland, [2]Zoological Institute, University of Bern, Baltzerstrasse 6, CH-3012, Bern, Switzerland, [3]Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader 88, 08003, Barcelona, Spain, [4]Centre National de Genotypage, 2 rue Gaston Cremieux, CP 5721 Evry Cedex, France, [5]Foundation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), 27 rue Juliette Dodu, 75010 Paris, France

Patterns of linkage disequilibrium (LD) and recombination rate variation along the human genome are key issues in the quest for a common haplotype map of the human genome. Thus, finding an association depends largely on the genomic LD structure. However, LD is known to vary considerably between genome regions and populations mostly due to the combined effects of mutation, recombination and past demographic history. Here, we present the distribution of LD and recombination rate variation at the pharmacogenetically relevant CYP2C and CYP2D gene subfamily regions for the first time in a set of 11 European and one African population. Our results show significantly different patterns of genetic diversity and LD for these regions among Europeans. Moreover, significant frequency differences of a CYP2C19*2 A-19154 allele associated with variable drug reactions were observed among Europeans. These results further suggest that the CEPH sample population may not be the optimal sample population for association studies of common complex traits within Europe, despite its selection for the HapMap project. The assumed small and constant size population of Saami demonstrates a highly extended LD compared to CEPH and other European sample populations. This unique structure of LD in the Saami may offer a substantial advantage for further association mapping of cytochrome P450 variants relevant to complex pharmacogenetic traits.

143: Impact of multi functionality in shaping molecular evolution of human housekeeping and tissue-specific interactome over connectivity

Soumita Podder, Pamela Mukhopadhyay, T. C. Ghosh

Bose Institute, Bioinformatics Centre, P 1/12, C.I.T. Scheme VII M, Kolkata-700054, India

Discrepancy in amino acid substitution rate across eukaryotic proteome has always been a study of interest. It has been claimed that proteins with many interacting partners tend to evolve more slowly than fewer interacting proteins. In the present communication, we have analyzed human housekeeping and tissue specific interactome to study evolutionary rate differentiation between similar connected proteins. We observed that there is significant difference in evolutionary rate between housekeeping and tissue specific proteome of equal connectivity. To explain the above observation we tested for an association between multifunctionality and evolutionary rate of the human protein interactome as multifunctional genes are expected to evolve at lower rates. We have demonstrated that the evolutionary rate difference between housekeeping and tissue specific hub and non hub proteins is guided by the protein multifunctionality. Delving deeper insight we even observed that multi interface hubs (more than two interacting interfaces) are evolutionary conserved between both sets of genes as their functionality is overall parallel. On contrary, single interface hubs in housekeeping genes evolve more slowly as compared to tissue specific genes owing to their involvement in a number of unique biological processes. Finally we provide evidence that the correlation between multifunctionality and evolutionary rate of the interacting proteins is strong enough to substantially constrain the rate of adaptation of a protein.

144: Adaptive human-specific gain of function in a developmental enhancer

[1,3]**Shyam Prabhakar**, [1]Axel Visel, [1]Jennifer A. Akiyama, [1]Malak Shoukry, [1,4]Keith D. Lewis, [1]Amy Holt, [1]Ingrid Plajzer-Frick, [5]Harris Morrison, [5]David R. FitzPatrick, [1]Veena Afzal, [1,2]Len A. Pennacchio, [1,2]Edward M. Rubin, [1,6]James P. Noonan

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [2]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA, [3]Computational and Mathematical Biology, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore, [4]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA, [5]MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK, [6]Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA

Changes in gene regulation are thought to have contributed to the evolution of human development. However, in vivo evidence for uniquely human developmental regulatory function has remained elusive. We show in transgenic mice that a conserved noncoding sequence that evolved extremely rapidly in humans is a developmental enhancer that has gained a strong limb expression domain compared to the orthologous elements from chimpanzee and rhesus macaque. This gain of function is consistent across two developmental stages (embryonic days 11.5 and 13.5) and includes the presumptive anterior wrist and proximal thumb. In vivo analyses using synthetic enhancers, in which human-specific substitutions were introduced into the chimpanzee enhancer sequence or reverted in the human enhancer to the ancestral state, indicate that 13 substitutions clustered in an 81-basepair module otherwise highly constrained among terrestrial vertebrates are sufficient to confer the human limb expression pattern. Human-specific functional change in this enhancer

is highly likely to have been driven by positive selection, since the sequence has evolved in the human lineage at greater than four times the neutral substitution rate (P value = 1.3e-6). The enhancer is located within an intron of the putative cytoskeletal remodelling gene CENTG2, and ~300 kb downstream of the essential developmental transcription factor GBX2. Our study provides experimental evidence that human-specific nucleotide substitution can alter the activity of developmental enhancers and identifies a cis-regulatory element that potentially contributed to adaptive human morphological evolution.

145: Unique signatures of the natural background radiation on the human Y chromosomes from Kerala (India)

Sanjay Premi, Jyoti Srivastava, Sebastian Padinjarel Chandy, Safdar Ali, Rana Samad, Saima Wajid, Sher Ali

National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi-110067, India

Most frequently observed macroscopic consequences of ionizing radiation affecting the entire genome are chromosomal lesions and cancers. Owing to the haploid status and absence of recombination, the human Y-chromosome is an ideal candidate to be assessed for the possible genetic alterations induced by ionizing radiations. We studied the human Y chromosome in 390 males from the South Indian state, Kerala, where the level of natural background radiation (NBR) is ten-fold higher compared to that of the worldwide average, along with the 790 control males. We observed random micro deletions in the AZFa, b and c regions in >90%, tandem duplication and Copy Number Polymorphism (CNP) of 11 different Y-linked genes in about 80% radiation exposed males, whereas the autosomal homologues of Y-linked CDY largely remained unaffected. Multiple polymorphic copies of the Y-linked genes with a single Y specific localization correlated with their tandem duplication. Some NBR exposed males also evidenced unilocus duplication of the DAZ genes resulting in six copies. Notably, approximately 25% NBR exposed males showed exclusive deletion of the DDX3Y gene, leaving behind other AZFa candidates USP9Y and UTY genes unaffected. All the Y chromosome alterations here observed were exclusive to the somatic tissues (blood) and were not detected in the germline (sperm). The exposure to high level of natural background radiation correlates with several interstitial polymorphisms of the human Y chromosome. The CNPs and enhanced transcription of the SRY gene after duplication observed is envisaged to compensate for the loss of Y chromosome in small population of the cells. The above mentioned changes confined to the peripheral blood lymphocytes, suggest a possible innate mechanism protecting the germline DNA from the NBR. A detailed genome analysis of the much larger population exposed to NBR may provide new insights into the mechanisms and risks of the resultant genetic damages.

