Title: From North Asia to South America: Tracing the longest human migration through genomic sequencing

Authors: Elena S. Gusareva,[1,2,3]† Amit Gourav Ghosh,[1,2,3]† Vladimir N. Kharkov,[3,4]† Seik-Soon Khor,[2,3,5] Aleksei Zarubin,[3,4] Nikita Moshkov,[3,6] Namrata Kalsi,[2,3] Aakrosh Ratan,[3,7] Cassie E. Heinle,[2] Niall Cooke,[3,8] Claudio M. Bravi,[9] Marina V. Smolnikova,[3,10] Sergey Yu. Tereshchenko,[3,10] Eduard W. Kasparov,[3,10] Irina Yu. Khitrinskaya,[4] Andrey Marusin,[4] Magomed O. Razhabov,[4,11] Maria V. Golubenko,[3,4] Maria Swarovskaya,[4] Nikita A. Kolesnikov,[4] Ksenia V. Vagaitseva,[4] Elena R. Eremina,[12,13] Aitalina Sukhomyasova,[14,15] Olga Shtygasheva,[16] Deepa Panicker,[2] Poh Nee Ang,[2] Choou Fook Lee,[2] Yanqing Koh,[2] See Ting Leong,[2] Changsook Park,[2] Sachin R. Lohar,[2] Zhei Hwee Yap,[2] Soo Guek Ng,[2] Justine Dacanay,[2] Daniela I. Drautz-Moses,[2] Nurul Adilah Binte Ramli,[2,3] Katsushi Tokunaga,[3,5] Ian McGonigle,[3,17,18] Inaho Danjoh,[19] Andrés Moreno-Estrada,[3,20] Atsushi Tajima,[3,21,22] Hideyuki Tanabe,[3,23] Yukio Nakamura,[3,24] Shigeki Nakagome,[3,8] Tatiana V. Tatarinova,[3,25] Vadim A. Stepanov,[3,4]†† Stephan C. Schuster,[2,3,18]†† Hie Lim Kim,[1,2,3]*.

Affiliations:

[1]The Asian School of the Environment, Nanyang Technological University; Singapore, 637459, Singapore.

[2]Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, 637551, Singapore.

[3]GenomeAsia 100K consortium; Singapore, 059818, Singapore.

[4]Research Institute of Medical Genetics, Tomsk National Medical Research Center, Russian Academy of Science; Tomsk, 634050 Russian Federation.

[5]National Center for Global Health and Medicine; Tokyo, 162-8655, Japan.

[6]Synthetic and Systems Biology Unit, Biological Research Centre (BRC); Szeged, H-6726, Hungary.

[7]Center for Public Health Genomics, University of Virginia School of Medicine; Charlottesville, Virginia, 22908, USA.

[8]School of Medicine, Trinity College Dublin, Dublin, D02 PN40, Ireland.

[9]Instituto Multidisciplinario de Biología Celular (IMBICE), CCT La Plata CONICET-CICPBA-Universidad Nacional de La Plata; La Plata, B1906APO, Buenos Aires, Argentina.

[10]Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences", Research Institute of Medical Problems of the North; Krasnoyarsk, 660022, Russian Federation.

[11]Sector of Biophysics, Institute of Physics, Dagestan Federal Research Center; Mahachkala, 367015, Russian Federation.

[12]Buryat State University; Ulan-Ude, 670000, Buryat Republic, Russian Federation.

[13]Buryat Republican Perinatal Center; Ulan-Ude, 670031, Buryat Republic, Russian Federation.

[14]Research laboratory for Molecular Medicine and Human Genetics, Medical Institute, M.K. Ammosov North-Eastern Federal University, Yakutsk, 677027, Russian Federation.

[15]Republican Hospital #1 - National Medical Centre, Ministry of Public Health of the Sakha Republic; Yakutsk, 677019, Russian Federation.

[16]Katanov State University of Khakass; Abakan, 655017, Russian Federation.

[17]School of Social Sciences, Nanyang Technological University; Singapore, 637459, Singapore.

[18]School of Biological Science, Nanyang Technological University; Singapore, 637551, Singapore.

[19]Tohoku Medical Megabank Organization, Tohoku University; Miyagi, 980-8573, Japan.

[20]Advanced Genomics Unit, Center for Research and Advanced Studies of the IPN (Cinvestav); Irapuato, Guanajuato, 36821, Mexico.

[21]Department of Bioinformatics and Genomics, Graduate School of Advanced Preventive Medical Sciences, Kanazawa University; Kanazawa, Ishikawa, 920-8640, Japan.

[22]Sapiens Life Sciences, Evolution and Medicine Research Center, Kanazawa University; Kanazawa, Ishikawa, 920-8640, Japan.

[23]Research Center for Integrative Evolutionary Science, The Graduate University for Advanced Studies, SOKENDAI, Shonan Village; Hayama, Kanagawa, 240-0193, Japan.

[24]RIKEN BioResource Research Center, Ibaraki, Japan.

[25]University of La Verne; La Verne, CA 917 50, USA.

†, †† These authors contributed equally to this work.

* Corresponding author. E-mail: HLKIM@ntu.edu.sg (H.L.K.)

Abstract: Genome sequencing of 1,537 individuals from 139 ethnic groups reveals the genetic characteristics of understudied populations in North Asia and South America. Our analysis demonstrates that West Siberian ancestry, represented by the Kets and Nenets, contributed to the genetic ancestry of most Siberian populations. West Beringians, including the Koryaks, Inuit, and Luoravetlans, exhibit genetic adaptation to Arctic climate, including medically relevant variants. In South America, early migrants split into four groups – Amazonians, Andeans, Chaco Amerindians, and Patagonians – ~13,900 years ago. Their longest migration led to population decline, while settlement in South America's diverse environments caused instant spatial isolation, reducing genetic and immunogenic diversity. These findings highlight how population history and environmental pressures shaped the genetic architecture of human populations across North Asia and South America.

The late Pleistocene saw the expansion of humans into the frigid lands of Eurasia. The earliest known presence of modern humans in northern Eurasia at latitudes greater than 50°N was around 45,000 years ago (ya) in West Siberia (*1*), and by 31,600 ya, humans had migrated far east towards Beringia, north of the Arctic Circle at 71° N (*2*). The earliest human remains identified in this region are two Yana RHS individuals that, despite their extreme Northeast Siberian geographical location, show substantial genetic relatedness to early West Eurasian hunter-gatherers (*3*). The Upper Palaeolithic people who initially populated Northeast Siberia were then replaced by arrivals from East Asia. The Kolyma1 remains, excavated near the Chukotka region and dated 9,800 ya, demonstrates greater affinity to East Asians and present-day west Beringian populations, such as Koryaks and Luoravetlans (also known as Chukchi), as well as to Native Americans (*3*). The linguistic and cultural diversity of present-day Indigenous Siberian populations (*4*) is mirrored by the complex patterns of admixture, as shown by genome-wide genotype data analysis (*5*-*7*). This genetic structure in Siberians, comprised of several ancestral components, is estimated to have emerged within the past 10,000 to ~3,400 years (*5, 8*). The Western Eurasian ancestry component presented in majority of Indigenous Siberian populations is not the result of postcolonial Russian admixture but one of the ancient components (*5, 8*) dating back to 12,500 to 25,000 ya in different Siberian populations (*8*). Among the present-day populations of Northeast Eurasia, the Koryaks from the Kamchatka Peninsula (*3, 9*) and Inuit from Chukotka (*10, 11*) show the closest genetic relatedness to Native North Americans.

The migration of humans to the Americas occurred when the Bering Land Bridge was still open (*12*), with the earliest human remains in North America found in the Clovis burial site in western Montana dating back to around 12,700 ya (*13*). However, recent evidence suggests a human presence in North America at least 23,000 ya (*14*). By the time the Ice-Free Corridor opened up and became suitable for travel around 13,300 ya (*15*), humans were already widely dispersed in North America (*16*), likely due to Pacific coastal migration routes (*17*). The divergence between northern and southern Native American populations is estimated to have occurred between 17,500 and 14,600 years ago, south of the North American ice sheets, according to modern and ancient genomic analyses (*9, 18*). The rapid dispersal of humans in South America is suggested by archaeological records, which date the earliest human presence in North Patagonia, the southernmost tip of the Americas, to 14,500 ya (*19*). However, the number of basal divergences, founding populations, admixture, and the degrees of isolation among Native South American populations remain a subject of debate (*20-29*), with most of the current understanding coming from analyses of genome-wide genotyping or ancient DNA data. Additionally, fine-scale population genetic studies based on high-coverage whole genome sequencing datasets for contemporary populations of North Eurasia and South America have not been performed to date.

In this study, we aim to improve our understanding of the prehistoric population dynamics and shaping of the contemporary populations of North Eurasia and the Americas, as well as the history of adaptation to the diverse environments encountered by humans during their migration and settlements. We use large-scale whole-genome sequencing datasets representing populations from Northeast Europe, Siberia, and the Russian Far East (west Beringia), as well as Native Americans, to identify population structure for reconstructing demographic history. We also aim to clarify the role of past environments and lifestyles in the diversification of human populations in North Eurasia and South America. Finally, we aim to demonstrate the importance of incorporating population history and ancestry information into present biomedical research.

Population structure and admixture

The GenomeAsia 100K consortium has created a high-resolution whole-genome dataset of 1,537 individual genomes representing 139 ethnic groups from 27 countries across North Eurasia and Native America [see supplementary text S1, tables S1.1-S1.4 (*30*)]. The dataset, GA100K:NENA, was generated using the Illumina sequencing platform and processed with a uniform pipeline for mapping and variant calling [see supplementary text S2 (*30*)]. In total, the dataset includes 72,207,507 biallelic single nucleotide polymorphisms (SNPs) and 40,821 small insertions and deletions (INDELs) compared to the GRCh37 reference genome. For this study, the dataset was filtered to a subset of 50,557,893 high-quality autosomal SNPs in 1,477 unrelated individuals [see supplementary text S3, S4 (*30*)]. To further explore the genetic history of North America, the GA100K:NENA dataset was combined with an open-source genotyping data from modern (*31, 32*) and ancient Native North Americans (*11, 25, 28, 33-38*). These data provide a valuable resource for understanding the population structure and ancient history of North Eurasia and South America, as well as the adaptation of these populations to diverse environments [see supplementary text S5 (*30*)].

The genetic diversity of West and North Eurasians is influenced by six ancestral components that reflect the geographical distribution of these populations, according to Admixture (*39*) and local ancestry analysis (*40, 41*) (Fig. 1), [see supplementary text S5.1, S5.2 (*30*)]. Indigenous groups in Northeast Europe, such as the Finno-Ugric speakers, share common ancestry with Northeast Europeans. However, Finno-Ugric speakers living further to the east, such as the Komi, Udmurts, Mansi, and Khants, share an ancestry component most enriched in West Siberian populations, such as Kets and Nenets living in the Yenisei River basin. West Siberian ancestry is prevalent among contemporary Siberian populations, particularly the Kets and Nenets. Most West Siberians are traditionally nomadic or semi-nomadic [table S1.3 (*30*)] and show admixture patterns with Yeniseians, Northeast Europeans, Caucasians, and East Asians, reflecting frequent multi-directional migrations across North Eurasia in the past. East Siberians are related to West Siberians and East Asians, while European admixture is generally minor. The west Beringian populations of Chukotka and the Kamchatka Peninsula are less admixed, likely due to their isolation in remote and frigid areas. These populations represent a distinct ancestry, which is substantially shared with contemporary Indigenous Canadian Chipewyan and is present in several ancient Native North Americans such as Alaskan USR1 and Athabaskan, Anzick, Lovelock and Spirit Cave remains [fig. S5.11B (*30*)]. In the principal component analysis (*42*) (PCA), Beringians exhibit a linear cline towards the Indigenous Canadian Chipewyan and the clusters of Native Meso- and South Americans (Fig. 2A), [see supplementary text S5.3, S5.5 (*30*)]. As described above, the variants generated from our high coverage genomes were intersected with previously published genotyping data (*3, 10, 11*), and confirms that within North Eurasian populations, Koryaks, Luoravetlans (aka Chukchi), and Inuit are the closest to Native Americans.

