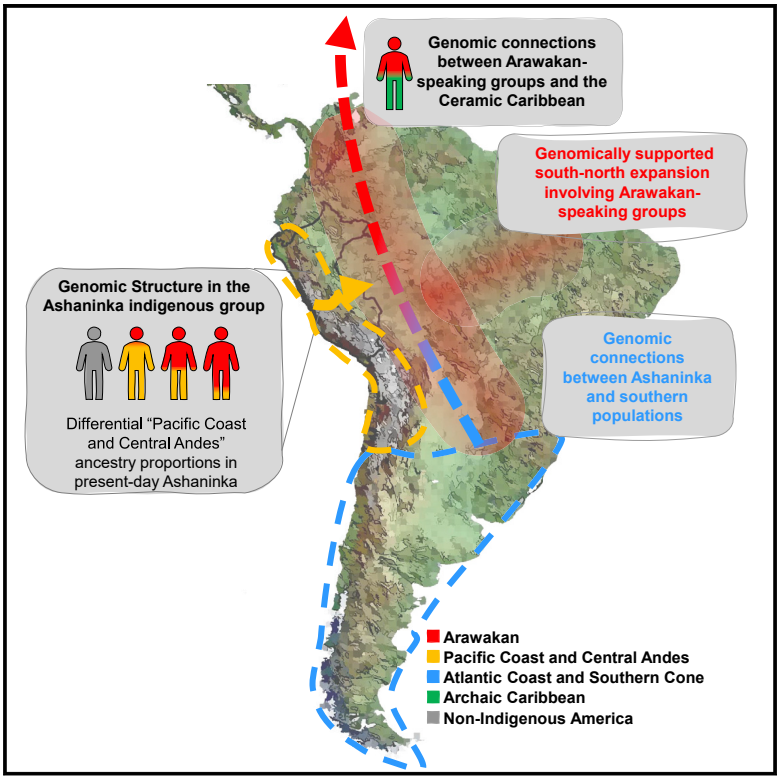
Report

Current Biology

The multifaceted genomic history of Ashaninka from Amazonian Peru

Graphical abstract



Highlights

- Genome-wide profiles of Ashaninka show a genetic structure inside the group
- Ashaninka ancestors derive from a south-north migration across South America
- Ashaninka differentially admixed with Andean and Pacific Indigenous populations
- Ashaninka testify to an Arawakan Caribbean genetic link in the Early Ceramic Age

Authors

Marco Rosario Capodiferro, Ana María Chero Osorio, Nicola Rambaldi Migliore, ..., Walther Parson, Leonor Gusmão, Alessandro Achilli

Correspondence

marcorosario.capodiferro@gmail.com (M.R.C.), alessandro.achilli@unipv.it (A.A.)

In brief

Capodiferro et al. reveal extensive genomic variation in Ashaninka from Amazonian Peru. Their gene pool testifies to south-north migration(s) across South America linked to the spread of the Arawakan language family, which reached the Caribbean in the Early Ceramic Age. Ashaninka differentially interacted with Andean and Pacific Indigenous groups.

Capodiferro et al., 2023, Current Biology *33*, 1573–1581
April 24, 2023
https://doi.org/10.1016/j.cub.2023.02.046

CellPress

Current Biology

CellPress

Report

The multifaceted genomic history of Ashaninka from Amazonian Peru

Marco Rosario Capodiferro,[1,2,*] Ana María Chero Osorio,[1] Nicola Rambaldi Migliore,[1] Dean Herman Tineo Tineo,[3] Alessandro Raveane,[4] Catarina Xavier,[5,6] Martin Bodner,[5] Filipa Simão,[7] Linda Ongaro,[2] Francesco Montinaro,[8,9] John Lindo,[10] Emilia Huerta-Sanchez,[2,11] Gustavo Politis,[12] Chiara Barbieri,[13,14] Walther Parson,[5,15] Leonor Gusmão,[7] and Alessandro Achilli[1,16,*]

[1]Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, 27100 Pavia, Italy
[2]Smurfit Institute of Genetics, Trinity College Dublin, D02 CX56 Dublin 2, Ireland
[3]Laboratorio de Biología Forense, Instituto de Medicina Legal y Ciencias Forenses, Ministerio Público, Lima 15033, Perú
[4]Human Technopole, 20157 Milan, Italy
[5]Institute of Legal Medicine, Medical University of Innsbruck, 6020 Innsbruck, Austria
[6]I3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4099-002 Porto, Portugal
[7]Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Rio de Janeiro 23968-000, Brazil
[8]Department of Biology-Genetics, University of Bari, 70125 Bari, Italy
[9]Institute of Genomics, University of Tartu, 51010 Tartu, Estonia
[10]Department of Anthropology, Emory University, Atlanta, GA 30322, USA
[11]Ecology and Evolutionary Biology and Center for Computational and Molecular Biology, Brown University, Providence, RI 02906, USA
[12]INCUAPA-CONICET, Facultad de Ciencias Sociales, Universidad Nacional del Centro de la Provincia de Buenos Aires, Olavarría 7400, Argentina
[13]Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland
[14]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[15]Forensic Science Program, Pennsylvania State University, State College, PA 16801, USA
[16]Lead contact
*Correspondence: marcorosario.capodiferro@gmail.com (M.R.C.), alessandro.achilli@unipv.it (A.A.)
https://doi.org/10.1016/j.cub.2023.02.046

SUMMARY

Despite its crucial location, the western side of Amazonia between the Andes and the source(s) of the Amazon River is still understudied from a genomic and archaeogenomic point of view, albeit possibly harboring essential information to clarify the complex genetic history of local Indigenous groups and their interactions with nearby regions,[1–8] including central America and the Caribbean.[9–12] Focusing on this key region, we analyzed the genome-wide profiles of 51 Ashaninka individuals from Amazonian Peru, observing an unexpected extent of genomic variation. We identified at least two Ashaninka subgroups with distinctive genomic makeups, which were differentially shaped by the degree and timing of external admixtures, especially with the Indigenous groups from the Andes and the Pacific coast. On a continental scale, Ashaninka ancestors probably derived from a south-north migration of Indigenous groups moving into the Amazonian rainforest from a southeastern area with contributions from the Southern Cone and the Atlantic coast. These ancestral populations diversified in the variegated geographic regions of interior South America, on the eastern side of the Andes, differentially interacting with surrounding coastal groups. In this complex scenario, we also revealed strict connections between the ancestors of present-day Ashaninka, who belong to the Arawakan language family,[13] and those Indigenous groups that moved further north into the Caribbean, contributing to the early Ceramic (Saladoid) tradition in the islands.[14,15]

RESULTS AND DISCUSSION

The low amount of genomic data on the Indigenous communities living in Amazonian Peru points to a higher homogeneity than in the Andean and coastal groups, likely due to extended isolation periods.[1] On the other hand, the expansion of the Arawakan language family, widely spoken in the area, has been characterized as a diaspora,[16,17] which could also be connected to the Saladoid pottery tradition in the Caribbean in the Early Ceramic Age,[14,15] while others explain the linguistic diffusion with cultural processes mediated by trade.[18] To give a fine-grained description of the demographic dynamics in the Amazon, and to search for demographic signatures of expansion behind linguistic and cultural packages, we genotyped Indigenous individuals from the largest Arawakan-speaking group from Amazonian Peru, the Ashaninka.

Datasets

After quality control and kinship analyses (STAR Methods), the genome-wide profiles (obtained with the Affymetrix Human Origin 600K chip) of 44 unrelated Ashaninka individuals were

