Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Runs of Homozygosity Analysis Reveals 1,900-Year-Old ECEL1 Arthrogryposis Mexican Founder Variant

Paul McKay

paul.mckay@mail.mcgill.ca

Shriners Hospital for Children - Canada https://orcid.org/0009-0004-0927-4224

Daniel Ramírez-García

Shriners Children's Mexico

Sena Tavukcu

Shriners Hospital for Children - Canada

Ghalib Bardai

Shriners Hospital for Children - Canada

Anxhela Gjyshi Gustafson

Shriners Children's Genomics Institute

Philip Giampietro

University of Illinois-Chicago

Noémi Dahan-Oliel

Shriners Hospital for Children - Canada

Frank Rauch

Thania Ordaz-Robles

Article

Keywords: Genome-sequencing, AMC, arthrogryposis, haplotype, founder effect, ECEL1

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11516027/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is no duality of interest

Abstract

ECEL1-related distal arthrogryposis type 5D is characterized by distal joint contractures, ocular abnormalities, and absence of central nervous system involvement. We investigated the genetic basis of arthrogryposis multiplex congenita (AMC) and explored the ancestry and origin of recurrent ECEL1 variants. The study included 93 individuals with AMC residing in Mexico. Genome sequencing identified biallelic *ECEL1* variants in 37 individuals (40%). Twenty unrelated individuals were homozygous for a 46-bp frameshift deletion, c.110_155del (p.Phe37CysfsTer151), and eight additional individuals were compound heterozygous for c.110_155del; the remaining nine individuals carried other pathogenic *ECEL1* variants.

Overall, c.110_155del accounted for 50 of 74 disease-causing ECEL1 alleles (68%). Analysis of population genomic data from the Mexico City Prospective Study estimated a carrier frequency of 1 in 600 among individuals with Indigenous Mexican ancestry, or approximately 30-fold enrichment relative to European and African ancestries. To investigate a potential founder effect, runs of homozygosity analysis, haplotype dating, and ancestry inference were performed. Eighteen of 20 homozygotes shared a 272-kb ancestral haplotype encompassing *ECEL1*, and the most recent common ancestor was estimated to have lived ~ 1,900 years ago. ADMIXTURE and local ancestry inference demonstrated predominantly Indigenous Mexican ancestry among homozygous individuals (median ancestry proportion 0.81), while all 40 deletion-bearing chromosomes carried Indigenous Mexican ancestry at the c.110_155del locus.

ECEL1-related distal arthrogryposis is a common genetic cause of AMC in Mexico. The high frequency of c.110_155del, extensive haplotype sharing, and uniform Indigenous Mexican local ancestry at the deletion locus support an Indigenous Mexican founder effect.

1. Introduction

Arthrogryposis multiplex congenita (AMC) comprises a diverse group of disorders characterized by congenital joint contractures, including Amyoplasia, distal arthrogryposis (DA), and syndromic forms, affecting 1 in 3,000–4,300 live births (1) (2). According to the Online Mendelian Inheritance in Man (OMIM), DA comprises at least 18 clinically and genetically distinct subtypes (OMIM Phenotypic Series # PS108120; accessed October 1, 2026), characterized predominantly by congenital contractures affecting the distal joints of the upper and lower limbs. DA type 5D (DA5D; OMIM #615065) is a rare autosomal recessive subtype caused by pathogenic variants in *ECEL1* (3). *ECEL1* (endothelin converting enzyme-like 1) encodes a highly conserved transmembrane metallopeptidase expressed primarily in motor neurons, where its loss of function impairs motor nerve terminal development, axonal branching, and neuromuscular junction formation, resulting in reduced skeletal muscle innervation, decreased fetal movement and congenital joint contractures (3, 4). Typical DA5D clinical features include congenital hand and foot contractures accompanied by characteristic craniofacial findings, such as ptosis and central tongue furrowing, as well as musculoskeletal manifestations including scoliosis and

camptodactyly (3, 5) (6, 7). DA type 5D is an ultra-rare disorder; a 2024 review identified fewer than with 60 cases reported worldwide (8).

The Shriners AMC Registry (9) currently includes 93 individuals with AMC from Mexico, all of whom underwent genome sequencing. Of these, 21 were homozygous for the *ECEL1* c.110_155del frameshift variant, suggesting a founder effect. Founder effects for the *ECEL1* c.110_155del variant had previously been observed in Israel and India (5, 10).

A founder event is a form of genetic drift, defined as stochastic changes in allele frequencies across generations, that occurs when small numbers of ancestral individuals contributes disproportionately to the descendants of a population, reducing genetic variation and increasing the frequency of rare alleles, including those associated with recessive diseases (11). A population bottleneck, an extreme form of genetic drift resulting from a severe reduction in population size due to factors such as famine, epidemic disease, or natural disasters, can further amplify the genetic consequences of a founder event when the surviving individuals contribute disproportionately to the genetic makeup of subsequent generations (11, 12). In founder populations, individuals affected by recessively inherited disorders often share a common ancestral haplotype that can be identified through runs of homozygosity (ROH) (13), which are long stretches of homozygous genotypes inherited identically by descent from a common ancestor (14). As meiotic recombination progressively fragments ancestral haplotypes over successive generations, the size and extent of shared ROH can be used to estimate the time to the most recent common ancestor (TMRCA), with longer shared segments indicating a more recent common ancestor (12).

Given the unexpectedly high frequency of the *ECEL1* c.110_155del variant in our Mexican AMC cohort, we investigated whether this variant represents a founder allele in the Mexican population. Demonstrating a founder effect would have important implications for diagnosis, carrier screening, genetic counselling, and family planning, while also providing broader insight into the population-specific genetic architecture of AMC in Mexico. We therefore combined runs of homozygosity analysis, haplotype dating, ancestry inference, and mutation rate estimation to determine whether the *ECEL1* c.110_155del allele is a founder variant in the Mexican population and to characterize its origin and age.

