

Ancestry, admixture & selection in Colombian genomes

King Jordan

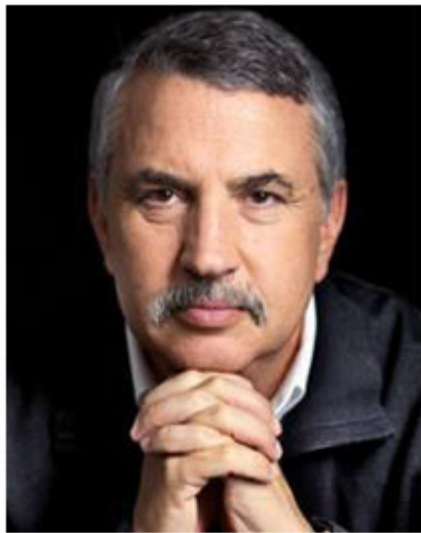
School of Biology, Georgia Tech
BIOS Centro de Bioinformática y Biología Computacional



The New York Times

The Opinion Pages

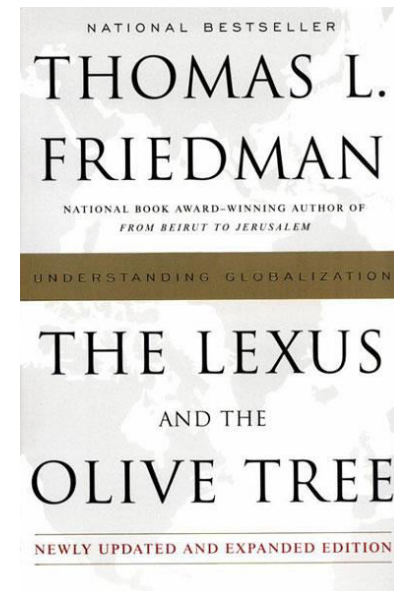
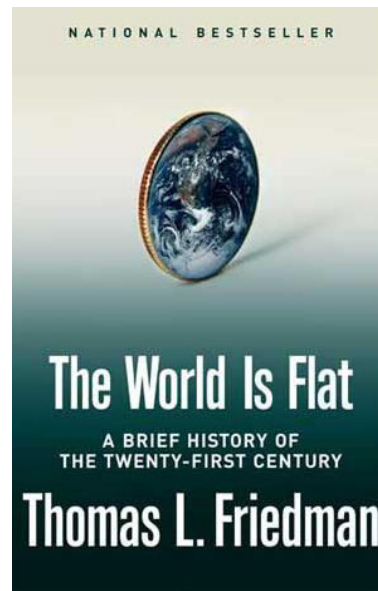
Thomas L. Friedman



Josh Haner/The New York Times

“The biggest trend in the world today is ... the **merger of globalization and the information technology revolution.**”

New York Times Feb 25, 2014

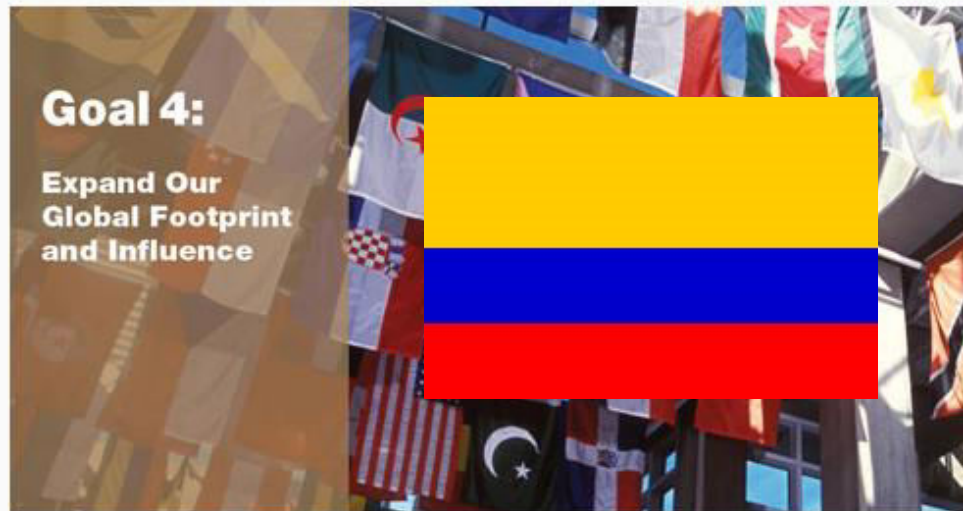


Strategic Vision Home

Implementation

The Planning Process

A STRATEGIC VISION FOR GEORGIA TECH



THE STRATEGIC PLAN

[Download the Strategic Plan](#)
(9MB Adobe PDF file)

[Download the Strategic Plan Goals and Institutional Initiatives Brochure](#)
(987KB Adobe PDF file)

STRATEGIC PLAN UPDATES

Review how Georgia Tech is acting on the goals outlined in the strategic plan.

[Learn more >](#)

1. Expand the world's footprint at Georgia Tech
2. Extend and leverage Georgia Tech's impact around the world
3. Embrace and support globally trained students

We invite you to:

- Review [the Plan](#) (PDF) & [Institutional Initiatives](#) (PDF)

THE LATEST NEWS

[Planning for 'X Degree' Moves Forward](#)

October 2, 2011

[Strategic Plan Steering Committee Forms](#)

August 21, 2011

[Technology and Law Task Force Makes Progress](#)

August 7, 2011

[Faculty Invited to Participate in First Day Initiative](#)

August 7, 2011

[More News >](#)



STRATEGIC VISION GOALS

Be Among the Most Highly Respected Technology-Focused

PanAmerican Bioinformatics Institute

Leveraging biotechnology for public health and economic development in the Americas

HOME

MISSION

FACULTY

STUDENTS

PROJECTS

PUBLICATIONS

COURSES

STUDENT EXCHANGES



December 2009 PABI is born in Santa Marta, Colombia



The PABI mission

To facilitate public health and economic development in the Americas through education, research and outreach in bioinformatics and genomics.

[Learn more about our mission](#)

We are

A Panamerican network of bioinformatics and genomics researchers committed to the use of science in service of humanity. PABI was originally founded as a collaboration between Universidad Nacional de Colombia and Georgia Tech in the USA.

[Learn more about our members](#)

Activities

PABI members engage in collaborative research efforts in bioinformatics and genomics, workshops and short courses for students and faculty, and student exchanges between host institutions in Latin America and the USA.

[Learn more about our work](#)

FULBRIGHT

COLOMBIA

mentes que **mueven** el mundo





The mission of our Genome Research Center is to apply education, training and research in genome sciences towards the development of a knowledge-based economy in Colombia.



PARTICIPANTS





CENTRO DE BIOINFORMÁTICA
Y BIOLOGÍA COMPUTACIONAL

Conocimiento e innovación para el desarrollo



Georgia Tech & BIOS Project: Human Clinical and Population Genomics

- Genomic applications to health & medicine hold great promise
- Most clinical genomic studies are based on US & European populations
- Need regional (country)-specific surveys of Latin American populations
- Need to develop local capacity in genome sequence analysis & bioinformatics
- Research, development & training in human genomics & bioinformatics

People of Colombia



Colombian demography & admixture

<i>Colombian Demographics</i>	
<i>Ethnic Group</i>	<i>Percent</i>
Mestizo (Amerindian and European)	58%
White (European)	20%
Mulatto (Black/African and European)	14%
Black (African)	4%
Zambo (Amerindian and Black/African)	3%
Amerindian	1%

- The Colombian population is highly admixed
- Ancestry from Africa, the Americas and Europe

Strong Amerind/White Sex Bias and a Possible Sephardic Contribution among the Founders of a Population in Northwest Colombia

Luis G. Carvajal-Carmona,^{1,4} Iván D. Soto,¹ Nicolás Pineda,¹ Daniel Ortíz-Barrientos,¹ Constanza Duque,¹ Jorge Ospina-Duque,² Mark McCarthy,⁵ Patricia Montoya,^{1,3} Victor M. Alvarez,³ Gabriel Bedoya,¹ and Andrés Ruiz-Linares^{1,4}

OPEN ACCESS Freely available online

PLoS GENETICS

Geographic Patterns of Genome Admixture in Latin American Mestizos

Sijia Wang¹, Nicolas Ray², Winston Rojas³, Maria V. Parra³, Gabriel Bedoya³, Carla Gallo⁴, Giovanni Poletti⁴, Guido Mazzotti⁵, Kim Hill⁶, Ana M. Hurtado⁶, Beatriz Camrena⁷, Humberto Nicolini⁷, William Klitz^{8,9}, Ramiro Barrantes¹⁰, Julio A. Molina¹¹, Nelson B. Freimer¹¹, Maria Cátira Bortolini¹², Francisco M. Salzano¹², Maria L. Petzl-Erler¹³, Luiza T. Tsuneto¹³, José E. Dipierri¹⁴, Emma L. Alfaro¹⁴, Graciela Bailliet¹⁵, Nestor O. Bianchi¹⁵, Elena Llop¹⁶, Francisco Rothhammer^{16,17}, Laurent Excoffier², Andrés Ruiz-Linares^{1*}

Hum Genet (2003) 112: 534–541
DOI 10.1007/s00439-002-0899-8

ORIGINAL INVESTIGATION

Luis G. Carvajal-Carmona · Roel Ophoff
Susan Service · Jaana Hartiala · Julio Molina
Pedro Leon · Jorge Ospina · Gabriel Bedoya
Nelson Freimer · Andrés Ruiz-Linares

Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica

Genome-wide patterns of population structure and admixture among Hispanic/Latino populations

Katarzyna Bryc^{a,1}, Christopher Velez^{b,1}, Tatiana Karafet^c, Andres Moreno-Estrada^{a,d}, Andy Reynolds^a, Adam Auton^{a,2}, Michael Hammer^c, Carlos D. Bustamante^{a,d,3,4}, and Harry Ostre^{b,3,4}

Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate

Gabriel Bedoya¹, Patricia Montoya², Jenny Garcia³, Ivan Soto¹, Stephane Bourgeois⁵, Luis Carvajal, Damian Labuda⁶, Victor Alvarez², Jorge Ospina⁵, Phillip W. Hedrick^{1†}, and Andrés Ruiz-Linares^{1†*}

COMPOSICIÓN GENÉTICA DE UNA POBLACIÓN

del suroccidente de Colombia

LILIANA CÓRDOBA^A (lilianac27@yahoo.com),

JHARLEY JAIR GARCÍA^A, LUZ STELLA HOYOS^B, CONSTANZA DUQUE^A,

WINSTON ROJAS^A, SILVIO CARVAJAL^B, LUISA F. ESCOBAR^{B,C}, INGRID REYES^B,
NOHELIA CAJAS^B, ADALBERTO SÁNCHEZ^D, FELIPE GARCÍA^D, GABRIEL BEDOYA^A,

ANDRÉS RUIZ-LINARES^{A,E}

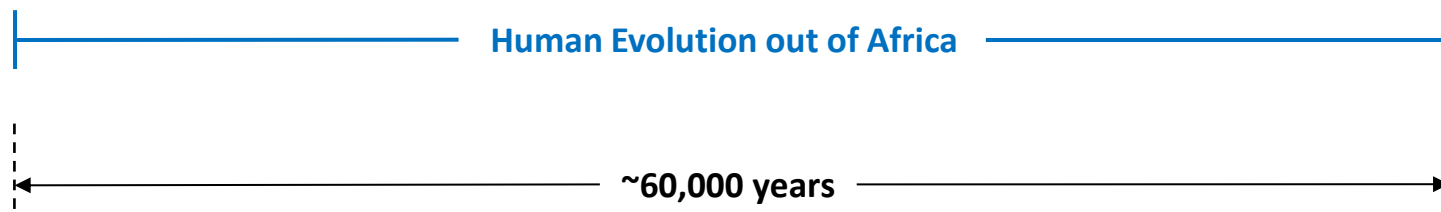
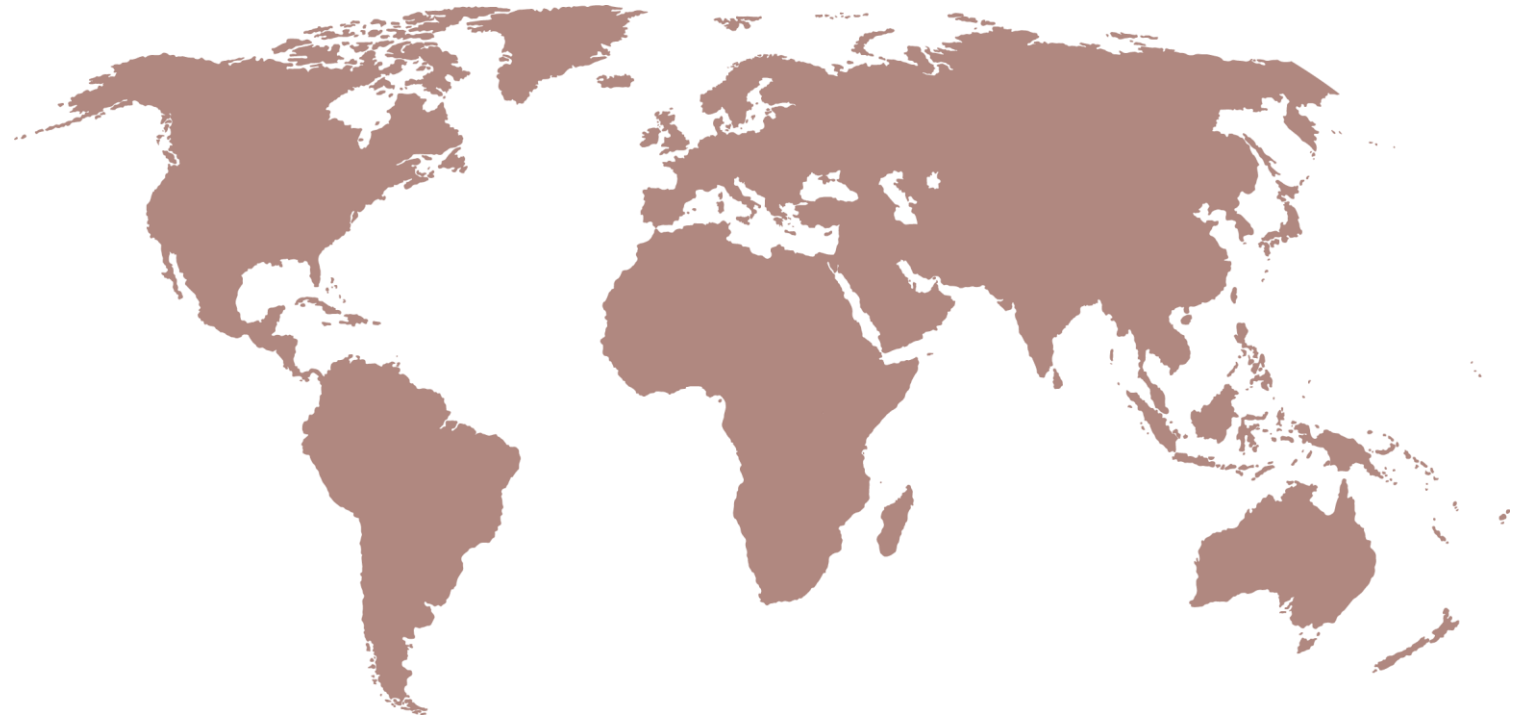
Revista Colombiana de Antropología