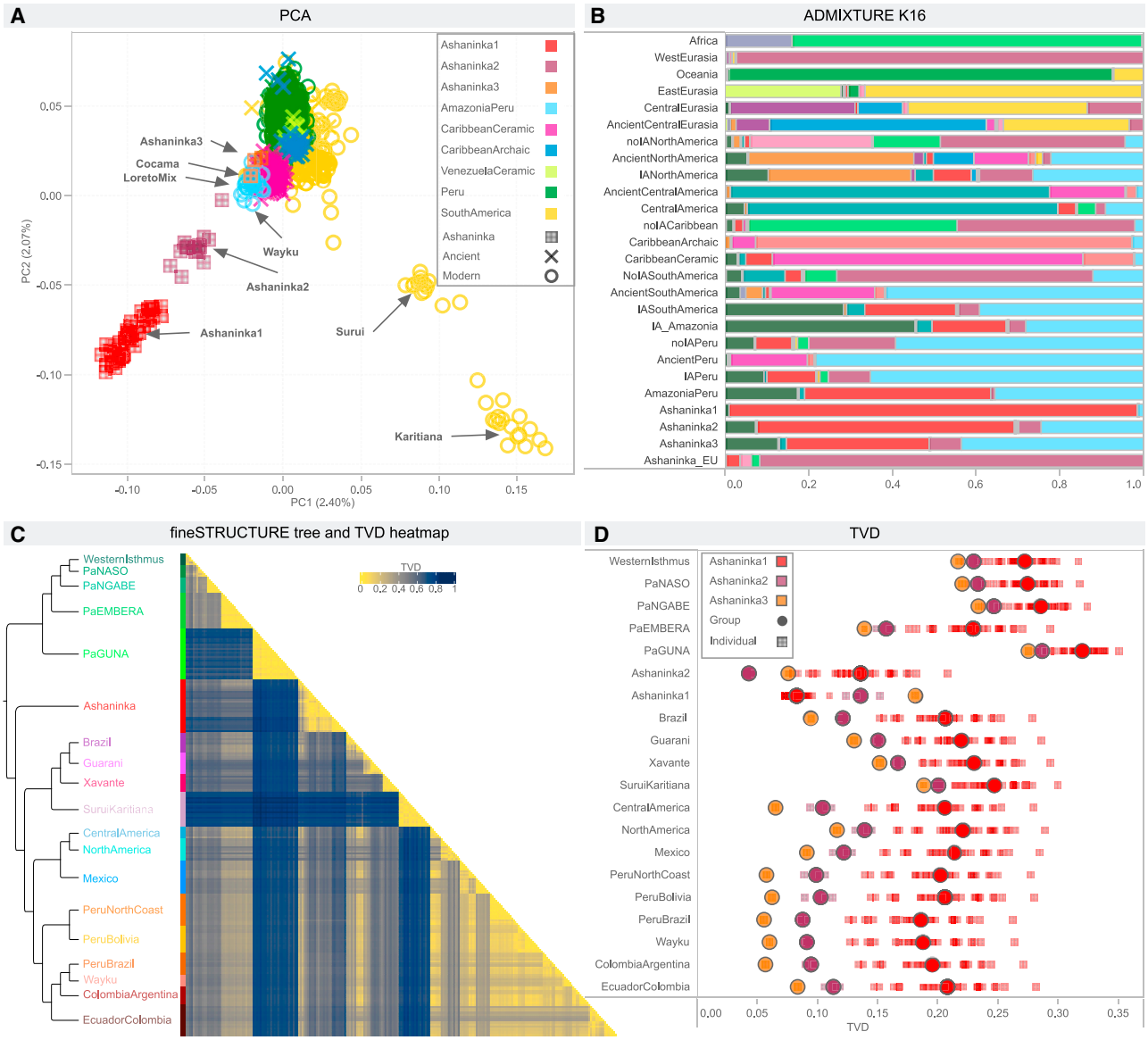


Figure 1. Population genetic structure and clustering analyses

(A) PCA of modern and ancient South American Indigenous people. Ancient and masked data were projected on the variability of the uIA245 dataset.

(B) ADMIXTURE plot on the rWD1604 and aDNA552 datasets at K16; the average value is reported for each group.

(C) fineSTRUCTURE tree and TVD heatmap based on the results of CHROMOPAINTER analysis on the uIA245 dataset. In the tree, the specific Ashaninka branch includes only the *Ashaninka1* and *Ashaninka2* subgroups, while *Ashaninka3* individual is part of the PeruNorthCoast cluster, as detailed in Figure S1D and Data S1A.

(D) TVDs of Ashaninka individuals (squares) and average values (circles) are shown against each of the clusters identified in the fineSTRUCTURE tree.

obtained (Data S1A) and merged with available genomic data from modern and ancient individuals (Data S1B) to create different datasets. Starting from a worldwide dataset of 1,604 modern individuals (rWD1604), we extracted two subsets: uIA245, which included 245 individuals with more than 95% Indigenous American (IA) components, and uIA95, with 95 individuals with less than 1% African and 2% European ancestries. The non-Indigenous components of the Indigenous individuals excluded from these two subsets were then masked while the remaining SNPs were used to build a third dataset with 579 individuals (mIA579). Finally, a fourth dataset (aDNA552) included genomic data from 552 ancient Siberian and American individuals to better dissect the genetic history of the Ashaninka within the Americas (see STAR Methods for further details).

Genomics insights into the Ashaninka Indigenous group

Population genetic structure

The 44 genome-wide profiles of Ashaninka are stretched along a specific cline of variation in the principal component analyses (PCAs) of allele frequencies and haplotypes from the uIA245

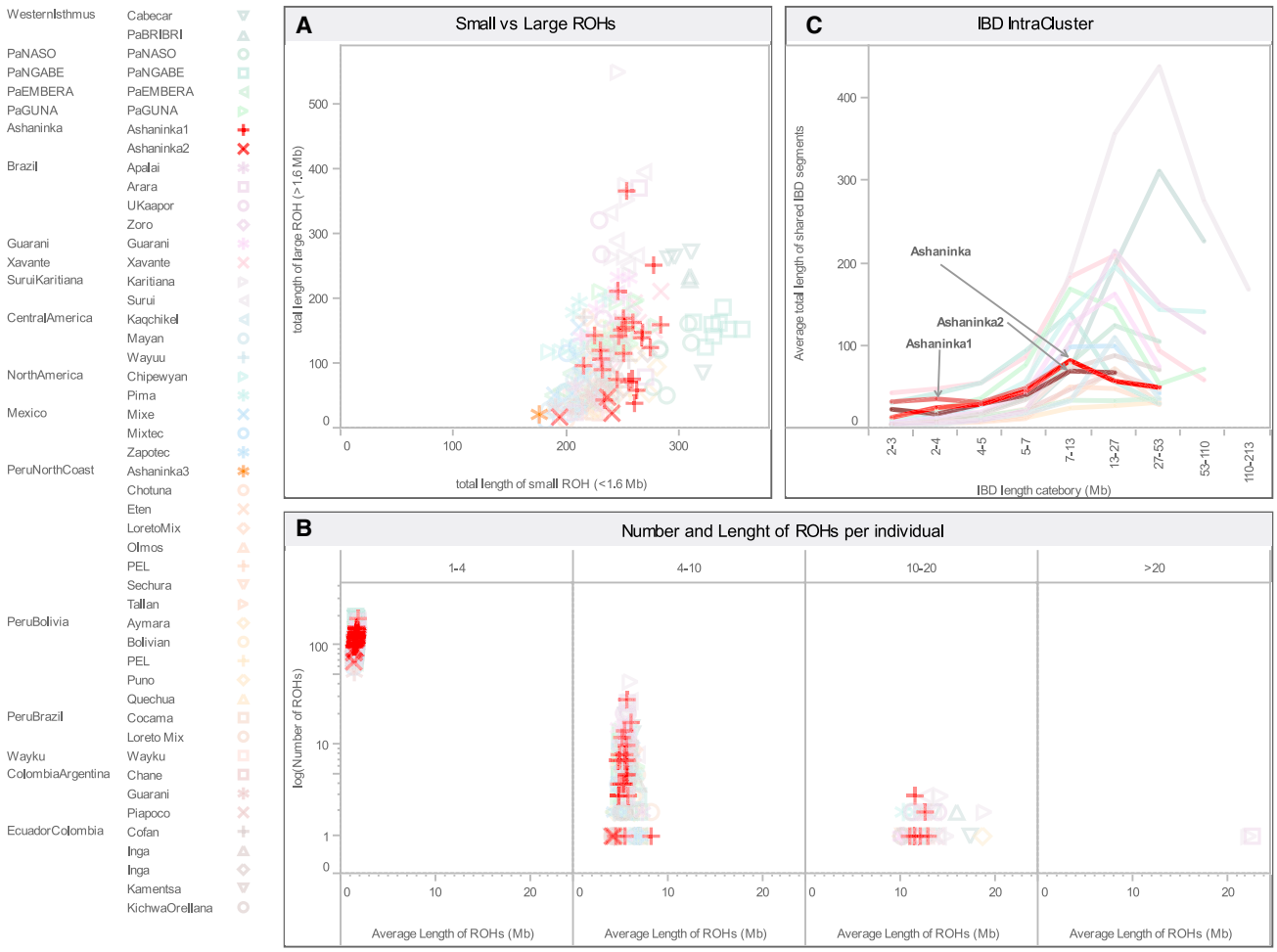


Figure 2. Population demographic inferences

These analyses were performed on the ulA245 dataset.

(A) Proportion of the total length of small (<1.6 Mb) and large (>1.6 Mb) ROHs for each individual.

(B) Number and average length of long ROH (>1 Mb) per each individual considering four bins of average ROH lengths.

(C) Density of the intrapopulation average total length of shared IBD blocks in Indigenous groups, considering nine bins of IBD lengths.

dataset (Figures 1A, S1A, and S1B) and reveal a less genetically homogeneous group than expected. It is possible to identify at least two major genetic subgroups, numerically ordered (based on the numerosity of individuals) as "*Ashaninka1*" and "*Ashaninka2*." These two main groups have different proportions in the ADMIXTURE components (Figures 1B and S1C), and despite clustering together in the fineSTRUCTURE tree (Figures 1C and S1D), they show a different distance (measured as total variation distance [TVD] on the "*chunkcount*" output of CHROMOPAINTER) toward the other clusters, especially in South America (Figure 1D). Two additional individuals, labeled as "*Ashaninka3*," are characterized by a genetic profile more related to other Indigenous groups from South America and cluster within the Peru North Coast group (Figures 1C and S1D).