The population structure of Native South Americans is characterized by four ancestral lineages (Fig. 1) that reflect geographical and environmental boundaries, such as the Andes Mountains, the hot and semiarid lowland Dry Chaco region, the Amazon basin with its moist tropical rainforest jungle, and the cold polar climate of the Patagonia region. The Andean and Chaco Amerindian ancestries are prevalent among contemporary Native South Americans. Andean ancestry is typical

for highlanders in western South America, but some of these populations, such as Aymara, Atacama, and the population in the Puna region, also show a significant relatedness to Chaco Amerindians and a minor admixture with Amazonians. Colombians and Guahibo mainly share ancestry with Chaco Amerindians and show recent admixture with Europeans and Africans (Fig. 1A), [supplementary text S5.1, S5.6 (*30*)]. The four Native South American ancestral lineages are identified in separate clusters in the PCA, diverging in a star-like pattern with Mesoamericans located at the center (Fig. 2C). Similarly, ancient Native North American samples exhibit all four ancestral components in varying proportions [fig. S5.11B (*30*)] and cluster with contemporary Native Mesoamericans in PCA [fig. S5.14 (*30*)]. This suggests that Native Meso- and North Americans are genetically closest to the common ancestor of the four Native South American ancestries. The Andean component is predominant in the ancient samples from Peru, while the Patagonian ancestral component prevails mostly in the ancient samples from Chile and Argentina (figs. S5.11B and S5.14), indicating genetic continuity in South America.

Demographic history and its effect on genetic diversity

Analysis of the dynamics of population sizes and population splits (Fig. 2D, 2E), [see supplementary text S7 (*30*)] showed that nomadic hunter-gatherer West Siberians, represented by the Kets in our dataset, were among the largest populations in North Eurasia around 10,000-13,900 ya, as indicated by Relate (*43*) (Fig. 2D). The MSMC-IM (*44*) and qpGraph (*45*) results suggest that the Kets were shaped by admixture between East Asians (77%) and Northeast Europeans (23%) (Fig. 2E), likely due to frequent migrations in North Eurasia ~8,000-15,000 ya [fig. S7.5 (*30*)]. This admixture is also supported by the qpGraph analysis including ancient genomes, as described in supplementary text S7.3 (*30*). However, the population of Kets declined by 73.6% since 10,000 ya (effective population size *Ne* from 4,448 to 1,194). Similarly, the Arctic hunter-gatherer Koryaks also saw a population decline by 64.4% since 10,000 ya (*Ne* from 3,021 to 1,075). In contrast, during the same period, agricultural populations of Northeast Europeans and East Asians expanded by 176.8% (*Ne* from 4,641 to 12,844) and 91.9% (*Ne* from 3,029 to 5,813), respectively.

Our estimates of population split times suggest that a deep divergence occurred between North Eurasians and Native Americans between 26,800 and 19,300 ya during the Last Glacial Maximum (Fig. 2D), confirming previous estimates (*3, 9, 14, 18, 31*). After the split, gene flow from the Americas back to Beringia is detected in Koryaks and Inuit, with estimates of 5% and 28% Native American ancestry, respectively, according to qpGraph (Fig. 2E), [see supplementary text S5.6, S7.3 (*30*)]. The population split time estimates also suggest that the divergence of the four Native South American lineages occurred over a short period, from 13,900 to 10,000 ya (Fig. 2D), [fig S7.6, fig, S7.8 (*30*)]. All four lineages show a continuous population decline. However, the Andean highlanders managed to maintain their population size during the rise of maize horticulture around 5,200-3,700 ya (*46*). It has declined by 45.1% since then (*Ne* from 1,771 to 972), while Chaco Amerindians has declined by 46.89%, *Ne* from 1,448 to 769 since 10,000 ya (Fig. 2D). Amazonians and especially Patagonians have seen a dramatic decrease in population size over the

last 10,000 years, with declines of 66.59% (*Ne* from 1,368 to 457) and 79.68% (*Ne* from 1,171 to 238), respectively.

To assess the impact of population decline on genetic diversity, we estimated genome-wide runs of homozygosity (ROHs) segments (*47*) [see supplementary text S8 (*30*)]. In Native South Americans, the average number and length of ROHs segments estimated across all populations were 10.5 times and 1.3 times higher than those in Africans (Yoruba), and 3.75 times and 1.2 times higher than those in Northeast Europeans, respectively. The highest abundance of extended ROHs was observed in Amazonians, Patagonian Kawésqar, and Chaco Amerindians (Fig. 3A), and was similar to that seen in isolated island populations like the Andamanese and Baining (Fig. 3B). This high homozygosity is likely the result of the founder effect due to long-distance migration and/or population isolation. The strong correlation between the average total number of ROHs and the average nucleotide diversity ($R_{Pearson} = –0.78$) supports the idea that the extended homozygosity is a result of population history (Fig. 3A).

To evaluate the impact of population history on immune genes, we analyzed the diversity of human leukocyte antigen (*HLA*) genes in different population groups [see supplementary text S9 (*30*)]. Maintaining high diversity of *HLA* genes is important for host defense mechanisms, enabling the immune system to present a wide range of antigens to effector cells (*48*). We calculated the average total number of unique *HLA* alleles across eight genes in each population group and its correlation with the average total number of ROHs. Although the number of unique *HLA* alleles is expected to be maintained under natural selection, we observed variation among different populations (Fig. 3B, 3C). Populations with higher homozygosity had a smaller average total number of unique *HLA* alleles ($R_{Pearson} = –0.7$, Fig. 3B), particularly the isolated populations such as the Baining, Amazonians, and Patagonian Kawésqar. This suggests that genetic drift, driven by bottlenecks and founder effects, may have been sufficiently strong to suppress diversifying selection on *HLA* genes and reduce the *HLA* diversity in these populations.

Adaptation to the cold environment in Indigenous Beringian populations

We performed a genome-wide scan for selection sweep loci (using XP-EHH and iHS analyses) in Chukotka and the Kamchatka Peninsula populations that have adapted to living in a frigid Arctic climate. We identified selection signals in genes involved in lipid metabolism and thermogenesis, sensory perception such as olfaction and vision, and the regulation of reproductive functions and the immune system (Fig. 4A), [supplementary text S11 (*30*)].

Our analysis identified a strong signal for positive selection in the carnitine palmitoyltransferase 1 A (*CPT1A*) gene, which plays a crucial role in the import of long-chain fatty acids into mitochondria for fatty acid oxidation and energy production. We used the iSAFE (*49*) method and identified a missense mutation rs80356779 (Pro479Leu) favored by positive selection see supplementary text S11.1 (*30*)]. While this variant has been associated with carnitine palmitoyltransferase I deficiency (*50*) and spinal muscular atrophy (*51*), it is common in Arctic populations (Inuit of North America (*52-55*)) and does not appear to have any clinical manifestations in carriers (*54*). In our study, the frequency of the Leu479 allele in Beringian

populations ranged from 75% to 90% (Fig. 4B), and it may help carriers to maintain body heat by keeping certain fats unmetabolized.

Our study also identified signals of positive selection on the lysophosphatidic acid receptor 1 gene (*LPAR1*). The iSAFE analysis highlighted a variant rs1043128 in the 3' UTR of LPAR1, a cis eQTL for the gene (*56*). The frequency of the G allele is very high in Inuit and Luoravetlans (90%), common in Koryaks (41%), but much lower in Siberians and East Asians (Fig. 4C). *LPAR1* is involved in the regulation of smooth muscle cell chemotaxis, bioactive lipid receptor activity, and cellular responses to oxygen levels, as indicated by gene ontology analysis [see supplementary text S11.4 (*30*)]. These findings suggest that *LPAR1* may be involved in thermoregulation processes.

Adaptation to hypoxia in Andean highlanders

We screened for selection signals in Andean highlanders and identified a selection sweep in the gene hypoxia-inducible transcription factor 2α (*HIF-2α*, also known as *EPAS1*). This gene plays a role in the cellular and systemic responses to hypoxia (*57, 58*), including the stimulation of new blood vessel formation and the production of red blood cells. The selection signal in *EPAS1* has also been identified in Tibetan highlanders (*59-61*) and Colla from Northwest Argentina (*62*). Our iSAFE analysis identified a missense mutation rs570553380 (His>Arg) as a top-ranked variant, which differs from the ones previously described in Tibetans [see supplementary text S11.10 (*30*)]. The Arg allele has a frequency of 28-45% among Quechua speakers and 33% among Peruvians living in highlands but is absent in other Native South Americans (Fig. 4D).

Pathogenic and adverse drug response variants

In the GA100K:NENA datasets, we identified 67,252 variants that are clinically relevant (ClinVar v.20190305, https://www.ncbi.nlm.nih.gov/clinvar/) and calculated their allele frequencies per population [table S12.1 (*30*)]. Among these, 529 variants are classified as pathogenic, including 349 (66%) non-synonymous protein-altering variants in 280 genes and 93 (18%) stop-gain or stop-loss variants in 87 genes [see supplementary text S12.1, table S12.2 (*30*)]. We counted the number of pathogenic variants in each individual for both heterozygous and homozygous forms and presented their distributions across 47 population groups (Fig. 5A). West Eurasians show a slightly higher load of pathogenic heterozygous variants (on average 6 to 8), likely reflecting bias in variant discovery in European populations (*63*). The accumulation of pathogenic variants exceeding two homozygotes was observed exclusively in the Nivkhs and Andamanese and Baining islanders. Most individuals (99.1%) carry at least one pathogenic allele. On average, individuals harbor five heterozygous variants (ranging from 0 to 13) and one homozygous variant (ranging from 0 to 6) (Fig. 5C).

Additionally, we identified 77 variants associated with adverse drug reactions (DrugBank, https://go.drugbank.com) and reported their allele frequencies [table S12.3, supplementary text S12.2 (*30*)]. Compared to pathogenic variants, the number of variants associated with adverse drug reactions per individual is higher (Fig. 5B). On average, individuals carried 19 such variants in a heterozygous form (ranging from 7 to 34) and 7 in a homozygous form (ranging from 1 to 19) (Fig. 5C). West Eurasians again show slightly higher numbers of heterozygous variants, whereas African Yoruba and Baining populations exhibit more homozygous variants.

Discussion

Using the GA100K:NENA dataset, we have shown that the genetic composition of Siberian populations is defined by six ancestral lineages (Fig. 1), and that West Siberian ancestry is shared by all contemporary Siberians, as well as Northeast Europeans and Central Asians. Large-scale genomic sequencing has allowed for more precise refinement of genetic ancestry that was previously studied using genotyping data (*6, 7*). The West Siberian ancestors were numerous around 10,000 ya and then gradually declined (Fig. 2D). Their population decrease during the early Holocene may have been due to global warming and the extinction of northern megafauna (e.g., mammoth (*64*)), which would have had a significant impact on nomadic hunter-gatherer populations in the north. In contrast, western and eastern populations that were expanding during the Neolithic Revolution in the early Holocene (*65*) likely benefited from the transition to agricultural and sedentary lifestyles.

Our analysis of whole-genome datasets also allowed us to infer the split time between North Eurasians and Native Americans, which occurred between 26,800 and 19,300 ya (Fig. 2D, 2E). This finding is consistent with estimates based on the recently published paleontological discovery of human footprints in North America (south-central New Mexico) dating back to 23,000 and 21,000 ya (*14*), as well as with other genetic studies, despite differences in the cohorts that were investigated (*3, 9, 18, 31*). A previous study of ancient genomes suggests limited genetic continuity in Beringia, as the most recent Arctic colonization occurred 6,000 ya (*10*). Therefore, it is likely that the first ancestors of the Native Americans in this region were replaced by the most recent wave of migration. We could not identify a specific Siberian group as direct Native American ancestors among the contemporary Indigenous populations in our dataset. However, we show that west Beringian populations, such as Inuit, Luoravetlans, and Koryaks, are genetically the closest to Native Americans (Fig. 2A). Moreover, we reveal the gene flow from Native Americans back to Inuit and Koryaks in Chukotka and the Kamchatka Peninsula between 700 to 5,100 ya [fig. S5.18, and supplementary text S7.3 (*30*)]. Our analyses also demonstrate the shared ancestry between the west Beringian populations and contemporary Native North Americans, particularly the Chipewyan from Canada [fig. S5.11 (*30*)]. This genetic relatedness is consistent with the PCA results (Fig. 2A). These findings are in line with previous reports that describe multiple waves of Northeast Asian gene flow into North Americans (*66*), including Neo-Inuit lineages (*10*).