2. Materials and Methods

Participant Enrollment

Participants were enrolled through the Shriners Children's AMC RE^{2}GISTRY study at Shriners Children's Mexico City, a tertiary pediatric orthopedic referral center. Individuals aged 0–18 years with congenital joint contractures involving two or more body regions were eligible for enrollment (9). Clinical data collected included medical and family history, physical examination findings, phenotypic characterization, clinical photographs, radiographic imaging, and patient-reported outcomes. Genome Sequencing was offered to probands and their biological parents and, when available, affected siblings. At the time of analysis, genome sequencing had been completed in 93 unrelated families from Mexico

City. The present study focused on individuals who were homozygous for the *ECEL1* c.110_155del variant. The study was approved by the Ethics Research Committee at Shriners Children's Mexico City (Approval No. CEI-2025-10). Written informed consent was obtained from all participants and/or their legal guardians.

Genome Sequencing

Saliva samples were collected from probands and available parents. DNA extraction and sequencing was performed at the Shriners Children's Genomics Institute (Tampa, FL, USA). Following genomic DNA extraction, paired-end sequencing was conducted on the Illumina NovaSeq X Plus platform, achieving a mean coverage depth of 28-35x. Raw sequencing data were converted to FASTQ format using bcl2fastq2 and aligned to the GRCh38 (hg38) human reference genome using the DRAGEN pipeline (v3.2.8) (15). Variant call format (VCF) files, phenotypic data, and pedigree information were subsequently uploaded to Emedgene (Illumina) for variant prioritization and annotation using public and proprietary genomic databases (16). Parental relatedness and sample quality metrics were assessed using Peddy (17). Variants were classified according to American College of Medical Genetics and Genomics guidelines (18).

ROH Analysis

ROH were identified from VCF files of individuals homozygous for *ECEL1* c.110_155del. VCFs were processed using bcftools (v.1.22) (19) and analyzed with AutoMap (v1.3) (20), which employs an error-tolerant hidden Markov model (HMM) to infer autozygous segments from genotype patterns and population allele frequencies. For burden analyses, only ROH ≥ 1 Mb were considered. However, the HMM was configured to detect autozygous segments ≥ 100 kb while tolerating isolated heterozygous or discordant genotype calls within otherwise homozygous regions, thereby preventing artificial fragmentation of ROH due to genotyping errors.

Individual ROH burden was quantified as the cumulative length of ROH segments across all autosomes and specifically across chromosome 2, which harbours *ECEL1*. ROH lengths were calculated from DRAGEN-derived BED intervals and summarized in base pairs and megabases. Individual autozygosity was quantified as the fraction of the autosomal genome contained within runs of homozygosity (F_{ROH}).

To distinguish patterns of shared ancestry, ROHs were stratified into four thresholds: <1 Mb (background linkage disequilibrium), 1–5 Mb (historical shared ancestry or population bottlenecks), 5–10 Mb (intermediate ancestral sharing), and ≥ 10 Mb (recent consanguinity) (13). Autosomal lengths were normalized to a standard human reference genome length to enable comparison across individuals.

Haplotype Dating

Shared ancestral haplotypes among *ECEL1* c.110_155del homozygotes were identified by aligning ROH encompassing the *ECEL1* locus and defining shared haplotypes as overlapping regions of at least 200 kb. TMRCA was estimated using the Gamma method (12). The hg38 genomic coordinates

corresponding to the *ECEL1* c.110_155del locus and shared ancestral haplotype intervals were lifted over to hg19 coordinates prior to variant-age estimation using the Gamma method (12). Analyses were performed under a correlated genealogy with a confidence coefficient of 0.95.

Ancestry Inference

To assess population structure and estimate ancestry proportions, variant identifiers were harmonized using PLINK v2.0 (21) and merged with reference datasets using PLINK v1.9 (22). Reference populations included African (AFR; n = 661 from seven populations) and European (EUR; n = 503 from five populations) individuals from the 1000 Genomes Project (23), together with Indigenous Mexican (IAM; n = 34) individuals from the Human Genome Diversity Project, comprising Maya (n = 21) and Pima (n = 13) populations (24). Individuals belonging to the admixed American (AMR) superpopulation from the 1000 Genomes Project were excluded to minimize confounding from pre-existing admixture (25).

Following dataset merging, quality control filtering was implemented using genotype missingness (--geno 0.02), minor allele frequency (--maf 0.01), and Hardy–Weinberg equilibrium (--hwe 1e-6, mid-p) thresholds. Multiallelic variants were removed, and linkage disequilibrium pruning was performed using a sliding-window approach (--indep-pairwise 200 50 0.2). Population structure was assessed via unsupervised principal component analysis (PCA) (26) implemented in PLINK v2.0. Principal components were derived from AFR, EUR, and IAM reference individuals, and study participants were subsequently projected onto these axes.

Global ancestry proportions were estimated using supervised ADMIXTURE (v1.3.0) (27) with K = 3 ancestral populations (AFR, EUR, and IAM). Approximately 100,000 autosomal single nucleotide polymorphisms (SNPs) restricted to A/C/G/T variants (--thin-count 100000) were analyzed to improve computational efficiency while retaining a genome-wide set of markers for ancestry inference. Additional ADMIXTURE analyses using K = 4 and K = 5 models, as well as a pan-Indigenous American reference panel including all Native American populations from the Human Genome Diversity Project, were explored. Because PCA and ADMIXTURE results demonstrated that participants clustered most closely with the Human Genome Diversity Project Indigenous Mexican populations (Maya and Pima), the broader pan-Indigenous framework was not pursued further. The final analyses therefore used a K = 3 model (AFR, EUR, and IAM), which best represented the ancestry composition of the study cohort.