As for the impact of colonialism, the ADMIXTURE plots (on unmasked datasets) show that the Ashaninka have a low non-Indigenous component in their genomes (Figure 1B). We estimated that 77% of *Ashaninka1* individuals, 33% of A*shaninka2*, and one *Ashaninka3* have more than 95% Indigenous components. It is worth mentioning that one individual, self-identified as Ashaninka (AD165; Data S1A), showed an almost complete European genomic profile (labeled as "*Ashaninka_EU*" in Figure 1B) and was not considered for further analyses on the Indigenous component(s).

Demographic analyses

Demographic reconstructions further mark different profiles within the Ashaninka population. The runs of homozygosity (ROHs; DNA fragments with homozygous genotypes; Figure 2A) show a characteristic profile for the *Ashaninka3* individual (the only one retained in the ulA245 dataset), with a few short ROHs and almost no long ROHs, which might suggest very recent admixture events.[19] Considering these signals and the low representation in our datasets, *Ashaninka3* data were excluded from further analyses. As for the other two genetic subgroups, *Ashaninka1* shows a higher number of long ROH fragments than *Ashaninka2*. In three *Ashaninka1* individuals, the total length of large ROHs is very high, comparable to populations who experienced prolonged isolation and consanguinity, such as those from Amazonian Brazil and eastern

Panama[9,20] (Figure 2A). This is also confirmed by the distribution of ROH fragments longer than 1 Mb (Figure 2B), where *Ashaninka1* individuals show a higher number of fragments between 4 and 10 Mb than *Ashaninka2*, and some of them also retained ROHs in the next bin (10–20 Mb). This testifies to a high level of consanguinity in *Ashaninka1* and suggests a different demographic history for *Ashaninka2*.

The IBD (identical by descent) fragments shared within each Indigenous fineSTRUCTURE cluster do not show very high peaks in the Ashaninka, confirming a high genetic variation (Figure 2C), with *Ashaninka1* driving the shape of the IBD distribution of the overall Ashaninka cluster. This analysis excludes a significant recent bottleneck for the Ashaninka, while it is still detectable for the Brazilian Amazonian and Isthmo-Colombian populations, thus suggesting the origin of the subgroups identified in this work is not very recent.

Genetic connections of Ashaninka on a continental scale

In the South American PCA (Figure 1A), the Ashaninka subgroups show an outlier position, plotting at one edge of the first principal component along an Amazonian cline driven on the other end by Brazilian Amazonian groups. It is worth mentioning that the Ashaninka pattern is also confirmed when excluding the Brazilian groups. The populations closest to the Ashaninka are from northern Peru and Colombian Amazonia (Loreto, Cocama, and Wayku). Those from Brazilian Amazonia (i.e., Surui and Karitiana) are separated by the first component and show a different genetic pattern when increasing the number of clusters in the ADMIXTURE analysis (Figure S1C). At K16 (Figure 1B), the Amazonian component (in red), present since K13 and modal in the *Ashaninka1*, is lost in the Amazonian Brazil groups (Karitiana and Surui), which are characterized by a different (dark green) component. Both Amazonian components are poorly represented among the ancient American individuals. Such a genetic distinctiveness of the Ashaninka is also confirmed by the fineSTRUCTURE Indigenous tree (Figure 1C), where the Ashaninka cluster (including *Ashaninka1* and *Ashaninka2*) branches before the other American clusters, with the exclusion of the Isthmo-Colombian ones. The TVD of the Ashaninka individuals (and the subgroup average) from the fineSTRUCTURE clusters (Figure 1D) shows the proximity of Ashaninka with PeruBrazil (which includes Loreto and Cocama) followed by Wayku and ColombiaArgentina. These connections are further supported by the high level of IBD sharing (Figure S2A).

The OrientAGraph maximum likelihood tree, the outgroup-f3-based neighbor-joining tree, and the multidimensional scaling (MDS) plot (Figures S2B–S2D) also point to genetic closeness among the populations from Amazonian Peru. Moreover, in the outgoup-f3-based analyses, which also include ancient DNA data, these connections are extended to individuals associated with the Caribbean Ceramic culture[10,11] to form an "AmazoniaPeruCeramic" macro-group, separated from the other Peruvian groups. In the MDS, Peruvian and non-Peruvian groups living in territories eastern to the Andes plot along the left bottom "eastern" cline, which is eventually led by the Caribbean Ceramic populations.

This trend is confirmed by formal f-statistics approaches. The highest level of shared genetic history of both *Ashaninka1* and *Ashaninka2* is with the modern Amazonian Peru groups of Loreto and Wayku, and with the ancient Ceramic Caribbean (Figure 3; Data S2). As for the genetic relationships with populations living on the two sides of the Andes, *Ashaninka2* shows greater proximity than *Ashaninka1* with the western side. On the other hand, *Ashaninka1* has a stronger proximity to the eastern part of the continent, reaching the coast of Brazil (i.e., Piapoco and UKaapor). The f4-statistics, in the form (*Ashaninka2*, *Ashaninka1*; *Modern*/*AncientIA*, *Mbuti*), confirms this trend (Figure 3) mainly when only ancient individuals are considered, as shown, for instance, by the significant relationship between *Ashaninka1* and the Argentinian individuals from the Laguna Chica site. The stronger eastern connections of *Ashaninka1* with modern populations and ancient individuals from Colombia, Amazonian Brazil, the Atlantic coast, and the Southern Cone are also confirmed with respect to other Amazonian Peru groups (Figure S3A). On the other hand, some ancient individuals from the Pacific coast contributed more to the current gene pool of *Ashaninka2*.

Another interesting difference between the two Askaninka subgroups is their relationships with ancient Caribbeans (Figure S3B). *Ashaninka1* shows more shared drift than *Ashaninka2* with the Ceramic Caribbean individuals, but not with the Archaic Caribbean ones. It is worth mentioning that the *VenezuelaCeramic* genetic cluster shows stronger proximity with the populations of the Isthmo-Colombian area, thus suggesting gene flows from Central America to Venezuela.

Taking into account the genetic findings reported so far, the origins and relationships of the two Ashaninka subgroups, as genetically defined here, were further explored and compared to South American and Caribbean ancient genomes through admixture graph modeling with qpGraph (Figures 4 and S4). The best fitting models, statistically supported by the worst *f4 Z* score below |3|, describe the *Ashaninka1* as deriving from a branch in common with the ancient genomes from the Southern Cone and separated by those from the Pacific Coast and the Central Andes. Ancient individuals from the Caribbean islands also derive from the same branch. This model reveals a slight (2%–5%) peculiar contribution from a node that precedes Ancient Beringia (USR_11500) in the Archaic Caribbeans. A signal from North America in the Archaic Caribbean has already been highlighted and discussed in previous studies.[10,11] We tested different models, including the most ancient groups from North, Central, and South America, to verify if this contribution could be associated with a North American migration to the Caribbean Island. We have not found any preferential trend among the different ancestries of North America (Early San Nicolas and Spirit Cave) and with respect to Central America (Ancient Panama) and South America (Lapa Do Santos) groups (Figure S4C). Our results suggest that this signal can be associated with the first peopling of the Americas[9] rather than a specific North American ancestry (see also STAR Methods for further details). Concerning the Ashaninka subgroups, we revealed a significant gene flow from the ancestors of *Ashaninka1* into the Caribbean Ceramic Age genomes (Figure 4). We propose that it might represent a migration from southern South America that contributed to the ancestral gene pool of the Ashaninka and other non-Brazilian populations of Amazonia and ended up in the Caribbean, eventually admixing with the local Archaic populations and contributing to the Ceramic transition on the islands. The legacy of this gene flow is primarily evident in the *Ashaninka1* individuals, while a different

