Population Genomics Research of Human Admixture History and Local Adaptation

Max Planck-CAS Research Group on Population Genomics

Shanghai-Leipzig, 2012.01.01 – 2014.12.31

The Max Planck Junior Research Group on Population Genomics at the Partner Institute for Computational Biology. This photo was taken in Xiangshan, Ningbo in June 2013. People in the photo include: (from left to right) YUAN Kai, HUANG Xin, WANG Minxian, TIAN Lei, FENG Qidi, ZHANG Chao, ZHOU Ying, LI Ran, YANG Xiong, FU Ruiqing, PU Jing, QIN Pengfei, Pankaj Kumar, DENG Lian, SHI Meng, LI Jing, YUAN Yuan, WANG Yuchen, LU Yan, XU Shuhua, LOU Haiyi, LU Dongsheng, XU Hongyang.

Officially set up in January 2011 by Prof. Dr. XU Shuhua at the Partner Institute for Computational Biology, the Population Genomics Group focuses on population genomics research of human admixture history and biological adaptation to the local environment. Population Genomics is a disciplinary to infer population genetic and evolutionary parameters from genome-wide data sets. The ultimate goal of this research group is to understand microevolution mechanisms in human, while genetic admixture is taken as a cut-in point to pursue this ambition. The partner group at MPG in cooperation with this team is the Research Group on Human Population History led by Prof. Mark Stoneking at the Max Planck Institute for Evolutionary Anthropology (Leipzig).

Reported by Group Leader XU Shuhua

credit: Photos by courtesy of Prof. XU Shuhua

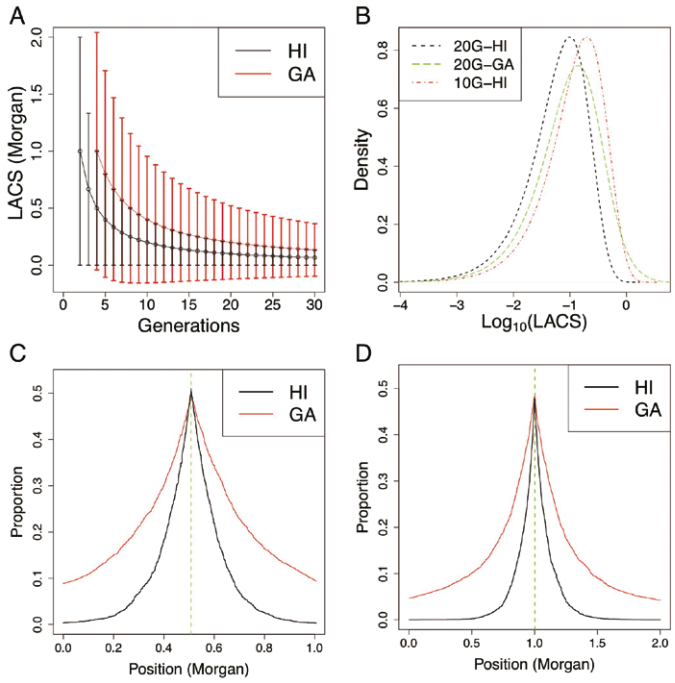

Figure 1: **Admixture Dynamics of European and African Ancestry in African Americans.** For the continuous gene flow (CGF) model, the case in which Europeans continually served as genetic donors was considered as CGF1 model, whereas the one in which Africans as genetic donors was considered as CGF2 model. To find the model that fits the empirical distribution the best, earth mover's distance (EMD) between empirical data and that of each model was calculated. The model showing the lowest EMD with the empirical data was considered as a best fit. (A) Distribution of EMDs for the African ancestral component between empirical data and each model. (B) Empirical distribution of the chromosomal segments of distinct ancestry (CSDA) length for the African ancestral component, and simulated distributions when the number of generations was set to 14. (C) Distribution of EMDs for the European ancestral component between empirical data and each model. (D) Empirical distribution of CSDA length for the European ancestral component and the simulated distributions when the number of generations was set to 14. *American Journal of Human Genetics* (2012, 91:849–862).