Volumen 48 (1), enero-junio 2012, pp. 21-48

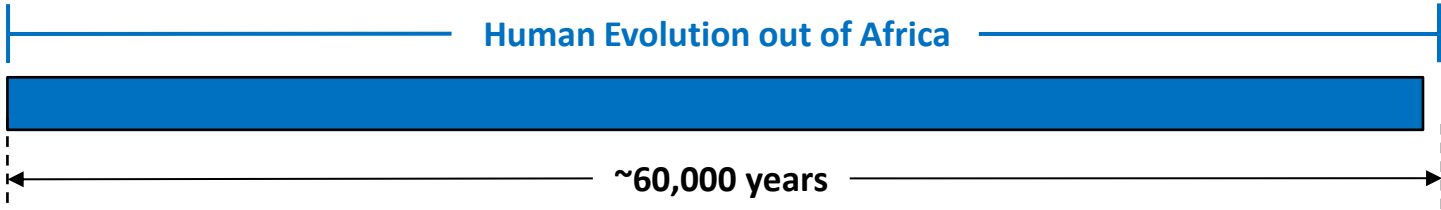
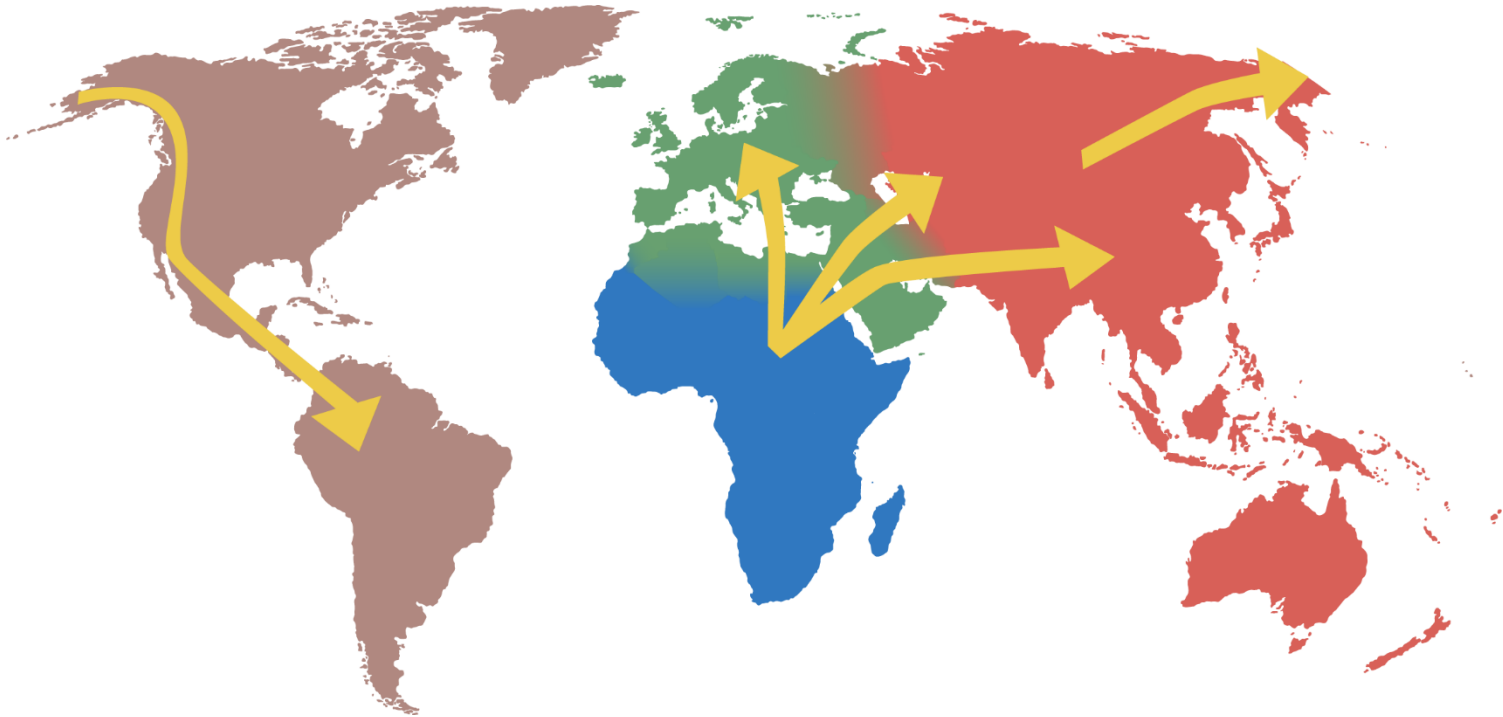
Outline & Approach

- Characterize ancestry & admixture patterns of Colombian genomes
 - Continental & sub-continental levels
 - Genome-wide & locus-specific patterns
 - Overall & sex-specific admixture
- Relate admixture patterns to health & natural selection
 - Admixture enrichment analysis
 - Interrogation of anomalously admixed genomic regions

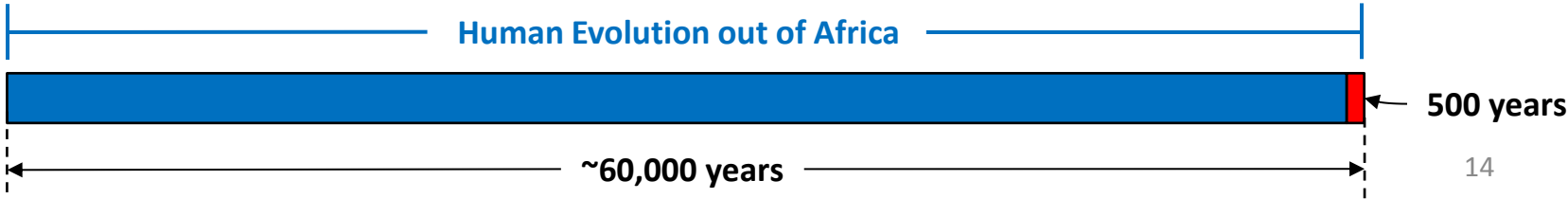
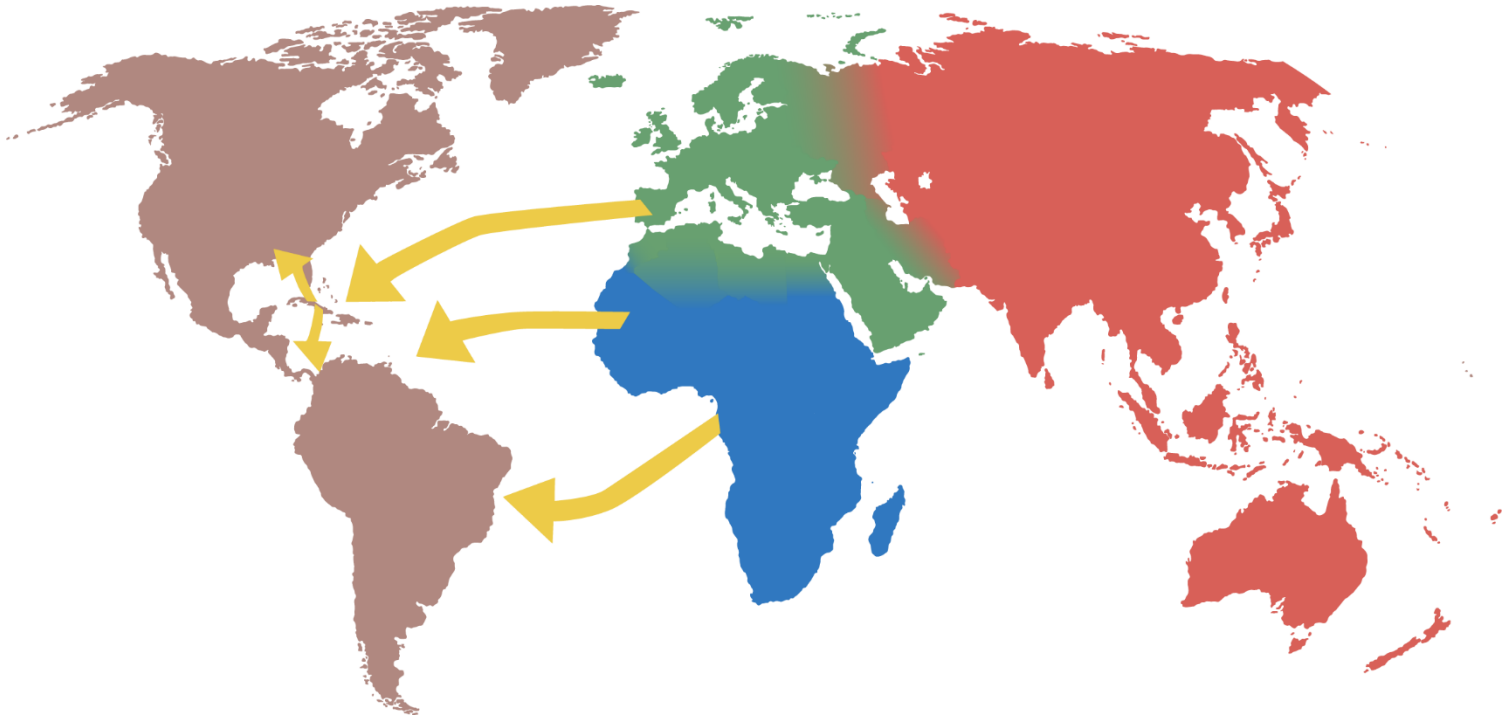
Human evolution: diversification & isolation (99%)



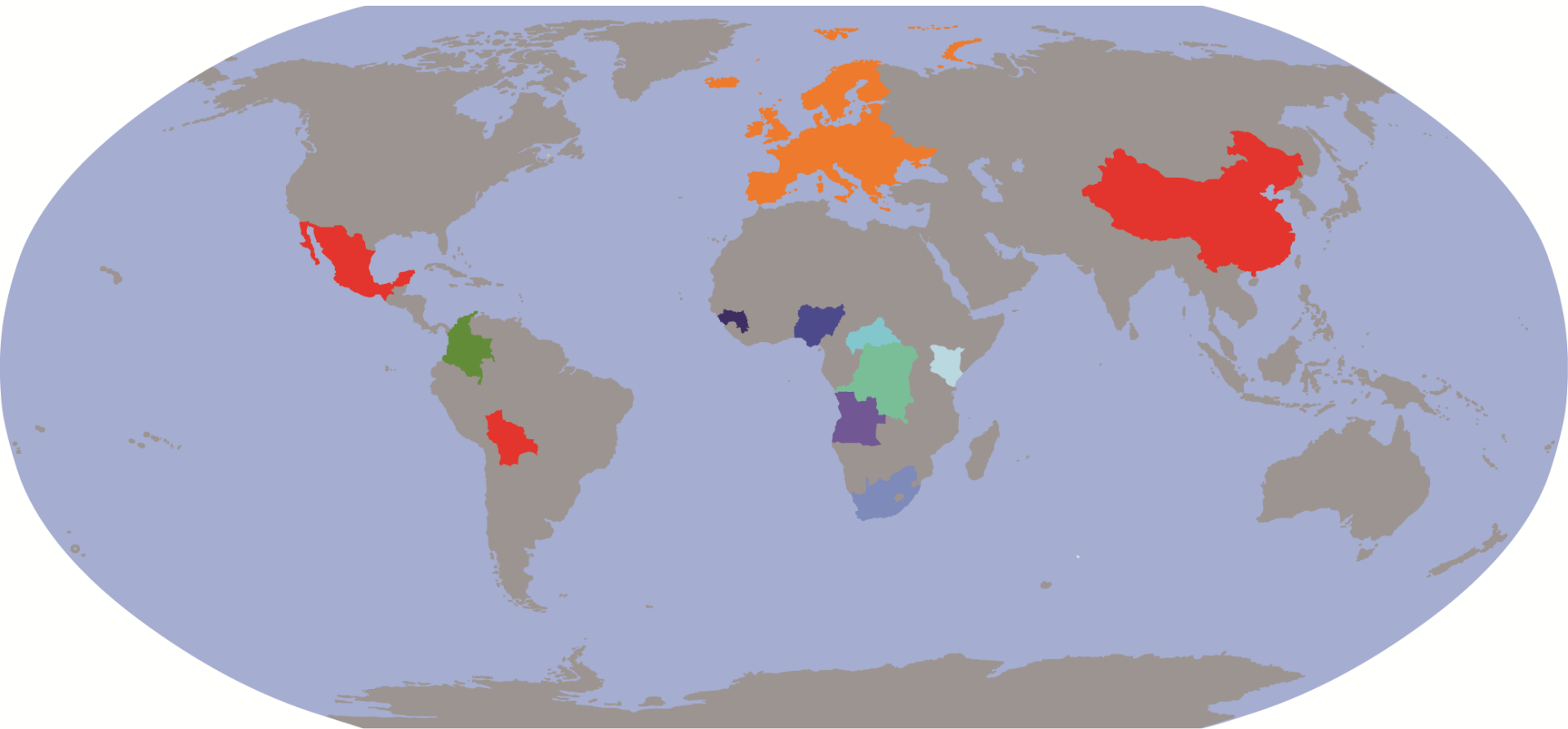
Human evolution: diversification & isolation (99%)



Human evolution: convergence & admixture (<1%)



Genome sequences sampled for this study



Ancestral Populations from Around the World

Name	Ancestral Origin	n ¹	Source (PMID) ²	Platform ³
Bantu	Kenya	11	18292342	SNP microarray
Bantu	South Africa	8	18292342	SNP microarray
Biaka Pygmy	Central African Republic	22	18292342	SNP microarray
CEU	Northern & Western Europe	85	3498066	Whole genome sequencing
CHB	Beijing, China	92	3498066	Whole genome sequencing
Mandenka	Senegal	22	18292342	SNP microarray
Mbuti Pygmy	Congo	13	18292342	SNP microarray
San	Namibia	5	18292342	SNP microarray
Totonac & Bolivian	Mexico & Bolivia	45	3432609	SNP microarray
Yoruba	Nigeria	22	18292342	SNP microarray
YRI	Nigeria	87	3498066	Whole genome sequencing

Admixed Populations from the Americas

Name	Country	n ¹	Source (PMID) ²	Platform ³
CLM	Medellin, Colombia	60	3498066	Whole genome sequencing
Colombian	Colombia	26	20445096	Illumina 610K array
Dominican Republican	Dominican Republic	27	20445096	Illumina 610K array
Ecuador	Ecuador	20	20445096	Illumina 610K array
Mexican	Mexico	112	20445096	Affymetrix GeneChip 500K Array Set
Puerto Rican	Puerto Rico	27	20445096	Illumina 610K array

mtDNA and Y DNA Haplotype

Dataset	Ancestral Origin	n ¹	Source (PMID) ²	Platform ³
mtDNA	Gambia, Senegal, Guinea Bissau, Guinea, Sierra Leone	844	21253579	Mixed
mtDNA	Ghana, Togo, Benin, Nigeria, Cameroon, Sao Tome & Principe	1310	21253579	Mixed
mtDNA	Congo, DRC	204	21253579	Mixed
Y DNA	Gambia, Senegal, Guinea Bissau, Guinea, Sierra Leone	388	21253579	Mixed
Y DNA	Ghana, Togo, Benin, Nigeria, Cameroon, Sao Tome & Principe	755	21253579	Mixed
Y DNA	Congo, DRC	178	21253579	Mixed
mtDNA & Y DNA	Colombia	26	20445096	Illumina 610K array
mtDNA & Y DNA	Dominican Republic	27	20445096	Illumina 610K array
mtDNA & Y DNA	Ecuador	20	20445096	Illumina 610K array
mtDNA & Y DNA	Mexico	112	20445096	Affymetrix GeneChip 500K Array Set
mtDNA & Y DNA	Puerto Rico	27	20445096	Illumina 610K array