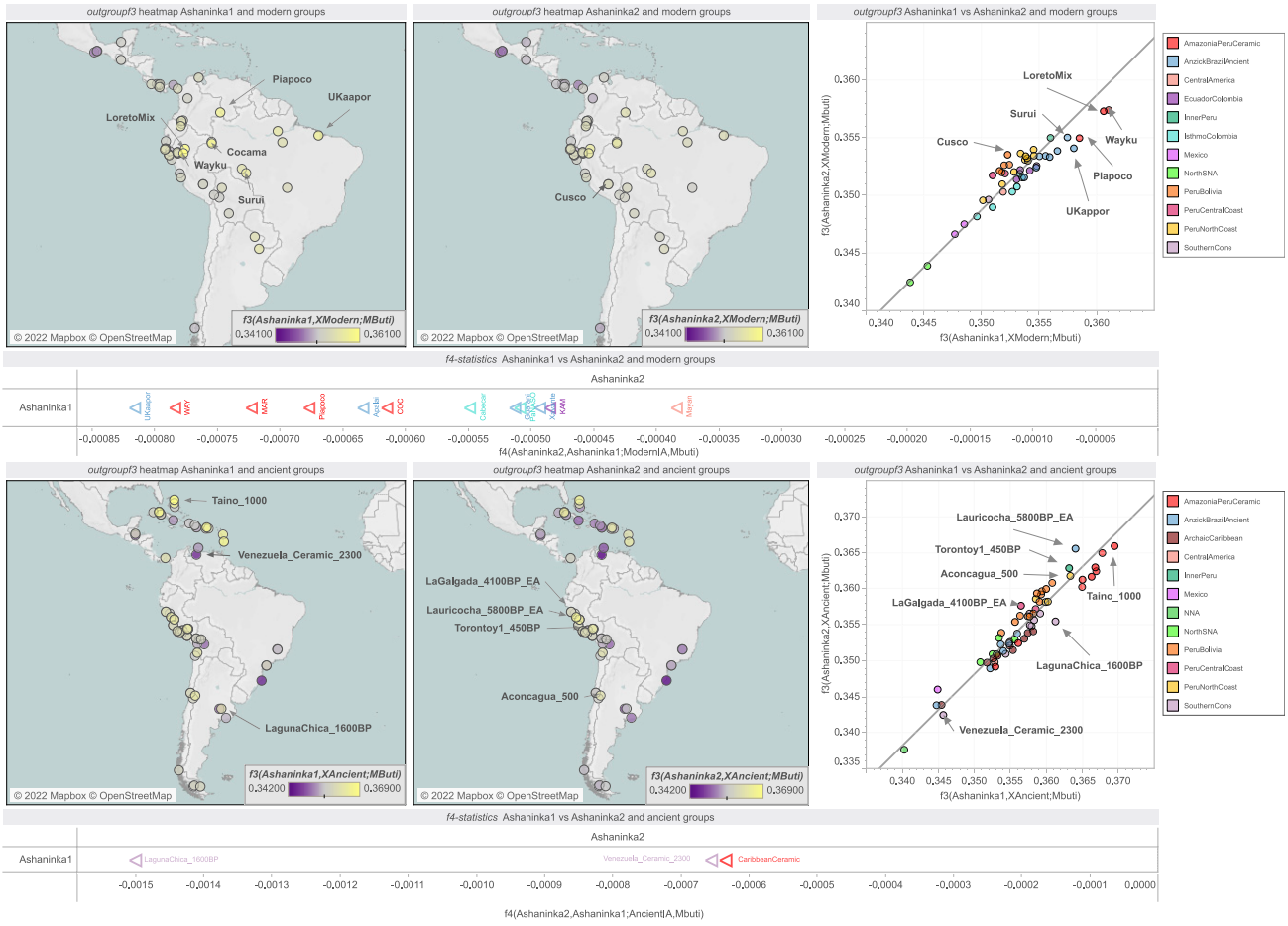


Figure 3. *F-statistics* on the Ashaninka genetic subgroups

A comparison of the different genetic relationships of *Ashaninka1* and *Ashaninka2* with other IA groups. The upper panels include only modern groups (uIA95 plus mIA579), and all SNPs were used. In the lower comparisons the ancient groups were included (uIA95 plus aDNA552), and only transversions were used. All outgroup-f3 comparisons have a *Z* score > 3 and were calculated with at least 30,000 SNPs. Population groups are colored according to the genetic clusters of Figure S2C. See also Data S2 for details.

(more recent) contribution (16%–38%) eventually shaped the genetic makeup of *Ashaninka2*. The proposed scenario was also tested using modern groups that speak an Arawakan language (Ashaninka, Chane, Piapoco, and Wayuu) or that have a different geographic origin (Chipewyan from North America and Puno from the Andes) (Figure S4D). In this model, the early split of the Argentinian group from a node in common between *Ashaninka1* and other Arawakan groups confirms our hypothesis of a migration from the south. Moreover, among the modern groups, the Ashaninka1 shows the highest shared genetic drift with the Ceramic Caribbean groups (taken separately or merged into one group; Figure S3B). Therefore, the *Ashaninka1* ancestors were probably involved in a south-north migration toward the Caribbean in the Early Ceramic Age that was previously suggested by other genomic studies[10–12] and supported by archeological/linguistic findings.[13,16,21]

Conclusions

The genetic structure of the Americas has been shaped by multiple waves of migration, leading to admixture events challenging our ability to reconstruct its genetic history.[22–24] Besides a macrogeographic approach, studies focused on specific Central and South American regions and/or IA groups have been paramount to adding knowledge on this issue.[9,25] As for Peru, where cultural and linguistic diversity is very high, genetic studies with extant Indigenous and rural communities have contributed to disentangling fine-scale population dynamics[1–3] complementing ancient DNA records that cover more than 10,000 years.[4–6] However, most available data are from the coast, leaving the Peruvian side of Amazonia, east to the Andes, still understudied from a genomic and archaeogenomic point of view, albeit possibly harboring fundamental information to clarify the complexity of South American genetic history and the origin and variability of its present-day populations. In this work, we analyzed the DNA of self-identified Ashaninka individuals from the Peruvian Amazon region of Pasco. Our genome-wide analyses unveiled at least two different genetic subgroups within the (relatively isolated) Ashaninka population, testifying to the heterogeneity of Indigenous groups, each possibly retaining the legacy of different genetic histories and wide interactions. The two

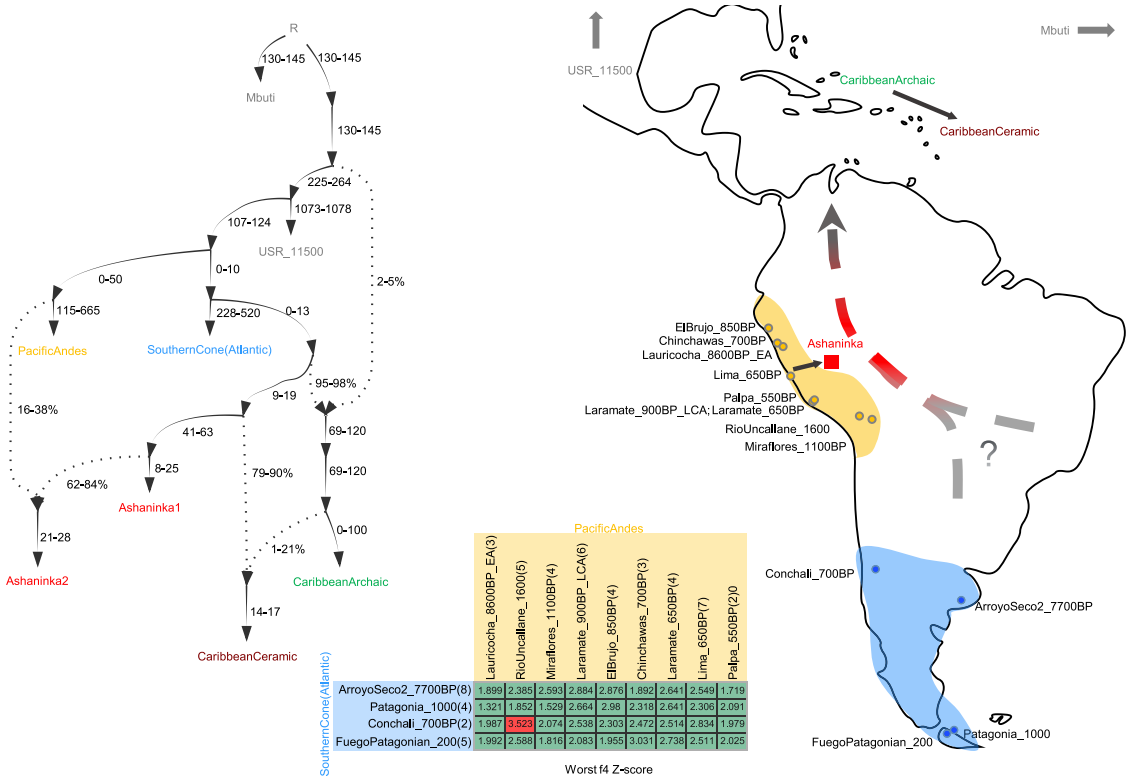


Figure 4. Admixture graph model

Final qpGraph model obtained using only the transversions retained after merging uIA95, mIA579, and aDNA552 datasets. The genetic groups reported in the graph include all populations with at least 55% transversions and two samples in the datasets (Tables S1 and S2). "PacificAndes" and "AtlanticSouthernCone" (representing nine and four populations, respectively) are defined in Figure S2C. The model has also been verified by replacing these two groups with others (Figure S4B). The matrix reports the worst f4 *Z* scores for each combination, while minimum and maximum values for each branch have been added in the admixture graph. The map on the right summarizes the proposed model of a south-north migration of Indigenous groups moving into the Amazonian rainforest and contributing to the early Ceramic (Saladoid) tradition in the Caribbeans.