Using our genome sequencing data from diverse Native South Americans, we have discovered that the simultaneous split of the four Native South American ancestral lineages occurred between 13,900 and 10,000 ya, from a common ancestral population in Mesoamerica (Fig. 2C-2E). This rapid radial dispersal and the establishment of sedentary settlements across South America are supported by previous genetic studies (*18, 67*) and the archaeological findings of early technologies (such as stone tools) that indicate regional cultural diversification in South America by at least 13,000 ya (*19*). This divergence occurred shortly after the split of the ancestral Native American lineages into northern and southern branches, which happened between 17,500 and

14,600 ya south of the North American ice sheets (*9, 18*). By the time the Ice-Free Corridor was fully opened 14,300-13,300 ya (*15*) during the abrupt warming, humans were already widely dispersed in North America (*16*).

Our study shows that the human migration across South America resulted in population splits with a loss of genetic diversity due to founder effects. Geographical and environmental boundaries caused population isolation and further enhanced the genetic homogenization, similar to islander populations (Fig. 3). The demographic history has significantly influenced the Patagonian Kawésqar, whose ancestors migrated the farthest distance out of Africa. They have the smallest effective population size (Fig. 2D) and one of the smallest genetic distances between community members (Fig. 3C). It has been reported that contemporary Native Patagonians (including the Kawésqar) show the highest genetic affinity to ancient Patagonian maritime individuals that lived one thousand years ago, indicating genetic continuity in the region (*18, 28, 29*). Our study cannot provide evidence for the reported back migration from the Southern Cone along South America's Atlantic coast (*20*), due to a lack of data on east coastal Native South American populations.

Our study also suggests that close genetic relatedness in Indigenous populations, along with reduced heterozygosity in *HLA* genes, may impact antigen recognition ability to new unexposed pathogens. In combination with socioeconomic factors and limited access to medical care, this could pose a potential health risk (*68*). High pathogen load regions, such as Southeast Asia, tend to have a higher diversity of promiscuous *HLA-DRB1* alleles, which allows to respond to a wider range of extracellular pathogens (*69*). However, emerging evidence that divergent allele advantage (a mechanism where the *HLA* genotypes present a broader set of epitopes) (*70*) and increase in *HLA* alleles promiscuity level (*69*) may counterplay the effect of loss of heterozygosity in *HLA* genes. Our work highlights an important implication for future research in population-based disease cohorts: epitope-binding repertoire studies are essential for identifying the dynamic effects of limited *HLA* diversity on disease susceptibility.

Access to the vastness of South American continent was constrained by the relatively small landmass of the Isthmus of Panama. Consequently, migrating groups could only inhabit the continent from a singular direction, significantly limiting the genetic diversity of human individuals. This ultimately led to the emergence of the four ancestries described in our analysis. While Indigenous groups managed to maintain their populations for over 13 millennia with minimal interaction with other groups, their endurance faced a critical challenge with the arrival of the initial colonists in the 1600s.

It is important for public health authorities to develop special measures of protection and interaction with Indigenous populations to minimize the spread of infections and improve medical support. We also show that the frequency of variants associated with adverse drug events and susceptibility to disease can substantially differ even among genetically related populations [see supplementary text S13 (*30*)]. Understanding these patterns of diversity is important to future genome-wide association studies (GWAS) and medical programs.

It is essential to emphasize the need for conserving the natural environment and respecting traditional lifestyles and food systems in Indigenous communities. In our study, we identified several genes under natural selection in Beringian populations that are adapted to the cold polar climate and a specific diet with low carbohydrate intake, as well as hypoxia resistance genes in Andean highlanders (Fig. 4). The lack of access to land resources, including environmental degradation (*71, 72*) with sustained pollution (*73, 74*), as well as decreasing biodiversity (*75, 76*), along with deprivation from traditional diets and lifestyle (*77*) and cultural loss, including the loss of traditional languages (*78*), have put some Indigenous communities at risk of extinction (*79*). Through this research, we aim to emphasize the special needs of Indigenous peoples and the importance of conserving their environment in the modern world.

Materials and methods summary

We sequenced 799 individual genomes and combined them with 738 genome data from the previous study (*80*) [see supplementary text S1 (*30*)]. Whole-genome sequencing libraries (Illumina TruSeq DNA Nano) of the genomes were sequenced on Illumina HiseqX platform [see supplementary Materials and Methods (*30*)]. The sequence reads were aligned to the human reference GRCh37 (human_g1k_v37_decoy.fasta). All genomes have been sequenced with an average coverage greater than 20X [table S1.2 (*30*)]. The final datasets GA100K:NENA includes 1,537 genomes and 52,663,159 variants: 52,589,813 SNPs (31,334,646 novel SNPs), including 21,217,063 singletons, and 26,398 INDELs (23,396 novel INDELs) [see supplementary text S2 (*30*)]. Three-field resolution HLA alleles were called using HLA-HD v.1.3.0 (*81*) based on IPD-IMGT/HLA database v.3.43.0 (*48*) [see supplementary Materials and Methods and supplementary text S9 (*30*)]. To identify cryptic relatedness between individuals, we used the identical-by-state (IBS) analysis (*82*) and excluded 60 first-degree relatives ending up with 1,477 unrelated individuals for further analyses [see supplementary text S3 (*30*)]. We phased the whole genomes using Shapeit v.4 (*83*) [see supplementary text S6 (*30*)]. We applied a panel of approaches, such as ADMIXTURE (*39*), local ancestry analysis (*40*), principal component analysis (PCA) (*42*), uniform manifold approximation and projection (UMAP) (*84*) methods [see supplementary text S5 (*30*)], as well as ALDER v.1.03 (*85*) and MALDER v.1.0 (*86*) metrics, and qpGraph (*45*) for demographic modelling [see supplementary text S7 (*30*)]. We also inferred the dynamics of the population sizes over time and population splits by Relate (*43*), MSMC (*87*), and MSMC-IM (*44*) [see supplementary text S7 (*30*)]. We calculated genome-wide runs of homozygosity (ROHs) segments by the observational genotype-counting algorithm (*47*) as implemented in PLINK v.1.9 (*88*) [see supplementary text S8 (*30*)]. The nucleotide diversity, π, was calculated using VCFtools v.0.1.17 (*89*). The genome-wide scan for detecting positive selection was performed using cross-population extended haplotype homozygosity analysis (*90*) (XP-EHH) and Integrated Haplotype Score (*91*) (iHS) [see supplementary text S11 (*30*)]. To identify specific variant (s) favoured by selection, we applied the integrated Selection of Allele Favoured by Evolution (iSAFE) (*49*) method [see supplementary text S11 (*30*)]. The annotation of the GA100K:NENA genetic variants to identify adverse drug reactions and clinically relevant pathogenic effects was performed using the DrugBank v.2020-11-02 (https://go.drugbank.com/, exported 2021-01-03) and ClinVar v.20190305 (https://www.ncbi.nlm.nih.gov/clinvar/) databases, respectively [see supplementary text S12, S13 (*30*)].

Figure captions

Fig. 1. Population structure and admixture. (**A**) Admixture (*39*) plot for 1,477 individuals showing ancestor lineages at optimal K=18 across geographical regions. (**B**) Local ancestry analysis (*40*) shows the distribution of the ancestral components at optimal K=18 across diverse geographical regions on the map. Representative ethnic groups are shown on the map. The geographical origin of genotyping reference data (open sources (*31, 32*)) is also indicated by grey dots.

Fig. 2. Demographic history and population dynamics. (**A**-**C**) Principal components analyses (*42*) (PCA) show the projection of individuals on the coordinates of the first two principal components (PC1 and PC2). The corresponding color indicates population groups. Shapes indicate the source of the data. Combined genome-wide GA100K:NENA sequencing data and reference genotyping data were used for global projection depicted in **A** (Africans were excluded from this representation). (**B**) PCA for Northeast European and North Eurasian populations (all admixed individuals were excluded from this representation). (**C**) PCA for Native Meso- and South Americans (all individuals with non-Native American admixture were excluded from this representation). (**D**) Relate (*43*) analysis depicted by the step histogram plot displays the dynamics of the effective population sizes between 100,000 ya to 1,000 ya. Shadow projections on the X-axis (Years ago) indicate corresponding population split periods. Colors indicate corresponding populations. (**E**) Best-fitting tree model of the relationships between contemporary populations, inferred by qpGraph (*45*). Tree edges (solid lines) are labeled with branch lengths in 1,000 times drift units, while admixtures (dotted arrows) are shown with their ancestral nodes (the ancestry proportion in %). The divergence time, as estimated by Relate, is indicated at the tree nodes. The admixture time is indicated in dotted circles, inferred by the MSMC-IM (*44*) and ALDER (*85*) analyses.

Fig. 3. Population demographic history and diversity of *HLA* genes. (A) The average total number of ROHs vs. the average nucleotide diversity, p, per population. The slope of the linear regression line (grey dotted line) indicates the relationship between the two variables. Pearson regression coefficient indicates a highly significant correlation between the two parameters ($R_{Pearson}$>|0.7|). (**B**) The average total number of ROHs vs. the average total number of unique *HLA* alleles across eight genes (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-DPB1*), per population. The slope of the linear regression line (grey dotted line) indicates the relationship between the two variables. Pearson's regression coefficient indicates a highly significant correlation between the two parameters. (**C**) Relative distribution of the *HLA* diversity to nucleotide diversity, p, in various populations across the continents. The range of the average total number of the unique *HLA* alleles across the eight genes is indicated by the colors of the circles. The range of the nucleotide diversity is depicted by the size of the circles.

Fig. 4. Adaptive variants. (**A**) Summary table of genes under natural selection identified by the XP-EHH (*90*), iHS (*91*), and iSAFE (*49*) analysis. (**B**) The selection signal in the *CPTA1* gene is related to cold adaptation. Allele frequencies of the rs80356779 (Pro479Leu) missense mutation in the *CPT1A* gene in Siberians, East Asians, and Native North Americans are shown as pie charts.

*Indicates reported literature data (*52-55, 92, 93*). **C,** The selection signal in the *LPAR1* gene is related to thermogenesis and cold adaptation. Allele frequency of the 3' UTR variant rs1043128 in Siberians, Indigenous Beringian populations, and East Asians are shown as pie charts. (**D**) The selection signal in the *EPAS1* gene is related to hypoxia adaptation in highlanders of South America. Allele frequencies of the rs570553380 (Pro479Leu) missense mutation in the *EPAS1* gene in Native South Americans are shown as pie charts. *Indicates reported literature data (*62*). Quechua groups: Al – Alota, COP/CAN – Copacabana/Candelaria, SJ – San Juan.

Fig. 5. Genetic load of medically relevant variants in GA100K:NENA. Distribution of the total number of genetic variants associated with pathogenic traits (**A**) and adverse medical drug reactions (**B**) per individual in 47 selected population groups across seven geographical regions, as reported in ClinVar v. 20190305 and DrugBank databases. Two bar plots for each population group indicate heterozygous (lighter color in left) and homozygous (solid color in right) forms. (**C**) The total number of pathogenic variants (n=529, ClinVar v.20190305) and variants associated with the adverse medical drug reactions (n=77, DrugBank) in heterozygous and homozygous forms per individual is presented in violin plots.