Human Population Genomics = Big Data

Data	# of Genomes	# of Bases
Ancestral Genomes	472	$\times 3.0 \times 10^8 = 1.4 \times 10^{11}$
Admixed Genomes	212	$\times 3.0 \times 10^8 = 6.4 \times 10^{10}$
Mitochondrial DNA	2570	$\times 1.6 \times 10^4 = 4.1 \times 10^7$
Y DNA	1530	$\times 5.0 \times 10^7 = 7.7 \times 10^{10}$
Total	4787	2.8×10^{11}*

*roughly equal to the number of stars in our Galaxy

Genome ancestry & admixture analysis

rs2814778

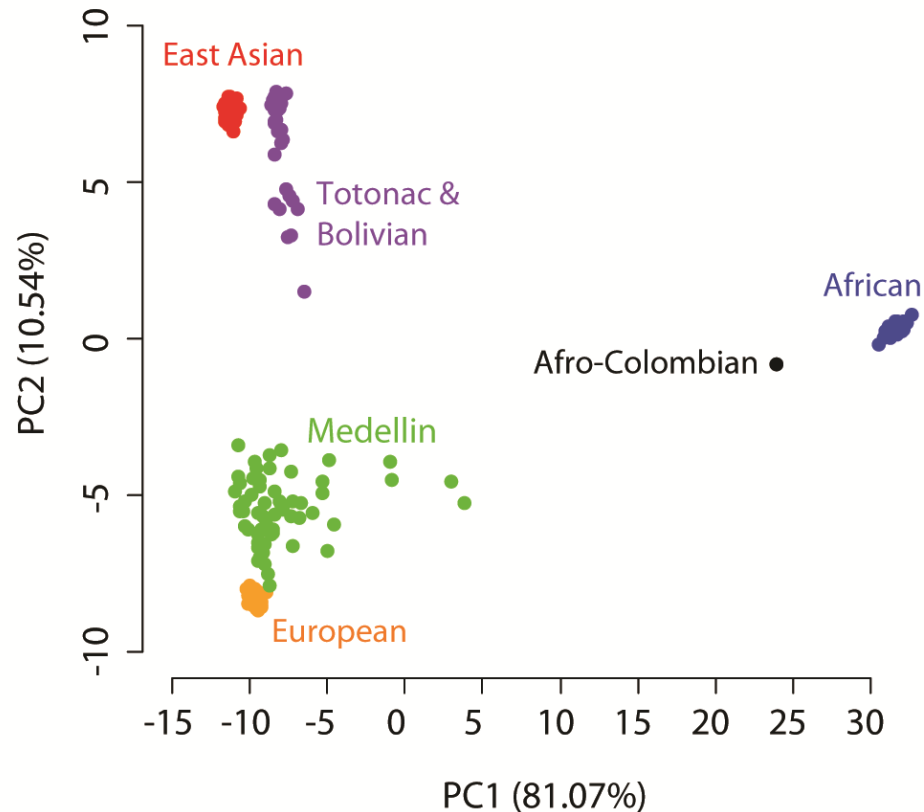
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**C**CTTGGA
● CTCTTA**C**CTTGGA
● CTCTTA**C**CTTGGA

- rs2814778 SNP segregates **African** from **Amerindian/European** populations
- **C** allele fixed in **African** virtually absent in **Amerindian/European** (< 1%)
- presence of **C** allele indicates **African** ancestry at the locus
- process repeated over millions of SNPs

Compare patterns of SNP variation of admixed **Colombians** with ancestral **African**, **Amerindian** and **European** populations

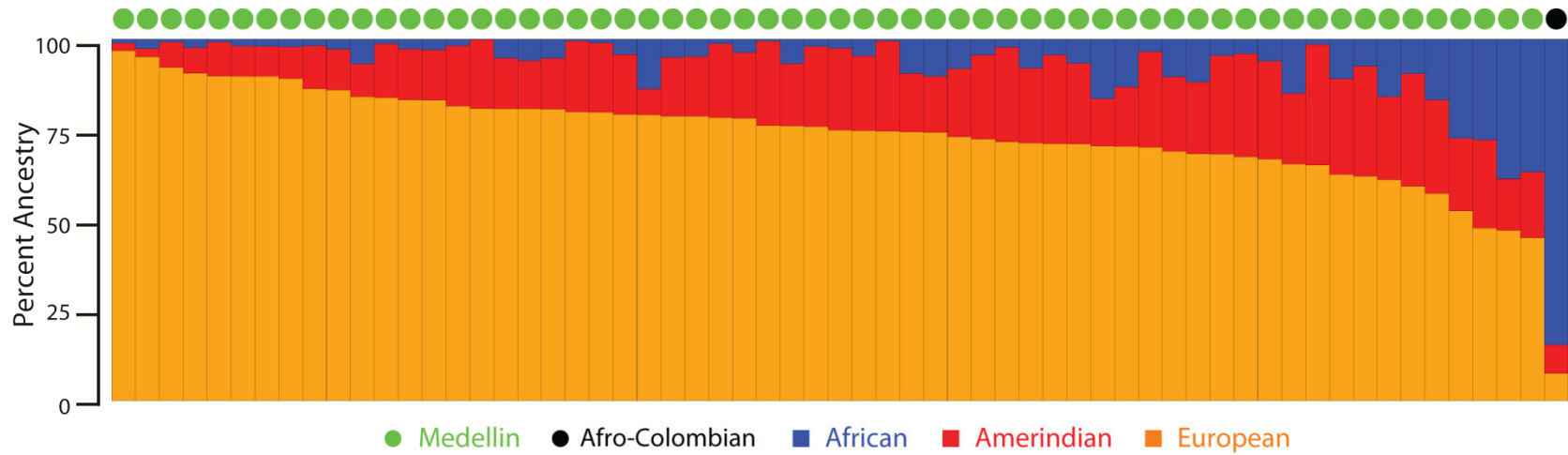
CONTINENTAL ANCESTRY & ADMIXTURE ANALYSIS

Colombian genome-wide ancestry & admixture



Colombian SNP variants were compared to **930,854** genotyped variants from 87 **African**, 47 **Amerindian** & 85 **European** genomes

Colombian genome-wide ancestry & admixture



SUB-CONTINENTAL ANCESTRY & ADMIXTURE ANALYSIS

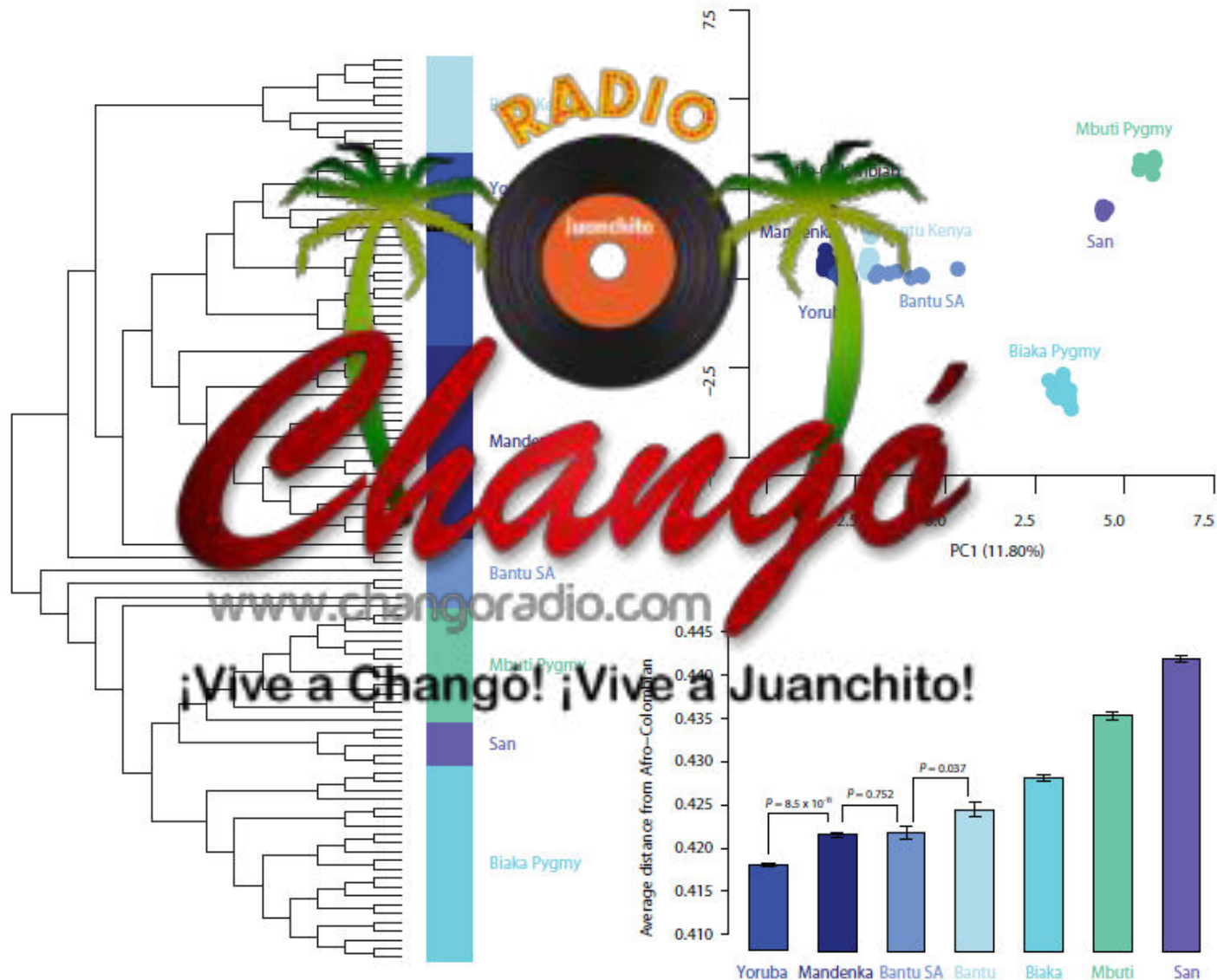
Sub-continental origins of an Afro-Colombian genome



Geographic Origin	Coordinates	Tribe	No. LCLs	No. Males
Central African Republic	4N, 17E	Biaka Pygmy relatives	36	33
Democratic Republic of Congo	1N, 29E	Mbuti Pygmy relatives	15	13
Senegal	12N, 12W	Mandenka relatives	24	16
Nigeria	6-10N, 2-8E	Yoruba relatives	25	13
Namibia	21S, 20E	San relatives	7	7
Kenya	3S, 37E	Bantu NE relatives	12	11
Bantu Speakers S. Africa			8	8
<i>S. Africa Bantu S.E.</i>	29S, 30E	Bantu S.E. Pedi	1	1
<i>S. Africa Bantu S.E.</i>	29S, 29E	Bantu S.E. Sotho	1	1
<i>S. Africa Bantu S.E.</i>	28S, 24E	Bantu S.E. Tswana	2	2
<i>S. Africa Bantu S.E.</i>	28S, 31E	Bantu S.E. Zulu	1	1
<i>S. Africa Bantu S.W.</i>	22S, 19E	Bantu S.W. Herero	2	2
<i>S. Africa Bantu S.W.</i>	19S, 18E	Bantu S.W. Ovambo	1	1
SUBSAHARAN AFRICA			127	109

Source: HGDP-CEPH Project - <http://www.cephb.fr/en/hgdp/table.php>

Sub-continental origins of an Afro-Colombian genome



SEX-SPECIFIC ANCESTRY & ADMIXTURE ANALYSIS

Afro-Colombian sex-specific admixture



Colombia has the most pronounced sex-specific admixture
Patterns in Latin America

LOCUS-SPECIFIC ANCESTRY & ADMIXTURE ANALYSIS

Genome ancestry & admixture analysis

rs2814778

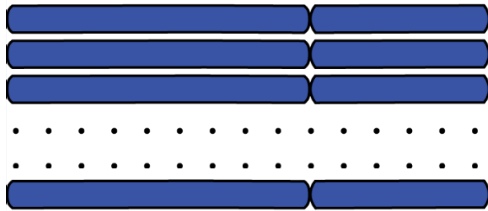
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**T**CTTGGA
● CTCTTA**C**CTTGGA
● CTCTTA**C**CTTGGA
● CTCTTA**C**CTTGGA

- rs2814778 SNP segregates **African** from **Amerindian/European** populations
- **C** allele fixed in **African** virtually absent in **Amerindian/European** (< 1%)
- presence of **C** allele indicates **African** ancestry at the locus
- process repeated over millions of SNPs

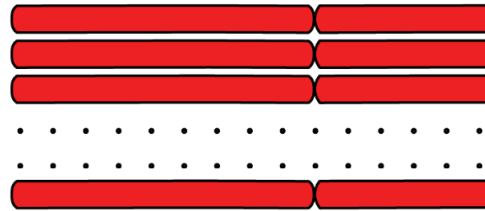
Compare patterns of SNP variation of admixed **Colombians** with ancestral **African**, **Amerindian** and **European** populations

Locus-specific admixture analysis (chromosome painting)

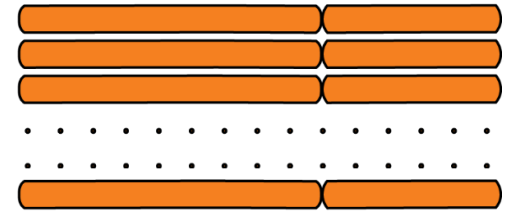
African (YRI, n = 87)