146: Allele frequencies at neutral loci suggest a recent bottleneck in southern India: Impact on susceptibility and co-morbidity in schizophrenia

[1]**Meera Purushottam**, [1]A. Ram Murthy, [1]H. B. Kiran Kumar, [1]A. Padmanabhan, [1]D. Subhashree, [2]Saurabh Ghosh, [1]Jagadisha Thirthalli, [3]Odity Mukherjee, [1]Sanjeev Jain

[1]National Institute of Mental Health and Neuro Sciences, Hosur Road, Bangalore, India, [2]Indian Statistical Institute, Kolkata, India, [3]National Centre for Biological Sciences, Bangalore, India

Allelic variations in populations may reflect recent bottlenecks, and explain certain aspects of co-morbidity and susceptibility to polygenic disorders. Parts of India have been subjected to repeated famines and a recent study reported that in a famine in southern India (19th century) almost 70% of the mentally ill and 20% of the general population died. Using BOTTLENECK program we show increased heterozygosity at several unlinked loci indicating a recent population bottleneck. Complex disorders such as obesity, cardiovascular disease, type II diabetes and schizophrenia involve multiple pathways sharing genetic susceptibility that may be influenced by past environmental effects. The Pro12Ala polymorphism of the PPARG gene has been linked to susceptibility to diabetes. We investigated this polymorphism in patients diagnosed with schizophrenia and matched controls, and observed a difference in allele frequencies of the 'thrifty allele' consistent with modest changes in susceptibility. In light of the above, we hypothesize that the prevailing allele frequencies at the PPARG locus in patients with schizophrenia that may have been influenced by variations in the frequency of the 'thrifty allele' during the population bottleneck could result in increased susceptibility to diabetes. Other candidate gene loci and biallelic unlinked markers were analysed to look for clues to the manifestation of this complex disease. The role of selection pressure on the manifestations of schizophrenia thus needs to be explored further.

147: Genomic studies on indigenous African populations

[1]**R. S. Ramesar**, [2]H. Soodyall, [3]W. Delport, [3]C. Seoighe

[1]MRC Human Genetics Research Unit, University of Cape Town, Anzio Rd, Cape Town, South Africa, [2]MRC/NHLS/WITS Human Genomic Diversity and Disease Research Unit, National Health Laboratory Service and University of the Witwatersrand, Braamfontien, Johannesburg, South Africa, [3]National Bioinformatics Node, Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, Anzio Rd, Cape Town, South Africa

Introduction: Africa is the origin of humankind, yet very little research is reported off this continent with respect to this phenomenon, and the related genomic diversity here. With the vast number of large-scale genome sequencing projects that have been embarked on, and even more initiatives to conduct large scale genome wide SNP-typing of populations around the world, it is worthwhile inventorising the progress with regards to genomic research on the African continent. With this as a backdrop we present our original research on five indigenous Southern African populations using high density genome-wide SNP (6) analysis. Methods and materials: This research involved (1) reviewing the expansive genetic and genomic research on the African continent, and (2) conducting a genome-wide SNP analysis of five indigenous Southern African population groups, using the Affymetrix SNP6 system. The subjects were chosen to represent Eastern and Western Bantu migrations, as well as the original San population.
Results and discussion: Genetic studies in Africa have dealt generally with monogenic disorders; e.g. haemaglobinopathies in central and North Africa, and neurodegenerative conditions in North Africa. More recently, expansive studies have begun looking at susceptibility to infectious diseases including HIV/AIDS, Tuberculosis, Malaria, Leishmaniasis, Leprosy, Schistosomiasis, and Trachoma. A growing number of studies are currently aimed at non-communicable diseases such as diabetes and hypertension. In addition to investigating predisposition to disease, several studies on worldwide human relationships using high density genome-wide SNP variations has supported evidence of a single human origin in sub-Saharan Africa. Our original research on a significant number of subjects from five indigenous populations from the Southernmost part of the African continent provides significant support for human origins south of the Sahara, and provides a means for dissecting population structure

much more incisively. There is genomic evidence of significant selection pressures correlating with population migration to the southern tip of Africa. Conclusion: The genetic dissection of indigenous African populations is growing at a significant pace. Useful information on population structure and linkage disequilibrium intervals will inform future studies here in the Southern part of Africa, aimed at mapping genotypic features against complex phenotypes.

148: Individual cancer genome sequencing using SOLiD-PET revealed extensive large SVs and provided architectural framework for completion

Yijun Ruan

Genome Institute of Singapore, 60, Biopolis Street, #02-01, Genome, Singapore 138672, Singapore