References

1. Q. Fu *et al.*, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).
2. V. V. Pitulko *et al.*, The Yana RHS site: humans in the Arctic before the last glacial maximum. *Science* **303**, 52-56 (2004).
3. M. Sikora *et al.*, The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182-188 (2019).
4. I. Sablin, M. Savelyeva, Mapping Indigenous Siberia: Spatial Changes and Ethnic Realities, 1900–2010. *Settler Colonial Studies* **1**, 77-110 (2011).
5. I. Pugach *et al.*, The Complex Admixture History and Recent Southern Origins of Siberian Populations. *Mol Biol Evol* **33**, 1777-1795 (2016).
6. K. Tambets *et al.*, Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biol* **19**, 139 (2018).
7. C. Jeong *et al.*, The genetic history of admixture across inner Eurasia. *Nat Ecol Evol* **3**, 966-976 (2019).
8. E. H. Wong *et al.*, Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res* **27**, 1-14 (2017).
9. J. V. Moreno-Mayar *et al.*, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203-207 (2018).
10. M. Raghavan *et al.*, The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
11. P. Flegontov *et al.*, Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* **570**, 236-240 (2019).
12. M. Jakobsson *et al.*, Post-glacial flooding of the Beringia Land Bridge dated to 11,000 cal yrs BP based on new geophysical and sediment records. *Clim. Past Discuss.* **13**, CP, 13, 991–1005 (2017).
13. M. Rasmussen *et al.*, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225-229 (2014).
14. M. R. Bennett *et al.*, Evidence of humans in North America during the Last Glacial Maximum. *Science* **373**, 1528-1531 (2021).
15. J. Clark *et al.*, The age of the opening of the Ice-Free Corridor and implications for the peopling of the Americas. *Proc Natl Acad Sci U S A* **119**, e2118558119 (2022).
16. L. Becerra-Valdivia, T. Higham, The timing and effect of the earliest human arrivals in North America. *Nature* **584**, 93-97 (2020).
17. S. K. Praetorius *et al.*, Ice and ocean constraints on early human migrations into North America along the Pacific coast. *Proc Natl Acad Sci U S A* **120**, e2208738120 (2023).
18. J. V. Moreno-Mayar *et al.*, Early human dispersals within the Americas. *Science* **362**, (2018).
19. T. D. Dillehay, Probing deeper into first American studies. *Proc Natl Acad Sci U S A* **106**, 971-978 (2009).
20. A. L. Campelo Dos Santos *et al.*, Genomic evidence for ancient human migration routes along South America's Atlantic coast. *Proc Biol Sci* **289**, 20221078 (2022).
21. E. S. M. A. Castro *et al.*, Population Histories and Genomic Diversity of South American Natives. *Mol Biol Evol* **39**, (2022).
22. J. R. Homburger *et al.*, Genomic Insights into the Ancestry and Demographic History of South America. *PLoS Genet* **11**, e1005602 (2015).
23. C. Barbieri *et al.*, The Current Genomic Landscape of Western South America: Andes, Amazonia, and Pacific Coast. *Mol Biol Evol* **36**, 2698-2713 (2019).
24. P. Skoglund *et al.*, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104-108 (2015).

25. C. Posth *et al.*, Reconstructing the Deep Population History of Central and South America. *Cell* **175**, 1185-1197 e1122 (2018).
26. G. A. Gnecchi-Ruscone *et al.*, Dissecting the Pre-Columbian Genomic Ancestry of Native Americans along the Andes-Amazonia Divide. *Mol Biol Evol* **36**, 1254-1269 (2019).
27. E. Arango-Isaza *et al.*, The genetic history of the Southern Andes from present-day Mapuche ancestry. *Curr Biol* **33**, 2602-2615 e2605 (2023).
28. C. de la Fuente *et al.*, Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc Natl Acad Sci U S A* **115**, E4006-E4012 (2018).
29. N. Nakatsuka *et al.*, Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. *Nat Commun* **11**, 3868 (2020).
30. Materials and methods and additional supplementary text sections are available as supplementary materials.
31. M. Raghavan *et al.*, POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
32. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206 (2016).
33. A. Margaryan *et al.*, Population genomics of the Viking world. *Nature* **585**, 390-396 (2020).
34. N. Nakatsuka *et al.*, A Paleogenomic Reconstruction of the Deep Population History of the Andes. *Cell* **181**, 1131-1145 e1121 (2020).
35. J. Lindo *et al.*, The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci Adv* **4**, eaau4921 (2018).
36. D. M. Fernandes *et al.*, A genetic history of the pre-contact Caribbean. *Nature* **590**, 103-110 (2021).
37. H. Schroeder *et al.*, Origins and genetic legacies of the Caribbean Taino. *Proc Natl Acad Sci U S A* **115**, 2341-2346 (2018).
38. K. Nagele *et al.*, Genomic insights into the early peopling of the Caribbean. *Science* **369**, 456-460 (2020).
39. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664 (2009).
40. E. Elhaik *et al.*, Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun* **5**, 3513 (2014).
41. N. Moshkov, A. Smetanin, T. V. Tatarinova, Local ancestry prediction with PyLAE. *PeerJ* **9**, e12502 (2021).
42. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
43. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* **51**, 1321-1329 (2019).
44. K. Wang, I. Mathieson, J. O'Connell, S. Schiffels, Tracking human population structure through time from whole genome sequences. *PLoS Genet* **16**, e1008552 (2020).
45. M. Lipson, D. Reich, A Working Model of the Deep Relationships of Diverse Modern Human Genetic Lineages Outside of Africa. *Mol Biol Evol* **34**, 889-902 (2017).
46. D. J. Kennett *et al.*, South-to-north migration preceded the advent of intensive farming in the Maya region. *Nat Commun* **13**, 1530 (2022).
47. T. J. Pemberton *et al.*, Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* **91**, 275-292 (2012).
48. J. Robinson *et al.*, IPD-IMGT/HLA Database. *Nucleic Acids Res* **48**, D948-D955 (2020).
49. A. Akbari *et al.*, Identifying the favored mutation in a positive selective sweep. *Nat Methods* **15**, 279-282 (2018).

50. N. F. Brown *et al.*, Molecular characterization of L-CPT I deficiency in six patients: insights into function of the native enzyme. *J Lipid Res* **42**, 1134-1142 (2001).
51. Y. J. Li, T. H. Chen, Y. Z. Wu, Y. H. Tseng, Metabolic and Nutritional Issues Associated with Spinal Muscular Atrophy. *Nutrients* **12**, (2020).
52. S. Zhou *et al.*, Genetic architecture and adaptations of Nunavik Inuit. *Proc Natl Acad Sci U S A* **116**, 16012-16017 (2019).
53. C. Rajakumar *et al.*, Carnitine palmitoyltransferase IA polymorphism P479L is common in Greenland Inuit and is associated with elevated plasma apolipoprotein A-I. *J Lipid Res* **50**, 1223-1228 (2009).
54. C. R. Greenberg *et al.*, The paradox of the carnitine palmitoyltransferase type Ia P479L variant in Canadian Aboriginal populations. *Mol Genet Metab* **96**, 201-207 (2009).
55. Q. Li *et al.*, The distribution of three candidate cold-resistant SNPs in six minorities in North China. *BMC Genomics* **19**, 134 (2018).
56. R. Jansen *et al.*, Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet* **26**, 1444-1451 (2017).
57. M. C. van Patot, M. Gassmann, Hypoxia: adapting to high altitude by mutating EPAS-1, the gene encoding HIF-2alpha. *High Alt Med Biol* **12**, 157-167 (2011).
58. J. Xu *et al.*, EPAS1 Gene Polymorphisms Are Associated With High Altitude Polycythemia in Tibetans at the Qinghai-Tibetan Plateau. *Wilderness Environ Med* **26**, 288-294 (2015).
59. C. M. Beall *et al.*, Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci U S A* **107**, 11459-11464 (2010).
60. A. W. Bigham *et al.*, Andean and Tibetan patterns of adaptation to high altitude. *Am J Hum Biol* **25**, 190-197 (2013).
61. T. S. Simonson, D. A. McClain, L. B. Jorde, J. T. Prchal, Genetic determinants of Tibetan high-altitude adaptation. *Hum Genet* **131**, 527-533 (2012).
62. C. A. Eichstaedt *et al.*, Evidence of Early-Stage Selection on EPAS1 and GPR126 Genes in Andean High Altitude Populations. *Sci Rep* **7**, 13042 (2017).
63. G. Sirugo, S. M. Williams, S. A. Tishkoff, The Missing Diversity in Human Genetic Studies. *Cell* **177**, 1080 (2019).
64. D. Nogues-Bravo, J. Rodriguez, J. Hortal, P. Batra, M. B. Araujo, Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biol* **6**, e79 (2008).
65. E. Borzenkova *et al.*, in *Second Assessment of Climate Change for the Baltic Sea Basin* T. B. I. A. Team, Ed. (Springer, 2015), pp. 25-49.
66. D. Reich *et al.*, Reconstructing Native American population history. *Nature* **488**, 370-374 (2012).
67. T. Ferraz *et al.*, Genomic history of coastal societies from eastern South America. *Nat Ecol Evol* **7**, 1315-1330 (2023).
68. N. Abd El-Baky, A. A. Amara, E. M. Redwan, HLA-I and HLA-II Peptidomes of SARS-CoV-2: A Review. *Vaccines (Basel)* **11**, (2023).
69. M. Manczinger *et al.*, Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol* **17**, e3000131 (2019).
70. F. Pierini, T. L. Lenz, Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. *Mol Biol Evol* **35**, 2145-2158 (2018).
71. J. E. Fa *et al.*, Importance of Indigenous Peoples' lands for the conservation of Intact Forest Landscapes. *Front Ecol Environ* **18**, 135-140 (2020).
72. L. Etchart, The role of indigenous peoples in combating climate change. *Palgrave Communications* **3**, 1-4 (2017).
73. J. Siqueira-Gay, L. E. Sánchez, The outbreak of illegal gold mining in the Brazilian Amazon boosts deforestation *Reg Environ Change* **21:28**, (2021).

74. S. Villen-Pérez, L. Anaya-Valenzuela, D. Conrado da Cruz, P. M. Fearnside, Mining threatens isolated indigenous peoples in the Brazilian Amazon. *Global Environmental Change* **72**, (2022).
75. E. Dinerstein *et al.*, A Global Deal For Nature: Guiding principles, milestones, and targets. *Sci Adv* **5**, eaaw2869 (2019).
76. I. Leiper *et al.*, Quantifying current and potential contributions of Australian indigenous peoples to threatened species management. *Conserv Biol* **32**, 1038-1047 (2018).
77. S. Lemke, T. Delormier, Indigenous Peoples' food systems, nutrition, and gender: Conceptual and methodological considerations. *Matern Child Nutr* **13 Suppl 3**, (2017).
78. E. L. Vandehey, University Honors College, (2021).
79. R. S. Walker, D. C. Kesler, K. R. Hill, Are Isolated Indigenous Populations Headed toward Extinction? *Plos One* **11**, e0150987 (2016).
80. K. C. GenomeAsia, The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106-111 (2019).
81. S. Kawaguchi, K. Higasa, M. Shimizu, R. Yamada, F. Matsuda, HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat* **38**, 788-797 (2017).
82. E. L. Stevens *et al.*, Inference of relationships in population data using identity-by-descent and identity-by-state. *PLoS Genet* **7**, e1002287 (2011).
83. O. Delaneau, J. Marchini, J. F. Zagury, A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181 (2012).
84. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, (2018).
85. P. R. Loh *et al.*, Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233-1254 (2013).
86. J. K. Pickrell *et al.*, Ancient west Eurasian ancestry in southern and eastern Africa. *Proc Natl Acad Sci U S A* **111**, 2632-2637 (2014).
87. S. Schiffels, K. Wang, MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol Biol* **2090**, 147-166 (2020).
88. C. C. Chang *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
89. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
90. J. K. Pickrell *et al.*, Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**, 826-837 (2009).
91. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).
92. S. Zhou *et al.*, Increased missense mutation burden of Fatty Acid metabolism related genes in nunavik inuit population. *Plos One* **10**, e0128255 (2015).
93. S. Y. Tereshchenko, M. V. Smolnikova, A pilot study of inherited carnitine palmitoyltransferase deficiency as an ethnogenetic risk factor of infant mortality in indigenous populations of the Far North. *Human Physiology* **42**, 145-149 (2016).
94. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
95. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
96. G. G. Faust, I. M. Hall, SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505 (2014).
97. A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, P. Prins, Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034 (2015).
98. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

