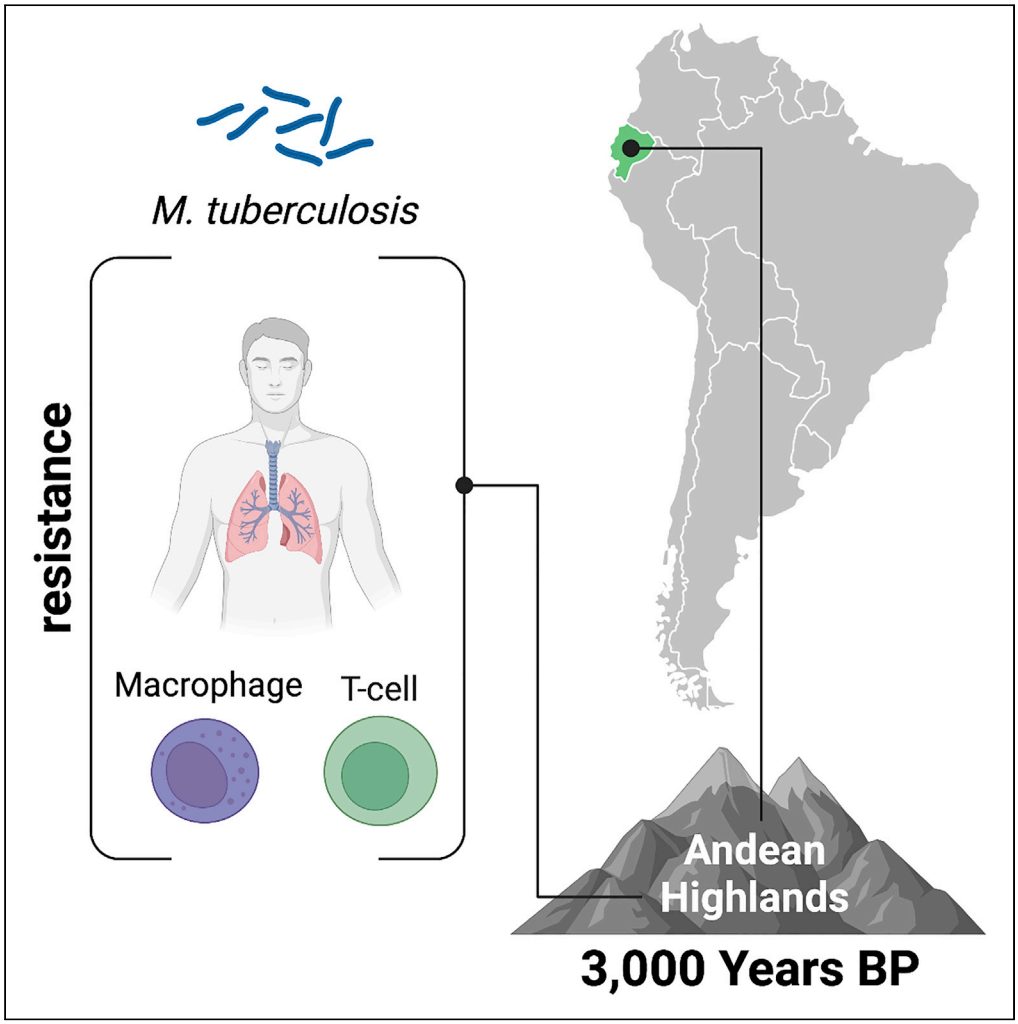
iScience

CellPress
OPEN ACCESS

Article

Genomic evidence for adaptation to tuberculosis in the Andes before European contact



Sophie K. Joseph, Nicola Rambaldi Migliore, Anna Olivieri, ..., Fabricio González-Andrade, Alessandro Achilli, John Lindo

alessandro.achilli@unipv.it (F.G.-A.)
fabriciogonzaleza@gmail.com (A.A.)
jlindo@emory.edu (J.L.)

Highlights

Indigenous people of Ecuador may have adapted to tuberculosis over 3,000 years ago

Selection in cardiovascular and hypoxia pathways distinct from Peruvian populations

Population collapse at arrival of Europeans more severe than other areas of the Andes

Joseph et al., iScience 26, 106034
February 17, 2023
https://doi.org/10.1016/j.isci.2023.106034

iScience

CellPress
OPEN ACCESS

Article

Genomic evidence for adaptation to tuberculosis in the Andes before European contact

Sophie K. Joseph,[1] Nicola Rambaldi Migliore,[2] Anna Olivieri,[2] Antonio Torroni,[2] Amanda C. Owings,[3] Michael DeGiorgio,[4] Wladimir Galarza Ordóñez,[5] J.J. Ortiz Aguilú,[6] Fabricio González-Andrade,[7,*] Alessandro Achilli,[2,*] and John Lindo[1,8,*]

SUMMARY

Most studies focusing on human high-altitude adaptation in the Andean highlands have thus far been focused on Peruvian populations. We present high-coverage whole genomes from Indigenous people living in the Ecuadorian highlands and perform multi-method scans to detect positive natural selection. We identified regions of the genome that show signals of strong selection to both cardiovascular and hypoxia pathways, which are distinct from those uncovered in Peruvian populations. However, the strongest signals of selection were related to regions of the genome that are involved in immune function related to tuberculosis. Given our estimated timing of this selection event, the Indigenous people of Ecuador may have adapted to *Mycobacterium tuberculosis* thousands of years before the arrival of Europeans. Furthermore, we detect a population collapse that coincides with the arrival of Europeans, which is more severe than other regions of the Andes, suggesting differing effects of contact across high-altitude populations.

INTRODUCTION

According to archeological evidence, people have inhabited the Andes region of South America for over 12,000 years, including areas encompassing parts of modern-day Ecuador, Peru, Bolivia, and Chile. For millennia, the Andes have been home to many complex and interconnected societies.[1–3] After the arrival of Europeans in the Andes, approximately 500 years before present (BP), Indigenous populations of the region suffered severe population contractions as a result of warfare, slavery, cultural changes, and the introduction of novel pathogens—notably, smallpox and influenza.[3,4] In addition, European colonizers carried strains of tuberculosis (TB) to the Americas. Genetic and bioarchaeological evidence suggest the possible presence of other TB strains in the Americas long before contact.[5–7] However, its prevalence throughout South America before European contact is still not well understood.

Another poorly understood aspect of the Andes is the genetic underpinnings of high-altitude adaptation. Relevant studies have thus far been focused on the Quechua and Aymara populations from Peru and Bolivia, which represent most of the genetic research in this area of the world.[8–10] Other high-altitude populations, such as Tibetans and Ethiopians, exhibit population-specific alleles within genes involved with the hypoxia inducible factor (HIF) and nitric oxide pathways.[11,12] However, previous studies in high-altitude Peruvian and Bolivian populations generally lack strong evidence for selection on population-specific alleles relating to the adaptation to hypoxic conditions in the Andean Highlands.[8–10]

Currently, genomic studies are lacking in collaboration with Indigenous peoples from Ecuador, including those living in the Highlands. Here we aim to investigate the genetic underpinnings of hypoxia adaptation by utilizing genomic data from two Indigenous Ecuadorian populations who are closely related. We report the results of scans for positive selection using allele-frequency, haplotype-based, and admixture-aware methods with whole genome sequence data from individuals of greater than 98% estimated Indigenous ancestry from Ecuador. We find signatures of selection detected at loci related to hypoxia and cardiovascular pathways, which differ from other populations in the Andean highlands. However, the strongest signals of selection were not associated with high-altitude. Instead, we find surprising evidence of selection on

[1]Department of Anthropology, Emory University, Atlanta, GA 30322, USA

[2]Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, Pavia 27100, Italy

[3]Department of Biology, University of Iowa, Iowa City, IA 52242, USA

[4]Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

[5]Centro de Estudios, Investigaciones y Trabajos Arqueológicos (CEITA), Cuenca, Ecuador

[6]Universidad Técnica de Manabí, 130105 Manabí, Ecuador

[7]Translational Medicine Unit, Central University of Ecuador, Faculty of Medical Sciences, Iquique N14-121 y Sodiro-Itchimbia, Sector El Dorado, 170403 Quito, Ecuador

[8]Lead contact

*Correspondence: alessandro.achilli@unipv.it (F.G.-A.), fabriciogonzaleza@gmail.com (A.A.), jlindo@emory.edu (J.L.)

https://doi.org/10.1016/j.isci.2023.106034

loci which may be linked to the immune response to TB, occurring thousands of years before the arrival of Europeans and highlighting questions regarding the pre-contact prevalence of TB in the ancient Andes.