It is known that cancer genomes include extensive variations from small base pair mutations to large structural variations (SV). However, very minimal is known about the scale, diversity, and function of such variations, particularly large SVs. To fully understand the contribution of SVs to cancer biology, it is necessary to reveal and compare individual cancer genome sequences. We have developed a robust strategy that combines the paired end ditag (PET) approach and ultra-high throughput SOLiD sequencing. In this SOLID-PET strategy, cancer genomic DNA is sheared in 10 Kb size range, from which PETs are constructed for SOLiD sequencing. The PET sequences are mapped to human reference genome, with the accordant PETs (correct tag mapping orientation in expected genomic distance) to define contig regions that are in good matches with the reference genome and genomic breakpoints, and the discordant PETs (paired tags mapped in discordant orientation, strands, distance, or in between different chromosomes) to characterize the breakpoints and the types of SV events. Each SOLID-PET experiment is capable of generating >100 million PETs for mapping analysis. With this amount of PET sequences, we are able to re-constitute the architecture of individual cancer genome, including mapping of all normal and amplicon regions, as well as various forms of SVs. Using this strategy, we have analyzed a number of breast cancer cell lines and solid tumor specimens. We demonstrated that SOLID-PET sequences derived from larger fragments are much more robust than shorter DNA for mapping breakpoints at bp resolution. We have also shown that the SOLID-PET data is exceptionally powerful for characterization of amplified genomic regions by measuring accurate copy numbers and furthermore by locating the sites where the amplified copies to reside in cancer genomes (either transposition or tandem repeat). Comparing the re-constructed individual cancer genome structures and normal genome sequences, we identified cancer specific SVs and are assessing the impact on altering gene structures (fusion genes) and transcription activities. The architectures of the re-constructed cancer genomes by SOLID-PET data provide the framework for further completion of cancer genome sequencing.

149: Meiotic recombination in the human Xp/Yp pseudoautosomal region 1

Shriparna Sarbajna, Celia May, Matthew Denniff, Alec Jeffreys

University of Leicester, University Road, Leicester LE1 7RH, UK

All heritable variations in human DNA arise from two processes: mutation, which introduces changes in our DNA sequence, and recombination, which re-shuffles these changes into the endless allelic combinations seen in contemporary human populations. Recent genomics programmes have provided an unprecedented view of human DNA diversity. What is lacking is a full understanding of the dynamics and processes of mutation and recombination that create this diversity. To address this, techniques pioneered at the University of Leicester are being used to analyse sperm DNA molecules that have undergone meiotic exchange within the Xp/Yp pseudoautosomal region 1, a region of obligate recombination in male meiosis with a rate greater than 10-fold above the genome average. The sperm approaches allow characterization of the distribution of exchange events at the subkilobase level and are allowing us to identify some of the most intense recombination hotspots in the genome. These in turn will provide a resource for investigating recombination initiation and processing as well as allow us to identify DNA sequence and epigenetic factors that influence human recombination activity. Ultimately, such knowledge will not only provide powerful insight into this very fundamental process but also aid interpretation of genome-wide association studies and thus assist in the identification of genetic determinants of common disease.

150: G6PD, natural selection and the evolution of india-specific deficiency variants

Somosree Sarkar, Nidhan K. Biswas, Badal Dey, Partha P. Majumder

Indian Statistical Institute, Human Genetics Unit, 203 B.T. Road, Kolkata-700108, India

Introduction: G6PD deficiency is the most common enzymopathy among humans. It causes severe hemolytic disease, but also confers protection against malaria. More than 400 biochemical variants of G6PD—including some that are predominantly India-specific—have been reported globally, indicating vast genetic heterogeneity. This is the first systematic study on molecular diversity and evolution of G6PD gene in India. Aims: To (a) estimate the nature and extent of genetic heterogeneity in the G6PD gene in India, (b) draw inferences on the evolution of known India-specific deficiency variants, and (c) test whether balancing selection has determined the population structure of this locus. Methodology: DNA samples collected from 10 ethnically, linguistically and geographically diverse populations were analyzed for G6PD by (a) resequencing in a random subset of 80 males, and (b) genotyping all polymorphic variants in the remaining subset ($n = 400$) that included both males and females. Additionally, data were also generated from two >10 kb gene-desert regions. Statistical analyses included allele frequency estimation, HWE test, haplotype determination, diversity estimation, M-J network reconstruction and tests of neutrality. Results: (a) Resequencing revealed 11 variant (including one novel non-synonymous Met > Ile) and three polymorphic sites. (b) The number of segregating sites per kb in G6PD (2.25) is even lower than two large gene-desert regions (~ 3.4), possibly due to strong selective effects. (c) Haplotypes reconstructed on the basis of five loci revealed low diversity (0.673) and two modal haplotypes—separated by four mutational steps—with frequencies (0.46 and 0.455). There are no major differences in the frequencies of these two modal haplotypes across Indian populations. The India-specific G6PD deficiency variants—Kalyan-Kerala and Orissa—have evolved on separate haplotype backgrounds. Comparison of our Indian data with available global data indicates strong haplotypic sub-structuring that weakly resembles balancing selection effect. (d) However, the standard statistical tests did not reveal statistically significant evidence of balancing selection. Conclusion: There is a very strong structuring of haplotypes at the G6PD locus in India, indicating strong selective effects. Evidence of balancing selection is not clearly discernible. The common India-specific deficiency variants have evolved independently on distantly-separated haplotype backgrounds.

151: An evaluation of tSNP portability across isolated population groups

[1]**Neeta Sarkar Roy**, [2]Shabana Farheen, [2]Nilanjana Roy, [2]Sanghamitra Sengupta, [1,2]Partha Pratim Majumder

[1]TCG-ISI Centre for Population Genomics, Bengal Intelligent Park Ltd., Building B, 3rd Floor, Block EP and GP, Sector V, Salt Lake Electronics Complex, Kolkata 700091, India, [2]Indian Statistical Institute, Human Genetics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