99. W. Haak *et al.*, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).
100. RCoreTeam, *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2017).
101. C. A. Maclean, N. P. Chue Hong, J. G. Prendergast, hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets. *Mol Biol Evol* **32**, 3027-3029 (2015).
102. T. Hamala, O. Savolainen, Genomic Patterns of Local Adaptation under Gene Flow in Arabidopsis lyrata. *Mol Biol Evol* **36**, 2557-2571 (2019).
103. S. Sonoda, H. C. Li, K. Tajima, Ethnoepidemiology of HTLV-1 related diseases: ethnic determinants of HTLV-1 susceptibility and its worldwide dispersal. *Cancer Sci* **102**, 295-301 (2011).
104. S. Ijichi *et al.*, Identification of human T cell leukemia virus type IIb infection in the Wayu, an aboriginal population of Colombia. *Jpn J Cancer Res* **84**, 1215-1218 (1993).
105. S. Horai *et al.*, Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* **10**, 23-47 (1993).
106. T. Miura *et al.*, Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. *J Mol Evol* **44 Suppl 1**, S76-82 (1997).
107. T. Fujiyoshi *et al.*, Characteristic distribution of HTLV type I and HTLV type II carriers among native ethnic groups in South America. *AIDS Res Hum Retroviruses* **15**, 1235-1239 (1999).
108. H. C. Li *et al.*, The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. *Nat Med* **5**, 1428-1432 (1999).
109. E. D. Roberson, J. Pevsner, Visualization of shared genomic regions and meiotic recombination in high-density SNP data. *Plos One* **4**, e6711 (2009).
110. T. V. Tatarinova, Schumitzky A., *Nonlinear Mixture Models: A Bayesian Approach*. s. Edition, Ed., (Imperial College Press, ed. Kindle Edition).
111. T. Tatarinova, J. Bouck, A. Schumitzky, Kullback-Leibler Markov chain Monte Carlo--a new algorithm for finite mixture analysis and its application to gene expression data. *J Bioinform Comput Biol* **6**, 727-746 (2008).
112. K. Kozlov *et al.*, Differential Evolution approach to detect recent admixture. *BMC Genomics* **16 Suppl 8**, S9 (2015).
113. M. Raghavan *et al.*, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91 (2014).
114. A. Jimenez-Kaufmann *et al.*, Imputation Performance in Latin American Populations: Improving Rare Variants Representation With the Inclusion of Native American Genomes. *Front Genet* **12**, 719791 (2021).
115. M. Rasmussen *et al.*, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762 (2010).
116. R. Maier, P. Flegontov, O. Flegontova, P. Changmai, D. Reich, On the limits of fitting complex models of population history to genetic data. *bioRxiv* (2022).
117. S. Gravel, Population genetics models of local ancestry. *Genetics* **191**, 607-619 (2012).
118. B. Yunusbayev *et al.*, The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet* **11**, e1005068 (2015).
119. L. Ongaro *et al.*, The Genomic Impact of European Colonization of the Americas. *Curr Biol* **29**, 3974-3986 e3974 (2019).
120. A. S. Malaspinas *et al.*, A genomic history of Aboriginal Australia. *Nature* **538**, 207-214 (2016).
121. O. Delaneau, J. F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat Commun* **10**, 5436 (2019).

122. P. C. Sabeti *et al.*, Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918 (2007).
123. D. R. Schrider, F. K. Mendes, M. W. Hahn, A. D. Kern, Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics* **200**, 267-284 (2015).
124. D. J. Lemas *et al.*, Genetic polymorphisms in carnitine palmitoyltransferase 1A gene are associated with variation in body composition and fasting lipid traits in Yup'ik Eskimos. *J Lipid Res* **53**, 175-184 (2012).
125. L. Skotte *et al.*, CPT1A Missense Mutation Associated With Fatty Acid Metabolism and Reduced Height in Greenlanders. *Circ Cardiovasc Genet* **10**, (2017).
126. F. J. Clemente *et al.*, A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. *Am J Hum Genet* **95**, 584-589 (2014).
127. M. K. Andersen *et al.*, Identification of Novel Genetic Determinants of Erythrocyte Membrane Fatty Acid Composition among Greenlanders. *PLoS Genet* **12**, e1006119 (2016).
128. J. Xu, A. C. Donepudi, J. E. Moscovitz, A. L. Slitt, Keap1-knockdown decreases fasting-induced fatty liver via altered lipid metabolism and decreased fatty acid mobilization from adipose tissue. *Plos One* **8**, e79841 (2013).
129. L. Fagerberg *et al.*, Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**, 397-406 (2014).
130. N. M. Drake, L. M. DeVito, T. A. Cleland, P. D. Soloway, Imprinted Rasgrf1 expression in neonatal mice affects olfactory learning and memory. *Genes Brain Behav* **10**, 392-403 (2011).
131. P. G. Hysi *et al.*, A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. *Nat Genet* **42**, 902-905 (2010).
132. R. Nielsen, I. Hellmann, M. Hubisz, C. Bustamante, A. G. Clark, Recent and ongoing selection in the human genome. *Nat Rev Genet* **8**, 857-868 (2007).
133. E. H. Baker, Ion channels and the control of blood pressure. *Br J Clin Pharmacol* **49**, 185-198 (2000).
134. G. Galarza-Munoz, S. I. Soto-Morales, M. Holmgren, J. J. Rosenthal, Physiological adaptation of an Antarctic Na+/K+-ATPase to the cold. *J Exp Biol* **214**, 2164-2174 (2011).
135. J. Korfanty *et al.*, Identification of a new mouse sperm acrosome-associated protein. *Reproduction* **143**, 749-757 (2012).
136. F. E. Mackenzie *et al.*, Upregulation of PKD1L2 provokes a complex neuromuscular disease in the mouse. *Hum Mol Genet* **18**, 3553-3566 (2009).
137. A. Li, X. Tian, S. W. Sung, S. Somlo, Identification of two novel polycystic kidney disease-1-like genes in human and mouse genomes. *Genomics* **81**, 596-608 (2003).
138. F. Rothhammer, C. Silva, Peopling of Andean South America. *Am J Phys Anthropol* **78**, 403-410 (1989).
139. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
140. D. Ellinghaus *et al.*, Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med* **383**, 1522-1534 (2020).
141. E. Pairo-Castineira *et al.*, Genetic mechanisms of critical illness in Covid-19. *Nature*, (2020).
142. D. J. Downes *et al.*, Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat Genet* **53**, 1606-1615 (2021).
143. J. E. Huffman *et al.*, Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19. *Nat Genet*, (2022).
144. G. H. L. Roberts *et al.*, AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci. *medRxiv*, (2020).
145. J. F. Shelton *et al.*, Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet* **53**, 801-808 (2021).

146. B. A. Schaefer *et al.*, Genetic Modifiers of White Blood Cell Count, Albuminuria and Glomerular Filtration Rate in Children with Sickle Cell Anemia. *Plos One* **11**, e0164364 (2016).
147. K. Poulton *et al.*, A role for human leucocyte antigens in the susceptibility to SARS-Cov-2 infection observed in transplant patients. *Int J Immunogenet* **47**, 324-328 (2020).
148. E. Schindler *et al.*, HLA genetic polymorphism in patients with Coronavirus Disease 2019 in Midwestern United States. *HLA* **98**, 370-379 (2021).
149. Y. L. Yung *et al.*, Association of HLA-B22 serotype with SARS-CoV-2 susceptibility in Hong Kong Chinese patients. *HLA* **97**, 127-132 (2021).
150. W. Wang, W. Zhang, J. Zhang, J. He, F. Zhu, Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). *HLA* **96**, 194-196 (2020).
151. R. Littera *et al.*, Human Leukocyte Antigen Complex and Other Immunogenetic and Clinical Factors Influence Susceptibility or Protection to SARS-CoV-2 Infection and Severity of the Disease Course. The Sardinian Experience. *Front Immunol* **11**, 605688 (2020).
152. A. Amoroso *et al.*, HLA and AB0 Polymorphisms May Influence SARS-CoV-2 Infection and COVID-19 Severity. *Transplantation* **105**, 193-200 (2021).
153. M. Lin *et al.*, Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* **4**, 9 (2003).
154. C. Y. Ng *et al.*, Impact of the HLA-B(*)58:01 Allele and Renal Impairment on Allopurinol-Induced Cutaneous Adverse Reactions. *J Invest Dermatol* **136**, 1373-1381 (2016).
155. N. Keicho *et al.*, Association of human leukocyte antigen class II alleles with severe acute respiratory syndrome in the Vietnamese population. *Hum Immunol* **70**, 527-531 (2009).
156. S. F. Wang *et al.*, Human-leukocyte antigen class I Cw 1502 and class II DR 0301 genotypes are associated with resistance to severe acute respiratory syndrome (SARS) infection. *Viral Immunol* **24**, 421-426 (2011).
157. T. Ahmad, S. E. Marshall, D. Jewell, Genetics of inflammatory bowel disease: the role of the HLA complex. *World J Gastroenterol* **12**, 3628-3635 (2006).
158. E. L. Barnes, E. V. Loftus, Jr., M. D. Kappelman, Effects of Race and Ethnicity on Diagnosis and Management of Inflammatory Bowel Diseases. *Gastroenterology* **160**, 677-689 (2021).
159. R. Malhotra, K. Turner, A. Sonnenberg, R. M. Genta, High prevalence of inflammatory bowel disease in United States residents of Indian ancestry. *Clin Gastroenterol Hepatol* **13**, 683-689 (2015).

Acknowledgements: We would like to express our deepest gratitude to the Indigenous community members who participated in this study. Every effort will be made to communicate the results of this research to the Indigenous communities that participated in this study. We thank Kazuo Tajima, Shunro Sonoda, and Satoshi Horai for their efforts on sample collections. We also thank the Asian DNA Repository Consortium (ADRC) for their sincere cooperation and approval for using the sample collections they have stored for this study. Public reports summarizing the outcomes of this research will be released on the GenomeAsia 100K official website (https://genomeasia100k.org/) and will be available in several languages used in Asia and South America. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg). We would like to thank Prabu Sekar and Rikky W. Purbojati from the Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, for the administration of the computing resources to ensure data analysis. We thank Yukinori Okada for his sincere reading and comments. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg).

Funding: Russian Science Foundation grant № 22-64-00060, https://rscf.ru/project/22-64-00060/. (V.A.S, V.N.Kh, M.V.G., A.M., M.S.). The Ministry of Education, Singapore, Academic Research Fund Tier 1 grant (RG100/20) (H.L.K., N.A.B.R., A.G.G., and E.S.G.)

Author contributions: H.L.K., S.C.S., V.A.S. designed the study; S.N., V.A.S., V.N.K., A.M-E., C.M.B., M.S., S.Yu.T., E.W.K., M.O.R., M.V.S., I.Yu.K., N.A.K., K.V.V., E.R.E., A.S., O.S., H.L.K., E.S.G., I.D., K.T. A.T., H.T., Y.N., I.McG. coordinated, collected and/or provided samples, consulted on samples; D.I.D-M., D.P., P.N.A., C.F.L., Y.K., S.T.L., C.P., S.R.L., Z.H.Y., S.G.N., J.D. performed the library preparation and sequencing; C.E.H., N.K., K.L.K., H.L.K., S-S.K., N.A.B.R. managed, transferred, and stored the data under security and administrated the computing resources to ensure data analysis; N.K., A.R., H.L.K., S-S.K. carried out the data processing and quality control; E.S.G., A.G.G, A.Z., T.V.T., N.M., S.N., N.C., V.N.K., M.V.G., H.L.K., S-S.K., N.A.B.R., V.A.S. carried out data analyses; E.S.G., A.G.G, H.L.K., S.C.S., V.N.K., V.A.S. wrote the first draft of the paper; All authors approved the results and contributed writing of the final version of the manuscript.

Competing interests: The authors declare no competing interests.