Local ancestry inference was performed using RFMix v2.03 (28) on haplotypes phased with Eagle v2.4.1 (29) including 20 unrelated individuals homozygous for *ECEL1* c.110_155del (40 phased haplotypes), leveraging the Mexican Ancestry in Los Angeles (MXL) recombination map from the 1000 Genomes Project Phase 3 with SHAPEIT5-phased reference panels (30). Local ancestry was inferred across all autosomes, and the ancestry composition at the *ECEL1* c.110_155del locus was subsequently quantified across all 40 cohort haplotypes.

Mutation Rate Estimation at the Variant Locus

To investigate whether *ECEL1* c.110_155del may represent a mutational hotspot, a locus-specific mutation rate was estimated using an approach adapted from published population-genetic mutation-rate estimators (31). The estimator incorporates the population allele frequency (q), the proportion of independent mutation events (f_{new}), and the number of generations since the most recent common ancestor (g). The locus-specific mutation rate (u) was calculated as:

$$u = \frac{q \times f_{new}}{g}$$

where q represents the population allele frequency, f_{new} denotes the proportion of independent mutation events, and g is the number of generations since the estimated TMRCA. The gnomAD AMR allele frequency was used as a proxy for the Mexican population frequency, and the TMRCA estimate was derived from haplotype-based dating analyses.

Statistical analyses were performed in R (v4.5.0), with two-sided p values < 0.05 considered statistically significant. Genomic data were visualized using Integrative Genomics Viewer (IGV v2.19.7) (32), and graphical outputs were generated in R.

3. Results

Among 93 individuals with AMC from Mexico, 21 were homozygous for the *ECEL1* c.110_155del pathogenic variant. After excluding one individual to account for sibling relatedness, 20 homozygotes were included in the ROH analysis (12 females, median age 7 years).

ROH Analysis

To assess whether the recurrent *ECEL1* c.110_155del variant occurred in the context of broader genomic homozygosity, autosomal ROH burden was quantified using a minimum ROH length threshold of 1 Mb. The median autosomal F_{ROH} derived from ROH segments ≥ 1 Mb was 3.49%. Examination of ROH length distributions further contextualized this burden (Fig. 1). Across individuals, F_{ROH}≥1 Mb ranged from approximately 2% to 6%, indicating generally modest levels of autozygosity consistent with background population structure. The majority of ROH burden was attributable to short segments (1–5 Mb), indicative of distant shared ancestry. In contrast, long ROH (≥ 10 Mb), which are typically associated with recent parental relatedness, were uncommon. Only five individuals (P3, P4, P14, P15, and P20) exhibited non-zero F_{ROH}≥10 Mb values, with P4 showing the greatest burden. The lack of long ROH suggests that recurrence of *ECEL1* c.110_155del homozygosity is unlikely to be driven by recent consanguinity and is instead consistent with inheritance of a shared ancestral haplotype.

Haplotype Dating

Alignment of ROH across chromosome 2 revealed that 18 of 20 *ECEL1* c.110_155del homozygotes (90%) shared a common ~ 272 kb ROH (chr2:232,345,576–232,617,350) encompassing a locus that

contained 878 informative SNP markers flanking the deletion (Fig. 2). We estimated the age of the variant using the gamma method under a correlated genealogy model (12), which was estimated to have arisen 76 generations ago (95% CI: 17–140) or approximately 1,900 years assuming a 25-year generation time. This estimated age, together with the presence of a shared ancestral haplotype in 90% *ECEL1* c.110_155del homozygotes from this Mexican cohort, strongly supports a common ancestral origin for the variant.

Ancestry Inference

ADMIXTURE models incorporating three, four, and five ancestral reference populations (K = 3–5) produced highly concordant ancestry estimates (Fig. S1, Fig. S2). Additional analyses using a pan-Indigenous American reference panel yielded similar results; however, PCA consistently showed that participants clustered most closely with the Indigenous Mexican reference populations (Fig. 3, Fig. S1). Furthermore, the K = 4 and K = 5 ADMIXTURE models identified negligible East and South Asian ancestry (Fig. S2). Accordingly, downstream genetic ancestry analyses were performed using the K = 3 model, comprising Indigenous Mexican, European, and African reference populations.

Unsupervised PCA revealed that *ECEL1* c.110_155del homozygotes clustered along a narrow IAM-EUR admixture axis, with minimal separation attributable to AFR ancestry (Fig. 3A–C). Consistent with this pattern, PC1 and PC3 were strongly negatively correlated with IAM ancestry, whereas PC2 was positively correlated, reflecting a near-linear gradient of IAM-EUR admixture. Together, these findings indicate that genetic variation within the cohort is primarily driven by differing proportions of IAM and EUR ancestry, while AFR ancestry contributes comparatively little to overall population structure.

Supervised ADMIXTURE analysis supported these observations, demonstrating that *ECEL1* c.110_155del homozygotes were predominantly of Indigenous Mexican ancestry (median IAM proportion = 0.81) (Fig. 3D-E). Three individuals (P1, P3, and P4) were assigned entirely to the IAM reference component, and all participants carried at least 50% inferred IAM ancestry. AFR ancestry was detected in 17 of 20 individuals (85%), with a median contribution (0.10) similar to that of EUR ancestry (0.09). While AFR ancestry was broadly distributed across the cohort, EUR ancestry accounted for much of the variation observed along the principal component axes, with some individuals exhibiting substantial EUR admixture (Fig. 3).