Introduction: If tag SNPs (tSNPs) identified in a population can be ported to many other populations, then the conduct of association studies for disease gene mapping is simplified.
Aim: To examine the portability of tSNPs (a) across isolated population groups of India, and (b) identified in continental (European and African) populations (using data from Seattle SNPs database) to Indian populations. Methodology: Ten ethnically and geographically disparate population groups of India were selected. DNA samples were obtained from 160 individuals belonging to these groups. Three genomic regions spanning 12 Kb were resequenced in each individual. Based on genotype data at the polymorphic loci (MAF > 0.05), tSNPs were identified in each population using pair-wise tagging method with a cut-off for r^2 set at 0.8. An index of portability of tSNPs chosen in one population and applied on another was devised, based on the ratio of diversity of haplotypes reconstructed using the chosen tSNPs to the total haplotype diversity in the population to which the tSNPs are to be ported. Higher values of this index indicate greater portability. Results: (1) The proportions of shared SNPs between continental and Indian populations vary greatly across genomic regions. (2) There is significant variation in allele frequencies at shared SNP loci across populations. (3) The extent of portability of tSNPs selected in one Indian population to other Indian populations was moderate (mean of values of portability index $= 0.816$; against a proposed cut-off value of 0.947). (4) The extent of portability was not higher across populations within broad ethnic clusters (caste and tribe) than between clusters. (5) A similar feature was also noted with respect to within and between geographical zones, except for the north-eastern zone. (6) tSNP portability across (both from and to) the European population and the Indian ethnic groups was higher compared to the African population. (7) The Mantel test of correlation between tSNP portability and genetic affinities of populations was not significant ($P > 0.1$). Conclusion: Although isolated populations are considered useful in mapping genes for complex diseases, tSNP identified in one population isolate may not be profitably used in another isolate, at least in India. Therefore, it may be necessary to carry out resequencing in a small number of individuals to discover SNPs and identify tSNPs in the specific population in which the disease association study is to be conducted.

152: Genetic differentiation and ancestral histories of Indian population groups: inferences from Y-chromosomal markers

[1]**Sanghamitra Sengupta**, [2]Debabrata Sutradhar, [2]Shabana Farfeen, [2]Neelanjana Roy, [2]Partha Pratim Majumder

[1]Department of Biochemistry, University of Calcutta, 35 Ballygunge Circular Road, Kolkata 700 019, West Bengal, India, [2]Human Genetics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, West Bengal, India

The phenomenal socio-cultural, including linguistic, and geographical diversity of the ethnic groups of India is well documented. However, the nature of genetic differentiation among the Indian ethnic groups vis-à-vis linguistic, social and geographical diversity, and the ancestral histories of these groups remain a matter of controversy. Since many past studies have used restricted sets of genomic markers and/or populations with limited socio-cultural/geographical diversity to draw inferences, in this study we have sought to rectify this lacuna. We have assayed 71 endogamous Indian ethnic groups ($n = 1{,}407$ males) representing four linguistic groups and seven geographical regions. Each individual was screened for 75 (69 binary and 8 STR) informative Y-chromosomal markers. Based on the binary markers, the Indian male gene pool could be classified into 27 haplogroups, of which seven had frequencies >5%. Overall, genetic differentiation among the four linguistic affiliates was the highest [F(ST) = 0.351], reflecting diverse origins of the people speaking different languages. A similar high F(ST) [= 0.389] was also observed among geographical regions; perhaps due to the confounding of geographical habitat and language spread in India. The Austric (Austro-Asiatic) and Sino-Tibetan (Tibeto-Burman) speaking populations co inhabit eastern and north-eastern parts of India and also southeast Asia. Both these languages have supposedly evolved in southern China and therefore the speakers of these languages are expected to have been derived from a common ancestral gene pool. However, we have found the major haplogroups in these language groups to be different, possibly indicating that these language groups are derivatives of distinct ancestral gene pools; among Austro-Asiatics the major haplogroup is M95 (frequency = 55.6%; STR variance 0.30), while among Tibeto-Burmans it is M134 (69.1%; 0.25). A remarkable correspondence between haplogroups M52 (29.7%) and M89* (12.4%) with Dravidian language was observed. In each geographical region, tribes and castes are highly differentiated [F(ST) range: 0.410 (East)–0.113 (North)]. Between caste ranks, however, the extent of genetic differentiation is low [F(ST) = 0.094] indicating superimposition of a ranked social structure on a common gene pool.

153: Genetic affinities and admixture among Siddis, a recent migrant population from Africa

[1]**Anish M. Shah**, [2]Gururaj Kulkarni, [3]Tanmoy Bhattacharya, [2]P. B. Gai, [3]G. Mustaq, [1]Rakesh Tamang, [1]L. V. K. S. Bhaskar, [1]Alla G. Reddy, [1]K. Thangaraj, [1]Lalji Singh

[1]Centre for Cellular and Molecular Biology, Hyderabad, India, [2]Department of Applied Genetics, Karnataka University, Dharwad, India, [3]Mangalore University, Mangalore, India

The non-recombining, uniparentally transmitting human Y chromosome and mitochondrial DNA are useful tools for tracing population history. In this study we have tried to establish the genetic affinities between the Siddi, recent migrant population from African subcontinent, and the neighboring populations. We have analyzed the Y chromosome and mtDNA of 157 Siddis inhabited in Karnataka, and Gujarat and 178 individuals belonging to Medar, Gram Vokkal, Korova and Kare Vokkal tribal populations, who are residing at the close vicinity of Siddis. The results showed that ~80% Siddi fell in Y- haplogroups M182-B2, M130-C and E (including M2-E3a and M33-E2) lineages, while individuals from surrounding population showed all Indian specific lineages such as M82-H1, M11-L, M172-J2 and M17-R1 haplogroup. Analysis of mitochondria DNA showed the presence of deep rooted African specific L haplogroups, including L0, L0a2, L1, L2, L2a1, L2b1, L3d, L3d3 and L3e1f among Siddi, along with South Asia specific haplogroups such as M, M2, M3, M4, M5, M6, M35, N, R, R30, U2 and U7 lineages suggesting maternal gene flow from the neighboring tribal populations. The results were compared with published data of 1,361 Y chromosomes and 3,784 mitochondrial DNA (hypervariable sequence) of African individuals.

The multidimensional analysis showed close affinity of Siddi with Bantu speaking population from Kenya supporting the anthropological literatures on Indo-African trade relationships.