Ashaninka subgroups genetically defined here, even if retaining a certain degree of genetic similarity (e.g., clustering together in the fineStructure tree; Figure 1C), show specific features and different relationships with the surrounding populations. The genetic subgroup *Ashaninka1* probably remained more isolated, while the other subgroup, *Ashaninka2*, probably experienced more interactions with the populations that lived on the slopes of the Andes and on the Pacific coast. The gene pool of two additional individuals (*Ashaninka3*) probably derives from a recent admixture. These data show that the Indigenous groups from the inner part of South America, including the Ashaninka, are less homogeneous than previously assumed with uniparental markers[7,8] or with other datasets. This complexity is also evident when exploring the demographic evolution over time. *Ashaninka1* has longer and more abundant homozygosity fragments than *Ashaninka2*, suggesting a higher degree of consanguinity. On the other hand, the analogous pattern shown by the IBD fragments could be the result of a similar demographic impact of the European contact.

The subgroup *Ashaninka1* probably experienced fewer influences from the Pacific coast, largely retained its ancestral genetic makeup, and can be associated with the early Ceramic cultures in the Caribbean islands. Therefore, we can hypothesize that the Ashaninka ancestors, likely arriving from the south, moved further north and contributed to the gene pool of the Caribbean people since Ceramic times, without leaving detectable signs among other Ceramic groups in South America (e.g., in Venezuela), who instead seem to have been influenced by the Isthmo-Colombian area. The connection with Ceramic Caribbean is also detectable in another Arawakan group, the Piapoco from Colombia. A specific Arawakan-Caribbean link has been previously suggested by genetic studies,[12] but then extended to a wider range of South American populations.[10,11,26] Archaeological and anthropological studies propose that the Caribbean Ceramic Age began when Arawakan-speaking peoples from lowland South America introduced novel ceramic industries and intensive agricultural technology to the archipelago,[14] as attested by the first Antillean Ceramic cultures Saladoid and Barrancoid.[27] Briefly, our final model (summarized in Figure 4 and supported by other analyses, e.g., Figures 1D, 3, and S2C) describes the Ashaninka gene pool as the result of a gene flow from southeastern South America. This initial migration from the south probably involved a large part of the continent to the east of the Andes, from the Southern Cone to the Amazon basin. Eventually, the ancestral populations diversified in the variegated geographic regions of the interior, differentially interacting among themselves and admixing with local and neighboring groups. The legacy of this original gene flow is clearly detectable in the

present-day *Ashaninka1* subgroup, which also testifies to a genetic connection with the Caribbean Islands since the Ceramic times. This northward migration probably involved the subgroup(s) that diffused the Arawakan of the Arawakan Caribbean and Palikuran branches from the upper Negro and Orinoco River going northward into the Caribbean-Atlantic Area.[16] As we cannot point at a specific demographic signature exclusive of the Arawakan speakers, these demographic movements could have included interactions with other Amazonian language families. Some of the Ashaninka ancestors that remained in the Amazonian Peru had more recent interactions with Indigenous groups from the Andean and Pacific regions. These admixture processes, also confirmed by historical records on interactions before and during the Inca empire,[28] differentially shaped the current gene pools of the two genetic subgroups identified in this study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - The Ashaninka Indigenous group
  - Ethics and community engagement
- METHOD DETAILS
  - Genome-wide data from Ashaninka individuals
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Comparative datasets
  - PCA and f-statistics
  - qpGraph
  - OrientAGraph
  - ADMIXTURE
  - Runs of homozygosity
  - Phasing
  - Local ancestry
  - CHROMOPAINTER & FINESTRUCTURE
  - Identity by descent

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.02.046.

ACKNOWLEDGMENTS

We are grateful to the Ashaninka volunteers who generously participated in this survey and made this research possible. This research received support from the Italian Ministry of University and Research (MUR) for Progetti PRIN2017 20174BTC4R (to A.A.) and Dipartimenti di Eccellenza Program (2018–2022) – Dept. of Biology and Biotechnology "L. Spallanzani" University of Pavia (to A.A.); the European Research Council (ERC), Horizon 2018 Starting Grant ERC-2018-STG, no. 804994 (to M.R.C., L.O., and E.H.-S.); Fundação de Amparo à Pesquisa do Rio de Janeiro– FAPERJ, Brazil, process E-26/202.275/2019 and CNE-2022 (to F.S. and L.G.); Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, Brazil, ref. 306342/2019-7 (to L.G.); the "Fondazione Adriano Buzzati – Traverso" for the L.L. Cavalli-Sforza fellowship (to N.R.M.); the URPP Evolution in Action of the University of Zurich; the NCCR Evolving Language, Switzerland National Science Foundation agreement #51NF40_180888; and the SNSF Sinergia project "Out of Asia" (to C.B).

AUTHOR CONTRIBUTIONS

D.H.T.T. performed sample collections supervised by L.G.; D.H.T.T., C.X., and F.S. extracted DNA supervised by M.B., W.P., and L.G.; M.R.C. and N.R.M. produced modern genome-wide data supervised by A.A.; M.R.C., A.M.C.O., N.R.M., A.R., and L.O. analyzed the genomic data supervised by F.M. and A.A.; J.L., E.H.-S., and G.P. provided additional expertise for the interpretation of the results; M.R.C., A.M.C.O., N.R.M., and A.A. wrote the original draft with major input from A.R., F.M., and C.B. All authors discussed the results and contributed to the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: September 20, 2022
Revised: December 14, 2022
Accepted: February 14, 2023
Published: March 16, 2023

REFERENCES

1. Borda, V., Alvim, I., Mendes, M., Silva-Carvalho, C., Soares-Souza, G.B., Leal, T.P., Furlan, V., Scliar, M.O., Zamudio, R., Zolini, C., et al. (2020). The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. Proc. Natl. Acad. Sci. USA *117*, 32557–32565. https://doi.org/10.1073/pnas.2013773117.
2. Barbieri, C., Barquera, R., Arias, L., Sandoval, J.R., Acosta, O., Zurita, C., Aguilar-Campos, A., Tito-Álvarez, A.M., Serrano-Osuna, R., Gray, R.D., et al. (2019). The current genomic landscape of Western South America: Andes, Amazonia, and Pacific Coast. Mol. Biol. Evol. *36*, 2698–2713. https://doi.org/10.1093/molbev/msz174.
3. Harris, D.N., Song, W., Shetty, A.C., Levano, K.S., Cáceres, O., Padilla, C., Borda, V., Tarazona, D., Trujillo, O., Sanchez, C., et al. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc. Natl. Acad. Sci. USA *115*, E6526–E6535. https://doi.org/10.1073/pnas.1720798115.
4. Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R.A., Watson, J.T., Viviano Llave, C., Witonsky, D., Beall, C., et al. (2018). The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci. Adv. *4*, eaau4921. https://doi.org/10.1126/sciadv.aau4921.
5. Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., Posth, C., Harkins-Kinkaid, K., Ferry, M., Harney, É., et al. (2020). A paleogenomic reconstruction of the deep population history of the Andes. Cell *181*, 1131–1145.e21. https://doi.org/10.1016/j.cell.2020.04.015.
6. Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T.C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., et al. (2018). Reconstructing the deep population history of Central and South America. Cell *175*, 1185–1197.e22. https://doi.org/10.1016/j.cell.2018.10.027.
7. Simão, F., Xavier, C., Tineo, D.H., Carvalho, E.F., Parson, W., and Gusmão, L. (2019). The maternal inheritance of the Ashaninka native group from Peru. Forensic Sci. Int. Genet. *7*, 135–137. https://doi.org/10.1016/j.fsigss.2019.09.052.
8. Tineo, D.H., Loiola, S., Paredes, F.V., Noli, L.R., Amaya, Y.C., Simão, F., Carvalho, E.F., and Gusmão, L. (2015). Genetic characterization of 27 Y-STR loci in the native population of Ashaninka from Peru. Forensic

Sci. Int. Genet. *5*, e220–e222. https://doi.org/10.1016/j.fsigss.2015.09.088.