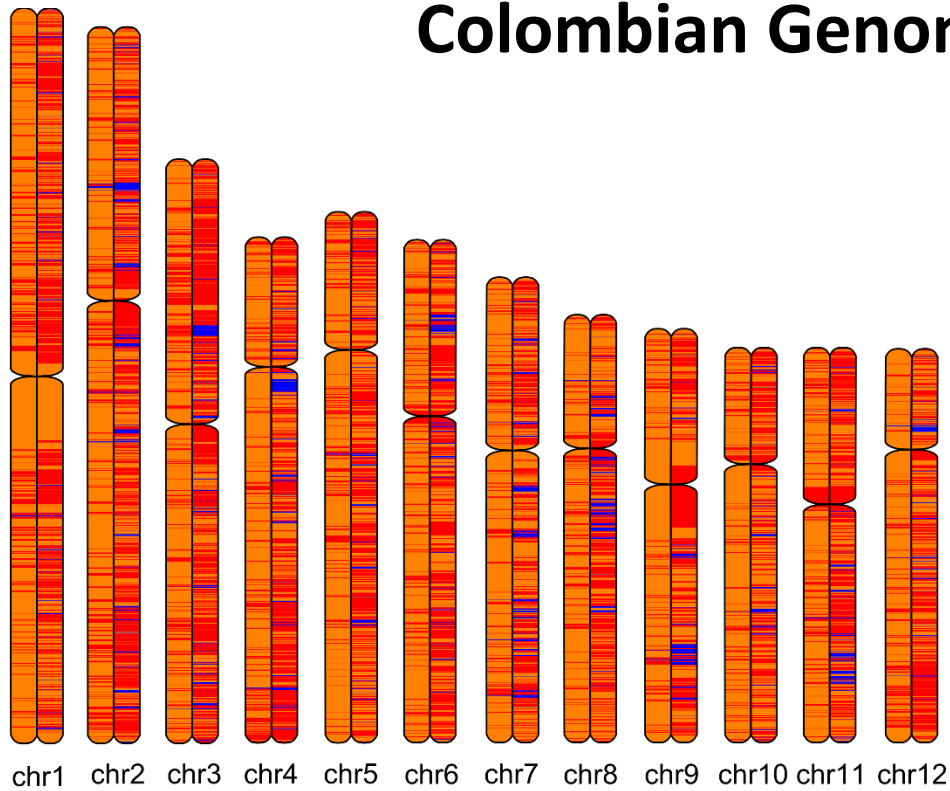
Amerindian (CHB, n = 92)



European (CEU, n = 85)

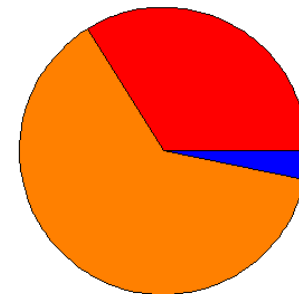
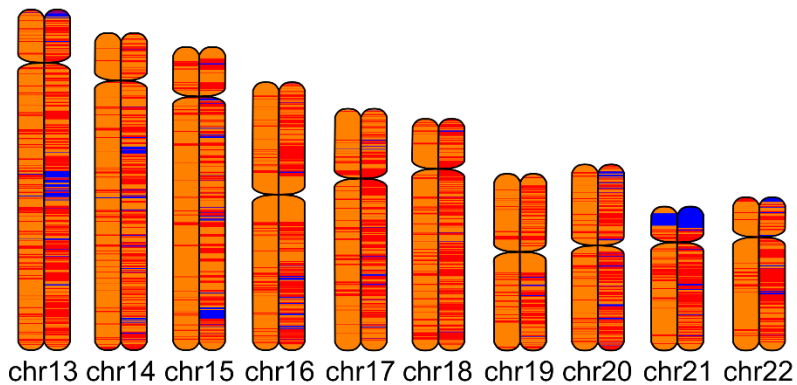
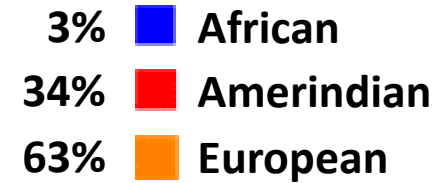


Colombian Genomic Ancestry

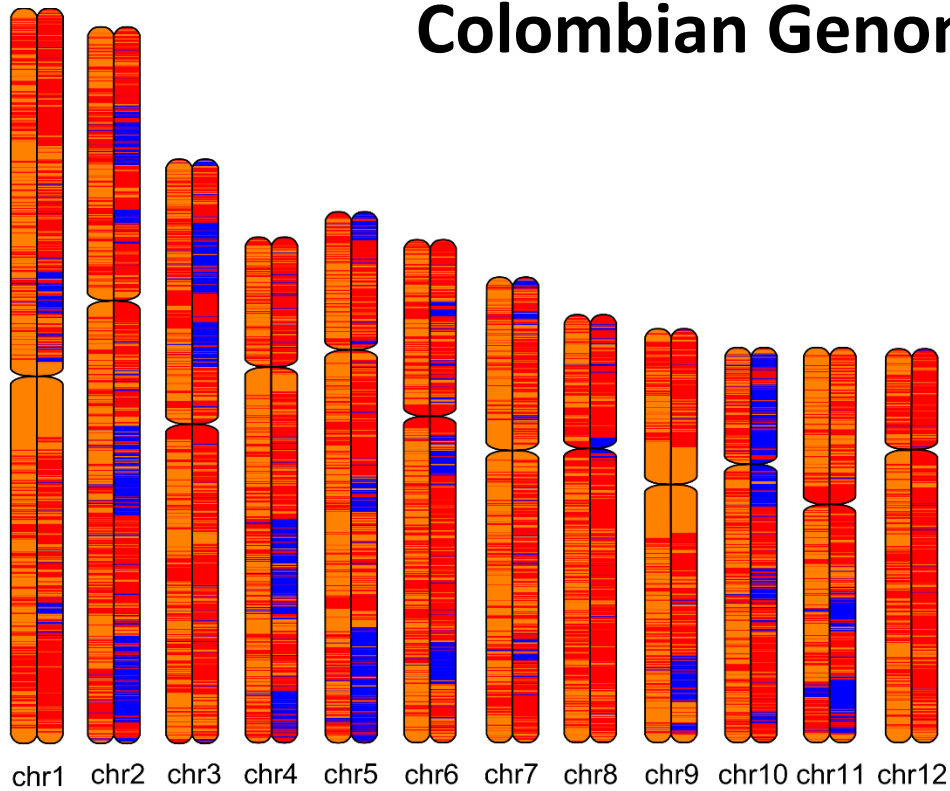


Individual #: 43

Admixture Percentages

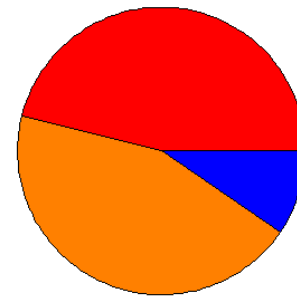
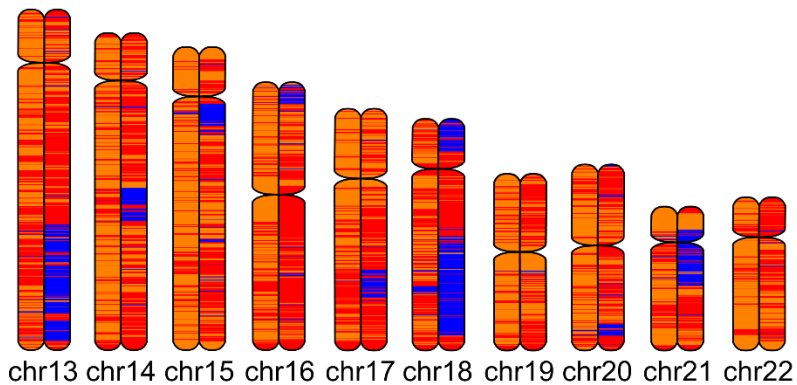
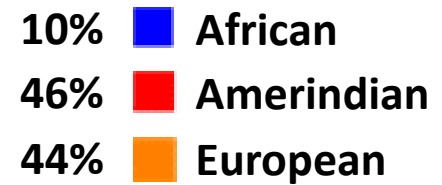


Colombian Genomic Ancestry

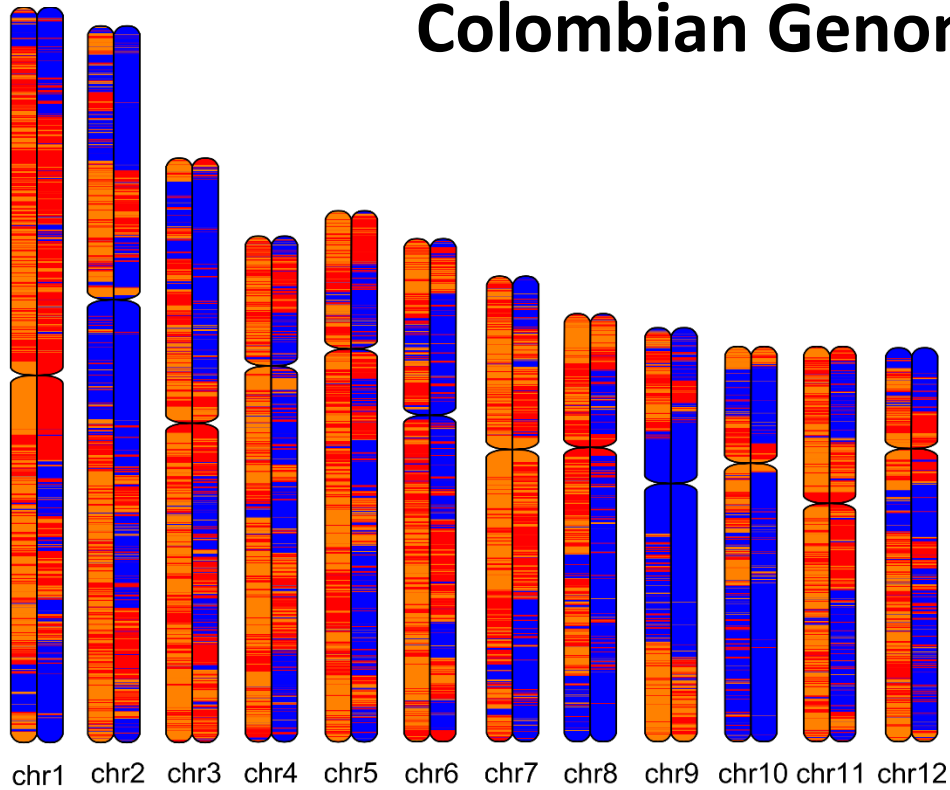


Individual #: 11

Admixture Percentages

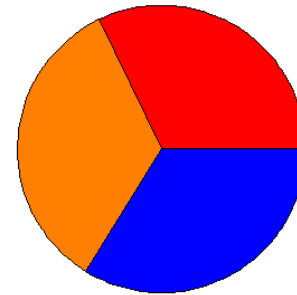
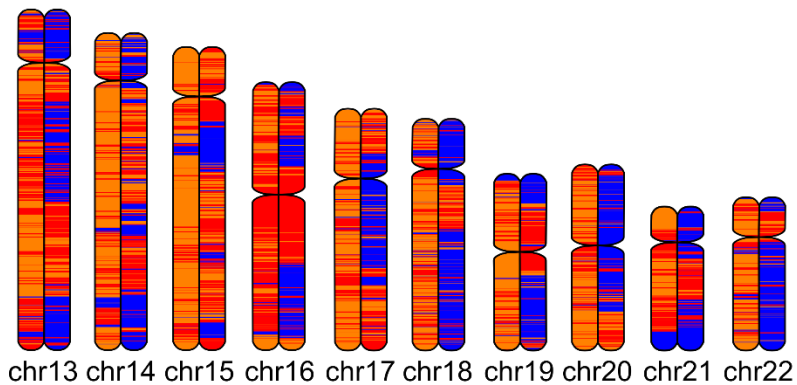
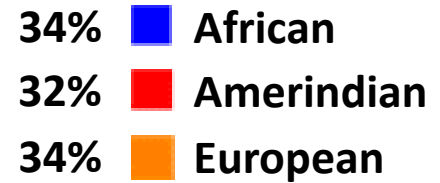


Colombian Genomic Ancestry

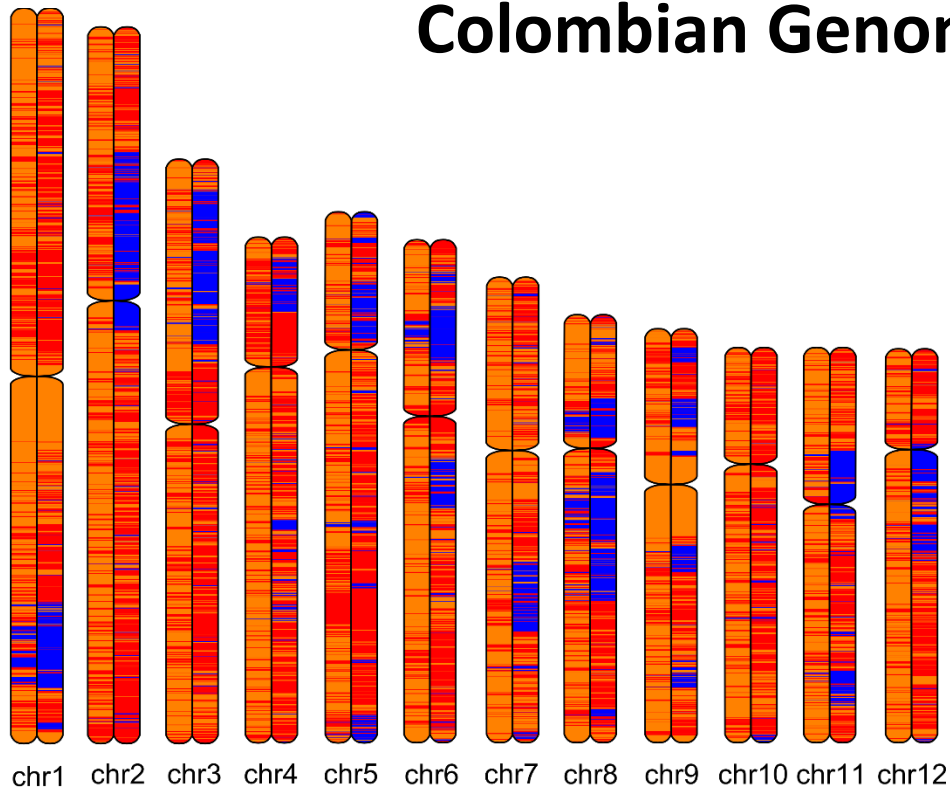


Individual #: 60

Admixture Percentages

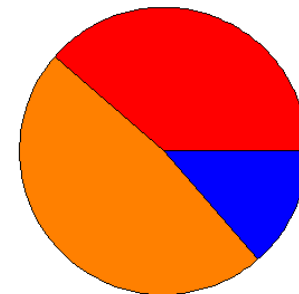
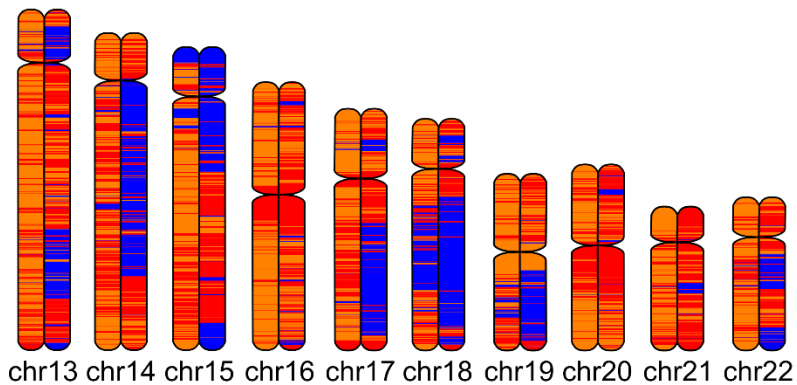
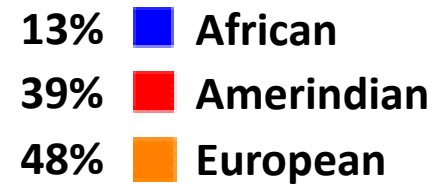


