

# Genome sequencing and analysis of admixed genomes of African and Mexican ancestry: implications for personal ancestry reconstruction and multi-ethnic medical genomics

Francisco M De La Vega<sup>1\*</sup>, Katarzyna Bryc<sup>2</sup>, Jeremiah D Degehardt<sup>2</sup>, Shaila Musharoff<sup>2</sup>, Jeffrey M Kidd<sup>3</sup>, Vrunda Seth<sup>4</sup>, Sarah Stanley<sup>1</sup>, Abra Brisbin<sup>2</sup>, Alon Keinan<sup>2</sup>, Andrew Clark<sup>2</sup>, Carlos D Bustamante<sup>2,3</sup>

*From* Beyond the Genome: The true gene count, human evolution and disease genomics  
Boston, MA, USA. 11-13 October 2010

## Background

Understanding the contribution of rare and common genetic variants to disease susceptibility will probably require multi- and trans- ethnic sequencing studies that compare the genomes of many individuals with and without a particular disease. Accounting for the role of population stratification at fine scales, both in terms of genomic and geographic location, will be important because rare alleles are likely to show more population stratification. Here, we present results from sequencing, assembly and genomic analysis of two genomes from the Phase 3 HapMap. The donor individuals are of Mexican and African ancestry and represent the first 'admixed' genomes to be sequenced to high coverage.

## Materials and methods

DNA was obtained from the Coriell Institute. We sequenced 2 x 25bp and 2 x 50bp mate-paired libraries with the Applied Biosystems SOLiD™ System. The total average haploid coverage achieved was ~20X. Mapping and SNP calling was performed with the SOLiD analysis pipeline. We developed methods to estimate 'admixture breakpoints' along the admixed genomes.

## Results

We used the distribution of admixture breakpoints to infer the personal admixture history of the samples and patterns of genomic diversity to reconstruct the demographic history of European, African and Native American continental populations. We compared the

distribution of functional and putatively neutral genetic variation among 12 sequenced genomes and found that difference in demographic history might account for statistically significant, differences in distributions of synonymous versus benign, possibly damaging, and probably damaging non-synonymous coding variants. Finally, we used the admixed genomes data together with 1000 Genomes to quantify the relative proportions of private, rare and common functional and neutral alleles within and among populations.

## Conclusions

Genomic sequencing provides finer resolution of admixture breakpoints based on allele frequency estimates from HapMap and the 1000 Genomes Projects. Our results suggest that there are many new variants to be discovered in populations from African and the American descent, some of potential disease relevance.

## Author details

<sup>1</sup>Genetic Systems R&D, Life Technologies, Foster City, CA 94404, USA. <sup>2</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. <sup>3</sup>Department of Genetics, Stanford University, Stanford, CA 94305, USA. <sup>4</sup>Genetic Systems R&D, Life Technologies, Beverly, MA 01915, USA.

Published: 11 October 2010

doi:10.1186/gb-2010-11-S1-O4

**Cite this article as:** De La Vega et al.: Genome sequencing and analysis of admixed genomes of African and Mexican ancestry: implications for personal ancestry reconstruction and multi-ethnic medical genomics. *Genome Biology* 2010 **11**(Suppl 1):O4.

<sup>1</sup>Genetic Systems R&D, Life Technologies, Foster City, CA 94404, USA  
Full list of author information is available at the end of the article