154: Diversity of mitochondrial and y-haplogroups in TB patients of Sahariya tribe of Central India

[2]**P. R. Sharma**, [1]Swarkar Sharma, [1]Sailesh Gochhait, [1]Mamata Jena, [2]P. K. Tiwari, [1]R. N. K. Bamezai

[1]Jawaharlal Nehru University, New Delhi, National Centre for Applied Human Genetics (NCAHG), School of Life Sciences, India, [2]Jiwaji University, Gwalior, Centre for Genomics, School of Studies in Zoology (CG-SSZ), India

Introduction: Sahariya is a primitive tribal group, populated in the North-West region of Madhya Pradesh, Chattisgarh and Rajasthan, (India). Despite being a significant part of our genetic lineage, no genetic study has yet been carried out, either on the origin of this tribe or the diseases. Objectives: We present a preliminary analysis on the diversity of mitochondrial and Y-chromosome haplogroups in TB patients in the studied tribe. In addition in a case-control comparison study, two known SNPs in the candidate genes, TLR1 and TLR2 are analyzed, to identify possible risk factor/s from mitochondrial or nuclear genome, towards the susceptibility of the disease. Methods: A demographic survey ($N = 400$) was made from ten Sahariya villages. Blood samples were collected after seeking ethical approval and consent of the donors. 55 Y-SNPs in 69 Y-chromosomes and 4Y-STRs in 100 Y-chromosomes were analyzed by PCR-based genotyping. Mitochondrial haplogroups were assigned to 200 samples after HVRI sequencing using ABI prism sequencer. TLR1 (rs4833095) and TLR2 (arg753gln) polymorphisms were studied as potential candidates for TB susceptibility. Results: The Y-SNP and STR analyses showed that the tribe has a high frequency (0.18) of R1a1 haplogroup. About 34% TB patients were from haplogroup R1a1, 19.1% from R1a*, 4.2% from O and H haplogroups of Y-chromosome. While in the controls, haplogroups L and O showed a relatively higher prevalence. Mitochondrial haplogroup M5, M3a/b, M3b and M4a were found significantly ($P < 0.05$) high in the TB patients of the tribe. The TLR1 SNP (rs4833095) analysis revealed that CC genotype of TLR1 gene is significantly ($P < 0.05$) high in cases than control subjects. GA genotype of TLR2arg753gln polymorphism showed a trend towards significance ($P < 0.06$) in cases, which may show up once a larger sample size is studied. However, AA genotype for TLR2 polymorphism was not found in cases or control subjects of the tribe. Conclusion: In this preliminary study, the relevance of R1a1 Y-chromosome haplogroup and sub-haplogroups of M lineage of mitochondrial genome backgrounds in TB patients of Sahariyas needs future investigation for their functional role, if any. The CC genotype of TLR1 and GA genotype of TLR2 gene polymorphisms in Sahariyas, may provide an increased susceptibility to tuberculosis, a conclusion which would need validation with a larger sample size. [Financial support to NCAHG from UGC and to NCAHG and CG-SSZ from DBT is acknowledged].

155: Phylogenomic study of a concealed Ladakh tribe of the Great Himalayas

Vishwas Sharma, Amrita Nandan, Varun Kumar Sharma, Kumaraswamy Thangaraj, Lalji Singh

Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India

Ladakh is cosseted by the world's two mightiest mountain ranges; the Great Himalayas and the Karakoram. This region is bordered by China and Pakistan. The high altitude ($\sim$14000 ft above mean sea level) and adverse living conditions has made this region unreachable. Therefore, the present study was endeavored to investigate the origin and historic expansion of a rare Ladakhi tribal population (Changpas). Fifty unrelated healthy male individuals were screened for variation in mtDNA and Y-chromosome binary, and STR markers. Complete mtDNA analysis showed the maternal gene pool of Ladakhi is highly diverse. In addition to the Indian specific haplogroups, we have also identified a few non-Indian specific haplogroups in this population. The Y-chromosome study demonstrates 86% of the Ladakhi paternal gene pool consists of D, N, O, P and R haplogroups. Presence of substantial number of non-Indian specific and a few rare haplogroups among Ladaki tribal population suggests their close affinities with South East Asian populations. More evidences on the close affinities would be provided at the time of presentation.

156: Genetic heritage of Indian Jews population

[1]**Kamayani Singh**, [1]Deepankar Pratap Singh, [2]Mini Kariappa, [1]Lalji Singh, [1]Thangaraj Kumarasamy

[1]Centre for Cellular and Molecular Biology, Habshiguda, Uppal, Hyderabad, Andhra Pradesh, India, [2]Department of Anatomy, Jubilee Mission Medical College, Trichur, Kerala, India

India played a major role in early human migration and it is evident from the existence of huge human diversity in this country. Attempts have been made to establish the origin and migration of some of the unique populations of India. In one such attempt, we are interested in understanding the entry of Jews in Indian and the status of their admixture with the local Indian populations, as there is no consensus of opinion about these. We have analyzed mitochondrial and Y chromosome DNA of 180 Jews from Kerala. Our study indicates higher frequency of Indian specific haplogroups; M2 (8.25%), R5 (5.5%), and R6 (9.9). In addition, we have also observed a noticeable frequency of Middle Eastern specific haplogroups HV, H2a, U2e, U4, and K. Y chromosome analysis revealed high frequency of R-M17 (24.4%) suggests possible gene flow from Eastern Europe. In addition to this, the Indian specific haplogroups such as M82-H1, M11-L was also observed in significant frequencies. Our study strongly supports that the Jews have assimilated into the wider non-Jewish society, inhabited in Kerala state. A detailed analysis with admixture ratio would be made available during presentation.

157: Identification and analysis of genetic signatures of selection in Indian populations

[1]**Prashant K. Singh**, [2]Amit Chaurasia, [1]Yasha Bhasin, [1,2]Arijit Mukhopadhyay, [1]Pankaj Jha, [2]Vinod Scaria, [3]Swapnil Sinha, [3]Saman Habib, [1,2]Mitali Mukerji, [4]Partha P. Majumder, [5]The Indian Genome Variation Consortium, [1,2]Samir K. Brahmachari

[1]Genomics and Molecular Medicine, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, [2]G.N. Ramachandran Centre for Genome Informatics, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India, [3]Central Drug Research Institute (CSIR), Lucknow, India, [4]Human Genetics Unit, Indian Statistical Institute, Kolkata, India, [5]Nodal Laboratory, Institute of Genomics and Integrative Biology (CSIR), Mall Road, Delhi, India