RESULTS

Demographic analyses

We sequenced high-coverage whole genomes from fifteen Indigenous individuals living in several Ecuadorian provinces from the high-altitude (above 2,500 m) Andean region (Figure 1A and Table S1). Seven of the participants live in neighboring communities from the northern provinces of Ecuador, which include Karanki, Kisapincha, Panzaleo, and Otavalo. Eight participants live in the Loja province in southern Ecuador, including individuals from the Oñacapac and Gañil communities. Community engagement is detailed under the STAR Methods and a detailed plan for the dissemination of the results is detailed in Figure S2.

To facilitate an evolutionary comparison between the different populations, we first conducted demographic analyses and found that the two population groups (geographically separated, Figure 1A) have a close ancestral relationship but remain distinct, despite being linguistically connected. We performed principal components analysis, generated maximum likelihood trees, and examined population substructure via cluster analysis (Figure 1).[13–15] Our cluster analysis revealed that the 15 high-altitude individuals from Ecuador exhibit greater than 98% estimated Indigenous American ancestry (Figure 1E) and share their own branch on the maximum likelihood tree, with the Quechua of Peru (also from the Andean highlands) forming an adjacent branch (Figure 1C). For the ancestry cluster analyses, we also included high-coverage ancient samples (Rio Uncallane) from Lake Titicaca (3,800m elevation), Peru, which date to ~1800 years BP.[3] Of interest, we find that the high-altitude Quechua populations share three ancestries, which include clusters seen in the ancient Andeans (purple) and contemporary populations from North and Central American populations (predominantly blue with some teal). The Northern Kichwa population shares the contemporary American ancestry clusters (predominantly teal with some blue), with a few individuals showing some of the ancient Andean cluster (purple), whereas the populations from Loja do not exhibit as much of this ancient ancestry. Combined with the maximum likelihood tree result, both the Ecuadorian highland populations seem somewhat differentiated from the high-altitude Quechua. This is a surprising result given their geographic proximity and their connection to the ancient empires of the region, including the Inca. This ancestral distinction may possibly reflect limited gene flow and perhaps differing selection pressures. Lastly, we investigated changes in effective population size in both Indigenous Populations using the program Relate.[16] We found evidence of a bottleneck that coincides with the first entry into the Americas ~20,000 years ago.[17] We also found a population collapse approximately 500 years ago, which coincides with the arrival of the Spanish to the Andean region.[18] Both populations suffer a decline of approximately 80% after contact (Figure 1D). Although this level of collapse has been reported in other Indigenous populations using genomic data,[19] it is more severe than that reported for high-altitude populations close to the Lake Titicaca region of Peru and Bolivia,[3] suggesting different dynamics and consequences of European contact in the Andean highlands.[16]

Selection scans

After establishing the ancestral connection between the two populations, we performed a series of scans for positive selection. We first considered any possible admixture that may have occurred between the groups, which could yield false-positive signals.[20] To accomplish this, we employed the Ohana Program, which is a robust approach for detecting positive selection in the presence of admixture, because of its incorporation of cluster analyses to determine the number of underlying ancestral populations.[20] This method predicted the genome-wide variance pattern of the two populations from Ecuador relative to outgroup populations (Figures 2A and 2B). If an allele in the Ecuadorian populations differed significantly from this predicted simulation pattern (using a likelihood ratio test), then the allele was inferred to be under putative positive selection. One of the strongest signals from the scan of the Loja population was associated with intergenic SNP near the *ANXA1* gene, which has an allele frequency of 1.0 in the Loja population (Table S3). This gene is important for triggering the cell death of TB-infected monocytes and generating antigens against *Mycobacterium tuberculosis* for presentation to $CD8^{+}$T-cells.[21] Another top candidate for selection is an intronic SNP on the *ANO7* gene. Data from Chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) assays indicate that this specific *ANO7* SNP is a transcription factor binding site for GATA-2, which is implicated in the HIF hypoxia response pathway.[22,23] In the scan of

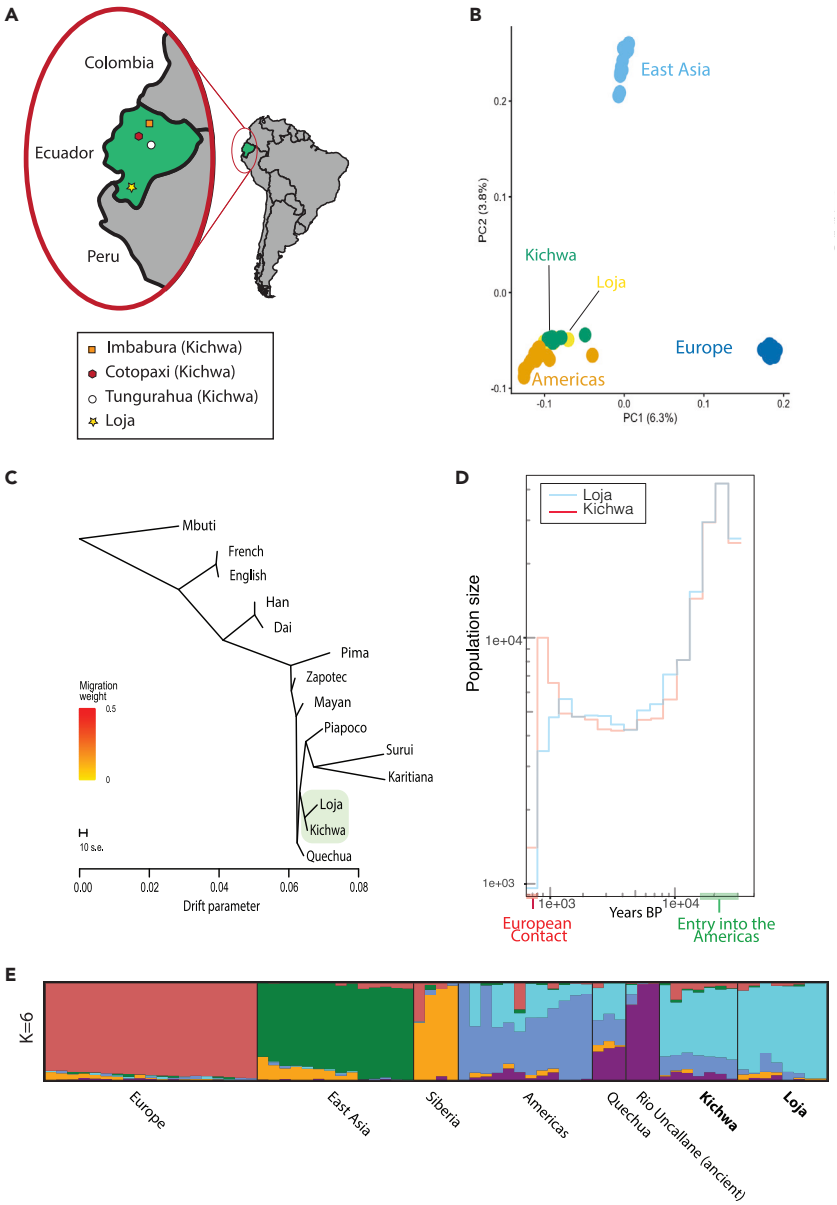


Figure 1. Demographic analyses

(A) Map of population locations. populations.