Data and materials availability: For each variant, allele frequency data per population group were calculated and will be available publicly at the GenomeAsia 100K website (https://browser.genomeasia100k.org). Individual level VCF data including the 799 newly sequenced genomes and other genomes from our previous study (*80*) are deposited to the NBDC human database under accession number JGAS000781.

Supplementary Materials

Materials and Methods

Supplementary Text S1 to S13

Figs. S1.1 to S12.2

Tables S1.1 to S13.6

References (1-159)

RESEARCH ARTICLE SUMMARY

INTRODUCTION

During the late Pleistocene, humans expanded across Eurasia and eventually migrated to the Americas. Those who reached Patagonia, at the southern tip of South America, completed the longest migration out of Africa.

RATIONALE

The extent of basal divergences, admixture, and degrees of isolation among Indigenous North Eurasian and Native South American populations remain debated, with most insights derived from genome-wide genotyping data. This study aims to deepen our understanding of the ancient dynamics that shaped contemporary populations in North Eurasia and the Americas. Using large-scale whole-genome sequencing of 1,537 individuals from 139 ethnic groups in these regions, we examine population structures, elucidate prehistoric migrations, and explore the influence of past environmental factors on the diversification of human populations.

RESULTS

Advances in large-scale genomic sequencing have significantly enhanced our understanding of the genetic ancestry of human populations across North Eurasia and South America. Our analysis reveals that all contemporary Siberians, as well as some Northeast Europeans and Central Asians, share ancestry with the West Siberian groups, represented by the Kets and Nenets. Their ancestors were widespread across Siberia 10,000 years ago, but now these groups face population decline by 73.6%, and are becoming a minority.

The populations of west Beringia, including the Koryaks, Inuit, and Luoravetlans, are the most genetically distinct from other Siberians. These groups have adapted to Arctic conditions with genetic variations related to lipid metabolism, thermogenesis, sensory perception, and the regulation of reproductive and immune functions.

We were not able to identify a specific Siberian group as the direct ancestors of Native Americans due to deep divergence and limited genetic continuity. However, west Beringian populations remain closely related to Native Americans. Koryaks and Inuit show 5% and 28% Native American ancestry, respectively, due to gene flow between 700 and 5,100 years ago.

We estimated the split time of Native South Americas into Amazonians, Andeans, Chaco Amerindians, and Patagonians to have occurred 13,900-10,000 years ago. Migration and settlement across the continent led to population isolations due to geographic boundaries and a reduction in their genetic diversity, particularly affecting immune genes like the human leukocyte antigen (*HLA*). Over the past 10,000 years, all four Native South American lineages have experienced population declines ranging from 38% to 80%. This dramatic decline, combined with the loss of traditional lifestyles, cultural practices, and languages, has pushed some Indigenous communities, such as the Kawésqar, to the brink of extinction.

CONCLUSION

The migration to an uninhabited continent of South America through the narrow Isthmus of Panama resulted in a founder effect among Native South Americans, leading to reduced genetic diversity compared to Indigenous populations of North Eurasia. Over 13,900 years, geographic barriers within the continent further isolated Indigenous groups, further reducing genetic diversity. These groups faced a profound challenge with the arrival of European colonists in the 1600s, who introduced new adversities that threatened their long-standing endurance.

Summary Figure. Genetic ancestry and nucleotide diversity. Colors represent genetic ancestries estimated by whole genome sequencing data of contemporary human populations. Countries having no data remained empty. The size of pies indicates the average nucleotide diversity of each population.

Summary Figure

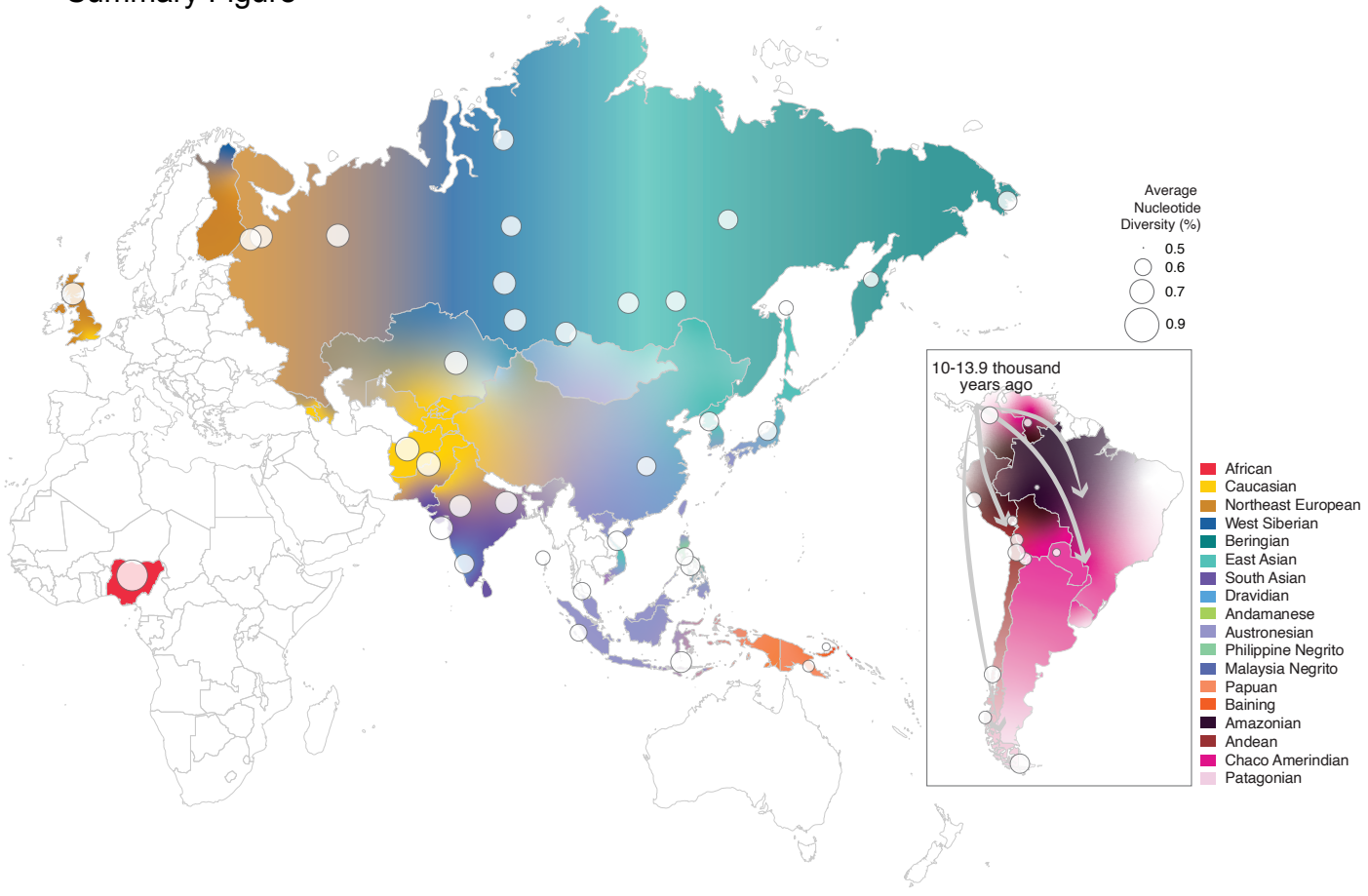


Figure 1

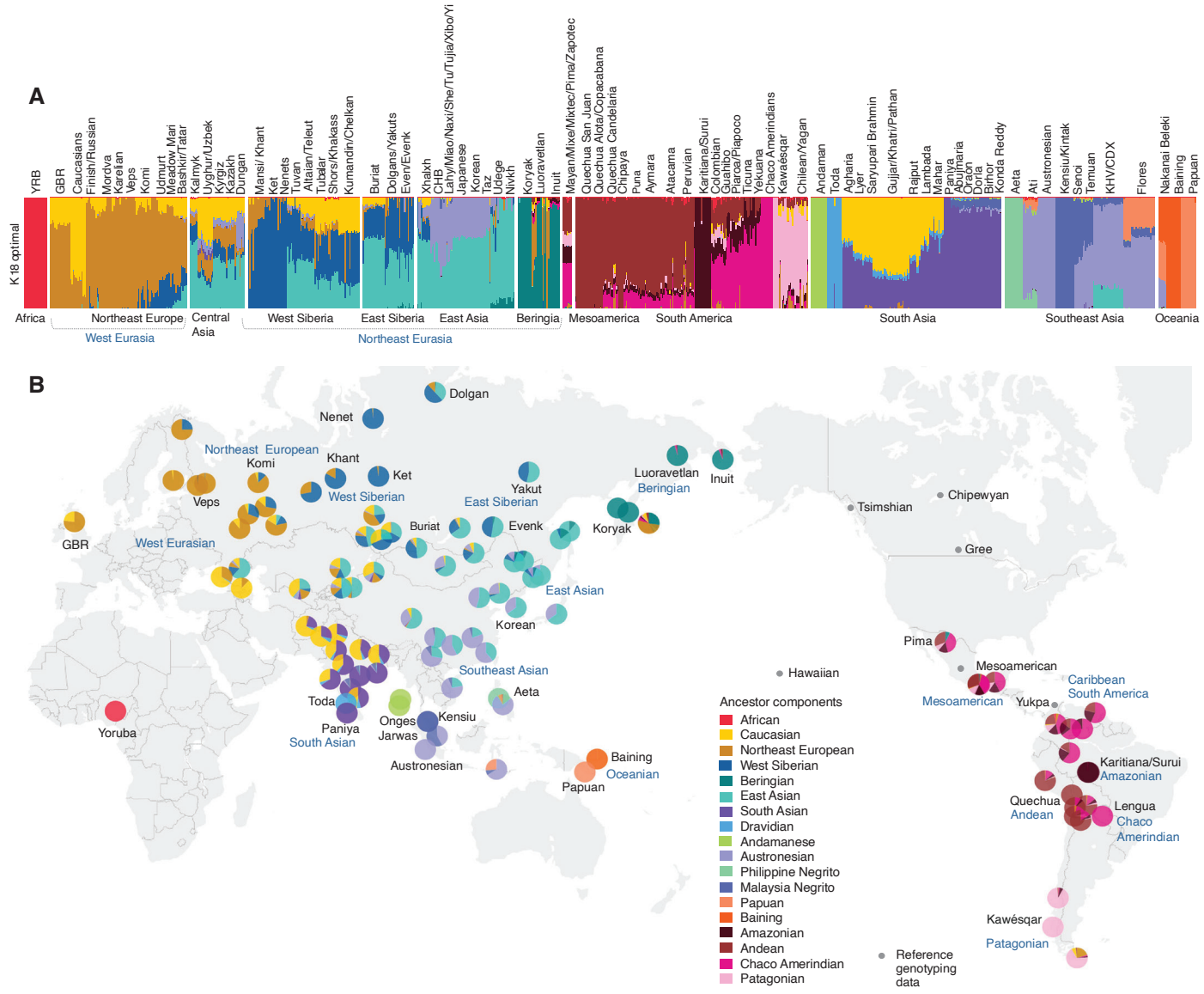

A
K 18 optimal
YRB
GBR
Caucasians
Finish/Russian
Mordva
Karelian
Veps
Komi
Udmurt
Meadow Mari
Bashkir/Tatar
Kalmyk
Uyghur/Uzbek
Kyrgiz
Kazakh
Dungan
Mansi/ Khant
Ket
Nenets
Tuvan
Altaian/Teleut
Tubalar
Shors/Khakass
Kumandin/Chelkan
Buriat
Dolgans/Yakuts
Even/Evenk
Xibo
CHB
Lahy/Miao/Naxi/She/Tu/Tujia/Xibo/Yi
Japanese
Korean
Taz
Udege
Nivkh
Koryak
Luoravetlan
Inuit
Mayan/Mixe/Mixtec/Pima/Zapotec
Quechua San Juan
Quechua Ajota/Copacabana
Quechua Candelaria
Chipaya
Puna
Aymara
Atacama
Peruvian
Karitiana/Surui
Colombian
Guahibo
Piaroa/Piapoco
Ticuna
Yekuana
Chaco Amerindians
Kawaésqar
Chilean/Yagan
Andaman
Toda
Agharia
Lepr
Saryupari Brahmin
Gujar/Khatri/Pathan
Rajput
Lambada
Mahar
Paniya
Adi/Hazra
Doria
Birhor
Konda Reddy
Aeta
Ati
Austronesian
Kensiu/Kintak
Semoi
Temuan
KHV/CDX
Flores
Nakanai Bileki
Baining
Papuan
Africa
Northeast Europe
West Eurasia
Central Asia
West Siberia
East Siberia
East Asia
Beringia
Northeast Eurasia
Mesoamerica
South America
South Asia
Southeast Asia
Oceania
B
Dolgan
Nenet
Northeast European
Komi
Khant
Ket
West Siberian
Veps
Yakut
East Siberian
Luoravetlan
Beringian
Inuit
Tsimshian
Chipewyan
GBR
West Eurasian
Buriat
Evenk
Koryak
Gree
East Asian
Korean
Pima
Southeast Asian
Mesoamerican
Hawaiian
Caribbean
South America
Aeta
Toda
Yukpa
Mesoamerican
Onges
Jarwas
Kensiu
Paniya
South Asian
Yoruba
Austronesian
Baining
Oceanian
Papuan
Karitiana/Surui
Amazonian
Quechua
Andean
Lengua
Chaco
Amerindian
Kawésqar
Patagonian
Ancestor components
African
Caucasian
Northeast European
West Siberian
Beringian
East Asian
South Asian
Dravidian
Andamanese
Austronesian
Philippine Negrito
Malaysia Negrito
Papuan
Baining
Amazonian
Andean
Chaco Amerindian
Patagonian
Reference genotyping data