The genetic association and linkage studies aim to identify genetic variants in different populations that predisposes to certain diseases. Identification of the genomic regions which have been targets of natural selection will help in the understanding of human

evolutionary history and also help in identifying the candidate genes for complex diseases. These signatures of selection are often confounded by population migration history and by variation in local recombination rates. Neutral regions of a genome are expected to have similar patterns of allele frequency variability and distributions. The patterns are expected to be different if the genomic region is influenced by biological selection. In the present study we selected ~1,000 genes known to be involved in various cellular and molecular processes, ~4,000 common single nucleotide polymorphisms genotyped across these genes in 24 diverse reference Indian populations and searched for the evidence of selection based on three tests: FST versus heterozygosity, derived allele frequency (DAF), extended haplotype homozygosity (EHH) and integrated haplotype scores (iHS). We have also genotyped ~58,000 neutral SNPs (50 K affymetrix array) in the same samples. Initially, we performed the analysis with a set of ~400 SNPs in a representative set of 75 genes. We first compared the distributions of average heterozygosity (HO) and levels of population differentiation (FST) for the various SNPs. Overall, for all SNPs there was a significant negative correlation between HO and FST (P less than 0.01). Among them, both exonic and intronic SNPs exhibited significant negative correlation (P less than 0.05), suggestive of balancing or local selection. Balancing selection was indicated for multiple SNPs of genes involved primarily in immune function and signaling; local selection was indicated for genes involved in transport, signaling and metabolism and positive selection was indicated for genes involved in cell growth and aging. Ancestral allele was ascertained from chimpanzee sequence revealed a wide range of ancestral allele frequencies (AAF). Nearly 40% of these SNPs were not flipped in any of the Indian populations whereas ~23% were flipped in majority (more than 50%) of the populations when compared to chimpanzee allele. Most of the highly differentiating flipped SNPs belonged to genesinvolved in signaling, transport and metabolism. In depth analysis of all 4,000 SNPs will be presented.

158: RISC—repeat induced sequence changes

Vipin Singh

Centre for Cellular and Molecular Biology, E-503, Bioinformatics, Centre for Cellular and Molecular Biology, Hyderabad, 500007, India

Mobile elements continue to be an enigma in the genome, with numbers close to millions and still, not assigned a well-defined function. Are these the agents for engineering subtle sequence changes by causing insertions, target site duplications (TSD), 3′ and 5′ flank transductions and occasionally effecting deletions upon transposition besides also providing room for illegitimate recombination? A comprehensive, comparative analysis of the sequence changes induced by transposons between evolutionarily close genomes might reveal the role of these repeat induced changes in evolution and speciation. To this end, I have designed RISC—Repeat Induced Sequence Changes—a comprehensive, comparative genomics based, in silico subtractive hybridization tool, to identify differential insertions (or polymorphism when genomes of the same species are compared), target site duplications, 3′ and 5′ flank transductions, and parallel insertions or deletions caused by transposons. Confirmation modules for each of the above mentioned changes are inbuilt in RISC. RISC allows both whole genome comparisons as well as locus specific comparisons with one or multiple genomes. It can delineate the exact boundaries of the differential transposition events. RISC also takes care of multiple complications that arise in the analysis of truncated repeats by providing a utility module for segregating true repeat truncation events from repeat disruption or miss-annotation events. Also, RISC allows a better resolution of the unaltered locus as also the appreciation of the transduced flank in its original context. Besides, RISC is not limited by the size of the TSD and can report TSDs as small as 1 bp with reasonable certainty. RISC is implemented as an automated pipeline and has been tested on full length L1HS sequences with reference human genome compared against reference chimpanzee genome and Celera human genome. Of the 300 full length L1HS elements (as identified by Repeat Masker in the reference human genome), RISC identified 247 and 42 Target Site duplications, 28 and 3 3′ flank transductions, 28 and 38 INDELS, in chimpanzee and Celera human genomes, respectively, when run on default parameters. None and 213 of the orthologous loci were predicted as occupied while no matches were found for 25 and 5 L1 s in chimpanzee and Celera genomes, respectively. The accuracy of RISC results depends on the quality of the genomes being compared. RISC results are presented as customized summary and can be exported into Excel sheets.

159: Population history of North Eurasia revealed by X-chromosome ZFX haplotypes

Vadim Stepanov, Irina Khitrinskaya, Vladimir Kharkov

Institute for Medical Genetics, Nab. Ushayky 10, Tomsk, Russia

To reconstruct the origin and evolution of human populations in North Eurasia we investigated the genetic diversity in 23 p opulation samples (about 1,300 individuals totally) using X chromosome lineages. SNP markers in a single 50 kb linkage disequilibrium region of ZFX gene was used to trace the X chromosomal population history. Forty one ZFX haplotypes were found in populations of North Eurasia, belonging to three major clusters (see cluster definitions below). North Eurasian populations are less divers with respect to X-chromosomal haplotypes (gene diversity within populations 0.65–0.80) and less differentiated (Fst = 3%) compared to Y lineages. But the regional differentiation demonstrates the same picture: the highest level of between-population genetic differences was observed in isolated populations of North East Asia, whereas European and Central Asian populations are almost undifferentiated by ZFX lineages. Three clusters of population are formed in the space of two first principal components. The most distant cluster consisted of African Yoruba population, and North Eurasian populations are joined into two clusters, one of which consisted mainly of Caucasoid populations, whereas Siberian and South-East Asian Mongoloids form the second. The positions of populations in the PC1/PC2 space strongly correlate with geographic distances between populations. The population clustering by X and Y gives, to a first approximation, a similar picture, and matrices of genetic distances between populations for X and Y haplotypes significantly correlates. Strong linkage disequilibrium was observed between all ZFX SNPs, which formed the single LD block in all studied populations. North Eurasian ZFX haplotypes belong to two main clusters, both originated from the same African root (cluster A). Two major haplotypes were observed within the cluster B, and the single founder haplotype of cluster C was found. The structure of root haplotype was estimated by comparing human and chimpanzee (Pan troglodytes) ZFX gene sequences. The founder of the whole tree was found with low frequency in African population, and haplotypes within 1–2 mutations distances (cluster A) from founder were observed mostly in Africans. The presence of two different clusters in Europeans and Asians probably provides the evidence of two population expansions outside Africa. This work is supported by RFBR grants ##06-04-48274 and 07-04-01629.