(B) Principal components analysis showing first two principal components, including individuals from this study and individuals from Europe, East Asia, and the Americas obtained from the SGDP dataset.

(C) Maximum likelihood trees generated by TreeMix showing ancestry relationships between Ecuadorian groups and individuals in the SGDP dataset.

(D) Changes in effective population size. According to the model, both the Kichwa and Loja populations suffer a population collapse that coincides with arrival of the Spanish to the Andean highlands.

(E) Visualization of cluster analysis at K = 6, which exhibited the lowest BIC value.

the Kichwa population, among the top hits was an intronic polymorphism of the *GALNT13* gene, which binds HIF and is known to be upregulated under hypoxic conditions.[24]

Next, we utilized the phased dataset to scan for positive selection using tests which rely on extended haplotype homozygosity (EHH).[25–27] The guiding principle of EHH holds that, when comparing two or more populations, the presence of longer haplotypes at a particular locus indicates positive selection,

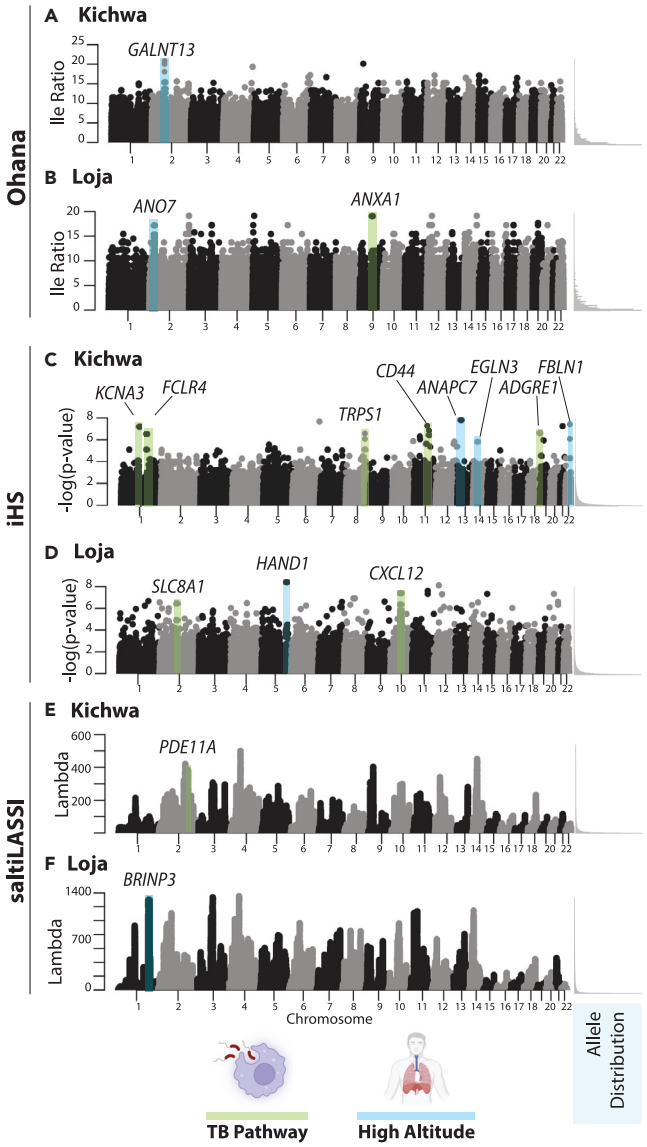


Figure 2. Selection scans highlighting the strongest signals of positive selection

Ohana selscan Manhattan plot and histogram of the allele distribution of the corresponding likelihood ratio (lle ratio) for the Kichwa (A) and Loja (B). iHS Manhattan plot and histogram of the allele distribution at the corresponding-$\log_{10}$(pvalue) for the Kichwa (C) and the Loja (D). Manhattan plot of the log composite likelihood ratio (Λ) from the saltiLASSI program and histogram of the allele distribution at the corresponding Λ for the Kichwa (E) and the Loja (F).

because those regions of the genome have not been broken down by recombination. Thus, the genomes obtained from these Indigenous populations are ideal for use with haplotype-based selection scans because these populations are relatively closely related, meaning that significant differences in haplotype length are likely because of recent positive selection (see Figure 1B for PCA plot comparing individuals from the present study to individuals from the Simons Genome Diversity Project (SGDP) dataset[28]).

We performed an integrated haplotype score (iHS) test, which detects signals of positive selection by utilizing a standardized log ratio of the integrals of observed decay of EHH in a focal population—the Kichwa individuals in the present study (Figure 2C).[25,27,29] Among the top candidates for selection within the high-altitude individuals is an intronic SNP on the *FBLN1* gene, which codes for a protein important for allowing hemostasis to occur, a critical process involving the coagulation of blood that is greatly affected by hypoxic conditions.[30,31] Top candidate SNPs for positive selection also include an intergenic SNP near the *EGLN3*

gene, which codes for an important regular stability of the HIFs in response to hypoxia.[32] However, it should be noted that this particular *EGLN3* SNP, unlike the *EGLN1* polymorphism characteristic of Tibetan populations, has yet to be characterized in the functional genomics literature, though the *EGLN3* gene was recently identified as being under selection in highland populations in Peru.[11,33] Furthermore, an intergenic SNP near the *ANAPC7* gene with significant probability of being under positive selection is a known transcription factor binding site for GATA-2, which has been implicated in the HIF hypoxia response pathway.[22,23] Also among the top hits were genes implicated in immune and anti-TB response pathways. These included intergenic SNP near the *TRPS1* gene, which, according to data from ChIP-seq, is a binding site for the CEBPB transcription factor, implicated in the early anti-TB response via its role in the expansion of macrophages.[22,34] In addition, an intergenic polymorphism near in the *CD44* gene was among the top candidates for positive selection. *CD44* codes for a well-known cell adhesion molecule involved in immune signaling, which has recently also been shown to serve as an important cell surface binding site for *M. tuberculosis* bacteria on macrophages.[35]*CD44* was recently identified as being under selection in highland populations in Peru.[33] Top hits also included intronic SNPs on the *ADGRE1* gene, an important and well-known marker of macrophages and the *FCRL4* gene, which codes for an immunoglobulin receptor important to the function of memory B-cells.[36–38] Furthermore, an intergenic polymorphism found to be under significant probability of positive selection located near the *KCNA3* gene, a regulator of antigen specificity in memory T-cells, is an eQTL for the *CD53* gene. *CD53* is important for activating and sustaining adaptive immunity through the course of bacterial and viral infections.[39–41]

We also performed the same iHS test for the Loja population, which yielded an intergenic SNP near the *SLC8A1* gene, also a binding site for the CEBPB transcription factor (Figure 2D).[25,29,34] The *SLC8A1* gene was recently identified as being under strong positive selection in highland populations in Peru.[33] An additional top candidate for positive selection was an intergenic SNP near the *CXCL12* gene, which codes for a chemokine known to be upregulated in severe TB infections that include pleurisy, the severe inflammation of thin layers which cover the lungs.[42] Relating to high-altitude, the top hit from this iHS scan was an intergenic SNP near the *HAND1* gene, which is a transcription factor regulated by hypoxic conditions.[43] When upregulated, it has been shown to have a protective effect against myocardial ischemia, which is the reduction in the heart muscle's ability to pump blood efficiently. It is an integral part of the regulatory connection between environmental oxygen levels and oxygen availability in the blood.