Colombian Genomic Ancestry

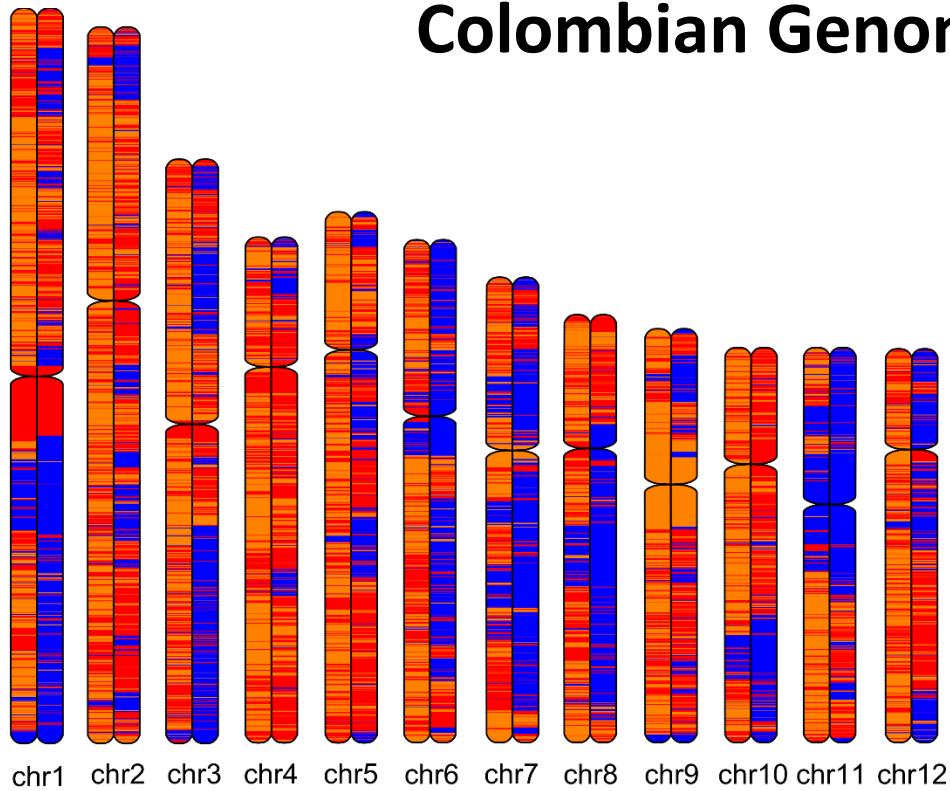


Individual #: 36

Admixture Percentages

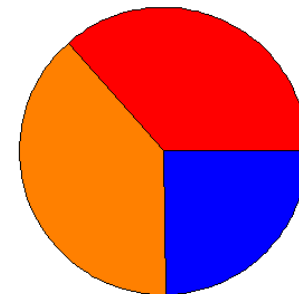
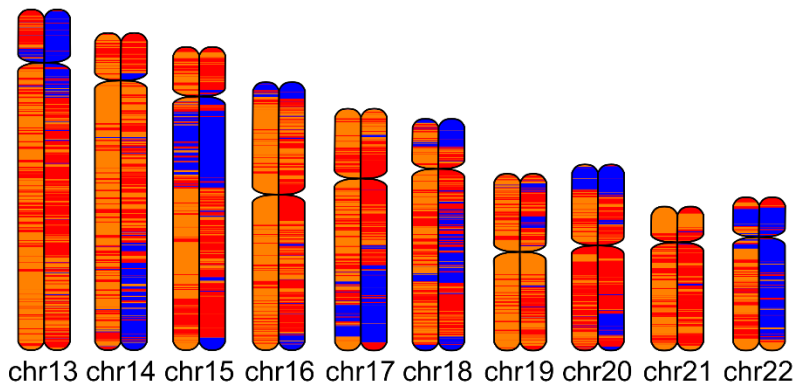
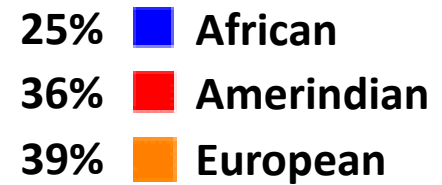


Colombian Genomic Ancestry



Individual #: 42

Admixture Percentages



ADMIXTURE & SELECTION IN COLOMBIAN GENOMES

Human evolution: convergence & admixture (<1%)

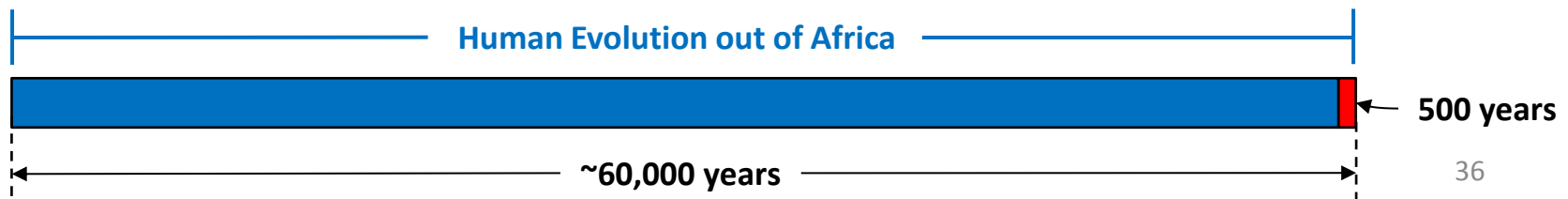
What does it mean when genomes separated for >60,000 are suddenly brought back together?

Is the process of admixture somehow related to health & selection?

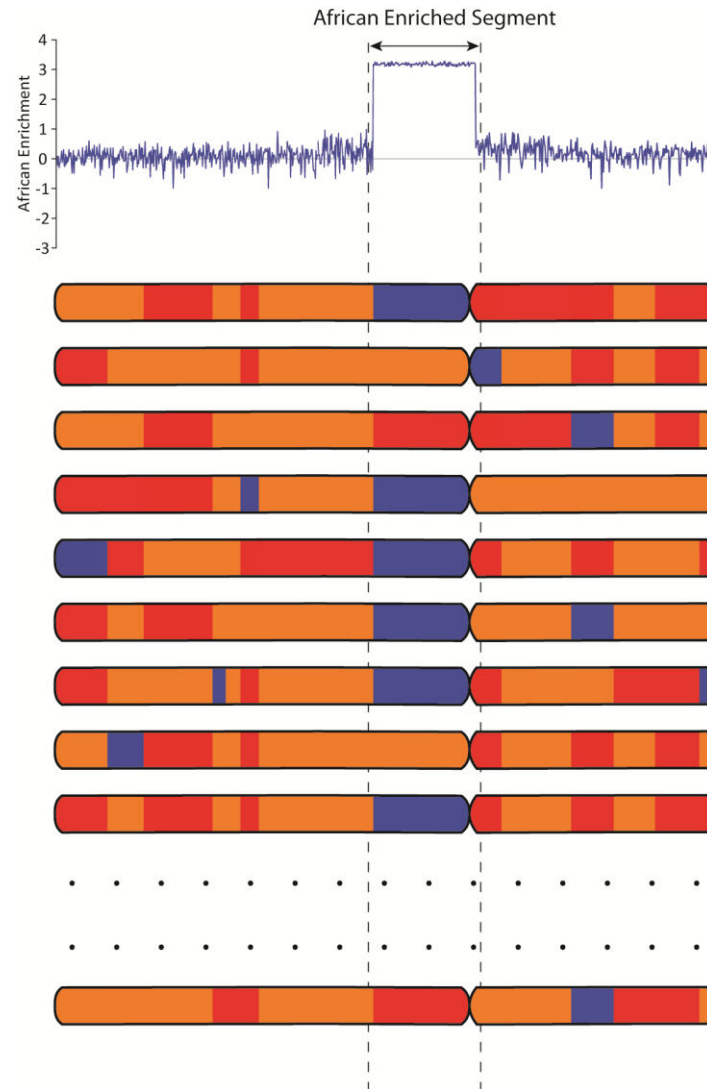
Working hypothesis:

Specific alleles (SNPs) evolved separately in different human populations based on their regional-specific utility (i.e. relationship to health & fitness)

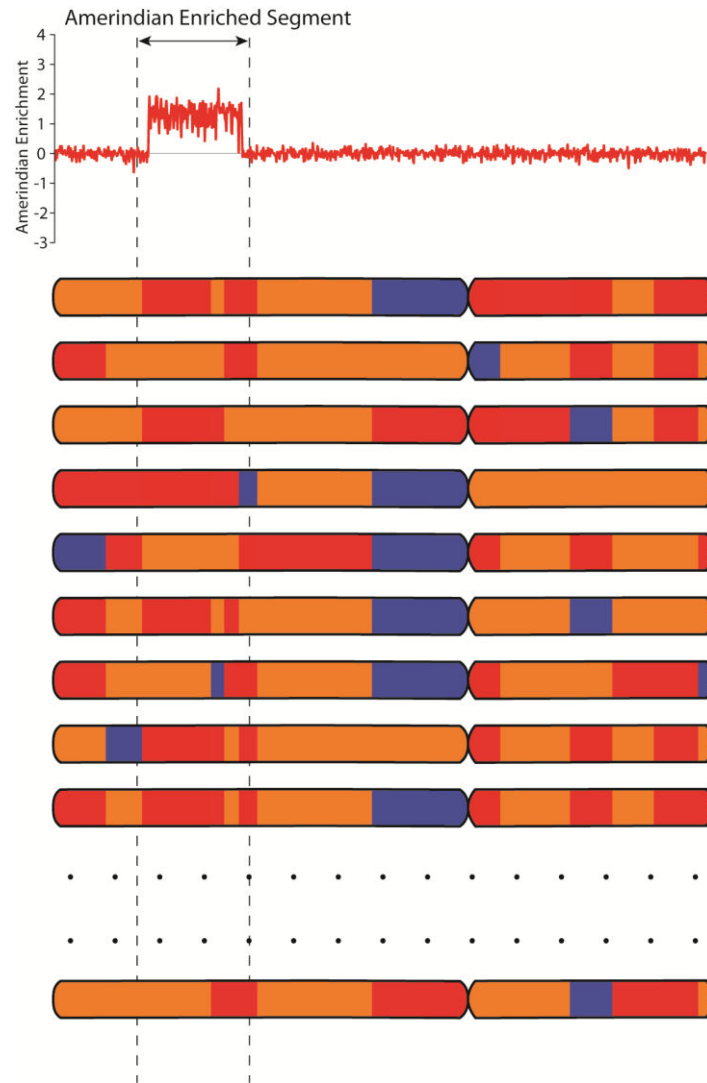
These pre-evolved ancestral population specific-alleles can be selected in the admixed population based on their utility in the new environment



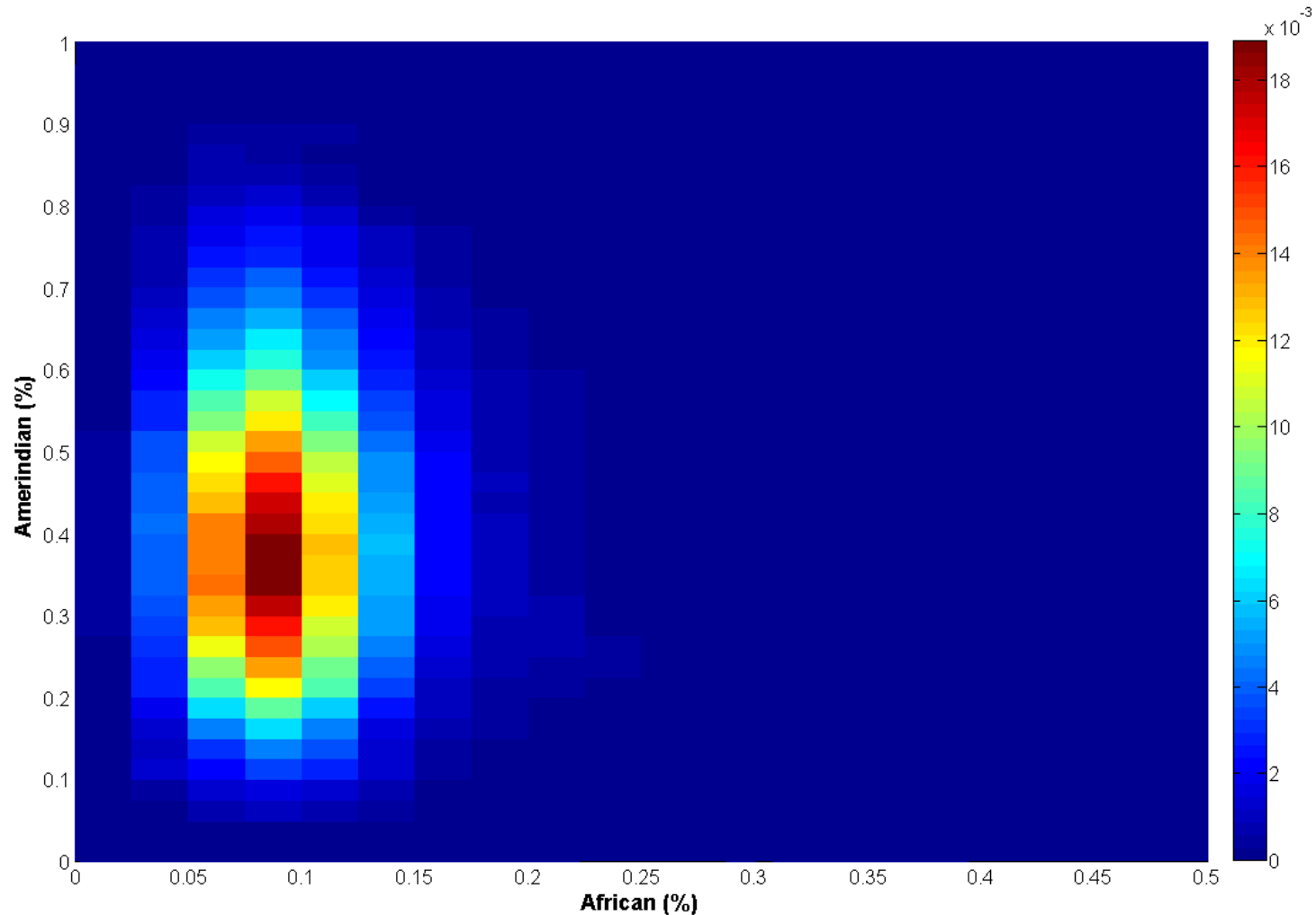
Population admixture enrichment approach