Admixture has been a common phenomenon throughout the history of modern humans, as previously isolated populations often come into contact through colonization and migration. It is important to conduct a full analysis of genetic structure and characterize the genetic make-up of admixed populations. On the one hand, this will shed light on the human genetic history; on the other, increased population admixture influences genome diversity, which in turn will affect phenotypes relevant to health; thus, genetic admixture has many implications for medical research. Dr. XU Shuhua's group is using computational approaches and developing new methods to dissect genetic architecture of human populations, quantitatively characterize their admixture features, and reveal their migration history and adaptive divergence. Specifically, Dr. Xu Shuhua's group is working on several projects, focusing on theoretical modeling of human population admixture, statistical inference of human migration history and detection of footprints of natural selection in human genome, respectively.

One major effort in Dr. XU's group is to elucidate population admixture dynamics of different human ethnic groups. The processes of genetic admixture determine the haplotype structure and linkage disequilibrium patterns of the admixed population, which is important for both medical and evolutionary studies. However, most previous studies do not consider the inherent complexity of admixture processes. Dr. XU and his team proposed two approaches to explore population admixture dynamics. His team applied the methods in the analysis of genome-wide data of 1,890 African Americans and revealed that a continuous gene flow model, in which the African American population continuously received a gene flow from European populations over about 14 generations, best explained the admixture dynamics of African Americans. In contrast, the admixture dynamics of Mexicans could be explained by a gradual admixture model, in which the Mexican population continuously received a gene flow from both European and Amerindian populations over about 24 generations. Their results also indicated that recent gene flows from Sub-Saharan Africans have contributed to the gene pool of Middle Eastern populations such as Mozabite, Bedouin, and Palestinian. In summary, this study not only provides approaches to explore population ad-

"I appreciate the support of Max Planck Society focusing on people-centered research and a safe budget over the past few years, which gave me a lot of freedom to explore my academic and research interests."
— Prof. Dr. XU Shuhua

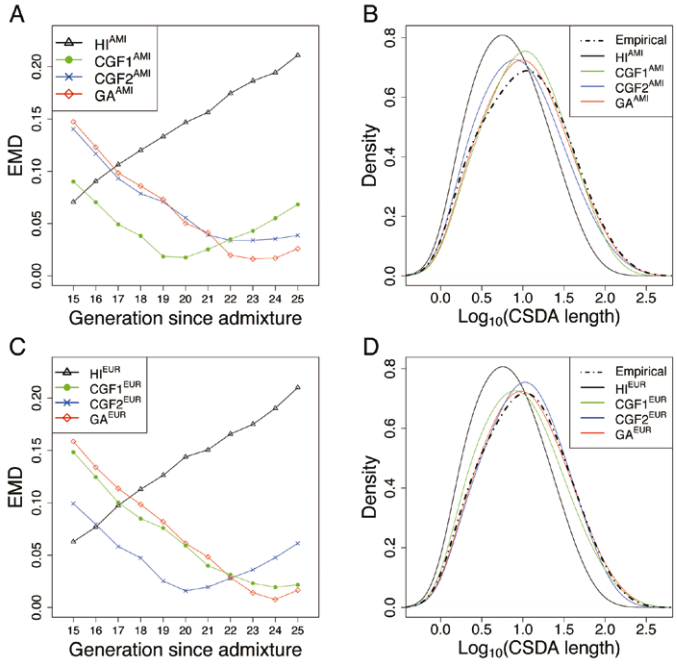

Figure 2: **Comparison of LACS distributions between HI model and GA model.** (A) Comparison of mean and SD of LACS between the HI model and the GA model. The error bars and circles represent SD and mean respectively. (B) Comparison of LACS distributions between HI model and GA model when generation is 10 or 20. (C) The change of genetic contribution that transmitted with the given locus along the chromosome in the HI and the GA models. (D) The change of genetic contribution transmitted with the given locus along the chromosome in the HI and the GA models in a longer chromosome. The green vertical dashed line represents the reference locus. *European Journal of Human Genetics* (2013, doi: 10.1038/ejhg.2013.265)

mixture dynamics, but also advances our understanding on population history of African Americans, Mexicans, and Middle Eastern populations. The work has been published in the *American Journal of Human Genetics* (2012, 91:849–862).