Figure 2

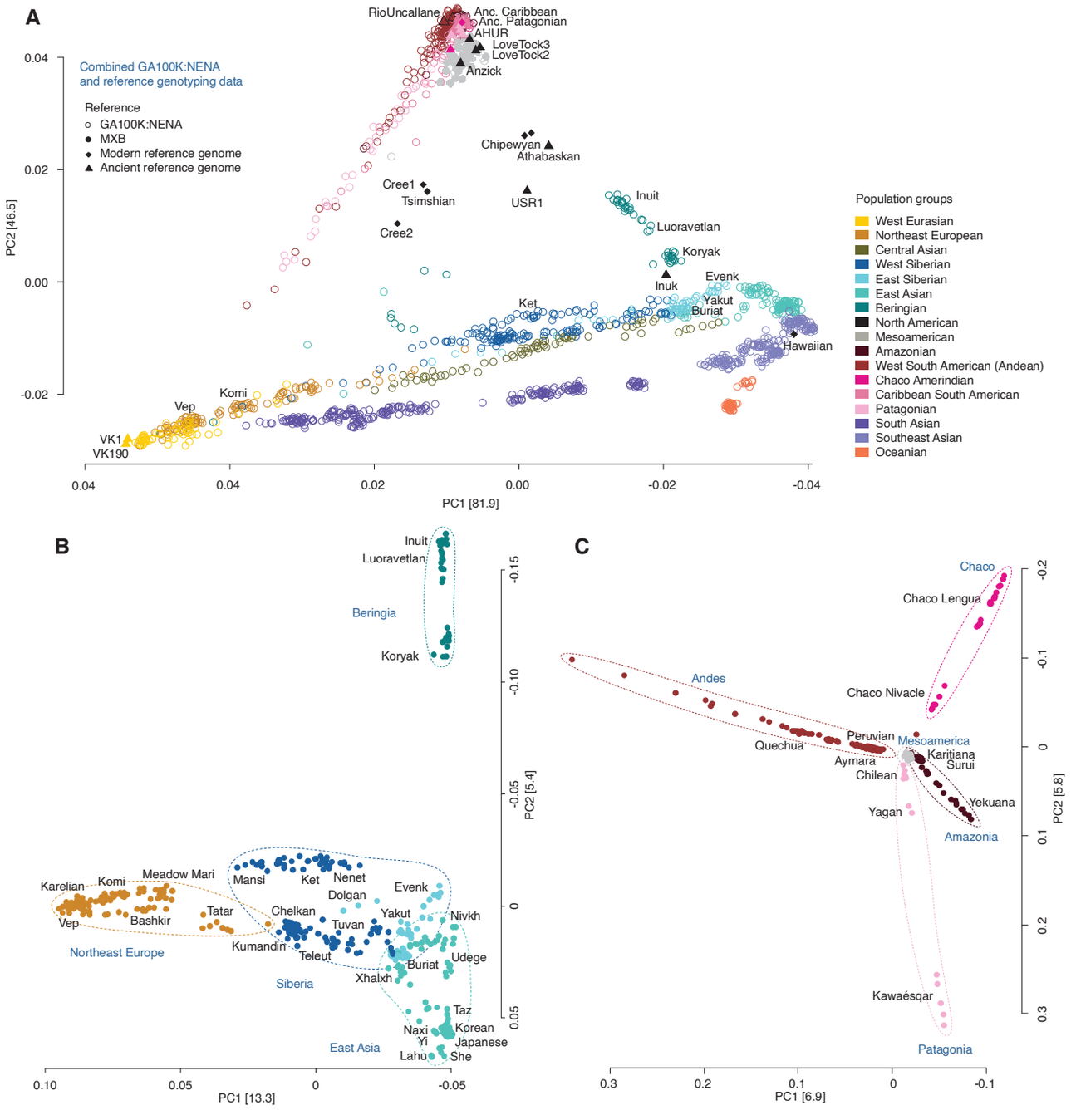

A
Combined GA100K:NENA
and reference genotyping data
Reference
GA100K:NENA
MXB
Modern reference genome
Ancient reference genome
RioUncallane
Anc. Caribbean
Anc. Patagonian
AHUR
LoveTock3
LoveTock2
Anzick
Chipewyan
Athabaskan
Cree1
Tsimshian
Cree2
USR1
Inuit
Luoravetlan
Koryak
Inuk
Evenk
Yakut
Buriat
Ket
Hawaiian
Komi
Vep
VK1
VK190
PC2 [46.5]
PC1 [81.9]
Population groups
West Eurasian
Northeast European
Central Asian
West Siberian
East Siberian
East Asian
Beringian
North American
Mesoamerican
Amazonian
West South American (Andean)
Chaco Amerindian
Caribbean South American
Patagonian
South Asian
Southeast Asian
Oceanian
B
Inuit
Luoravetlan
Beringia
Koryak
Karelian
Komi
Meadow Mari
Vep
Bashkir
Tatar
Mansi
Ket
Nenet
Dolgan
Evenk
Chelkan
Yakut
Nivkh
Tuvan
Kumandin
Teleut
Buriat
Udege
Xhalxh
Taz
Naxi
Yi
Korean
Japanese
Lahu
She
Northeast Europe
Siberia
East Asia
PC2 [5.4]
PC1 [13.3]
C
Chaco
Chaco Lengua
Andes
Chaco Nivacle
Quechua
Peruvian
Aymara
Chilean
Mesoamerica
Karitiana
Surui
Yekuana
Yagan
Amazonia
Kawaésqar
Patagonia
PC2 [5.8]
PC1 [6.9]
D
Effective population size
Time (years ago)
Diverging of
Northeast Europeans
(37.3 to 26.8 kya)
Split between
Northeast Asians
and Native Americans
(26.8 to 19.3 kya)
Split among
Native Americans
(13.9 to 10.0 kya)
E
Populations
Veps, Northeast Europeans
Kets, West Siberians
Koreans, East Asians
Koryaks, Beringians
Quechua (Copacabana), Andeans
Lengua, Chaco Amerindians
Karitiana, Amazonians
Kawésqar, Patagonian tribes
37.3-26.8 kya
26.8-19.3 kya
13.9-10.0 kya
15-8 kya
1.7-0.7 kya
23%
77%
95%
5%
Veps
Northeast
Europeans
Kets
West
Siberians
Koreans
East
Asians
Koryaks
Beringians
Quechua
Copacabana
Andeans
Lengua
Chaco
Amerindians
Karitiana
Amazonians
Kawésqar
Patagonians

Figure 3

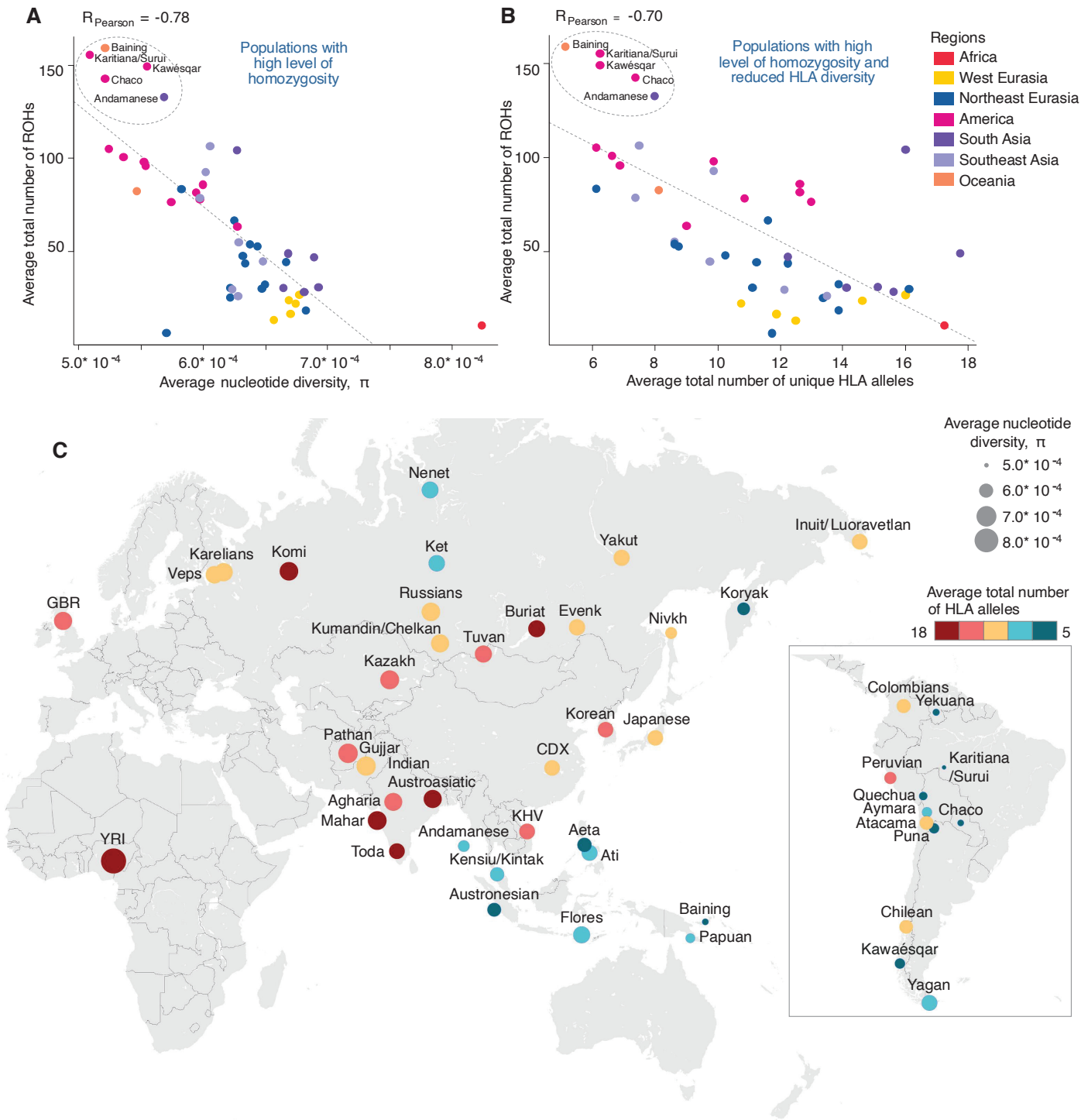

A
R Pearson = -0.78
Baining
Karitiana/Surui
Kawésqar
Chaco
Andamanese
Populations with high level of homozygosity
Average total number of ROHs
150
100
50
5.0* 10 -4
6.0* 10 -4
7.0* 10 -4
8.0* 10 -4
Average nucleotide diversity, π
B
R Pearson = -0.70
Baining
Karitiana/Surui
Kawésqar
Chaco
Andamanese
Populations with high level of homozygosity and reduced HLA diversity
Average total number of ROHs
150
100
50
6
8
10
12
14
16
18
Average total number of unique HLA alleles
Regions
Africa
West Eurasia
Northeast Eurasia
America
South Asia
Southeast Asia
Oceania
C
Average nucleotide diversity, π
5.0* 10 -4
6.0* 10 -4
7.0* 10 -4
8.0* 10 -4
Average total number of HLA alleles
18
5
Nenet
Inuit/ Luoravetlan
Yakut
Karelians
Veps
Komi
Ket
GBR
Russians
Koryak
Buriat
Evenk
Nivkh
Kumandin/Chelkan
Tuvan
Kazakh
Korean
Japanese
Pathan
Gujjar
Indian
CDX
Austroasiatic
Agharia
Mahar
KHV
Andamanese
Aeta
YRI
Toda
Ati
Kensiu/Kintak
Austronesian
Flores
Baining
Papuan
Colombians
Yekuana
Peruvian
Karitiana
/Surui
Quechua
Aymara
Chaco
Atacama
Puna
Chilean
Kawaésqar
Yagan