Population admixture enrichment approach



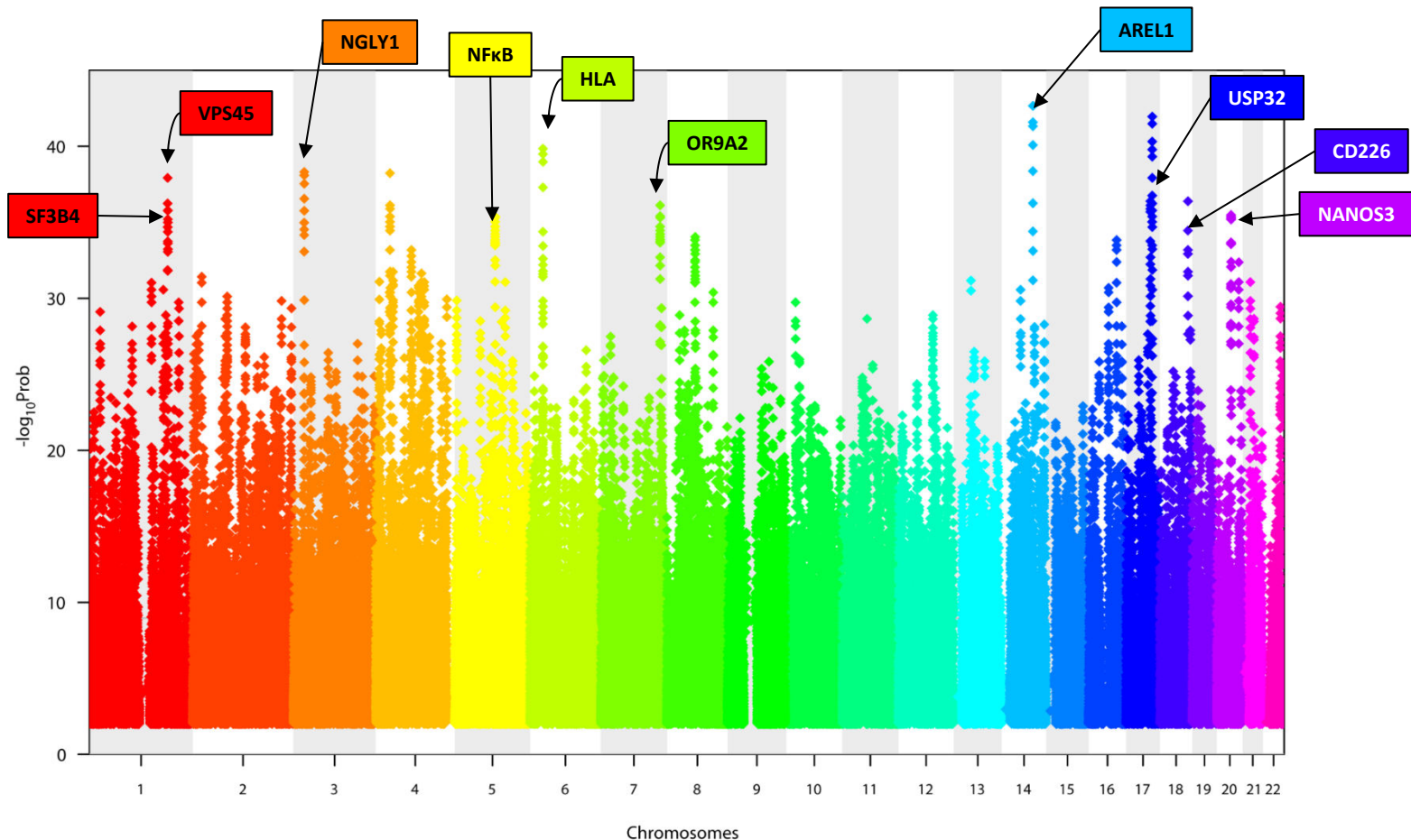
Population admixture enrichment results



Average Percent Contribution

African = 8.3%; Amerindian = 39.8%; European = 51.9%

Population admixture enrichment peaks



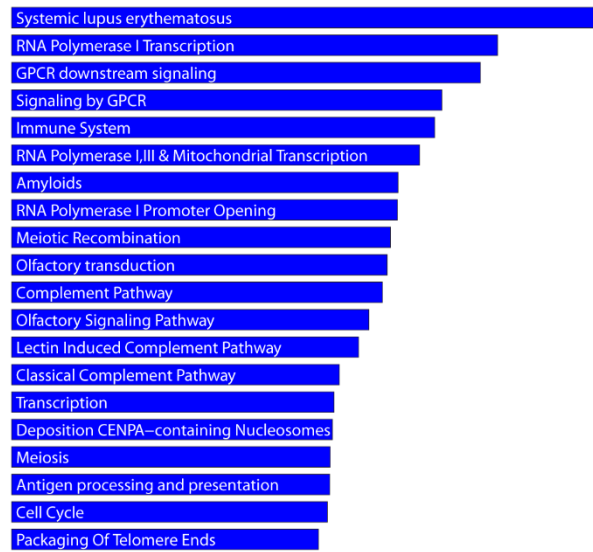
$$P(\text{CHB} = x_1 \text{ and YRI} = x_2 \text{ and CEU} = x_3) = \frac{120!}{x_1! \times x_2! \times x_3!} p_{\text{CHB}}^{x_1} \times p_{\text{CEU}}^{x_2} \times p_{\text{YRI}}^{x_3}$$

where,

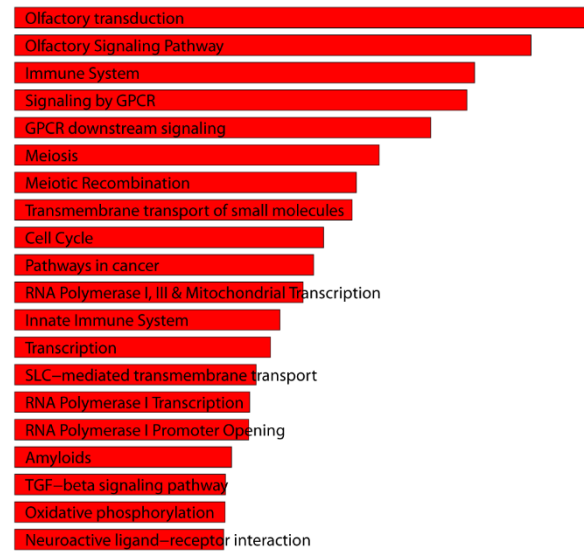
x_1, x_2, x_3 are the observed chromosomes of CHB, YRI and CEU ancestries respectively at a given LD, and $p_{\text{CHB}}, p_{\text{YRI}}, p_{\text{CEU}}$ are the expected probabilities that the given loci will belong to CHB, YRI or CEU ancestry respectively

Functional categories for ancestry-specific enriched genes

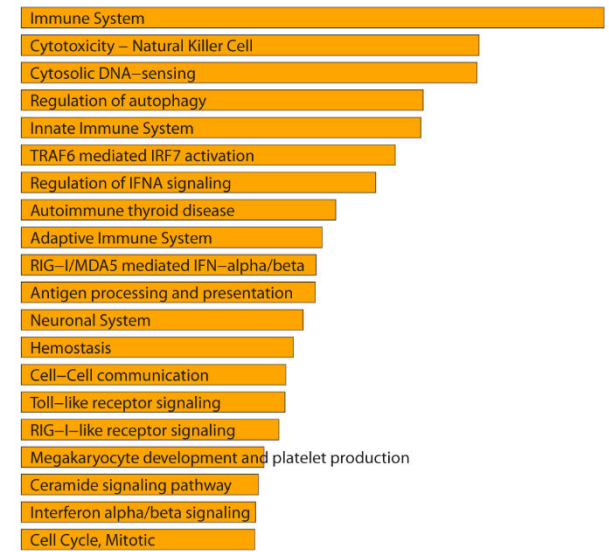
African-enriched genes



Amerindian-enriched genes



European-enriched genes



- abundant immune system related functions & pathways
- abundant cell surface/receptor functions along with downstream signaling pathways

Functional categories for ancestry-specific enriched genes

African-enriched genes

Immune System

Amerindian-enriched genes

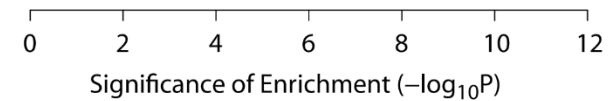
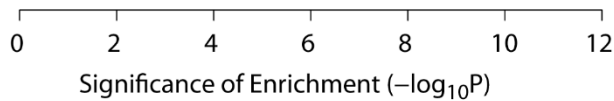
Immune System

Innate Immune System

European-enriched genes

Immune System

Innate Immune System



- abundant immune system related functions & pathways
- abundant cell surface/receptor functions along with downstream signaling pathways

Functional categories for ancestry-specific enriched genes

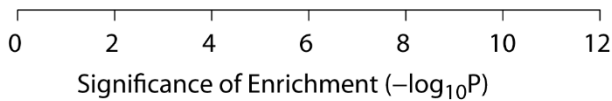
African-enriched genes

GPCR downstream signaling
 Signaling by GPCR

Complement Pathway

Lectin Induced Complement Pathway
 Classical Complement Pathway

Antigen processing and presentation



Amerindian-enriched genes

Signaling by GPCR
 GPCR downstream signaling

Transmembrane transport of small molecules

SLC-mediated transmembrane transport

TGF-beta signaling pathway

Neuroactive ligand-receptor interaction



European-enriched genes

Cytotoxicity – Natural Killer Cell
 Cytosolic DNA-sensing
 Regulation of autophagy

TRAF6 mediated IRF7 activation

Regulation of IFN-alpha signaling

Autoimmune thyroid disease

Adaptive Immune System

RIG-I/MDA5 mediated IFN-alpha/beta

Antigen processing and presentation

Hemostasis

Cell-Cell communication

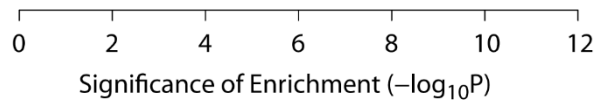
Toll-like receptor signaling

RIG-I-like receptor signaling

Megakaryocyte development and platelet production

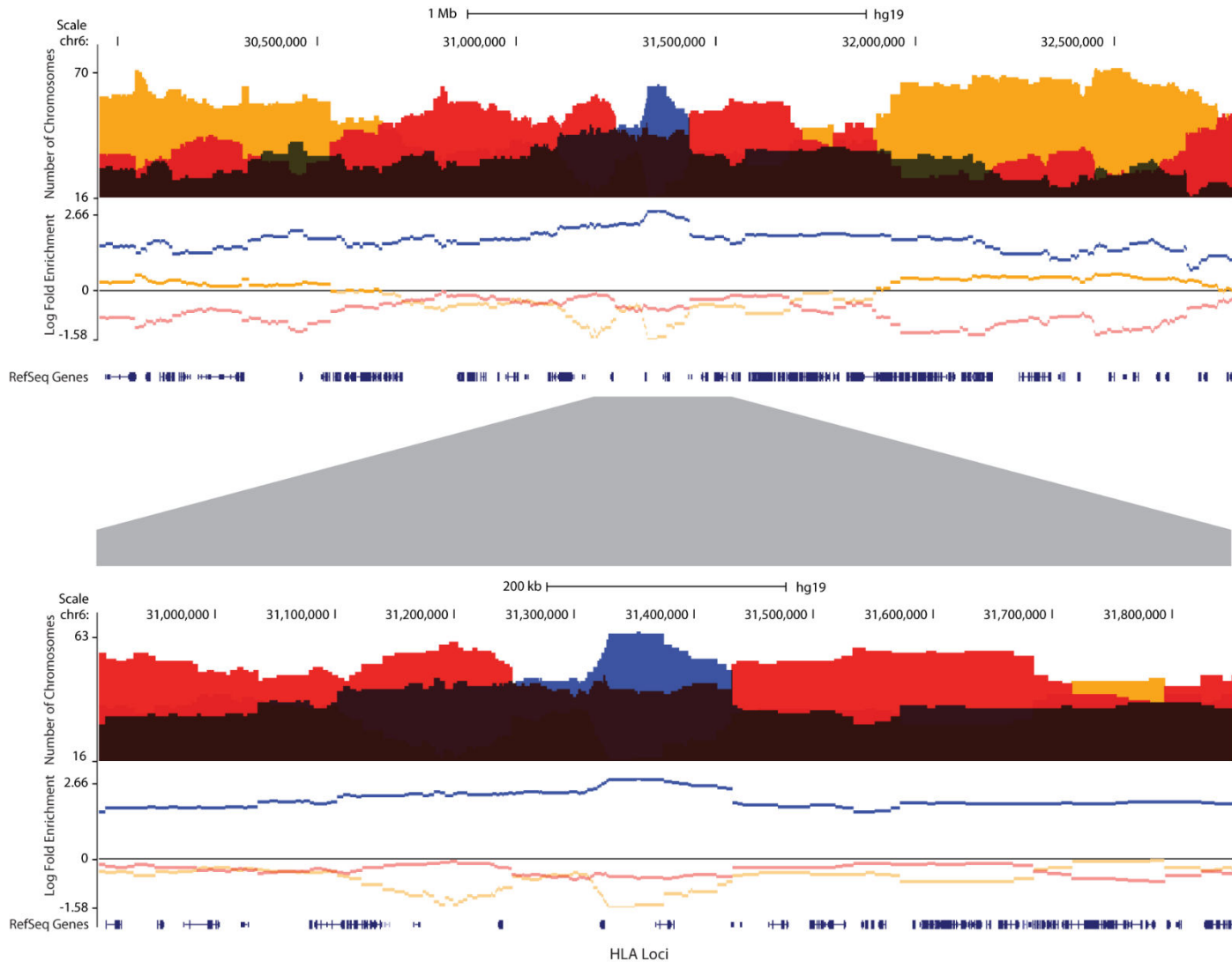
Ceramide signaling pathway

Interferon alpha/beta signaling



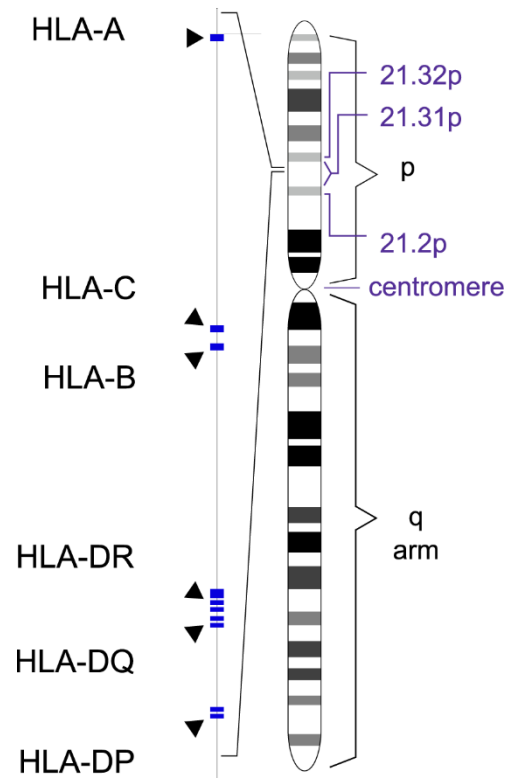
- abundant immune system related functions & pathways
- abundant cell surface/receptor functions along with downstream signaling pathways

