

## Deciphering signals of positive selection in Peruvian populations from three ecoregions

Rocio Caro-Consuegra<sup>1</sup>, Maria A. Nieves-Colón<sup>2,3</sup>, Erin Rawls<sup>3</sup>, Alexandra Obregón-Tito<sup>4</sup>, Raul Tito<sup>4</sup>, Cecil Lewis<sup>4</sup>, Karla Sandoval-Mendoza<sup>2</sup>, Carlos D. Bustamante<sup>5,6</sup>, Genevieve L. Wojcik<sup>7</sup>, Chris Gignoux<sup>8</sup>, Julie Baker<sup>6</sup>, Laura Fejerman<sup>9</sup>, Tatiana Vidaurre<sup>10</sup>, Beatriz Lizárraga<sup>11</sup>, Verónica Rubin de Celis<sup>12</sup>, Anne C. Stone<sup>3</sup>, Andrés Moreno-Estrada<sup>2</sup>, Elena Bosch<sup>1</sup>.

1 Institute of Evolutionary Biology (UPF-CSIC), Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Catalonia, Spain. 2 Laboratorio Nacional de Genómica para la Biodiversidad, Unidad de Genómica Avanzada (LANGEBIO-CINESTAV), Irapuato, Guanajuato, MX. 3 School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona. 4 Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK. 5 Department of Biomedical Data Science, Stanford University, Stanford, CA. 6 Department of Genetics, Stanford University, Stanford, CA. 7 Department of Genetic Epidemiology, Johns Hopkins University, Baltimore, MD. 8 Colorado Center for Personalized Medicine, University of Colorado, CO. 9 Department of Medicine, University of California San Francisco, San Francisco, CA. 10 Instituto de Enfermedades Neoplásicas, Lima, Perú. 11 Emeritus Professor, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Perú. 12 Laboratorio de Genómica Molecular Evolutiva, Instituto de Ciencia y Tecnología, Universidad Ricardo Palma, Lima, Perú.

**Abstract:** Perú hosts three extremely diverse ecoregions: The Pacific coast desert, the Andean highlands, and the Amazonian rainforest. Multiple analyses have already identified potential candidate genes for human adaptation to hypobaric hypoxia in highlands, but selection in the Peruvian coast and rainforest remains unexplored.

Genome-wide SNP data (Illumina Infinium<sup>®</sup> MEGA) from 198 individuals distributed across Perú were used to identify signals of recent positive selection in each ecoregion. Specifically, we computed population differentiation (PBS) and haplotype-based selection scans (iHS and XP-EHH). Across the top 50 candidate regions identified per scan, we explored for strong selective sweeps, as well as for signatures of polygenic selection using gene-set and SNP-trait enrichment approaches.

Among the top 10 genomic signatures of recent positive selection found in highlands, we replicate some previously known candidates (*TBX5*, *TGFA*). We also identify novel signals related to cardiac function, glucose metabolism and epidermal growth factors. In the coast, we identify genes related to the immune system and to vitamin D synthesis. In the rainforest, we detect genes linked to respiratory functions, immune system, heart development and gametogenesis. Furthermore, we identify multiple common candidate regions among ecoregions, including genes involved in lipid metabolism (*CPT2* and *LRP8*) and the immune system (*DUOX2*, *DUOXA1* and *DUOX1*). Gene-set and trait-associated SNP overrepresentation analyses also yield common terms linked to xenobiotic metabolism and insulin. In addition, highland and coastal populations are also enriched in cardiac muscle contraction and lipid metabolism categories, whereas several immune function terms are enriched in the coast and in the rainforest.

In summary, we find genes that provide adaptations specific to the local ecological conditions in the three ecoregions. However, we also detect shared adaptations among ecoregions that point to the recent common adaptive history of Peruvian populations.