Local ancestry inference using RFMix corroborated the ADMIXTURE estimates, yielding mean genome-wide ancestry proportions of IAM = 0.72, EUR = 0.24, and AFR = 0.04. Analysis of chromosome 2 further demonstrated that the recurrent *ECEL1* c.110_155del variant was carried on an Indigenous Mexican ancestral background in all 40 disease-associated chromosomes (Fig. 4). Together with the shared ancestral haplotype and variant age estimate, these findings support an Indigenous Mexican origin for the recurrent *ECEL1* c.110_155del pathogenic variant and indicate that its elevated frequency in this cohort is attributable to a historical founder effect rather than recent consanguinity.

Mutation Rate Estimation

Although local ancestry inference assigned the homozygous deletion locus to an Indigenous Mexican ancestral background in all affected chromosomes, two of the twenty *ECEL1* c.110_155del homozygotes (P19 and P20) from unrelated families did not share the common ancestral haplotype, raising the possibility that the variant arose independently on more than one occasion. Using the gnomAD AMR allele frequency together with the estimated TMRCA of 76 generations, we calculated a locus-specific mutation rate of 7.2×10^{-7}. This estimate is approximately 50-fold higher than the genome-wide mutation rate of $1.45 \pm 0.05 \times 10^{-8}$ mutations per base pair per generation (33), suggesting that recurrence of the *ECEL1* c.110_155del deletion may be promoted by a local mutational hotspot. As proposed by Cohen et al. (2023) (5), the presence of a 7-bp direct repeat sequence (CCGGGCT) flanking the 46-bp deletion provides a plausible mechanistic basis for the recurrent formation of this variant, whereby short stretches of flanking microhomology may promote deletion formation through microhomology-mediated end joining and related DNA repair processes (34).

4. Discussion

In this study, we investigated the population genetic basis of the recurrent *ECEL1* c.110_155del variant in a cohort of Mexican children with AMC, testing whether its elevated frequency reflects widespread consanguinity or a founder effect. The marked paucity of long ROH (> = 10 Mb) among c.110_155del homozygotes does not support widespread consanguinity in the study cohort (35). Instead, the predominance of 1–5 Mb ROH segments suggests older shared ancestry accompanied by historical bottlenecks (35). From aligning ROH of 20 *ECEL1* c.110_155del homozygotes, 18 of 20 (90%) shared a ~ 272 kb ancestral haplotype encompassing the variant locus on chromosome 2, supported by 878 dense SNP markers. Haplotype dating under a correlated genealogy model (12) estimated the variant's 1,900-years-old origin in the ancestral population. Combined historical and genomic evidence suggests that severe population bottlenecks following European colonization, including those associated with smallpox and cocoliztli epidemics, profoundly altered population genetic architecture in Mexico (36, 37). These demographic events likely contributed to the elevated present-day frequency of the variant in Mexico through founder effects and subsequent genetic drift (38, 39), paralleling documented founder events in Indigenous populations across the Americas (40–43). Genetic ancestry analyses, including unsupervised PCA, supervised ADMIXTURE, and local ancestry inference, supported an Indigenous Mexican founder origin for the recurrent *ECEL1* c.110_155del variant. Unsupervised K = 3 PCA (IAM, EUR, and AFR reference populations) showed that cohort individuals clustered along a near-linear IAM-EUR admixture axis, while supervised ADMIXTURE estimated a median Indigenous American ancestral component of 0.81. Consistent with this finding, local ancestry inference along chromosome 2 demonstrated that all 40 disease-associated chromosomes carrying c.110_155del were derived from an Indigenous Mexican ancestral background. The identification of an *ECEL1* c.110_155del founder effect in Mexico, in addition to those reported in Israel and India, suggests that this pathogenic allele may have undergone recurrent population-specific enrichment or may represent an older ancestral variant shared across populations (5, 10).

The identification of this *ECEL1* c.110_155del founder effect in Indigenous Mexicans adds to a growing number of documented founder effects in this population, including a possible *PEX6* Zellwger spectrum disorder founder effect identified in the Mixteco population originating from Oaxaca, Mexico, who were conquered first by the Aztecs, and later, the Spanish (44). Reyes et al. recently identified a novel pathogenic variant in *FANCG* in three individuals with Fanconi anemia in the *Mixe* community, also indigenous to Oaxaca, but with distinct language and history from the Mixteco people, where haplotype and ancestry analysis revealed the three individuals shared an ancestral haplotype supported by 30 SNP markers around *FANCG*, confirming the Mixteco *FANCG* Fanconi anemia founder effect (45).

From the 2/20 cohort individuals who did not share the 272kb ancestral haplotype encompassing the deletion locus, but still carry the recurrent *ECEL1* c.110_155del variant on an IAM genetic background, we estimated a 50-fold predisposition to deletion formation at the variant locus relative to the genome-wide mutation rate (31), possibly via microhomology-mediated repair (46), likely due to the presence of a 7-bp direct repeat sequence flanking the 46-bp deleted sequence, providing plausible explanation for the variant's recurrence in geographically isolated populations (5, 10, 47).

Sample size limitations reduce the accuracy of our TMRCA estimate as well as our variant locus mutation rate estimation. Genetic ancestry analyses were limited by lack of IAM reference panels available from the Human Genome Diversity Project, where only two Indigenous Mexican populations were represented. This may have reduced the accuracy of our analyses and precluded fine-scale localization of the founder variant to a specific Indigenous Mexican population.

Enrichment of the recurrent *ECEL1* c.110_155del variant in Mexico is attributable to 1,900-year-old Indigenous Mexican founder effect, pre-dating European contact (48). Although the variant likely arose long before colonization, the severe population bottlenecks and demographic restructuring that followed European contact may have amplified its frequency through genetic drift, resulting in its present-day enrichment in Indigenous Mexican and admixed Mexican populations. This finding adds to a growing list of Indigenous Mexican founder variants (44, 45, 49), underscoring the need for clinical vigilance in this historically underserved population (50). This work contributes to ongoing efforts to improve molecular diagnosis, variant interpretation, genetic counselling, and equitable implementation of precision medicine for Indigenous populations across the Americas.