African enrichment at the HLA locus

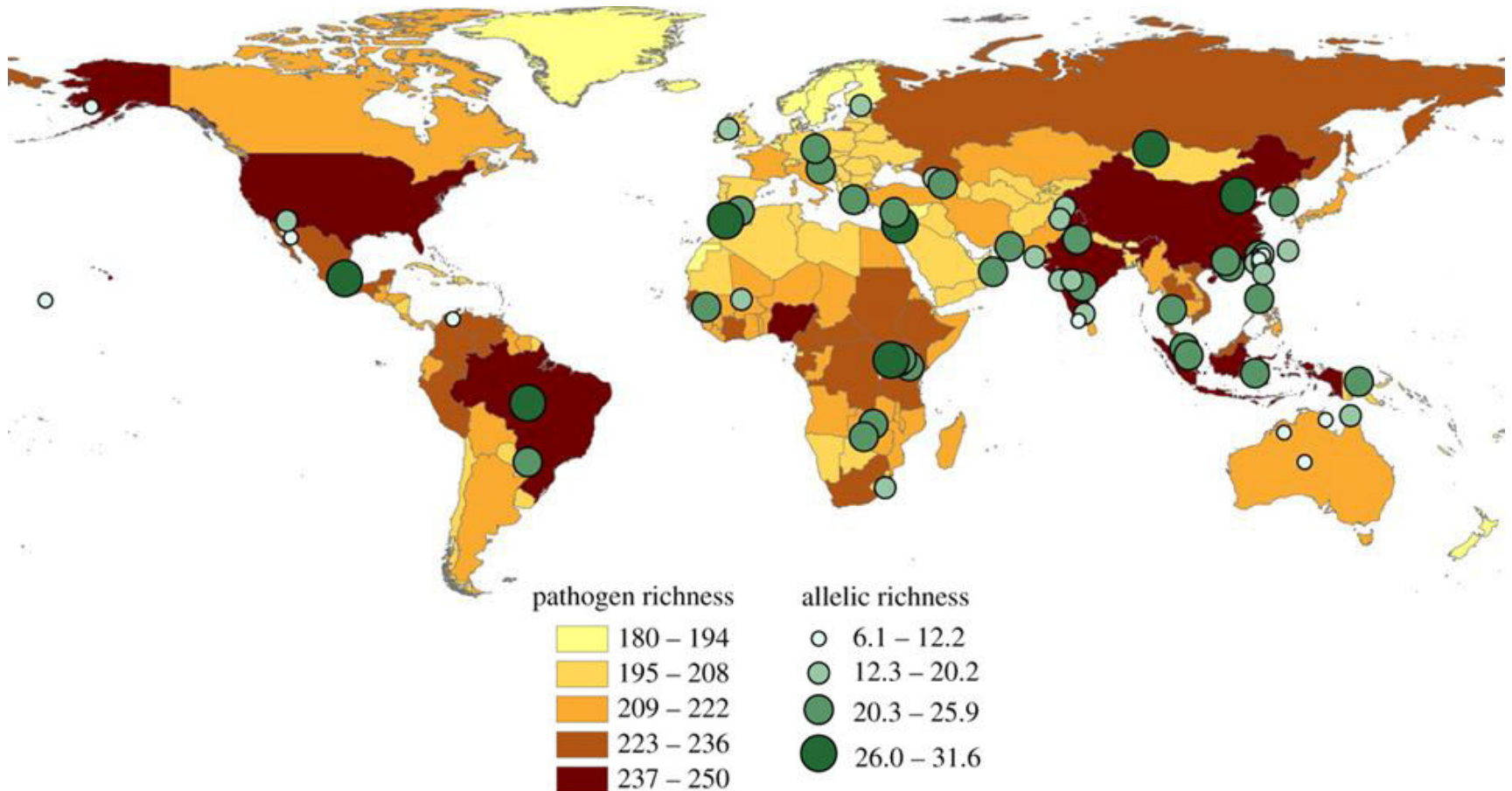


Admixture & HLA gene diversity

- Major histocompatibility complex (MHC) encodes receptors that present antigens to immune (T) cells
- Associated with over 100 different diseases from diabetes, arthritis to various autoimmune diseases.
- More diversity in HLA genes has been observed in African population owing to more pathogen richness
- This increased diversity provides an increased fitness advantage against pathogens
- The ability to mix and match diverse African HLA alleles could similarly provide a selective advantage in the pathogen rich environment of new world tropical regions

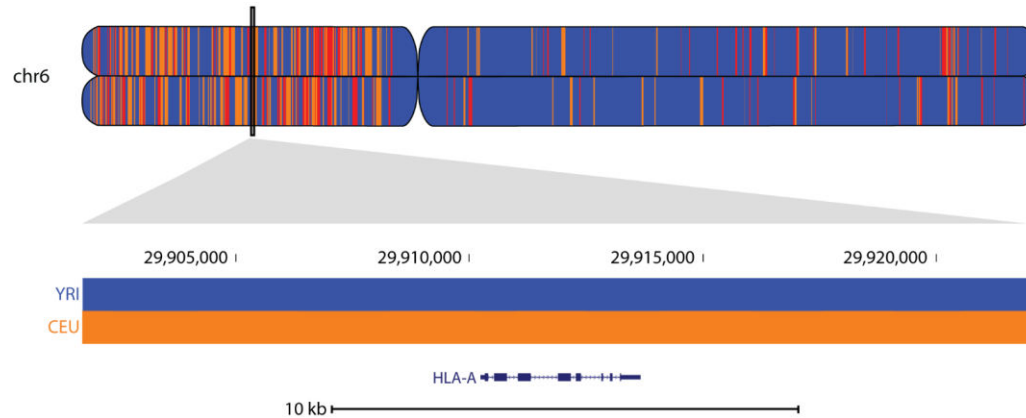


Global Diversity in HLA Loci

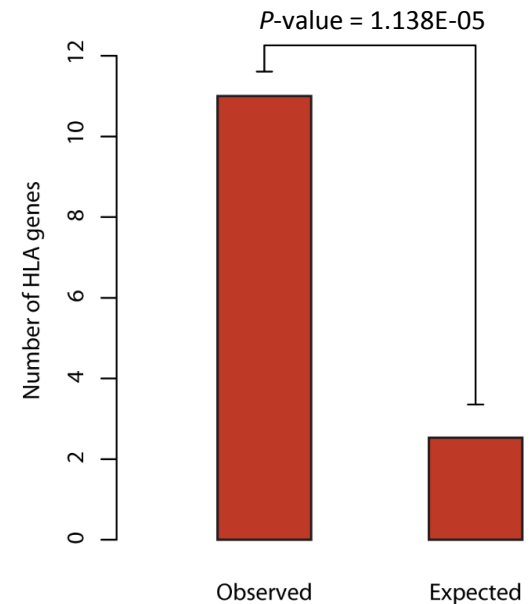


Sanchez-Mazas, A., Lemaître, J. F., & Currat, M. (2012). Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367:830-839.

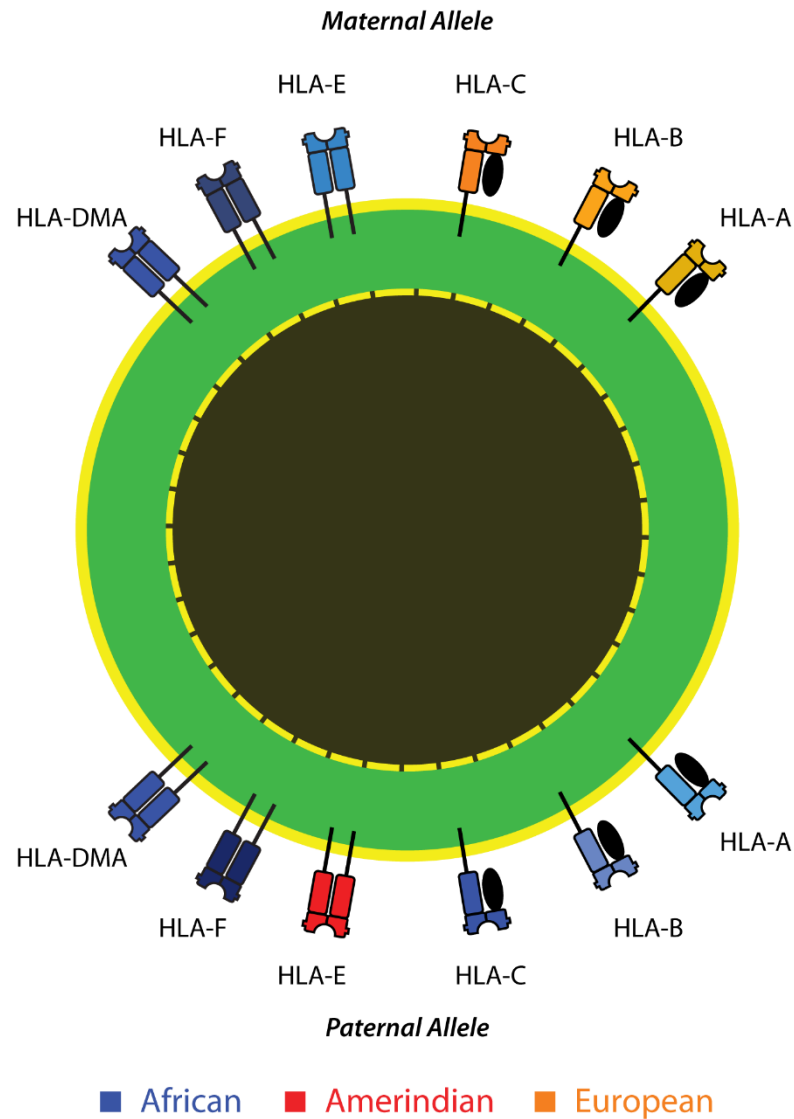
Admixture & HLA gene diversity



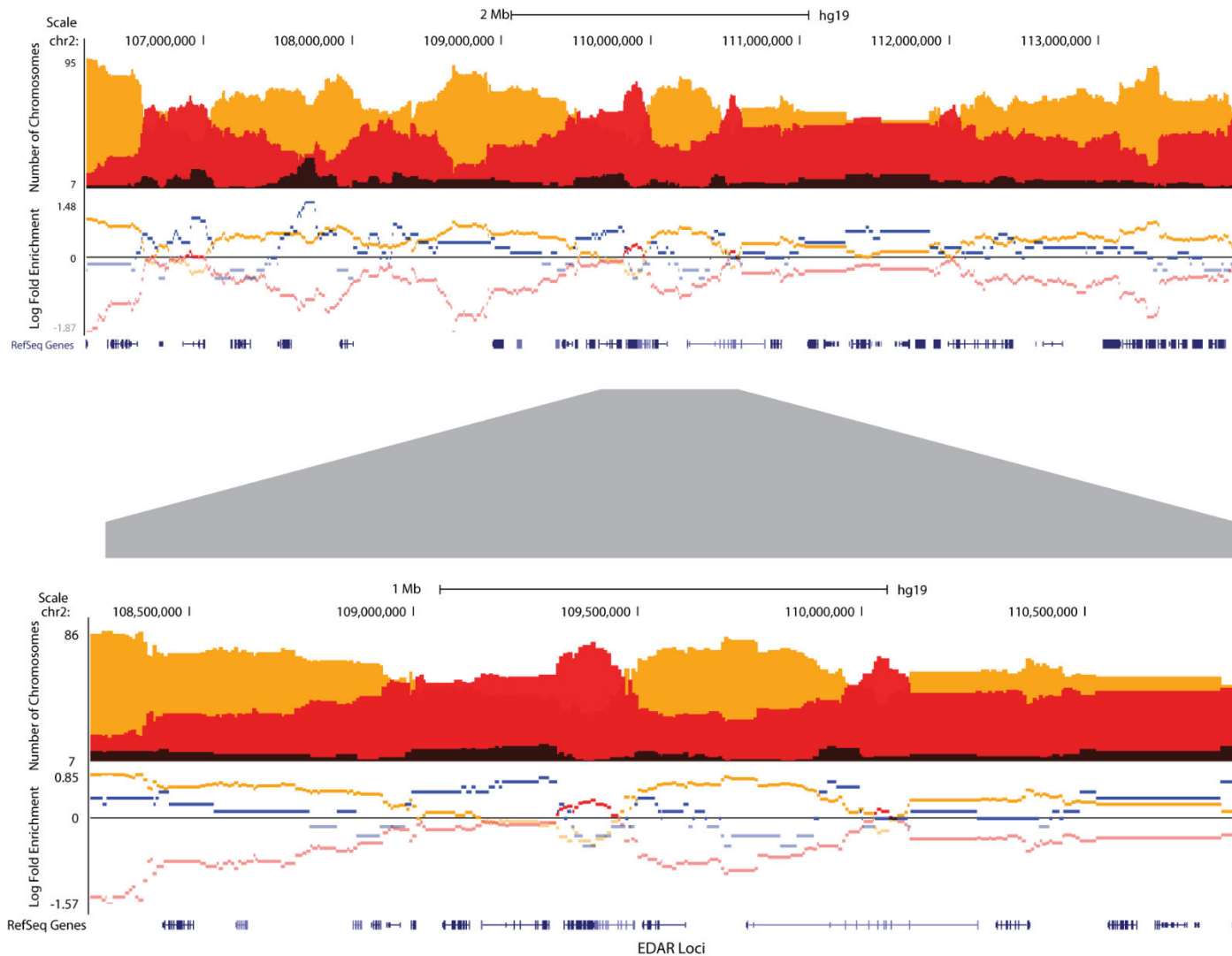
	<i>HLA Genes</i>	<i>Non-HLA Genes</i>	
<i>European/African</i>	11	2,393	2,404
<i>Non-European/African</i>	14	21,307	21,321
	25	23,700	23,725