As a follow-up study, Dr. XU's group introduced a theoretical framework on the distribution of the lengths of ancestral chromosomal segments (LACS) in two representative admixture models, *i.e.*, the hybrid isolation (HI) model and the gradual admixture (GA) model. Although the distribution of LACS in the GA model differs from that in the HI model, they demonstrated that the mean LACS in the HI model is approximately half of that in the GA model if both admixture proportion and admixture time in the two models are identical. The theoretical framework established by them greatly facilitated the inference and understanding of population admixture history by analyzing African-American and Mexican empirical data. In addition, they found that the peak of association signatures in the HI model was much narrower and sharper than that in the GA model, indicating that the identification of putative causal allele in the HI model is more efficient than that in the GA model. Therefore, theoretically speaking, admixture mapping with case-only data would be a reasonable and economical choice in the HI model due to the weak background noise. However, in reality, many populations are likely to be gradually admixed and have pretty high background linkage disequilibrium. Taken together, they suggested using a case-control approach rather than a case-only approach to search for disease-related genes, to retain the statistics power in recently admixed populations. The work has been published in the *European Journal of Human Genetics* (2013, doi: 10.1038/ejhg.2013.265).

Working together with his collaborators from Germany and Netherlands, Dr. XU applied admixture analysis to answer questions on human migration history. An example is the genetic dating of the Austronesian expansion, which had a major impact on the human genetic diversity and languages of Island Southeast Asia. The collaborative team led by Dr. XU applied admixture analysis and utilized recombination information estimated from admixed genomes to estimate the amount and time of admixture in Eastern Indonesian populations. Their analyses of two genome-wide datasets indicate an eastward progression of the Asian admixture signal in Eastern Indonesia beginning about 4000 to 3000 years ago, in excellent agreement with inferences based on Austronesian languages. This was the first genetic dating based on admixture analysis and recombination information of the Austronesian expansion. Their results significantly advanced our understanding of the biological origins of human populations in the Asia-Pacific region. This work has been published in

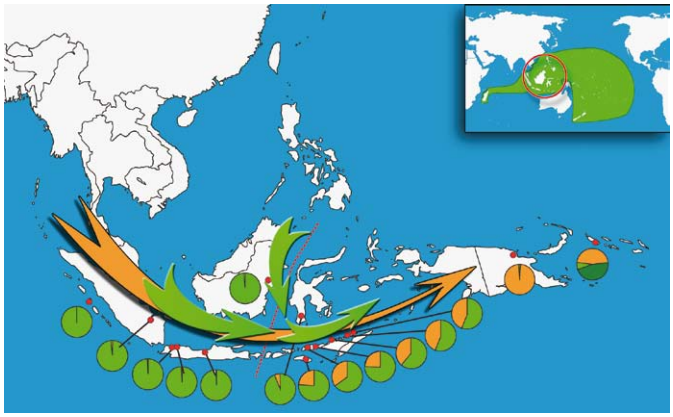

Figure 3: **Human population expansion and admixture in E. Indonesia.** Colored arrows depict eastward migration of Papuan (orange) and Austronesian (green). Red dots on the map are sampling locations. Each circle graph represents a population sample, with colored sectors showing the frequency of inferred genetic components of Papuan (orange) and Asian (green). Red dashed line denotes Wallace's biogeographic line. *PNAS* (2012, 109:4574–4579).