Declarations

Ethics Approval and Consent to Participate

The study was approved by the Ethics Research Committee at Shriners Children's Mexico City (Approval No. CEI-2025-10). Written informed consent was obtained from all participants and/or their legal guardians.

Competing Interests

The authors declare no competing interests.

Funding

This work was supported by Shriners Children's through the Shriners Children's AMC RE^{2}GISTRY study.

Acknowledgements

We gratefully acknowledge the patients and families who participated in the Shriners Children's AMC RE2GISTRY study and thank the staff of Shriners Children's Mexico City and the Shriners Children's Genomics Institute for their support of this work.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request. Individual-level genomic data are not publicly available because of participant privacy and consent restrictions.

References

1. Bamshad M, Van Heest AE, Pleasure D. Arthrogryposis: a review and update. The Journal of Bone and Joint Surgery American Volume. 2009;91(Suppl 4):40–6.
2. Lowry RB, Sibbald B, Bedard T, Hall JG. Prevalence of multiple congenital contractures including arthrogryposis multiplex congenita in Alberta, Canada, and a strategy for classification and coding. Birth Defects Res A Clin Mol Teratol. 2010;88(12):1057–61.
3. McMillin MJ, Below JE, Shively KM, Beck AE, Gildersleeve HI, Pinner J, et al. Mutations in ECEL1 cause distal arthrogryposis type 5D. Am J Hum Genet. 2013;92(1):150–6.
4. Nagata K, Kiryu-Seo S, Tamada H, Okuyama-Uchimura F, Kiyama H, Saido TC. ECEL1 mutation implicates impaired axonal arborization of motor nerves in the pathogenesis of distal arthrogryposis. Acta Neuropathol. 2016;132(1):111–26.
5. Cohen D, Sloma R, Pizem H, Fedida A, Kalfon L, Ovadia R, et al. Long term ophthalmic complications of distal arthrogryposis type 5D. Ophthalmic Genet. 2023;44(1):28–34.
6. Gowda M, Mohan S, Ramesh D, Chinta N. Distal arthrogryposis type 5D in a South Indian family caused by novel deletion in ECEL1 gene. Clin Dysmorphol. 2021;30(2):100–3.
7. Huddar A, Polavarapu K, Preethish-Kumar V, Bardhan M, Unnikrishnan G, Nashi S, et al. Expanding the Phenotypic Spectrum of ECEL1-Associated Distal Arthrogryposis. Children (Basel). 2021;8(10).
8. Jing S, Peng M, He Y, Hua Y, Li J, Li Y. A novel compound heterozygous variant of ECEL1 induced joint dysfunction and cartilage degradation: a case report and literature review. Front Neurol.

2024;15:1343025.

9. Dahan-Oliel N, van Bosse H, Darsaklis VB, Rauch F, Bedard T, Bardai G, et al. Epidemiology, aetiology, interventions and genomics in children with arthrogryposis multiplex congenita: protocol for a multisite registry. BMJ Open. 2022;12(10):e060591.
10. Endrakanti M, Sharma J, Ethayathulla AS, Kaur P, Khan SA, Kabra M, et al. ECEL1 related distal arthrogryposis 5D in an Indian cohort-Report of recognizable musculoskeletal phenotype and a possible founder variant. Am J Med Genet A. 2024;194(8):e63592.
11. Tournebize R, Chu G, Moorjani P. Reconstructing the history of founder events using genome-wide patterns of allele sharing across individuals. PLoS Genet. 2022;18(6):e1010243.
12. Gandolfo LC, Bahlo M, Speed TP. Dating rare mutations from small samples with dense marker data. Genetics. 2014;197(4):1315–27.
13. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19(4):220–34.
14. Bortoluzzi C, Bosse M, Derks MFL, Crooijmans R, Groenen MAM, Megens HJ. The type of bottleneck matters: Insights into the deleterious variation landscape of small managed populations. Evol Appl. 2020;13(2):330–41.
15. Behera S, Catreux S, Rossi M, Truong S, Huang Z, Ruehle M, et al. Comprehensive genome analysis and variant detection at scale using DRAGEN. Nat Biotechnol. 2025;43(7):1177–91.
16. Meng L, Attali R, Talmy T, Regev Y, Mizrahi N, Smirin-Yosef P, et al. Evaluation of an automated genome interpretation model for rare disease routinely used in a clinical genetic laboratory. Genet Med. 2023;25(6):100830.
17. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet. 2017;100(3):406–13.
18. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24.
19. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).
20. Quinodoz M, Peter VG, Bedoni N, Royer Bertrand B, Cisarova K, Salmaninejad A, et al. AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. Nat Commun. 2021;12(1):518.
21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

23. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.
24. Bergstrom A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020;367(6484).
25. Shriner D. Overview of admixture mapping. Curr Protoc Hum Genet. 2013;Chap. 1:Unit 1 23.
26. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190.
27. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. 2009;19(9):1655–64.
28. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013;93(2):278–88.
29. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48(11):1443–8.
30. Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. Nat Genet. 2023;55(7):1243–9.
31. Lipson M, Loh PR, Sankararaman S, Patterson N, Berger B, Reich D. Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. PLoS Genet. 2015;11(11):e1005550.
32. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24–6.
33. Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Xue Y, et al. Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. Nature Communications. 2017;8(1).
34. Wang H, Xu X. Microhomology-mediated end joining: new players join the team. Cell Biosci. 2017;7:6.
35. Palamara PF, Lencz T, Darvasi A, Pe'er I. Length distributions of identity by descent reveal fine-scale demographic history. Am J Hum Genet. 2012;91(5):809–22.
36. Acuna-Soto R, Stahle DW, Cleaveland MK, Therrell MD. Megadrought and Megadeath in 16th Century Mexico. Emerging Infectious Diseases. 2002;8(4):360–2.
37. Acuna-Soto R, Stahle DW, Therrell MD, Griffin RD, Cleaveland MK. When half of the population died: the epidemic of hemorrhagic fevers of 1576 in Mexico. FEMS Microbiol Lett. 2004;240(1):1–5.
38. Thompson EA, Neel JV. Probability of founder effect in a tribal population. Proceedings of the National Academy of Sciences of the United States of America. 1978;75(3):1442–5.
39. O'Fallon BD, Fehren-Schmitz L. Native Americans experienced a strong population bottleneck coincident with European contact. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(51):20444–8.
40. Karadimas CL, Vu TH, Holve SA, Chronopoulou P, Quinzii C, Johnsen SD, et al. Navajo neurohepatopathy is caused by a mutation in the MPV17 gene. Am J Hum Genet. 2006;79(3):544–8.

41. Laberge AM, Michaud J, Richter A, Lemyre E, Lambert M, Brais B, et al. Population history and its impact on medical genetics in Quebec. Clin Genet. 2005;68(4):287–301.
42. Fogli A, Wong K, Eymard-Pierre E, Wenger J, Bouffard JP, Goldin E, et al. Cree leukoencephalopathy and CACH/VWM disease are allelic at the EIF2B5 locus. Ann Neurol. 2002;52(4):506–10.
43. Zhou S, Xie P, Quoibion A, Ambalavanan A, Dionne-Laporte A, Spiegelman D, et al. Genetic architecture and adaptations of Nunavik Inuit. Proc Natl Acad Sci U S A. 2019;116(32):16012–7.
44. Slaton D, Chang A, Ahluwalia T, Alfaro S, Javed B, Greer R. Zellweger's Syndrome With PEX6 Gene Mutation in Mixteco Neonates Due to Possible Founder Effect. Cureus. 2023;15(9):e45162.
45. Reyes P, Garcia-de Teresa B, Juarez U, Perez-Villatoro F, Fiesco-Roa MO, Rodriguez A, et al. Fanconi Anemia Patients from an Indigenous Community in Mexico Carry a New Founder Pathogenic Variant in FANCG. Int J Mol Sci. 2022;23(4).
46. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics. 2008;1(1):4.
47. Zhang YL, Zhen L, Xu LL, Li DZ. Fetal akinesia: The need for clinical vigilance in first trimester with decreased fetal movements. Taiwanese journal of obstetrics & gynecology. 2021;60(3):559–62.
48. Lindo J, Huerta-Sanchez E, Nakagome S, Rasmussen M, Petzelt B, Mitchell J, et al. A time transect of exomes from a Native American population before and after European contact. Nat Commun. 2016;7:13175.
49. Zhen X, Betti M, Kars ME, Patterson A, Medina-Torres EA, Scheffler Mendoza SC, et al. Molecular and clinical characterization of a founder mutation causing G6PC3 deficiency. Res Sq. 2024.
50. Castro ESMA, Nunes K, Ribeiro MR, Passarelli-Araujo H, Lemes RB, Kimura L, et al. The evolutionary history and unique genetic diversity of Indigenous Americans. Nature. 2026;653(8113):134–45.