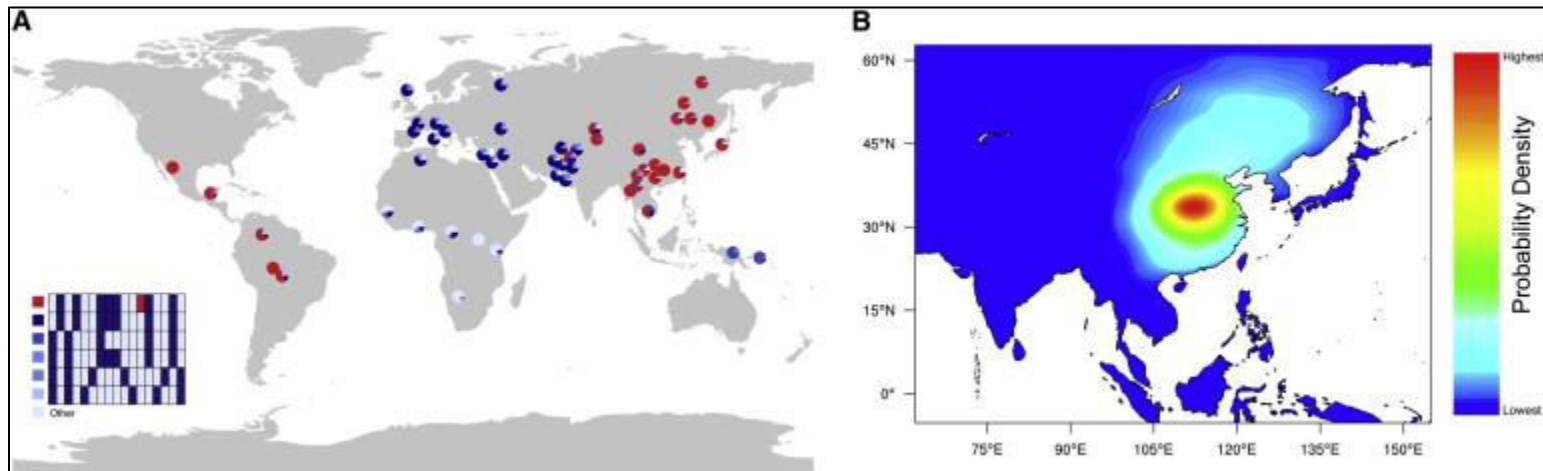
Admixture & HLA gene diversity



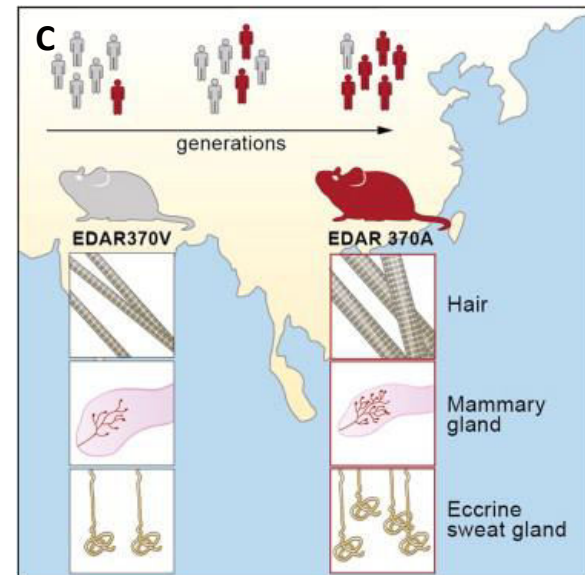
Amerindian depletion at the EDAR locus



Natural selection at the EDAR locus

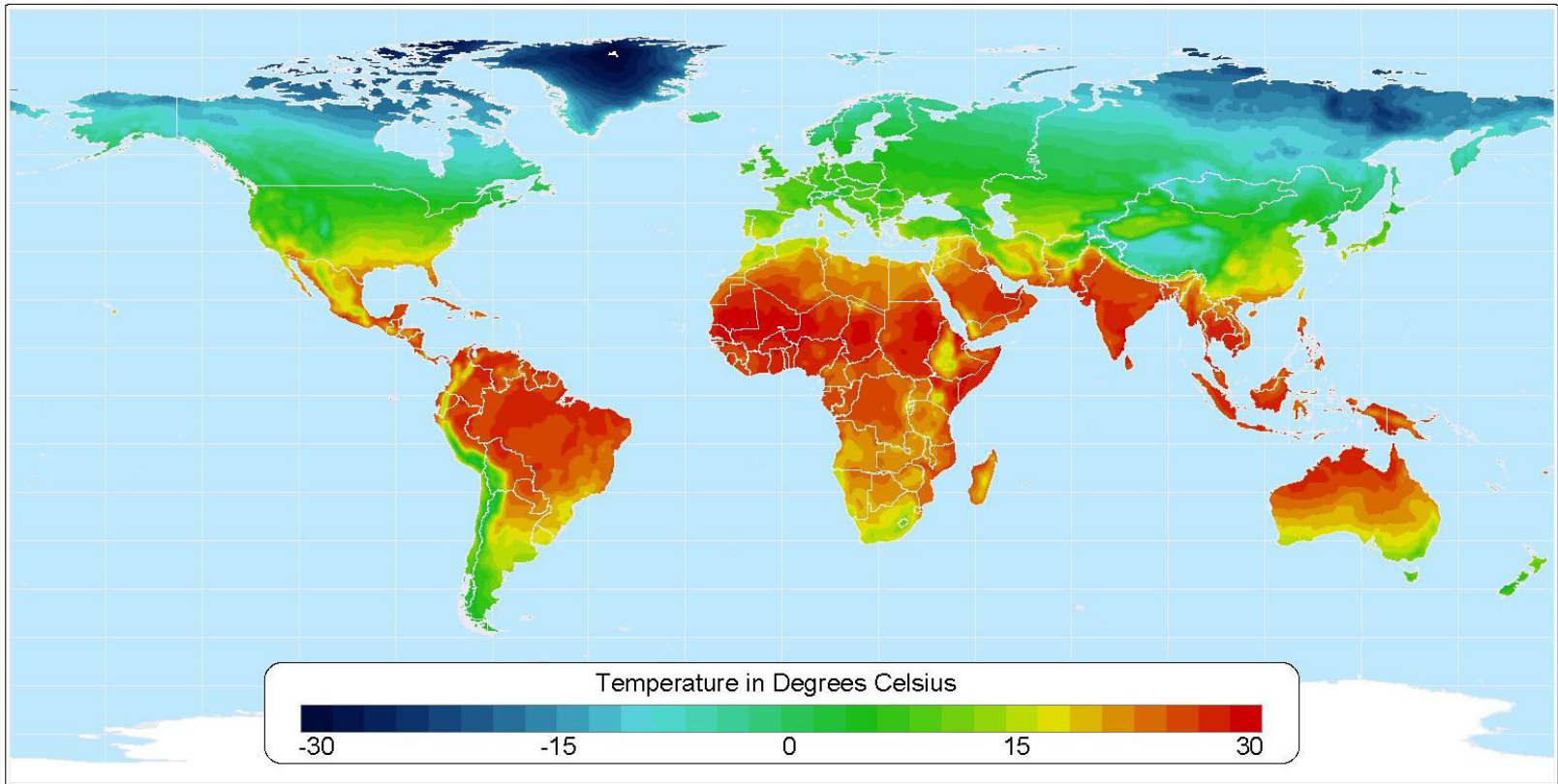


- EDAR is a cell surface receptor influencing the development of a number hair follicles and cutaneous glands.
- Genetic evidence supporting recent positive selection at this loci has been demonstrated in Asian and Native American populations
- The selected allele leads to increased glandular secretions, increases lubrication and reduces evaporation from exposed facial structures and upper airways, a helpful trait for survival in notably cold and dry environment
- Such an adaptation to cold and dry environments would be disadvantageous in the hotter more humid tropical environment of Colombia



Natural selection at the EDAR locus

Average Annual Temperature



Data taken from: CRU 0.5 Degree Dataset (New, et al.)

Atlas of the Biosphere

Center for Sustainability and the Global Environment
University of Wisconsin - Madison

Additional ancestry-enriched genes of interest

Gene	Associated trait
ADCY3	GPCR signaling
AIM2	Innate immune system
AREL1	Negative regulation of Apoptosis
ARFGEF1	Intracellular vesicular trafficking
ATM	Meiosis
CASP8	Innate immunity
CD226	Adaptive immune system, cell adhesion
HLA	Immune response
IFNA10	Regulation of autophagy
KEL	Knockdown confers resistance to HIV-1
LCT	Confers lactase persistence
MANBA	Decreases Colorectal cancer risk. Also associated with Mannosidosis
MAPK10	
NANOS3	Germ cell development
NFκB	Reduces breast cancer risk
NGLY1	Metabolism
OR10K1	Olfactory transduction
OR9A2	Olfactory receptor
RAC1	Natural killer cell mediated cytotoxicity
RANBP2	SLC mediated transmembrane transport
SF3B4	Nagar syndrome
SLC24A5	Decreases melanin pigmentation in skin
SLC44A2	Metabolism
TYROBP	Immune System
ULBP1	Cytosolic DNA sensing pathway
USP32	Overexpressed in breast cancer
VPS45	congenital neutropenia and primary myelofibrosis of infancy.

■ African
 ■ Amerindian
 ■ European

Conclusions

- Colombians have highly admixed genomes – African, Amerindian & European ancestry
- Ancestry can be assigned at continental and sub-continental levels
- Admixture patterns are highly sex-specific (Amerindian female & European male)
- There is substantial enrichment of ancestry-specific loci genome-wide
- Alleles pre-selected in ancestral regions may have been retained for their utility in the new world (immune function, HLA)



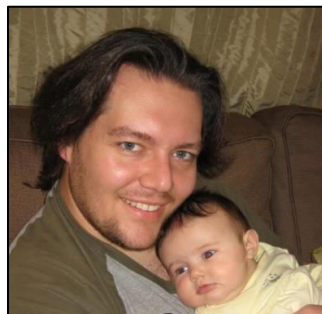
Augusto Valderrama



Mauricio Rodriguez



Fredrik Vannberg



Andrew Conley



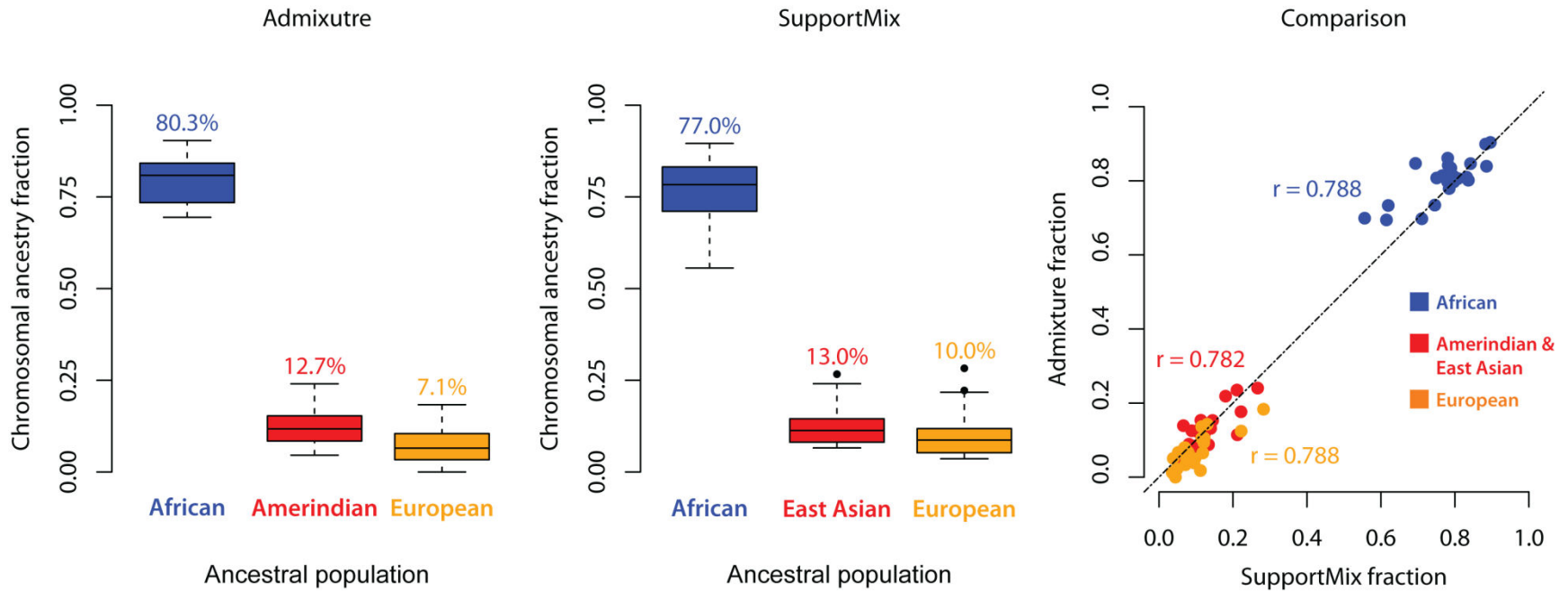
Charles Wigington



Lavanya Rishishwar

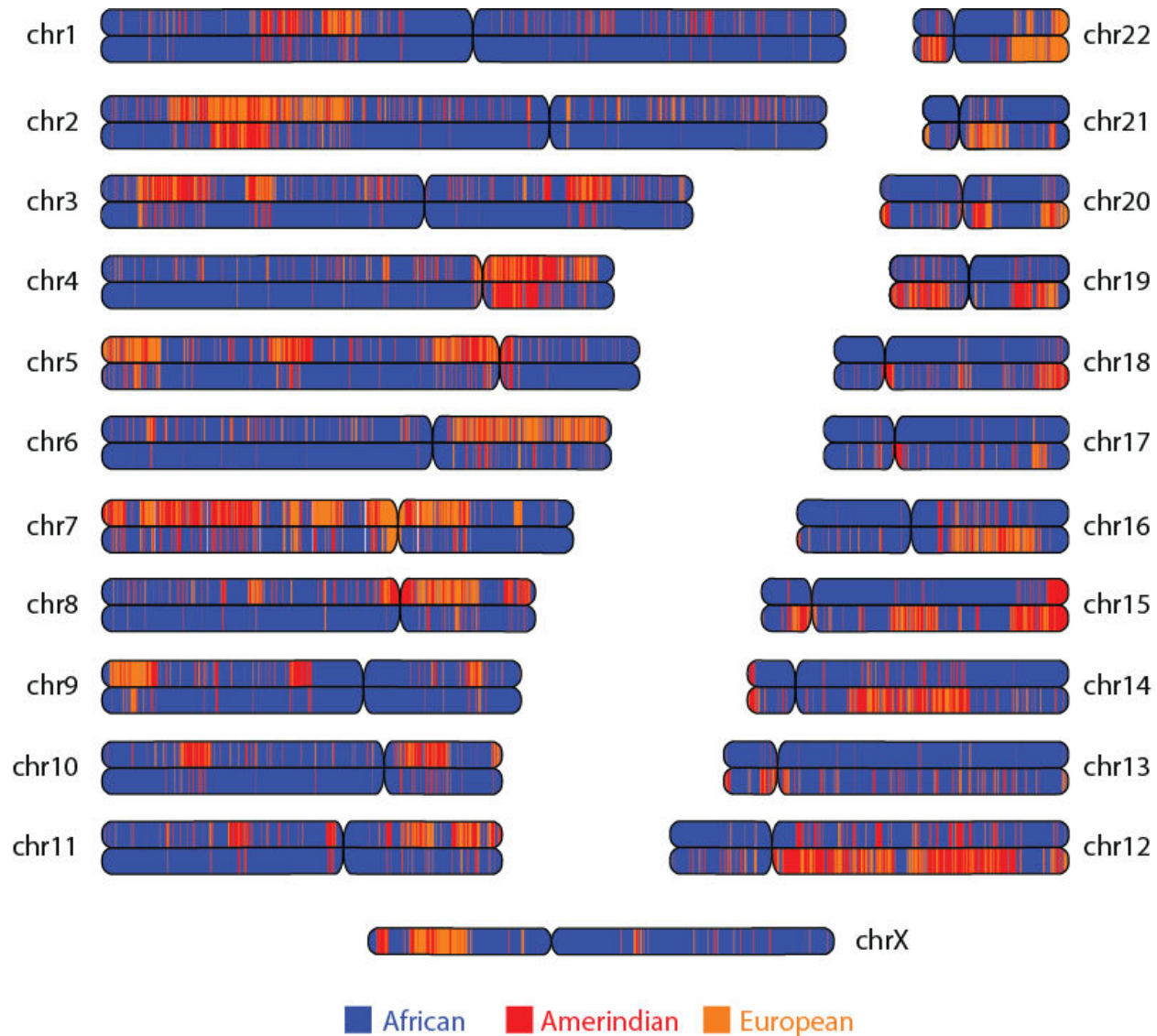


East Asian SNPs can be used as a surrogate for Indigenous genome admixture



- Use of complete Asian genome sequences allows for the inclusion of **37,939,312** SNP variants (>40x more)
- With this level of resolution genomic ancestry can be ascertained for individual chromosomes, sub-regions and down to the individual nucleotide level

Afro-Colombian chromosomal ancestry admixture



Afro-Colombian genetic heterozygosity