Furthermore, we conducted additional haplotype-based selection scans with the saltiLASSI, method, which uses a composite likelihood ratio test statistic that detects signals of positive selection based on the spatial distribution of the distortions of haplotype frequency spectra across a chromosome.[44] In particular, saltiLASSI has been shown to have high power to detect both hard and soft sweeps from positive selection of recent to moderate age. Applying saltiLASSI to the Kichwa analysis group, one of the regions identified with a high likelihood ratio score and evidence of being an outlier (Figure 2E), and thus likely under positive selection, is a polymorphism on the *PDE11A* gene. According to results from ChIP-seq assays, this SNP is a binding site for the CEBPB transcription factor, implicated in the early anti-TB response via its role in the proliferation of macrophages.[22,34] In the Loja analysis group, we found that an intergenic SNP near the *BRINP3* gene was likely under positive selection (Figure 2F). This gene has been previously identified in Andean Aymara populations in Bolivia as a candidate for adaptation to hypoxic conditions through adaptations of the cardiovascular system, rather than the HIF response pathway; however, this particular SNP was not identified in previous studies.[8] The SNPs representing approximately the top 1% for all scans are detailed in the Data S1 spreadsheet.

To better understand the timing and strength of selection on the alleles described above, we employed the CLUES approximate likelihood method that allows for detecting the allele trajectory, timing of selection, and selection strength.[45] We found evidence that alleles associated with genes thought to be involved in TB resistance, particularly the *PDE11A* (identified in the Kichwa analysis group) and *SLC8A1* (identified in the Loja analysis group) genes, underwent strong positive selection (Figure 3 and Table S4 and Figure S5). Furthermore, the allele frequency trajectories indicate that these immune-related alleles increased within the population many generations before the arrival of Europeans. Both the *PDE11A* and *SLC8A1* SNP trajectories began rapidly increasing ~3000 BP, and this period roughly corresponds with the transition from small hunter-gatherer groups to larger agricultural societies.[1] These societies were more compact and populous, which could have facilitated the spread of respiratory diseases, such as TB.

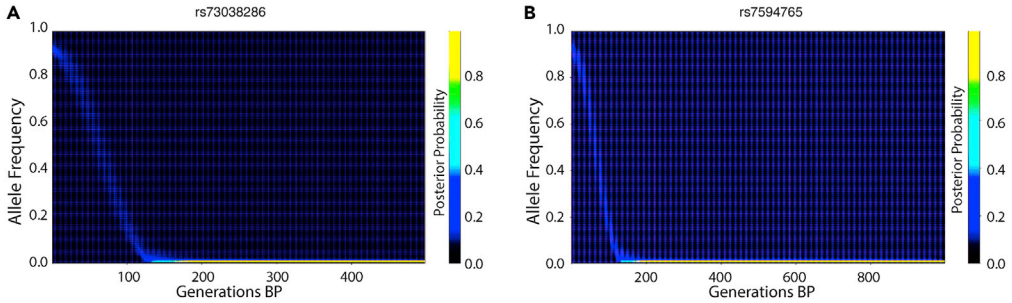


Figure 3. Allele trajectories and selection strength

(A) Estimated allele trajectory over time (generations before present) for the allele associated with the *PDE11A* gene putatively under selection in the northern Kichwa.

(B) Estimated allele trajectory for the allele associated with the *SLC8A1* gene putatively under selection in the Loja population. Both alleles are known binding sites for the CEBPB transcription factor, implicated in the early anti-TB response via its role in the proliferation of monocyte macrophages.

DISCUSSION

Life in the high-altitude Andes undoubtedly requires many complex adaptations, both cultural and genetic. In the Andean highlands, Indigenous people display higher than average aerobic capacity, known as VO_2 max, at high altitude, as well as many other physiological changes.[10] Here, we aimed to shed light on the genetic underpinnings of these physiological adaptations. Our analyses highlighted several alleles with a high probability of positive selection that are located in genes generally thought to be involved in important cardiovascular and hypoxia-related pathways; however, unlike studies from Tibet and Ethiopia, we did not identify specific functionally characterized polymorphisms.[11,12]

One similarity with studies of Tibetan and Ethiopian populations was our identification of an SNP near *EGLN3*, a member of the EGLN gene family, which is the essential regulator of the HIFs.[32] This gene family, including the *EGLN1* gene, has been identified in Tibetan and Peruvian populations as key for aerobic function in hypoxic conditions.[10,11] However, this particular *EGLN3* SNP from the present study has yet to be characterized in the functional genomics literature, though *EGLN3* was recently also identified as being under selection in highland populations in Peru, as was the *GALNT13* gene.[33] In addition, an SNP near the *BRINP3* gene was identified as a candidate for positive selection in the present study in the Loja analysis group. *BRINP3* is thought to be important for adaptation to hypoxic conditions via the cardiovascular system and has been identified for high probability of recent positive selection in Bolivian Aymara and highland Peruvian populations.[8,33] All other genes with SNPs under high probability of selection related to hypoxia or cardiovascular function in our study differed from those identified in other Andean and worldwide high-altitude population studies (Table S6).[8–10] This suggests there are more complex modes of genetic adaptation to hypoxic conditions in the Andes that are yet to be understood. Our demographic analyses indicate some divergence between Andean highland populations, which could explain why we see mostly different hypoxia-related selection signals (Figure 1). In other words, there may have been convergent evolution of the adaptation to hypoxic conditions, with different genes involved in each separate population. Alternatively, it could be that a polymorphism detected in our selection scans has just not yet been characterized in the functional genomics literature, or it could be that physiological adaptations are epigenetic in nature.

A surprising outcome of this study was the identification of alleles putatively under strong positive selection in genes that may be related to the anti-TB immune response that arose in the Ecuadorian populations thousands of years before European contact. There is considerable debate regarding the presence of TB in the pre-contact Americas. Although TB's prevalence and geographic spread is not fully understood, both genetic and bioarchaeological evidence has suggested its presence within humans in the ancient Americas at least 1,400 BP.[5–7] However, phylogenetic analyses of TB lineages in the present day show that the L4 phylogenetic lineage left Europe during colonization and remains dominant in the Americas today.[7] These works collectively suggest a possible presence of a pre-contact zoonotic lineage before the L4 European lineage brought by colonists. The SNPs putatively under selection identified in the present study were not only involved in direct bacterial antigen recognition by innate immune macrophages, as with *ANXA1*, *CD44*, and *ADGRE1*, but also included genes involved in the pro-inflammatory adaptive

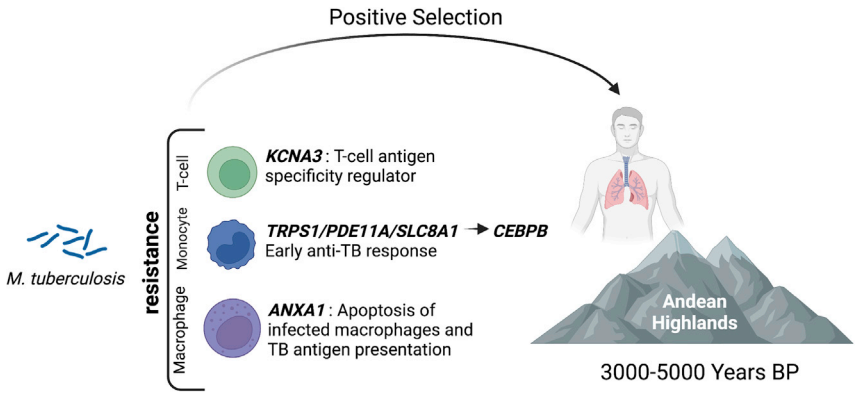


Figure 4. Visualization of major lines of evidence relating to alleles under a high probability of positive selection associated with anti-TB immune response