160: High altitude population stratification based on SNPs in genes of cardiopulmonary function: IGVdb application

[1]**Tsering Stobdan**, [1]Azim Nejatizadeh, [1]Tsering Norboo, [3]Gulam Mohammad, [3]Mohammad Iqbal, [3]Tashi Thinles, [4]The Indian Genome Variation Consortium, [1]M. A. Qadar Pasha

[1]Institute of Genomics and Integrative Biology, Functional Genomics Unit, Institute of Genomics and Integrative Biology, Delhi, India, [2]Ladakh Inst of Prevention, Ladakh Inst of Prevention, Leh, J and K, India, [3]SNM Hospital, SNM Hospital, Leh, J and K, India, [4]IGVdb, Nodal Laboratory, Institute of Genomics and Integrative Biology, Delhi, India

Introduction: The strong environmental selection pressure plays a major role in defining individual phenotype. The hypobaric hypoxia at high altitude (HA) is one such selection pressure which may play role in genetic selection of individuals. The major disorders at high altitude are related to cardiopulmonary function, such as high-altitude pulmonary edema. The present study aimed to identify the differential distribution of population on a principal component (PC) plot based on few genes having role in cardiopulmonary function. Methods: Eight indigenous populations were selected from four corners of India. The data were obtained from IGVdb (Indian Genome Variation database). Ninety one loci were selected from chromosome 22 and 51 loci from seven genes having role in cardiopulmonary functions. The population FST values were calculated using software Arlequin version 3.11 and PC plot was drawn by SPSS for Windows (version 10). Result: The PC plot from 91 loci of chromosome 22 depicts wide dispersal of the three Tibeto-Burmans (TB) populations while rest of the five populations consisting remaining major ethnic groups, i.e. Dravidian, Indo-European and Austro-Asiatic forms a cluster. The PC plot obtained from 51 loci of genes depicted the two TB populations residing at similar altitudes clusters together. The remaining populations of other ethnicity were widely dispersed on the plot. Conclusion: The genetic landscape based on 91 loci of chromosome 22 shows that TB are considerably different from rest of the ethnic groups of India. The candidate gene of cardiopulmonary function was under considerable selection pressure. There is a strong environmental selection of high altitude on gene involved in cardiopulmonary function. A large scale genetic marker of candidate gene may provide population stratification and the similar environment they share.

161: High frequency of rare Y chromosome and mtDNA haplogroup suggests unique origin of Darjeeling populations

[1]**Rakesh Tamang**, [1]Anish M. Shah, [1]A. Sharath, [2]Tikaram Sharma, [1]Alla G. Reddy, [1]K. Thangaraj, [1]Lalji Singh

[1]Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad, India, [2]Banaras Hindu University, Varanasi, India

Himalayas Mountain ranges plays important role in migration and the peopling of the nearby regions. Darjeeling, the place surrounded by extended mountainous region of the countries like Nepal and Bhutan, is a district of West Bengal in its northern part inhabited by different tribal population. The neighboring districts are mostly the plains including Jalpaiguri and Cooch Bihar districts. The place remains colloidal inhabitation for different linguistic groups such as Sino-Tibetian, Austro-Asiatic and Indo-European populations. Genetic origin and affinities of the populations of Darjeeling is very poorly understood, hence we intend to study the populations of Darjeeling using the Y chromosome and mtDNA markers. We have analyzed a total of 173 males from Sherpa, Subba and surrounding populations. The Y chromosome haplogroup analysis revealed that nearly 50% of the individuals fell in M122-O3 haplogroup followed by M174-D haplogroup (17%). Presence of high frequency of M174-D haplogroup, which is predominant among Andamanese and Japanese populations needs further attention and analysis of sub haplogroups and Y chromosome specific STRs would provide valuable information about the origin of the populations of Darjeeling. Analysis of mitochondrial DNA of these populations revealed North East Asian specific D4 lineage in approximately 17% of the individuals followed by A4 and C haplogroup in $\sim$10% individuals while rest all individuals fell in Indian specific M sublineages such as; M2, M3, M4, M5, M6, M8, M9 and G haplogroups. This is for the first time we are detecting the unique combinations of Y chromosome and mtDNA haplogroups among Indian populations. Further, comparative analysis and the probable origin of the Darjeeling populations would be made available at the time of presentation.

162: Tree of life constructed by genome-wide information

Yoshio Tateno

National Institute of Genetics, Yata, Mishima 411-8540, Japan

For the construction of a tree of life that consists of the three super-kingdoms, Archaea, Bacteria and Eukaryota, we should use not a limited number of genes but full-scale genome information. With this in mind we first developed a new method for constructing a tree of life based on the protein domain organizations of all proteins detected in the genome of a species. The protein domain organization is defined as the sequential order of domains in a protein. The new method is not in need of the identification of orthologs among the species in question and thus free from the burdensome and error-prone computation necessary for the identification. We then compared the repertoires of the protein domain organizations of 17 archaeal, 136 bacterial and 14 eukaryotic species/organisms, computed evolutionary distances among them and constructed a tree of life for them. Our tree shows that not only the three super-kingdoms but also each eukaryotic kingdom and most bacterial phyla are in monophyly. The branching pattern of the bacterial phyla in our tree is consistent with the widely accepted bacterial taxonomy and very close to other genome-based trees. However, a few inconsistent aspects between the traditional trees and ours would perhaps urge to revisit and revise the conventional view of tree of life, in particular on the phylogenetic positions of hyperthermophiles.