9. Capodiferro, M.R., Aram, B., Raveane, A., Rambaldi Migliore, N., Colombo, G., Ongaro, L., Rivera, J., Mendizábal, T., Hernández-Mora, I., Tribaldos, M., et al. (2021). Archaeogenomic distinctiveness of the Isthmo-Colombian area. Cell *184*, 1706–1723.e24. https://doi.org/10.1016/j.cell.2021.02.040.
10. Nägele, K., Posth, C., Iraeta Orbegozo, M., Chinique de Armas, Y., Hernández Godoy, S.T., González Herrera, U.M., Nieves-Colón, M.A., Sandoval-Velasco, M., Mylopotamitaki, D., Radzeviciute, R., et al. (2020). Genomic insights into the early peopling of the Caribbean. Science *369*, 456–460. https://doi.org/10.1126/science.aba8697.
11. Fernandes, D.M., Sirak, K.A., Ringbauer, H., Sedig, J., Rohland, N., Cheronet, O., Mah, M., Mallick, S., Olalde, I., Culleton, B.J., et al. (2021). A genetic history of the pre-contact Caribbean. Nature *590*, 103–110. https://doi.org/10.1038/s41586-020-03053-2.
12. Schroeder, H., Sikora, M., Gopalakrishnan, S., Cassidy, L.M., Maisano Delser, P., Sandoval Velasco, M., Schraiber, J.G., Rasmussen, S., Homburger, J.R., Ávila-Arcos, M.C., et al. (2018). Origins and genetic legacies of the Caribbean Taino. Proc. Natl. Acad. Sci. USA *115*, 2341–2346. https://doi.org/10.1073/pnas.1716839115.
13. Santos-Granero, F. (2002). The Arawakan Matrix: ethos, language, and history in native South America. In Comparative Arawakan Histories: Rethinking Language Family and Culture Area in Amazonia, J.D. Hill, and F. Santos-Granero, eds. (University of Illinois Press), pp. 25–50.
14. Keegan, W.F., and Hofman, C.L. (2017). The Caribbean before Columbus (Oxford University Press). https://doi.org/10.1093/acprof:oso/9780190605247.001.0001.
15. Ross, A.H., Keegan, W.F., Pateman, M.P., and Young, C.B. (2020). Faces divulge the origins of Caribbean prehistoric inhabitants. Sci. Rep. *10*, 1–9. https://doi.org/10.1038/s41598-019-56929-3.
16. Heckenberger, M.J. (2002). Rethinking the Arawakan diaspora: hierarchy, regionality, and the Amazonian formative. In Comparative Arawakan Histories: Rethinking Language Family and Culture Area in Amazonia, J.D. Hill, and F. Santos-Granero, eds. (University of Illinois Press), pp. 99–122.
17. Heckenberger, M.J. (2008). Amazonian mosaics: identity, interaction, and integration in the tropical forest. In The Handbook of South American Archaeology, H. Silverman, and W.H. Isbell, eds. (Springer), pp. 941–961.
18. Hornborg, A. (2005). Ethnogenesis, regional integration, and ecology in prehistoric Amazonia: toward a system perspective. Curr. Anthropol. *46*, 589–620. https://doi.org/10.1086/431530.
19. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., and Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. *19*, 220–234. https://doi.org/10.1038/nrg.2017.109.
20. Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the Americas. Nature *525*, 104–108. https://doi.org/10.1038/nature14895.
21. Michael, L., de Carvalho, F., Chacon, T., Rybka, K., Sabogal, A., Chousou-Polydouri, N., and Kaiping, G. (2023). Deriving calibrations for Arawakan using archaeological evidence. Interface Focus *13*, 20220049. https://doi.org/10.1098/rsfs.2022.0049.
22. Wade, L. (2018). Ancient DNA tracks migrations around Americas. Science *362*, 627–628. https://doi.org/10.1126/science.362.6415.627.
23. Waters, M.R. (2019). Late Pleistocene exploration and settlement of the Americas by modern humans. Science *365*, eaat5447. https://doi.org/10.1126/science.aat5447.
24. Willerslev, E., and Meltzer, D.J. (2021). Peopling of the Americas as inferred from ancient genomics. Nature *594*, 356–364. https://doi.org/10.1038/s41586-021-03499-y.
25. Joseph, S.K., Migliore, N.R., Olivieri, A., Torroni, A., Owings, A.C., DeGiorgio, M., Ordóñez, W.G., Aguilú, J.J.O., González-Andrade, F., Achilli, A., and Lindo, J. (2023). Genomic evidence for adaptation to tuberculosis in the Andes before European contact. iScience *26*, 106034. https://doi.org/10.1016/j.isci.2023.106034.
26. Nieves-Colón, M.A. (2022). Anthropological genetic insights on Caribbean population history. Evol. Anthropol. *31*, 118–137. https://doi.org/10.1002/evan.21935.
27. Roksandic, I., and Roksandic, M. (2018). Peopling of the Caribbean. In New Perspectives on the Peopling of the Americas, K. Harvati, G. Jäger, and H. Reyes-Centeno, eds. (Kerns Verlag), pp. 199–223.
28. Santos-Granero, F. (1992). Etnohistoria de la Alta Amazonía: siglos XV - XVIII (Abya Yala).
29. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.
30. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.
31. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinf. *15*, 356. https://doi.org/10.1186/s12859-014-0356-4.
32. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
33. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093. https://doi.org/10.1534/genetics.112.145037.
34. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020.
35. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190. https://doi.org/10.1371/journal.pgen.0020190.
36. Felsenstein, J. (1989). Mathematics vs. evolution: mathematical evolutionary theory. Science *246*, 941–942. https://doi.org/10.1126/science.246.4932.941.
37. Maier, R., Flegontov, P., Flegontova, O., Changmai, P., and Reich, D. (2022). On the limits of fitting complex models of population history to genetic data. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.08.491072.
38. Molloy, E.K., Durvasula, A., and Sankararaman, S. (2021). Advancing admixture graph estimation via maximum likelihood network orientation. Bioinformatics *37*, i142–i150. https://doi.org/10.1093/bioinformatics/btab267.
39. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. *15*, 1179–1191. https://doi.org/10.1111/1755-0998.12387.
40. Rosenberg, N.A. (2003). distruct: a program for the graphical display of population structure. Mol. Ecol. Notes *4*, 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x.
41. Delaneau, O., Marchini, J., and Zagury, J.-F. (2011). A linear complexity phasing method for thousands of genomes. Nat. Methods *9*, 179–181. https://doi.org/10.1038/nmeth.1785.
42. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. *8*, e1002453. https://doi.org/10.1371/journal.pgen.1002453.

43. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics *194*, 459–471. https://doi.org/10.1534/genetics.113.150029.

44. Base de Datos de Pueblos Indígenas u Originarios (2020). Lista de pueblos indígenas u originarios: Ashaninka (Ministerio de Cultura del Perú). https://bdpi.cultura.gob.pe/pueblos/ashaninka.

45. Espinosa, O. (2014). Los pueblos ashaninka, kakinte, nomatsigenga y yanesha (Ministerio de Cultura). http://repositorio.cultura.gob.pe/handle/CULTURA/48.

46. Instituto Nacional de Estadística e Informática (2018). Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1597/.

47. Landolt, G., Herrera, C.D., and Balaguer, A.; Programa de Formación de Maestros Bilingües (Peru) (2000). El Ojo Verde: Cosmovisiones Amazónicas (Fundación Telefónica).

48. Sarmiento Barletti, J.P. (2011). Kametsa asaiki: The pursuit of the 'good life' in an Ashaninka village (Amazonian Peru) (University of St Andrews). PhD thesis.

49. Rojas Zolezzi, E. (1994). Los asháninka, un pueblo tras el bosque: contribución a la etnología de los Campa de la Selva Central Peruana (Fondo Editorial PUCP).

50. Scheib, C.L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., Griffith, P.W., Mörseburg, A., Johnson, J.R., Potter, A., et al. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. Science *360*, 1024–1027. https://doi.org/10.1126/science.aar6851.

51. Gnecchi-Ruscone, G.A., Sarno, S., De Fanti, S., Gianvincenzo, L., Giuliani, C., Boattini, A., Bortolini, E., Di Corcia, T., Sanchez Mellado, C., Dàvila Francia, T.J., et al. (2019). Dissecting the pre-Columbian genomic ancestry of Native Americans along the Andes–Amazonia divide. Mol. Biol. Evol. *36*, 1254–1269. https://doi.org/10.1093/molbev/msz066.