Associated alleles are listed in Tables S3 and S4. Created with www.BioRender.com.

immune response, such as *KCNA3* and *FCRL4* (Figure 4).[21,35–40] Future microbiological studies could directly test the alleles identified here for functional consequences with respect to the TB immune response. However, in concert with previously published genetic and bioarchaeological evidence, the genomic evidence presented here represents a plausible scenario that TB may have been a selective pressure in the agricultural societies of the Andes, pre-European contact.[5–7]

Recent work on human immune response variation among differing TB strains suggests that immunity is not necessarily robust to strain diversity.[46] Thus, in line with our findings, it is possible that Ecuadorian populations could have developed immunity to South American TB strains in the ancient past while still being susceptible to the European L4 lineage at the time of European contact. However, further functional studies on the identified alleles would be needed for definitive evidence. Furthermore, ancient Andean populations at high-altitude did typically suffer lower proportions of population collapse compared with coastal populations, perhaps due in part to the geographic barrier provided by the mountains.[3,4] Today, the Andes region has a higher proportion of TB L4 sublineage S (L4.4) as compared with other parts of Ecuador, which could be a product of this geographic isolation, though more epidemiological work is needed to understand disease outcomes of this particular sublineage in Ecuadorian populations and how this may relate to population genetic studies such as this one.[47]

European colonizers carried many diseases to the Americas, including the TB L4 lineage, which were unquestionably devastating to Indigenous peoples.[3,4] The findings presented here, in concert with other genetic and bioarchaeological evidence, suggest that TB could have been present to some degree for several thousand years in the Andes before Europeans arrived.[5–7] However, the prevalence of TB before European contact is poorly characterized through time, and the present study suggests that the details of human-pathogen coevolution with respect to TB in the Americas is an area of tremendous potential for future research.

Limitations of the study

Though tremendous advancements have been made in recent years, the size and complexity of the human genome often means that intronic and intergenic polymorphisms—such as those identified in the selection scans from the present study—are not prioritized for functional genomic research. Thus, although future microbiological work could directly test the alleles identified here for functional consequences, it is not possible to know their direct effects on TB immunity or high-altitude adaptation at this time. The present study joins a larger body of previously published genetic and bioarchaeological evidence that TB may have been present in the Andes pre-European contact[5–7] and that humans have adapted uniquely to high-altitude life in the Andes.[8–12,33]

STAR★METHODS

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

- KEY RESOURCES TABLE

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106034.

ACKNOWLEDGMENTS

This work was funded by National Geographic Society grant GR-000043627, by National Science Foundation grants BCS-1926075, BCS-1945046, BCS-2001063, DEB-1949268, and DBI-2130666, and by National Institutes of Health grant R35GM128590. This research also received support from the Italian Ministry of Education, University and Research (MIUR) for Progetti PRIN2017 20174BTC4R, Dipartimenti di Eccellenza Program (2018–2022)–Department of Biology and Biotechnology "L. Spallanzani," University of Pavia, and the "Fondazione Adriano Buzzati – Traverso" for the L. Luca Cavalli-Sforza fellowship.

AUTHOR CONTRIBUTIONS

F.G-A. engaged with and collected blood spot samples from communities which comprise the Kichwa analysis group from several northern provinces in Ecuador. A.Ow. extracted DNA and built libraries from dried blood spots from the Kichwa analysis group. A.Ol., A.T., A.A., and N.R.M. engaged with and collected buccal swab samples from communities which comprise the analysis group from the southern Ecuadorian province of Loja. A.A., N.R.M., A.Ol., and A.T. extracted DNA and built libraries from buccal swabs from the Loja analysis group, as previously published by Brandini et al.[48] S.K.J. and J.L. performed data analysis, including processing, demographic analyses, and selection scans. M.D. performed data analysis, including selection scans. W.G.O. and J.J.O.A. returned results to the Indigenous communities who have contributed to the present study via a series of community outreach workshops. All authors revised and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

INCLUSION AND DIVERSITY

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

Received: July 13, 2022
Revised: November 11, 2022
Accepted: January 17, 2023
Published: January 24, 2023

REFERENCES

1. Rademaker, K., Hodgins, G., Moore, K., Zarrillo, S., Miller, C., Bromley, G.R.M., Leach, P., Reid, D.A., Álvarez, W.Y., and Sandweiss, D.H. (2014). Paleoindian settlement of the high altitude Peruvian Andes. Science *346*, 466–469. https://doi.org/10.1126/science.1258260.

2. 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.

3. 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 through European contact. Sci. Adv. *4*, eaau4921. https://doi.org/10.1126/sciadv.aau4921.

4. Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J.S., Willerslev, E., DeGiorgio, M., and Malhi, R.S. (2016). A time transect of exomes from a Native American population before and after European contact. Nat. Commun. *7*, 13175. https://doi.org/10.1038/ncomms13175.

5. Nelson, E.A., Buikstra, J.E., Herbig, A., Tung, T.A., and Bos, K.I. (2020). Advances in the molecular detection of tuberculosis in

pre-contact Andean South America. Int. J. Paleopathol. *29*, 128–140. https://doi.org/10.1016/j.ijpp.2019.12.006.

6. Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., et al. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature *514*, 494–497. https://doi.org/10.1038/nature13591.

7. Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J., Alfsnes, K., Pettersson, J.O.-H., Kirkeleite, I., et al. (2018). Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci. Adv. *4*, eaat5869. https://doi.org/10.1126/sciadv.aat5869.

8. Crawford, J.E., Amaru, R., Song, J., Julian, C.G., Racimo, F., Cheng, J.Y., Guo, X., Yao, J., Ambale-Venkatesh, B., Lima, J.A., et al. (2017). Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. Am. J. Hum. Genet. *101*, 752–767. https://doi.org/10.1016/j.ajhg.2017.09.023.

9. Valverde, G., Zhou, H., Lippold, S., de Filippo, C., Tang, K., López Herráez, D., Li, J., and Stoneking, M. (2015). A novel candidate region for genetic adaptation to high altitude in Andean populations. PLoS One *10*, e0125444. https://doi.org/10.1371/journal.pone.0125444.

10. Brutsaert, T.D., Kiyamu, M., Elias Revollendo, G., Isherwood, J.L., Lee, F.S., Rivera-Ch, M., Leon-Velarde, F., Ghosh, S., and Bigham, A.W. (2019). Association of EGLN1 gene with high aerobic capacity of Peruvian Quechua at high altitude. Proc. Natl. Acad. Sci. USA. *116*, 24006–24011. https://doi.org/10.1073/pnas.1906171116.

11. Yang, J., Jin, Z.-B., Chen, J., Huang, X.-F., Li, X.-M., Liang, Y.-B., Mao, J.-Y., Chen, X., Zheng, Z., Bakshi, A., et al. (2017). Genetic signatures of high-altitude adaptation in Tibetans. Proc. Natl. Acad. Sci. USA. *114*, 4189–4194. https://doi.org/10.1073/pnas.1617042114.

12. Huerta-Sánchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H.E., Cavalleri, G.L., Robbins, P.A., et al. (2013). Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. Mol. Biol. Evol. *30*, 1877–1888. https://doi.org/10.1093/molbev/mst089.

13. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967.

14. Novembre, J., Williams, R., Pourreza, H., Wang, Y., Carbonetto, P., and Novembre, J. (2013). PCAviz: Visualizing Principal Components Analysis. R Package Version 0.3-37. http://github.com/NovembreLab/PCAviz.

15. Chiu, A.M., Molloy, E.K., Tan, Z., Talwalkar, A., and Sankararaman, S. (2022). Inferring population structure in biobank-scale genomic data. Am. J. Hum. Genet. *109*, 727–737. https://doi.org/10.1016/j.ajhg.2022.02.015.

16. Speidel, L., Forest, M., Shi, S., and Myers, S.R. (2019). A method for genome-wide genealogy estimation for thousands of samples. Nat. Genet. *51*, 1321–1329. https://doi.org/10.1038/s41588-019-0484-x.

17. Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.S., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. Science *349*, aab3884. https://doi.org/10.1126/science.aab3884.

18. Livi-Bacci, M. (2006). The depopulation of hispanic America after the conquest. Popul. Dev. Rev. *32*, 199–232. https://doi.org/10.1111/j.1728-4457.2006.00116.x.

19. García-Ortiz, H., Barajas-Olmos, F., Contreras-Cubas, C., Cid-Soto, M.Á., Córdova, E.J., Centeno-Cruz, F., Mendoza-Caamal, E., Cicerón-Arellano, I., Flores-Huacuja, M., Baca, P., et al. (2021). The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. Nat. Commun. *12*, 5942. https://doi.org/10.1038/s41467-021-26188-w.

20. Cheng, J.Y., Mailund, T., and Nielsen, R. (2017). Fast admixture analysis and population tree estimation for SNP and NGS data. Bioinformatics *33*, 2148–2155. https://doi.org/10.1093/bioinformatics/btx098.

21. Tzelepis, F., Gillard, J., Jaworska, J., Nishimura, T., Verway, M., Hassani-Ardakani, K., Remold, H., Vali, H., and Maziar, D. (2013). Critical role of Annexin A1 in protection against pulmonary. Mycobacterium tuberculosis infection *190*, P3320.

22. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831. https://doi.org/10.1101/gr.136184.111.

23. Tabata, M., Tarumoto, T., Ohmine, K., Furukawa, Y., Hatake, K., Ozawa, K., Hasegawa, Y., Mukai, H., Yamamoto, M., and Imagawa, S. (2001). Stimulation of GATA-2 as a mechanism of hydrogen peroxide suppression in hypoxia-induced erythropoietin gene expression. J. Cell. Physiol. *186*, 260–267. https://doi.org/10.1002/1097-4652(200002)186:2<260::aid-jcp1025>3.0.co;2-k.

24. Desai, A.A., Arana, N., Kafisanwo, O., Letsiou, E., Warda, A., Chen, J., Fahs, A., Garcia, J.G.N., and Machado, R.F. (2013). N-acetylgalactosaminyltransferase 13 (galnt13) is A novel candidate gene in pulmonary arterial hypertension and in pathobiologic responses to hypoxia. Am. J. Respir. Crit. Care Med. *187*, A4645.

25. Gautier, M., Klassmann, A., and Vitalis, R. (2017). REHH 2.0: a reimplementation of the R package REHH to detect positive selection from haplotype structure. Mol. Ecol. Resour. *17*, 78–90. https://doi.org/10.1111/1755-0998.12634.

26. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918. https://doi.org/10.1038/nature06250.

27. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832–837. https://doi.org/10.1038/nature01140.

28. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The simons genome diversity project: 300 genomes from 142 diverse populations. Nature *538*, 201–206. https://doi.org/10.1038/nature18964.

29. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72. https://doi.org/10.1371/journal.pbio.0040072.

30. Tran, H., Tanaka, A., Litvinovich, S.V., Medved, L.V., Haudenschild, C.C., and Argraves, W.S. (1995). The interaction of fibulin-1 with fibrinogen: a potential role in hemostasis and thrombosis. J. Biol. Chem. *270*, 19458–19464. https://doi.org/10.1074/jbc.270.33.19458.

31. Schobersberger, W., Hoffmann, G., and Gunga, H.-C. (2005). Interaction of hypoxia and haemostasis–hypoxia as a prothrombotic factor at high altitude? Wien. Wien Med. Wochenschr. *155*, 157–162. https://doi.org/10.1007/s10354-005-0163-7.

32. Ivan, M., and Kaelin, W.G. (2017). The EGLN-HIF O2-sensing system: multiple inputs and feedbacks. Mol. Cell *66*, 772–779. https://doi.org/10.1016/j.molcel.2017.06.002.

33. Caro-Consuegra, R., Nieves-Colón, M.A., Rawls, E., Rubin-de-Celis, V., Lizárraga, B., Vidaurre, T., Sandoval, K., Fejerman, L., Stone, A.C., Moreno-Estrada, A., and Bosch, E. (2022). Uncovering signals of positive selection in Peruvian populations from three ecological regions. Mol. Biol. Evol. *39*, msac158. https://doi.org/10.1093/molbev/msac158.

34. Delgobo, M., Mendes, D.A., Kozlova, E., Rocha, E.L., Rodrigues-Luiz, G.F., Mascarin, L., Dias, G., Patrício, D.O., Dierckx, T., Bicca, M.A., et al. (2019). An evolutionary recent IFN/IL-6/CEBP axis is linked to monocyte expansion and tuberculosis severity in humans. Elife *8*, e47013. https://doi.org/10.7554/elife.47013.

35. Leemans, J.C., Florquin, S., Heikens, M., Pals, S.T., van der Neut, R., and Van Der Poll, T.

(2003). CD44 is a macrophage binding site for Mycobacterium tuberculosis that mediates macrophage recruitment and protective immunity against tuberculosis. J. Clin. Invest. *111*, 681–689. https://doi.org/10.1172/jci16936.

36. Waddell, L.A., Lefevre, L., Bush, S.J., Raper, A., Young, R., Lisowski, Z.M., McCulloch, M.E.B., Muriuki, C., Sauter, K.A., Clark, E.L., et al. (2018). ADGRE1 (EMR1, F4/80) is a rapidly-evolving gene expressed in mammalian monocyte-macrophages. Front. Immunol. *9*, 2246. https://doi.org/10.3389/fimmu.2018.02246.

37. Liu, Y., Goroshko, S., Leung, L.Y.T., Dong, S., Khan, S., Campisi, P., Propst, E.J., Wolter, N.E., Grunebaum, E., and Ehrhardt, G.R.A. (2020). FCRL4 is an Fc receptor for systemic IgA, but not mucosal secretory IgA. J. Immunol. *205*, 533–538. https://doi.org/10.4049/jimmunol.2000293.

38. du Plessis, W.J., Walzl, G., and Loxton, A.G. (2016). B cells as multi-functional players during Mycobacterium tuberculosis infection and disease. Tuberculosis *97*, 118–125. https://doi.org/10.1016/j.tube.2015.10.007.

39. Chiang, E.Y., Li, T., Jeet, S., Peng, I., Zhang, J., Lee, W.P., DeVoss, J., Caplazi, P., Chen, J., Warming, S., et al. (2017). Potassium channels Kv1.3 and KCa3.1 cooperatively and compensatorily regulate antigen-specific memory T cell functions. Nat. Commun. *8*, 14644. https://doi.org/10.1038/ncomms14644.

40. Dunlock, V.E. (2020). Tetraspanin CD53: an overlooked regulator of immune cell function. Med. Microbiol. Immunol. *209*, 545–552. https://doi.org/10.1007/s00430-020-00677-z.

41. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

42. Kohmo, S., Kijima, T., Mori, M., Minami, T., Namba, Y., Yano, Y., Yoneda, T., Takeda, Y., Kitada, S., Yamaguchi, T., et al. (2012). CXCL12 as a biological marker for the diagnosis of tuberculous pleurisy. Tuberculosis *92*, 248–252. https://doi.org/10.1016/j.tube.2012.01.001.

43. Breckenridge, R.A., Piotrowska, I., Ng, K.-E., Ragan, T.J., West, J.A., Kotecha, S., Towers, N., Bennett, M., Kienesberger, P.C., Smolenski, R.T., et al. (2013). Hypoxic regulation of Hand1 controls the fetal-neonatal switch in cardiac metabolism. PLoS Biol. *11*, e1001666. https://doi.org/10.1371/journal.pbio.1001666.

44. DeGiorgio, M., and Szpiech, Z.A. (2022). A spatially aware likelihood test to detect sweeps from haplotype distributions. PLoS Genet. *18*, e1010134. https://doi.org/10.1101/2021.05.12.443825.

45. Stern, A.J., Wilton, P.R., and Nielsen, R. (2019). An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. PLoS Genet. *15*, e1008384. https://doi.org/10.1371/journal.pgen.1008384.

46. Tientcheu, L.D., Koch, A., Ndengane, M., Andoseh, G., Kampmann, B., and Wilkinson, R.J. (2017). Immunological consequences of strain variation within the Mycobacterium tuberculosis complex. Eur. J. Immunol. *47*, 432–445. https://doi.org/10.1002/eji.201646562.

47. Garzon-Chavez, D., Garcia-Bereguiain, M.A., Mora-Pinargote, C., Granda-Pardo, J.C., Leon-Benitez, M., Franco-Sotomayor, G., Trueba, G., and de Waard, J.H. (2020). Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. Sci. Rep. *10*, 6237. https://doi.org/10.1038/s41598-020-62824-z.

48. Brandini, S., Bergamaschi, P., Cerna, M.F., Gandini, F., Bastaroli, F., Bertolini, E., Cereda, C., Ferretti, L., Gómez-Carballa, A., Battaglia, V., et al. (2018). The paleo-Indian entry into South America according to mitogenomes. Mol. Biol. Evol. *35*, 299–311. https://doi.org/10.1093/molbev/msx267.

49. Walsh, P.S., Metzger, D.A., and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques *10*, 506–513.

50. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

51. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

52. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

53. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

54. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569. https://doi.org/10.1038/ng.608.

55. 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.

56. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097. https://doi.org/10.1086/521987.

57. 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.

58. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909. https://doi.org/10.1038/ng1847.

59. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

60. 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.

61. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823. https://doi.org/10.1093/bioinformatics/btw327.

62. 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.

63. Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics *27*, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
DNA from FTA bloodspot cards	This study	Kichwa analysis group
DNA from Buccal swabs	Brandini et al. (2018)[48]	Loja analysis group
Critical commercial assays		
Chelex-100 extraction protocol	Walsh et al. (1991)[49]	N/A
Phenol/chloroform extraction protocol	Brandini et al. (2018)[48]	N/A
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®	New England Biolabs	Cat# E7645L
Deposited data		
The whole genomes generated in this study are available on the NCBI Sequence Read Archive	NCBI SRA	NCBI SRA PRJNA883976
Software and algorithms		
bwa-mem 0.7.17	Li and Durbin, 2009[50]	N/A
BCFtools	Danecek et al., 2021[51]	N/A
ANNOVAR	Wang et al., 2010[52]	N/A
VCFtools	Danecek et al., 2011[53]	N/A
Relatedness (VCFtools)	Yang et al., 2010[54]	N/A
Relatedness2 (VCFtools)	Manichaikul et al., 2010[55]	N/A
Beagle 5.2	Browning and Browning, 2007[56]	N/A
EIGENSOFT	Patterson et al., 2006[57] Price et al., 2006[58]	N/A
Plink v1.9	Chang et al., 2015[59]	N/A
PCAviz R package v0.3-37	Novembre et al., 2013[14]	N/A
ADMIXTURE	Alexander et al., 2009[60]	N/A
Pong	Behr et al., 2015[61]	N/A
TreeMix	Pickerell and Pritchard, 2012[13]	N/A
Ohana	Cheng et al., 2017[20]	N/A
REHH	Gautier et al., 2017[25]	N/A
saltiLASSI	DeGiorgio and Szpiech, 2022[44]	N/A
Clues	Stern et al., 2019[45]	N/A
Relate	Speidel et al., 2019[16]	N/A
RFmix	Maples et al., 2013[62]	N/A
SCOPE	Chiu et al., 2022[15]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact John Lindo, jlindo@emory.edu.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The whole genomes generated in this study are available on NCBI Sequence Read Archive and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact John Lindo (jlindo@emory.edu) upon request.

METHOD DETAILS

Community engagement and sample collection

Samples were allocated to the Kichwa and Loja analysis groups based on their geographic location. As this is not an experimental biomedical research study, we do not report the collection of health or immune status, or the use of cell lines.

Kichwa individuals

Donor blood samples were collected in partnership with Indigenous communities in the northern Ecuadorian provinces of Cotopaxi, Tungurahua, and Imbabura. Adult donor samples and geographic locations were collected in accordance with the ethical standards of the 1964 Declaration of Helsinki and its subsequent amendments. All samples were provided voluntarily and were securely archived and de-identified immediately, and informed consent was offered in Spanish. All sample donors signed the Informed ConsentForm (Figure S7). Donor samples for this project were approved by the Translational Medicine Unit of the Faculty of Medical Sciences at the Central University of Ecuador.

Loja individuals

Donor buccal swab samples were collected with Indigenous communities in the southern Ecuadorian province of Loja, as previously published and described by Brandini and colleagues.[48] Adult donor samples and geographic locations were collected in accordance with the ethical standards of the 1964 Declaration of Helsinki and its subsequent amendments. All samples were provided voluntarily and were securely archived and de-identified immediately, and informed consent was offered in both Spanish and English. Donor samples were approved by the Ethic Committee for Clinical Experimentation of the University of Pavia.

Returning results to communities

To return results to the Indigenous communities who have contributed to the present study, a series of workshops will be held in several Kichwa-speaking communities located in the Cañar cantón, including Sid-Sid, Amanta Bayopungo, and Quilloac. In addition to describing the purpose and methodologies employed in the present study, the workshops will also be devoted to contextualizing how this population genetic research can be understood within their Andean cultural spaces by collaboratively developing a kind of Community Active Memory (see Figure S2 for detailed letter describing the methodology of the workshops).