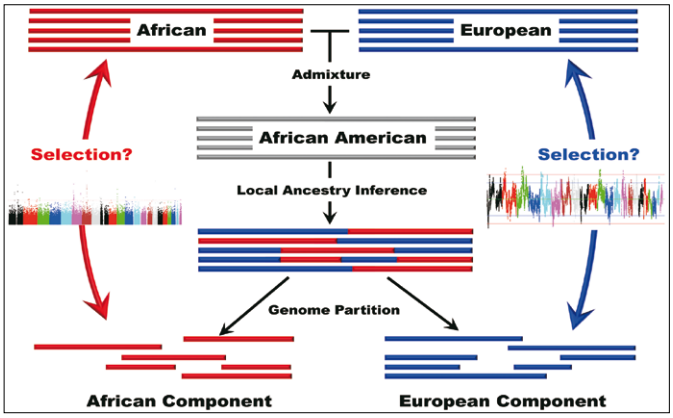

Figure 4: A Schematic of research framework for detecting natural selection signatures in African-American genomes. *Genome Research* (2012, 22: 519–527)

the *PNAS* (2012, 109:4574–4579).

Apart from modeling population admixture and inferring population history, identifying targets of positive selection in humans is another research focus of Dr. XU's group; it has also been a central topic in human evolutionary and population genetics. The difference of their approaches from others is that they search for footprints of natural selection in human genome using admixture analysis. Notably, Dr. XU's group developed a new strategy to detect natural selection both pre- and post-admixture, and they applied this new method to African American data. Interestingly, many selection-candidate genes identified by the new approach were associated with African American specific high-risk diseases, such as prostate cancer and hypertension, suggesting an important role these disease-related genes might have played in adapting to new environment. This work has been published in the *Genome Research* (2012, 22: 519–527).

Inspired by the findings from their analysis of African American data, i.e. many selection-candidate genes identified by the new approach were associated with African American specific high-risk diseases, Dr. XU's group specifically examined the evolutionary characters of human disease-related genes. They classified human genes into five categories: Mendelian and complex diseases (MC) genes, Mendelian but not complex diseases (MNC) genes, complex but not Mendelian diseases (CNM) genes, essential genes and OTHER genes. The most interesting finding from this analysis was that evidence from different aspects supported the hypothesis that MC genes, *i.e.* those genes associated with both Mendelian and complex diseases, have been subjected to both purifying and positive selection in different time of human history. These findings really advanced our understanding of the evolutionary mechanism of human diseases and microevolution of complex traits. The work has been published in the *Human Molecular Genetics* (2012, 21:1611–1624).

This paper was selected (marked with an "Exceptional Level" label) by the Faculty of 1000 (F1000), which placed this work in the library of the top 2% of published papers. As Dr. Thomas Mitchell-Olds from Duke University commented on this paper, "I found the results of this paper to be surprising and remarkable, both for human health and for functional evolutionary genetics"; "This paper also provides evidence for a continuum between Mendelian and complex disease – as our knowledge and study populations expand, we may see increasing recognition that Mendelian + complex genetic architecture is widespread, in humans and other organisms."

Dr. XU's group consists of researchers from around the country and outside. Currently, there are two Associated Professors (with experience of postdoctoral research at Max-Planck Institute in Germany and Ph.D. of La Trobe University in Australia), two Research Associates, two Postdoctoral researchers (from India and Pakistan), one Research Assistant, and 13 Ph.D. and Master students. Dr. XU is the recipient of a number of awards: Distinguished Young Scientist of Chinese Academy of Sciences (CAS); First Class Prize of Natural Science Award, Ministry of Education, China; Thesis Advisor for 2012 SIBS-Eli Lilly Outstanding Graduate Thesis Award Winner (US). In 2012, Dr. XU was appointed as adjunct professor by ShanghaiTech University. In 2013, he was accepted as "The National Top-notch Young Innovative Talent" by The "Ten-Thousand-Talents" Project, funded by the Organization Department of the Central Committee of the CPC. ◂