52. Ongaro, L., Scliar, M.O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., Gnecchi-Ruscone, G.A., Alarcón-Riquelme, M.E., Patin, E., Wangkumhang, P., et al. (2019). The genomic impact of European colonization of the Americas. Curr. Biol. *29*, 3974–3986.e4. https://doi.org/10.1016/j.cub.2019.09.076.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genotype data of 51 Ashaninka individuals	This study	EGA: EGAS00001006958
Software and algorithms		
PLINK v1.9	Purcell et al.[29]	https://www.cog-genomics.org/plink/1.9/
KING	Manichaikul et al.[30]	https://kingrelatedness.com/
ANGSD	Korneliussen et al.[31]	http://www.popgen.dk/angsd/index.php/ANGSD
ADMIXTURE v.1.23	Alexander et al.[32]	https://dalexander.github.io/admixture/
AdmixTools v.4.1	Patterson et al.[33]	https://github.com/DReichLab/AdmixTools
RFMix v.2	Maples et al.[34]	https://github.com/slowkoni/rfmix
EIGENSOFT v.7.2.0	Patterson et al.[35]	https://github.com/DReichLab/EIG
PHYLIP v.3.6	Felsenstein[36]	https://evolution.genetics.washington.edu/phylip.html
ADMIXTOOLS 2	Maier et al.[37]	https://uqrmaie1.github.io/admixtools/index.html
OrientAGraph v.1.0	Molloy et al.[38]	https://github.com/sriramlab/OrientAGraph
CLUMPAK	Kopelman et al.[39]	http://clumpak.tau.ac.il/
DISTRUCT	Rosenberg[40]	https://rosenberglab.stanford.edu/distruct.html
SHAPEIT v.2	Delaneau et al.[41]	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
CHROMOPAINTER v.2	Lawson et al.[42]	http://www.paintmychromosomes.com/
fineSTRUCTURE v.2	Lawson et al.[42]	http://www.paintmychromosomes.com/
Refined-IBD v.17Jan20.102	Browning and Browning[43]	http://faculty.washington.edu/browning/refined-ibd.html
Tableau	https://www.tableau.com/	https://www.tableau.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Alessandro Achilli (alessandro.achilli@unipv.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Genotype data have been deposited in the European Genome-phenome Archive (EGA; https://ega-archive.org/) and are available for download under the accession number listed in the key resources table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The Ashaninka Indigenous group

The Ashaninka is the largest among the 51 Indigenous groups currently living in Amazonian Peru.[44] Ashaninka people inhabit most of the *selva central*, from the eastern slope of the Andes mountains to Yurua in the Ucayali Region.[45] In 2017, according to the results of the national census of Indigenous communities, there were identified 520 Ashaninka communities throughout seven Peruvian regions,[46] including the Pasco region.[44] According to historical documents, the Ashaninka population bartered with Andean people before and during the Inca empire.[28] In addition to the documented trade exchanges, ancestral practices of Andean origin are evident in the textile work and use of wind instruments of the Ashaninka.[47] Their language belongs to the Arawakan family. Contact with the Europeans started in the 17th century[48] and produced a significant decrease in the population size, mostly due to diseases "imported" from Europe during the exploitation of the conquerors.[49] The resistance of the Ashaninka people and other Amazonian

populations headed by Juan Santos Atahualpa limited the Spanish invasion of their territories in 1742. More recently, the reduction of the Indigenous territories in favor of commercial activities and the Peruvian civil wars led to the displacement and dispersal of the Ashaninka groups.[45] A mitochondrial DNA (mtDNA) study on the 41 Ashaninka communities in the Pasco region revealed a low genetic input of non-Indigenous maternal lineages, with the macro-haplogroup D being the most represented one (32%).[7] As for the paternal lineages, the Y-STR profiles belong to the Indigenous American haplogroup Q and show a significant genetic distance from other Peruvian Indigenous groups.[8]

Ethics and community engagement

This multidisciplinary study was possible with the support of local authorities and Indigenous peoples of Peru and centrally involved Peruvian co-authors (AMCO and DHTT) with years of experience in population genetics analyses. The DNA samples were already available from previous works on uniparental markers.[7,8] Written informed consent for genomic studies was obtained from all voluntary participants under strictly confidential conditions. Community consent was also obtained from the Association of Ashaninka Communities of Valle del Río Pichis, Puerto Bermúdez District, Province of Oxapampa, Pasco region in Peru. The project was explained to Indigenous community leaders and designed to maximize opportunities for public engagement, as for our recent work on other Indigenous groups.[9] All experimental procedures were reviewed and approved by the Ethics Committee for Clinical Experimentation of the University of Pavia, Board minutes of October 5th, 2010, and April 11th, 2013.

METHOD DETAILS

Genome-wide data from Ashaninka individuals

Biological samples were collected in the region of Pasco (Peru) from healthy adults that self-identified as belonging to the Ashaninka Indigenous group. The DNAs had already been extracted for previous works on uniparental markers.[7,8] For this study, 96 DNA samples were genotyped with the Affymetrix Human Origin 600K chip at the Institute of Healthcare Research in Santiago de Compostela (CEGEN).

The genotypes of 51 individuals that passed the quality control (call rate >95%) were converted into PLINK files. Variants and individuals with a missing rate higher than 2% were filtered out. Kinship was inferred using KING[30] to exclude related individuals (relatedness cutoff of second degree). A final dataset of 44 individuals (460,589 SNPs) was used for downstream analyses (Data S1A).

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparative datasets

The new Ashaninka genome-wide data were compared with a published worldwide dataset (N = 1,560) of modern high-coverage sequences or genotypes obtained with the Human Origin chip (Data S1B), as in Capodiferro et al.[9] The Ashaninka and the published genomic profiles were merged using PLINK 1.9[29] and filtered for individuals and SNPs, as reported above. A total of 1,604 individuals and 409,084 SNPs were eventually kept (rWD1604).

The SNPs of the rWD1604 dataset were called from an ancient DNA (aDNA) dataset of 569 Siberian and American individuals (Data S1B) using ANGSD[31] with the *haplocall 1* option, which picks a random read starting from an input set of reads, and gives back pseudo-haploid calls, homozygous for each SNP called. The aDNA dataset was merged with the rWD1604 dataset and then SNPs having less than 60% of missing data *–geno* 0.60) and individuals with less than 98% of missing data (*–mind* 0.98) were retained.[9,50] These filters retained an ancient dataset of 552 (aDNA552) individuals with at least 10,000 SNPs covered (Data S1B).[6,9] The final dataset encompassed 2,156 individuals and 409,084 SNPs.

The Ashaninka individuals were genetically grouped in two genomic clusters numerically ordered as Ashaninka1 and Ashaninka2 based on the number of individuals that belonged to each cluster: 32 and nine, respectively, in the rWD1604 dataset. Genetic grouping, as well as the selection of individuals with low non-Indigenous components (uIA246 and uIA96), was initially based on ADMIXTURE analyses including only modern individuals. The Ashaninka clusters were confirmed by different analyses: several PCAs pointed to different subgrouping (Figures 1A, S1A, and S1B), the fineSTRUCTURE tree revealed specific sub-branches (Figure S1D); finally, TVD analyses confirmed different interactions with other population clusters (Figure 1D). This genetic clustering was also confirmed by an additional PCA and a Neighbor-Joining tree based on outgroup f3-statistics, which were performed considering only Ashaninka individuals (data not shown).

In our analyses, we excluded individuals genotyped with the Illumina array in order to not reduce the number of markers in our dataset. In fact, when including the previously published Ashaninka genotyped with the HumanOmniExpress 1.1 BeadChip,[51] the number of SNPs decreased from 409,084 to 92,884. However, we performed explorative analyses (PCA and ADMIXTURE) including these data, which confirm different genetic profiles in the Ashaninka, the most frequent one overlapping with Ashaninka 1, while only one individual shows similarity with Ashaninka 2.

"Nearly unadmixed" Indigenous American (uIA) datasets

Modern American populations contain heritage from several continents in their genomes due to the impact of European colonization and the associated African slave trades.[52] Therefore, the study of Indigenous American genomic history requires the selection of a subset of individuals containing a high amount of Indigenous American ancestry. To this end, three independent analyses were performed.

ADMIXTURE. We run the software ADMIXTURE v.1.23[32] both in an unsupervised and supervised manner (see 'ADMIXTURE' section) on the modern rWD1604 dataset. We considered the individuals of the previously reported *"uIA89"* dataset[9] as the IA panel in the supervised run.

f4 statistics. We applied f4 statistics[33] in the form: *f4(ancientIndigenous, X; Europe/Africa, Mbuti)*. The 'AncientIndigenous' used were five ancient high-coverage genomes representing different time periods and locations in the American continents (Anzick_1, Ayayema, Sumidouro5, SpiritCave, Taino).

Local ancestry (LA).

The LA analysis was performed using RFMix v. 2.[34] The individuals that passed the first two steps were included in the reference panel representing the Indigenous American ancestry, while African and European individuals from rWD1604 were used as a panel for the African and European ancestries, respectively.

Two datasets were selected, merging the results of all three analyses previously described.

- uIA245 encompassing individuals with less than 5% of non-Indigenous American ancestry in the Admixture and LA analyses and a Z-score lower than |3| in the f4 analysis. This dataset includes 28 Ashaninka individuals, 24 from Ashaninka 1, and three from Ashaninka2.
- uIA95 encompassing individuals with less than 1% of African and 2% of European components in the Admixture and LA analyses and a Z-score below |2| in the f4 analysis. This dataset includes 6 Ashaninka individuals.

Masked non-Indigenous American (mIA) datasets

The 579 individuals excluded from the uIA datasets have been masked for the non-IA components. The results of RFMix reported in the "viterbi" file were considered for masking. For each haplotype, the SNPs not classified as IA or with a probability of being IA <0.9 have been removed. The final dataset of masked individuals (m579) contains pseudo-haploid individuals where the haplotypes of each individual are kept separate (1,158 pseudo-haploid individuals).