163: Detection of mitochondrial DNA heteroplasmy in low quantity samples: Whole genome amplification (WGA) vs nested PCR

[1]**Binuja Varma**, [2]Garima Agrawal, [1]Sujatha Sunil

[1]Delhi Forensic Science laboratory, Delhi, India, [2]The Centre for Genomics Application, 254, Okhla Phase III, New Delhi, India

Mitochondrial DNA (mtDNA) sequence analysis of the hypervariable control region has been shown to be an effective tool in various fields like Forensics, Geriatrics and Mitochondrial diseases. High copy and maternal mode of inheritance make mtDNA analysis particularly useful when old samples or degradation of biological samples prohibits the detection of nuclear DNA analysis. Whole genome

amplification (WGA) promises to eliminate practical molecular genetic analysis limitations associated with genomic DNA (gDNA) quantity. It is being widely used to increase the quantity of starting gDNA and uses either multiple displacement amplification or isothermal strand displacement amplification. However, the sensitivity and specificity of this technology to detect heteroplasmy is under debate. In the present study, hypervariable region I (HV1) of the human mtDNA control region was studied using whole genome amplification of gDNA and compared with nested PCR technique. Samples were collected from three tissues—blood, hair and buccal swabs from 48 individuals and DNA extracted. Amplified DNA extracted from blood, hair and buccal swabs were compared with whole genome amplified (WGA) DNA from the respective tissues by amplifying the HV1 region and the amplicon subjected to sequencing. The standard protocol failed to yield successful PCR reactions in 95% (hair), 80% (buccal) of specimens, whereas WGA of blood, buccal and hair samples was markedly more efficient (2–5% failed PCR). However, when a secondary PCR was set up using the amplicons of primary PCR as templates, 97% of samples yielded a good product. In blood, buccal and hair specimens, length heteroplasmy was detected in four samples and point heteroplasmy in 13 samples respectively, while in WGA specimens, the percentage of length heteroplasmy was 5% and point heteroplasmy was only 13%. The study showed that whole genome amplification was not able to detect heteroplasmy accurately and cannot be used for studying heteroplasmy in mitochondrial DNA.

164: Going India to Australia: new genetic evidence on modern human colonization

[1]**Satish Kumar Yadav**, [2]V. R. Rao

[1]Anthropological Survey of India, S.R.C., Mysore, India,
[2]Anthropological Survey of India, 27. J.N.Road, Kolkata, India

The greatest ever reconstructed journey of Mankind, begins in Africa with a small group of hunter-gatherers and ends some 150–200 kiloyears (ky) later with their 6.5 billion descendants spread across the occupied world. However route(s) and time of such spread, undertaken by the anatomically modern humans to populate the world has been the greater untold story. Recent genetic studies suggest single 'southern' dispersal extended from Horn of Africa into Arabia and southern Asia some time before 50 ky. Subsequently with a rapid population growth around 50–69 kyBP modern humans expanded eastward along the coasts of southern and southeastern Asia into Australia at least by 45 kyBP. The major challenges to this scenario are: (1) Non overlapping distribution of the derived haplogroups within M, N and R in South Asia, Eastern Asia and Australasia. However, number of such derived haplogroups are of overlapping age to the said dispersal. (2) Despite the high haplotypic diversity, coalescent age estimate of M macrohaplogroup in India i.e. 54.1 ky is considerably low as compared to its East Eurasian counterparts. (3) Large area of both Arabia and India are at present largely blank on the archaeological map for the critical period from $\sim$50 to 60 kyBP. However Patne, Jwalapuram in India and Batadomba-lena in Sri Lanka provides some intriguing hints of early modern human occupation and striking resemblances in archaeological assemblages to those from eastern and southern Africa. But lack of similarly 'advanced' technologies to the east of the Indian subcontinent, especially in well-explored area of Australia and New Guinea. Our complete mtDNA sequencing of 900 individuals from the 27 relic populations identified six individual from Dravidian and Austro-Asiatic tribes who shares two basal coding region mtDNA mutations with Australian Aborigines specific M42 haplogroup. Thus, enforcing the phylogenetic inferences of 'single southern route'. With the given set of 900 complete mtDNA sequences, we estimated lineage wise past population size through time to draw inferences about human population history in India, using Bayesian coalescent approach. Our result suggest that the demographic event preceding to the initial growth phase, particularly during and immediately after the Last Glacial Maximum and in the last 3 ky (during the demic diffusion of agriculture and associated technologies) has played a major role in shaping phylogentic interrelationship of populations within and outside India.

165: Structure of Japanese population based on SNP genotypes from 7,003 individuals in comparison to other ethnic groups: effects on population-based association studies

[1]**Yumi Yamaguchi-Kabata**, [1]Kazuyuki Nakazono, [1]Atsushi Takahashi, [2]Susumu Saito, [2]Naoya Hosono, [2]Michiaki Kubo, [3]Yusuke Nakamura, [1]Naoyuki Kamatani

[1]Laboratory for Statistical Analysis, Center for Genomic Medicine, RIKEN, 4-6-1, Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan, [2]Laboratory for Genotyping, Center for Genomic Medicine, RIKEN, Yokohama 230-0045, Japan, [3]Laboratory for Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan, [4]Center for Genomic Medicine, RIKEN, Yokohama 230-0045, Japan, [5]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo 162-0054, Japan

Because population stratification can cause spurious associations in case-control studies, understanding the population structure is important. Here we examined the population structure of 7,003 Japanese individuals using the genotypes for 140,387 SNPs by 'Eigenanalysis', along with 60 European, 60 African and 89 East Asian individuals in the HapMap project. Most Japanese individuals fell into two main clusters, Hondo and Ryukyu clusters, which may reflect the descendents of the Yayoi and Jomon populations, respectively, in ancient Japan. The highly differentiated regions between Hondo and Ryukyu clusters were found in HLA region in chromosome 6, and particular regions in chromosomes four and nine. Two nonsynonymous SNPs, one of them is in the EDAR, which is associated with Asian hair thickness and the other in ABCC11, which is associated with dry ear wax, showed significantly different genotype frequencies between Hondo and Ryukyu clusters. Genetic differentiation was observed even among different regions in Honshu Island, the largest island of Japan. Simulation studies showed that the inclusion of different proportions of individuals from different regions of Japan in case and control groups can lead to an inflated rate of false positive results when the sample sizes are large.