Figure 4

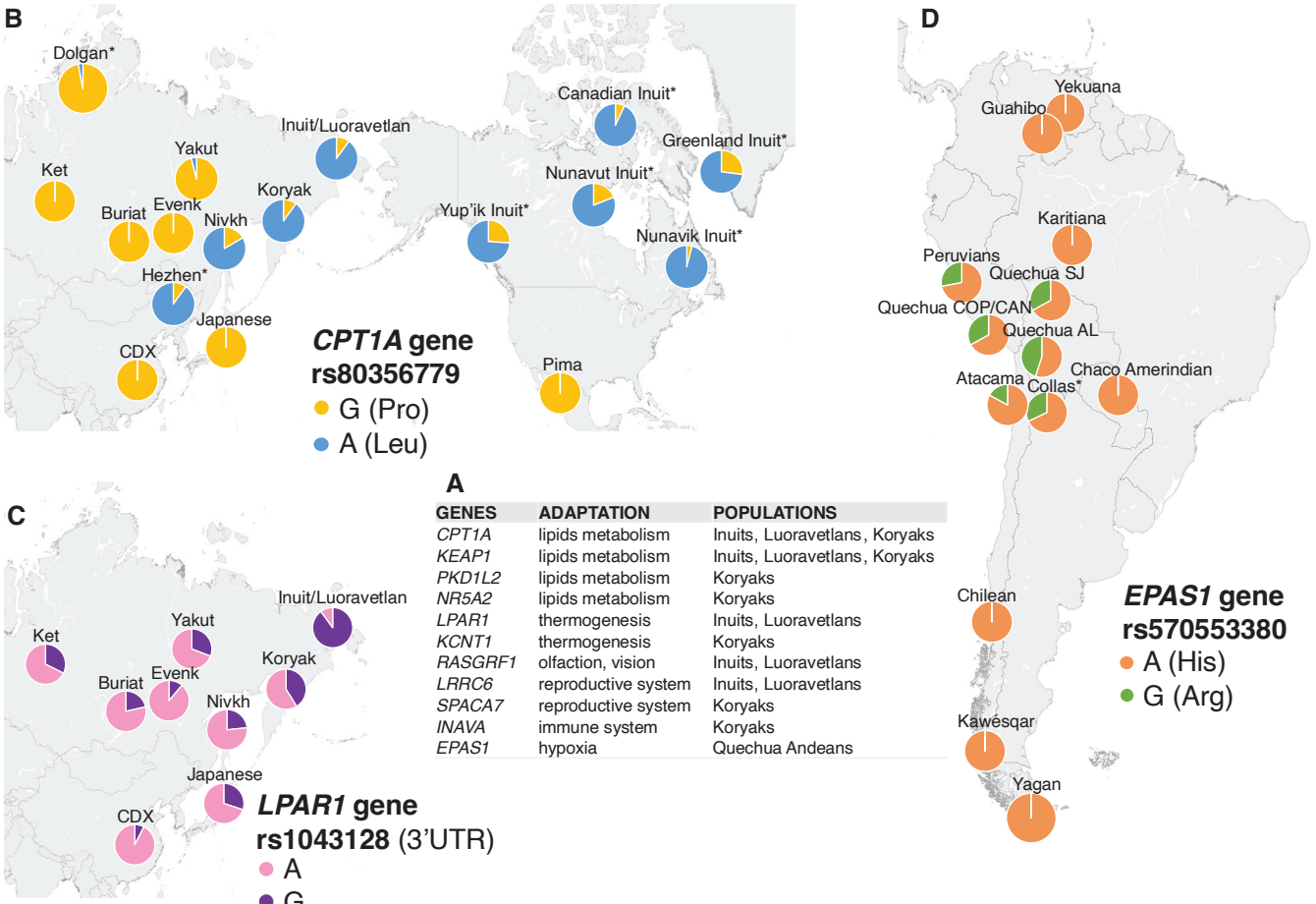
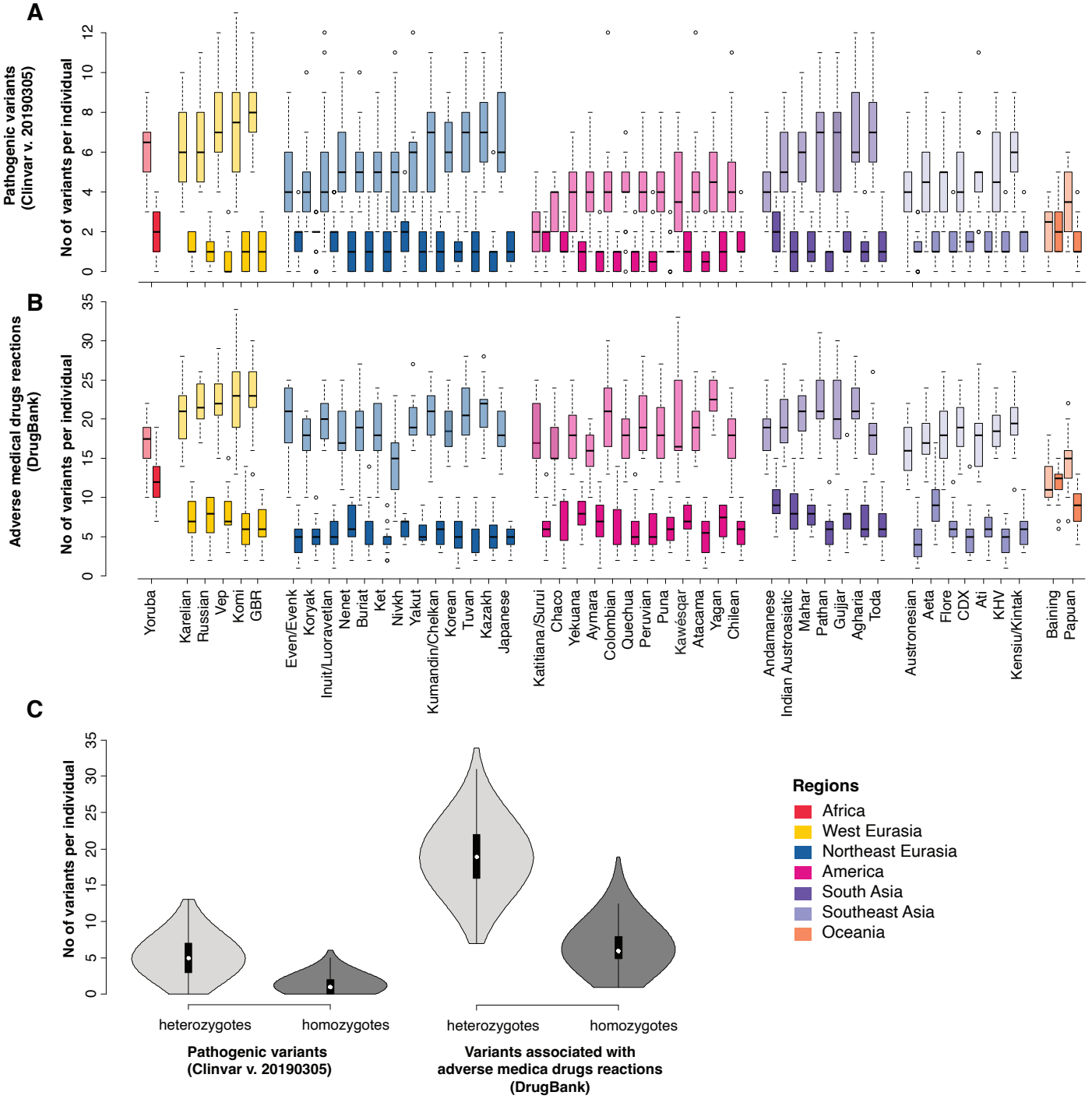

B
Dolgan*
Ket
Yakut
Buriat
Evenk
Nivkh
Koryak
Inuit/Luoravetlan
Hezhen*
Japanese
CDX
Canadian Inuit*
Greenland Inuit*
Nunavut Inuit*
Yup'ik Inuit*
Nunavik Inuit*
Pima
CPT1A gene
rs80356779
G (Pro)
A (Leu)
C
Ket
Yakut
Buriat
Evenk
Nivkh
Koryak
Inuit/Luoravetlan
Japanese
CDX
LPAR1 gene
rs1043128 (3'UTR)
A
G
A
GENES | ADAPTATION | POPULATIONS
CPT1A | lipids metabolism | Inuits, Luoravetlans, Koryaks
KEAP1 | lipids metabolism | Inuits, Luoravetlans, Koryaks
PKD1L2 | lipids metabolism | Koryaks
NR5A2 | lipids metabolism | Koryaks
LPAR1 | thermogenesis | Inuits, Luoravetlans
KCNT1 | thermogenesis | Koryaks
RASGRF1 | olfaction, vision | Inuits, Luoravetlans
LRRC6 | reproductive system | Inuits, Luoravetlans
SPACA7 | reproductive system | Koryaks
INAVA | immune system | Koryaks
EPAS1 | hypoxia | Quechua Andeans
D
Yekuana
Guahibo
Karitiana
Peruvians
Quechua SJ
Quechua COP/CAN
Quechua AL
Atacama
Collas*
Chaco Amerindian
Chilean
Kawésqar
Yagan
EPAS1 gene
rs570553380
A (His)
G (Arg)

Figure 5


A
Pathogenic variants (Clinvar v. 20190305)
No of variants per individual
0
2
4
6
8
10
12
B
Adverse medical drugs reactions (DrugBank)
No of variants per individual
0
5
10
15
20
25
30
35
Yoruba
Karelian
Russian
Vep
Komi
GBR
Even/Evenk
Koryak
Inuit/Luoravetlan
Nenet
Buriat
Ket
Nivkh
Yakut
Kumandin/Chelkan
Korean
Tuvan
Kazakh
Japanese
Katitiana/Surui
Chaco
Yekuana
Aymara
Colombian
Quechua
Peruvian
Puna
Kawésqar
Atacama
Yagan
Chilean
Andamanese
Indian Austroasiatic
Mahar
Pathan
Gujjar
Agharia
Toda
Austronesian
Aeta
Flore
CDX
Ati
KHV
Kensiu/Kintak
Baining
Papuan
C
No of variants per individual
0
5
10
15
20
25
30
35
heterozygotes
homozygotes
Pathogenic variants (Clinvar v. 20190305)
heterozygotes
homozygotes
Variants associated with adverse medica drugs reactions (DrugBank)
Regions
Africa
West Eurasia
Northeast Eurasia
America
South Asia
Southeast Asia
Oceania