DNA extraction

DNA from samples obtained from Kichwa individuals was extracted from FTA card blood spots using DNA samples were extracted using a modified Chelex-100 protocol from where samples were rehydrated in 125uL of Chelex-100 and incubated at 56°C overnight.[49] DNA from samples obtained from Loja individuals was extracted from buccal swabs following standard phenol/chloroform protocol by Brandini and colleagues.[48]

QUANTIFICATION AND STATISTICAL ANALYSIS

Dataprocessing, alignments, phasing, and demographic analyses

Data processing and alignments

Libraries were prepared with the NEB Ultra II Illumina library kit and corresponding unique dual indexes (New England Biolabs). The libraries underwent whole genome sequencing at a mean of 25x on an Illumina Novaseq 6000 (Dante Labs) (Table S1). The sequencing reads were aligned to the hg19 human reference genome using bwa-mem 0.7.17.[50] Variants were called with BCFtools, where a minimum of 10 reads

were required to call a variant, along with base and nucleotide quality set to a minimum of 30.[51] The combined VCF with the 15 individuals was then annotated using ANNOVAR.[52] The individuals were also assessed for relatedness utilizing both the relatedness and relatedness2 programs through the VCFtools software, and both programs confirmed there were no close familial relationships between individuals.[53–55]

Haplotype phasing

For use with haplotype-based selection scans, the genomes were phased without imputation (–impute=-FALSE) using the Beagle 5.2 program.[56] VCFtools was used to prepare the VCF by filtering sites out of Hardy-Weinberg equilibrium with a p-value below 10^{-4} and removing indels.[53] Default Beagle 5.2 parameters were used, and the SGDP phased dataset was employed as the reference.[28]

Principal components analysis

Eigenvectors were computed for individuals in Kichwa and Loja Ecuadorian groups from this study with outgroup samples from Europe, East Asia, and the Americas obtained from the SGDP dataset using the Smartpca function of the EIGENSOFT program after pruning whole genome dataset for SNPs in linkage disequilibrium and removing missing data using Plink v1.9.[28,57–59] The first two principal components were visualized using the PCAviz R package v0.3-37.[14]

Ancestry clustering for maximum likelihood trees

We performed model-based clustering analysis using the likelihood-free approach implemented in SCOPE, which utilizes latent subspace estimation and alternating least squares to estimate admixture proportions from individual allele frequencies.[15] To determine the number of underlying ancestral clusters, we assumed K = 2 to K = 15 ancestral clusters using the Adegenet R package v2.1.7.[63] For each K, 10,000 iterations were performed (find.clusters(), max.n.clust = 15, n.iter = 1e4, stat = c("BIC")). K = 6 had the lowest Bayesian information criterion (BIC) value, whereby the smallest BIC value indicated that the model with K = 6 ancestral clusters was the best fit for the data. We then ran the SCOPE program using K = 6 ancestral clusters (–k 6) and otherwise default parameters, and visualized with PONG (Figures S8 and S9).[15,61]

Maximum likelihood trees

TreeMix was applied to the dataset to generate maximum likelihood trees and admixture graphs from allele frequency data.[13] The Mbuti from the SGDP dataset was used to root the tree (–root Mbuti).[28] We accounted for linkage disequilibrium by grouping M adjacent sites (with the –k option), and we chose M such that a dataset with L sites will have approximately $L/M \approx 20{,}000$ independent sites. At the end of the analysis (i.e., number of migrations) we performed a global rearrangement (with the –global option). We considered admixture scenarios with $m = 0$ and $m = 1$ migration events. Each migration scenario was run with 100 replicates, and the replicate with the highest likelihood was chosen to represent the maximum likelihood tree or graph for the given migration scenario (Figure S8).

Estimation of changes in effective population size

The Relate program was also used to model changes in the effective population size of both populations using the phased dataset for both the Kichwa and Loja populations.[16] We utilized a mutation rate of 1.25×10^{-8}, seed set to 1, a generation time of 28 years, 100 iterations, and 0.5 fraction of trees to be dropped. Autosomal chromosomes 1-22 were used to estimate the population sizes.

Selection scans

The Ohana program uses a likelihood ratio test to detect alleles in each population that deviated strongly from a genome-wide covariance structure.[20] The Ohana program does not require the assignment of population labels to each individual's genome, nor does it utilize phased genomes for haplotype information. Instead, it relies on the genomic data to predict the ancestral population assignments of each genome based on allele frequencies and utilizes the results of cluster analysis to determine the number of these underlying ancestral populations. We conducted two Ohana selection scans, one for each population, and chose the outgroup for the scans to be comprised of East Asian and European individuals from the SGDP dataset (see East_Asia and Europe individuals in Table S10 for list of SGDP individuals used[28]). To estimate K, the number of underlying ancestral populations in the tree, we chose the K which exhibited the lowest cross-validation index from the ADMIXTURE program[60] after testing K=2 through K=15. We utilized K=3, which is also in accordance with the number of major population groups used as input for the

Ohana program—Kichwa or Loja plus the Europe and East Asia outgroup individuals from the SGDP dataset. VCFtools was used to prepare the VCF by filtering sites out of Hardy-Weinberg Equilibrium with a p-value below 10^{-4} and removing indels and missing data.[53] This dataset is comprised of high-coverage whole genomes, so we utilized the genotypes directly instead of genotype likelihoods. To estimate the correlation structure of each individual to a population tree, we downsampled the dataset at random to 5% of the original variants and then inferred component covariances using Ohana's qpas function to produce admixture-corrected allele frequencies. We then utilized Ohana's selscan function to detect the variants that deviated most strongly from the genome-wide covariance structure, which produces likelihood ratios for each locus in the dataset. A high likelihood ratio indicates a strong deviation from the genome-wide covariance structure, and therefore a strong selection signal. Because these were population-specific scans, not just a global estimate of covariance, the program requires a scalar addition of 10 (h=10) to one position of the covariance matrix corresponding to the focal population being scanned for selection.

We also tested for evidence of selection in our phased dataset using the REHH program, which relies on EHH to calculate the iHS within-population statistic (for each population group).[25–27,29] The iHS scan compares the integrated EHH profiles between alleles at focal SNPs within the same population. The default parameters were used, with the following modifications: using a frequency bin of 1 (freqbin = 1) to use the entire dataset as a single bin to perform standardization, applying the false discovery rate correction to resulting p-values (p.adjust.method = "FDR"), and not polarizing the data (polarized = F) in order to use major and minor alleles (reference genome hg19), instead of the default ancestral and derived. P-values are presented on a negative $\log_{10}$ scale.

We further explored signatures of positive selection using the composite likelihood ratio test for the saltiLASSI method implemented in the lassip software.[44] This test has high power to detect both hard and soft sweeps from positive selection of recent to moderate age by using the distribution of haplotype frequency spectra across the genome in a single population using phased whole genome data. For input to this software, we considered only biallelic polymorphic sites with no missing genotypes for any individuals. Using a window of 21 SNPs (–winsize) and a step size of 7 SNPs (–winstep), we computed the haplotype frequency spectrum, truncated at K=10 (–k) distinct haplotypes to obtain the haplotype frequency spectra at each window and an estimate of the genome-wide haplotype frequency spectrum through a mean across all such windows (–avg-spec). Using this mean spectrum as the null haplotype frequency spectrum (–null-spec), we then performed a genome-wide scan for positive selection in both Ecuador populations with the saltiLASSI (–salti) statistic within the lassip software. At each test location (center of each window), saltiLASSI computes a log composite likelihood ratio test statistic (Λ), for which values close to zero lend little support for positive selection and extreme positive values lend increased support.

CLUES is an approximate likelihood method that detects allele trajectory, timing of selection, and selection strength of a particular SNP of interest.[45] We ran the CLUES program for our 9 anti-TB response-related SNPs from the various scans above using a generation length of 28 years. We first estimated the genome-wide genealogies and coalescent rate using the Relate program on the phased dataset, with a mutation rate of 1.25×10^{-8}.[16] We then sampled the branch lengths with Relate, sampling each branch 1000 times (–num_Samples), and using the output for the selection inferences performed by CLUES (Table S4).

The individuals included in this study have very high (>98%) proportions of Indigenous ancestry, so we would not expect SNPs under high probability of selection to come from European haplotypes. Nevertheless, we still confirmed this for all sites of interest described in the results section using the RFmix local ancestry inference program, which produces global diploid ancestry estimates for each SNP based on haplotypes in a reference panel (see Table S10 for list of Europe reference panel individuals used).[62] Using a region of the genome 1 million base pairs on either side of the SNP of interest (–analyze-range=), we utilized the default program parameters, apart from specifying a higher number of generations to 12 (–G 12).