Figures

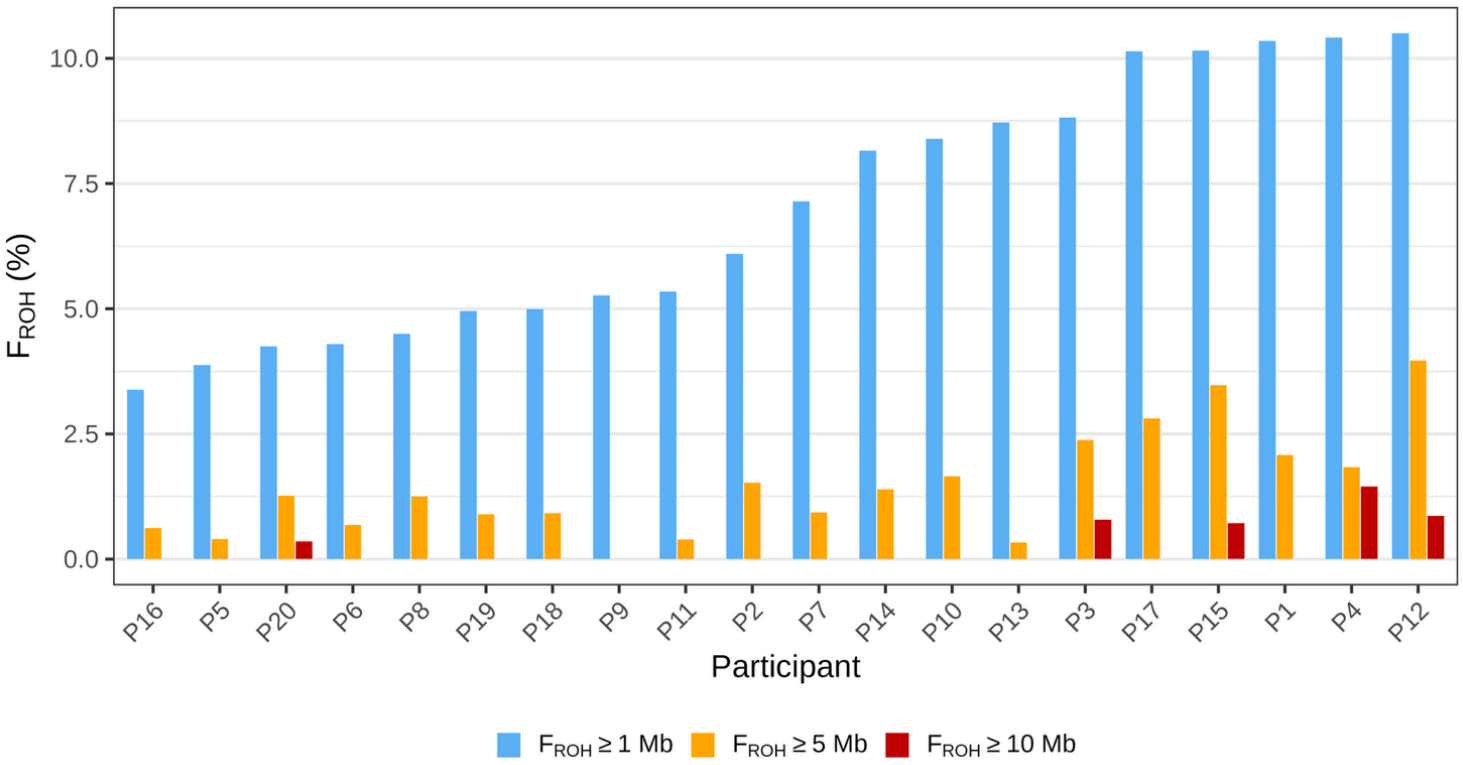


Figure 1

Autosomal runs of homozygosity (ROH) burden among the 20 of 37 DA5D-affected individuals homozygous for the recurrent *ECEL1* c.110_155del variant. Individual ROH burden was quantified as the fraction of the autosomal genome in runs of homozygosity (F_{ROH}) using length thresholds of ≥ 1 Mb (blue), ≥ 5 Mb (orange), and ≥ 10 Mb (red). Total ROH burden (F_{ROH} ≥ 1 Mb) ranges from approximately 3.5% to 10.5%, reflecting moderate background autozygosity. The predominance of shorter segments (1–5 Mb) across the cohort suggests older shared ancestry and historical population bottlenecks rather than recent parental relatedness. Long ROH segments (≥ 10 Mb) are rare or absent in most participants; the small subset of individuals exhibiting F_{ROH} ≥ 10 Mb values (0.3–1.5%) is consistent with distant parental relatedness at the second-cousin degree at the uppermost extreme.