PCA and f-statistics

PCA and f-statistics were performed with EIGENSOFT v7.2.0[35] and AdmixTools v4.1,[33] respectively, using default parameters if not explicitly reported. In the PCAs, ancient and masked data were projected onto the modern variation (uIA245) with the *lsqproject* and *autoshrink* options in *smartpca*. For the f-statistics, the uIA95, the ancient (aDNA552), and the masked (m579) datasets were used. Whenever ancient individuals were included in the analyses, only transversions were used, and to include the comparisons in the results, a minimum threshold of 30K SNPs was considered. The neighbor joining (NJ) tree was built using the program PHYLIP 3.6[36] on a distance matrix generated with the inverse of the outgroup-f3 statistics (1/outgroup-f3), while the Multidimensional Scaling was built with the R function *cmdscale* on a distance matrix based on the 1-outgroup-f3. For these analyses, only those populations that retained at least 55% of transversions were considered.

qpGraph

Considering the absence of ancient genomic data from individuals who have lived in the Amazonian rainforest, we selected a group of representative ancient genomes from the Caribbean and from the two coasts of South America. Population selection was based on findings from other analyses and the number of individuals and SNPs.

- Lauricocha_5600, which is closer to *Ashaninka2* than *Ashaninka1* in the outgroup-f3, has only one individual and a SNP missingness of 47% on the set of variants included in our dataset. For this reason, to represent the Pacific coast, the Lauricocha_8600 group, which has 3 individuals and 19% of SNP missingness, was chosen.
- Taking into account the availability of a few ancient genomic data from the Atlantic side and none from South America inland, we used the FuegoPatagonian_200 dataset, which includes one and four individuals from Patagonia and Tierra del Fuego, respectively, and has 21% of SNP missingness. In the outgroup-f3 analysis, LagunaChica_1600 is closer to *Ashaninka1* than *Ashaninka2* but contains one individual with 52% missingness, so it was excluded from the qpGraph analysis, as well as LagunaChica_6800 which has 3 individuals but a 63% of missingness.
- To represent the Caribbean Islands, we used all individuals already classified as belonging to Archaic and Ceramic periods.[11]
- USR_11500 has been selected as the ancestral source for the IA ancestries.

ADMIXTOOLS2[37] software was used to automatically generate a tree with a number of admixture events from 1 to 5 (Figure S4A). From the 100 replicates obtained the tree with the best score was chosen, which was then manually investigated using the classic qpGraph and considering the analyses already carried out (Figure S4B).

The red lines in Figure S4B are the waves modified to reach the final tree.

- A migration of 0% from a nonIA migration in Ashaninka
- Moving the migration from South America to Caribbean Ceramic closer to Ashaninka1
- Removing the migration of 1% from Archaic Caribbean to Ashaninka2

After this process, a very solid model (Figure 4) has been obtained (worst Z-score -1.321), which summarizes the information collected by various analyses in this work. This model was verified at the continental level by replacing the two representatives of the Pacific Coast and the Andes as well as the Southern Cone with other groups clustering together in the *outgroup-f3* NJ tree and having at least 2 samples. The Z-score matrix for each of these comparisons is reported in Figure 4.

In order to verify the origin of this signal from a node before Ancestral Beringia (USR) into Archaic Caribbean, we tried to construct a more complex qpGraph model that includes ancient groups from North (ASO-Ancient South Ontario, ESN-Early San Nicolas, SpiritCave), Central (Ancient Panama) and South (LapaDoSantos) America (Figure S4C) to test the possibility of a dispersal from North America into Archaic Caribbean.[10] It is worth reminding that ASO represents a northern Native American ancestry (NNA), while the others could be connected to an Indigenous ancestry that spread also to South America (SNA). Our results show a similar Z-score when the contribution starts from ESN or ASO. When we split NNA from SNA and branched the contribution to the Archaic Caribbean from SNA, we observed a consistent tree when considering the North (Spirit Cave) and Central (Ancient Panama) America. In a more complex tree including NNA and both ESN from North America and Lapa Do Santos from Brazil together, we do not need to add the additional branching event from North America in the Archaic Caribbean. Therefore, the extra contribution in the Archaic Caribbean in Figure 4 does not seem to be related to an early branching in North America but likely to the early peopling of the entire double continent, marked by the SNA1 ancestry defined in Capodiferro et al.[9]

Finally, to verify the model proposed in Figure 4 using comparative ancient genomes, we built a qpGraph tree based only on modern groups (Figure S4D) associated with Arawakan language (Ashaninka, Chane, Piapoco and Wayuu) or with a different origin (Chipewyan from North America and Puno from the Andes). This model, built including all SNPs, confirms the one obtained on ancient data, with the Chane group from Argentina diverging from a node in common with the other Arawakan groups, suggesting a migration from the south. It seems that the Wayuu need a further non-Arawakan contribution that should be further investigated due to the paucity of genetic data (only one individual) in our comparative dataset (Data S1B).

OrientAGraph

The maximum likelihood network orientation was obtained running OrientAGraph[38] on the pruned dataset uIA95 using TSI, CHB and YRI (Tuscans, Chinese Han and Yoruba) as outgroups. Five migration edges were tested (from 0 to 4) with noss and -global parameters active in TreeMix.

ADMIXTURE

The unsupervised Admixture analyses were performed using ADMIXTURE v.1.23[32] on different pruned (*–indep-pairwise* 200 25 0.4) datasets encompassing modern (rWD1604) and ancient (aDNA552) individuals. 10 independent runs were performed for each K (from 2 to 20).

In the supervised Admixture, at K3, the individuals previously reported having less than 1% of African and 2% of European components (uIA89 dataset in Capodiferro et al.[9]) have been used as IA reference, while the African and European individuals in the rWD1604 have been used as representative of African and European ancestries, respectively.

The runs were combined using CLUMPAK (Cluster Markov Packager Across K)[39] and the K aligned with DISTRUCT.[40] The lowest average value of the cross-validation (cv) error is at K13, but without any significant differences from K13 to K16 (Figure S1C). Moreover, considering that at K16 we can distinguish a specific component in the Amazonian Brazil group (in dark green), we have shown the plot at K16 in Figure 1C.

Runs of homozygosity

Run of Homozygosity (ROH) analyses were performed on the uIA245 dataset using PLINK 1.9[29] with default values. For each individual, the total length of the ROH fragments smaller than 1.6 Mega bases (Mb) was compared with the total length of fragments longer than 1.6 Mb.[18] For each individual, the long fragments of ROHs (>1 Mb) were divided into four length bins (1-4 Mb, 4-10 Mb, 10-20 Mb, >20 Mb). For each bin, the average length of ROH fragments for each individual was plotted against the total number of fragments.

Phasing

The rWD1604 dataset was phased using the Segmented Haplotype Estimation and Imputation tool SHAPEITv2[41] and the HapMap37 human genome build 37 recombination map.

Local ancestry

RFMix[34] was used to estimate the local ancestry for genomic fragments. As source populations, Bantu, Esan (ESN), Gambia (GWDwg), Mandenka, Mbuti and Yoruba (YRI) for Africa; Spanish (IBS), British (GBR), French, Icelandic and Tuscany (TSI) for Europe, and Chipewyan, Kichwa Orellana, PaGUNA, Puno, Surui and Karitiana for Indigenous ancestry were used. The options *PopPhased*, *-n 5* and *–forward-backward* were applied as recommended in the RFMix manual. Individuals with more than 2% of non-Indigenous American ancestry (see "nearly unadmixed" Indigenous American (uIA) datasets section above) were masked, creating a PLINK file with pseudo haploid data, in which, for each individual, the two phased haplotypes were separated (putatively called A and B) retaining only the fragment assigned as Indigenous (*forwardbackward* output with >0.9 of probability).

CHROMOPAINTER & FINESTRUCTURE

The painting profile of the individuals in the uIA245 dataset was obtained with CHROMOPAINTERv2[42] in an "all vs all" run. The recombination (*-n* 136.630) and mutation (*-m* 0.00034) parameters were estimated on five randomly selected chromosomes (3, 7, 10, 18 and 22). The *chunkcounts.out* matrix output was used as an input file for fineSTRUCTURE. The software was run with three million MCMC iterations, thinned every 10,000 and preceded by one million burn in iterations: *-x* 1,000,000; *-y* 3,000,000; *-z* 10,000; *-t* 1,000,000. The MCMC file (*.xml*) was used to build the tree that was cut considering the number of individuals in each cluster (less than five) and the Total Variation Distance (TVD <0.03) as elimination criteria.

Identity by descent

The IBD blocks shared in the uIA245 phased dataset were identified using Refined-IBD with default parameters.[43] The IBD blocks inside each cluster identified with fineSTRUCTURE were divided into 9 length bins,[9] and the cluster average of the pair's total length of IBD for each bin was plotted. To explore the IBD shared between clusters, only pairwise comparisons with more than one IBD block were retrieved and the IBD number was adjusted for sample size by dividing by the product of the number of individuals in the two clusters involved.[2]