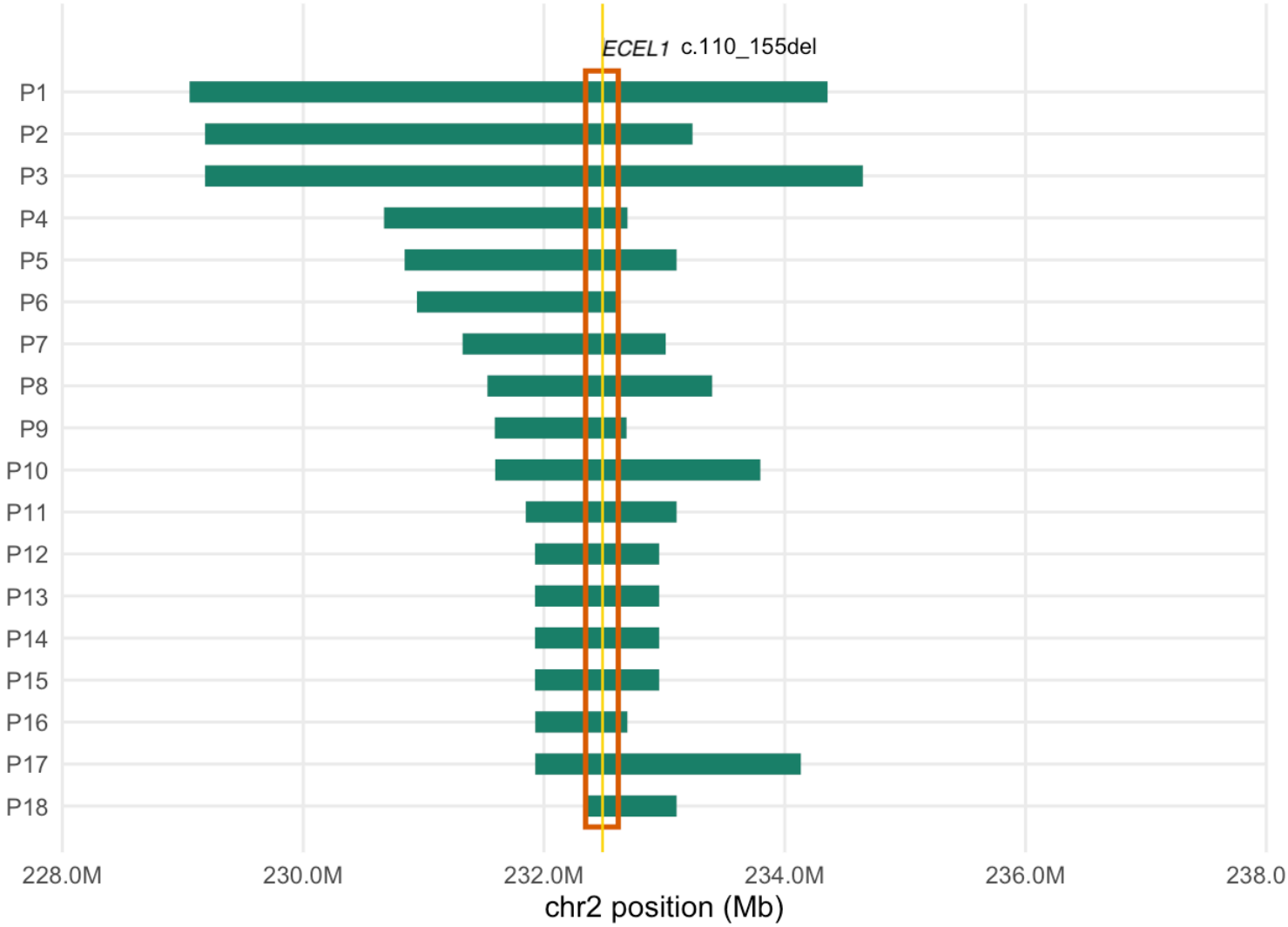


Figure 2

Shared ancestral haplotype at the *ECEL1*locus. Aligning ROH on chromosome 2 revealed a shared ~272 kb haplotype among 18 of 20 *ECEL1* c.110_155 homozygotes encompassing the deletion locus. Horizontal green bars represent individual ROH segments. The vertical yellow line indicates the position of the *ECEL1* c.110_155 variant. The shared ancestral haplotype is boxed in orange. The high degree of conservation of the ancestral haplotype across unrelated individuals from Mexico suggests a founder effect.

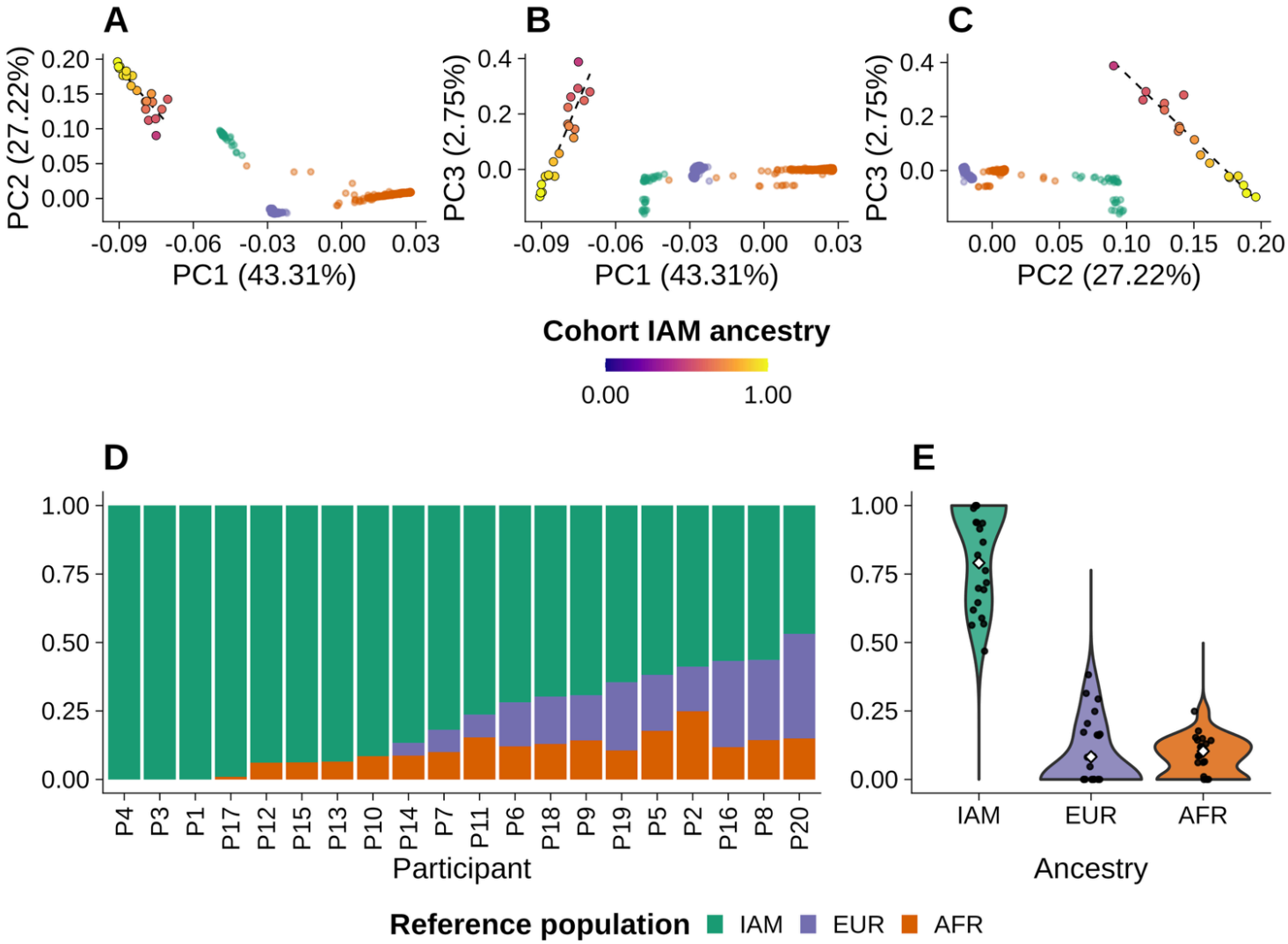


Figure 3

Principal component analysis (PCA) and ADMIXTURE of genetic ancestry. Unsupervised PCA was performed using AFR and EUR reference populations from the 1000 Genomes Project (1000GP) (AFR] and European [EUR]) and Human Genome Diversity Project (Indigenous Mexican [IAM]). Cohort individuals are represented by large, black-outlined circles, with color shading indicating the proportion of IAM ancestry as determined by supervised ADMIXTURE analysis. Panels show: **(a)** PC1 vs. PC2; **(b)** PC1 vs. PC3; **(c)** PC2 vs. PC3; **(d)** Individual ancestry proportion estimates from ADMIXTURE analysis as stacked bars with participants sorted in descending order according to IAM ancestry component; **(e)** Cohort ancestry proportion distributions as a violin plot to illustrate ancestry component spread, with ancestral components of individual participants plotted as black dots and median distributions of global cohort ancestry indicated by white diamonds.

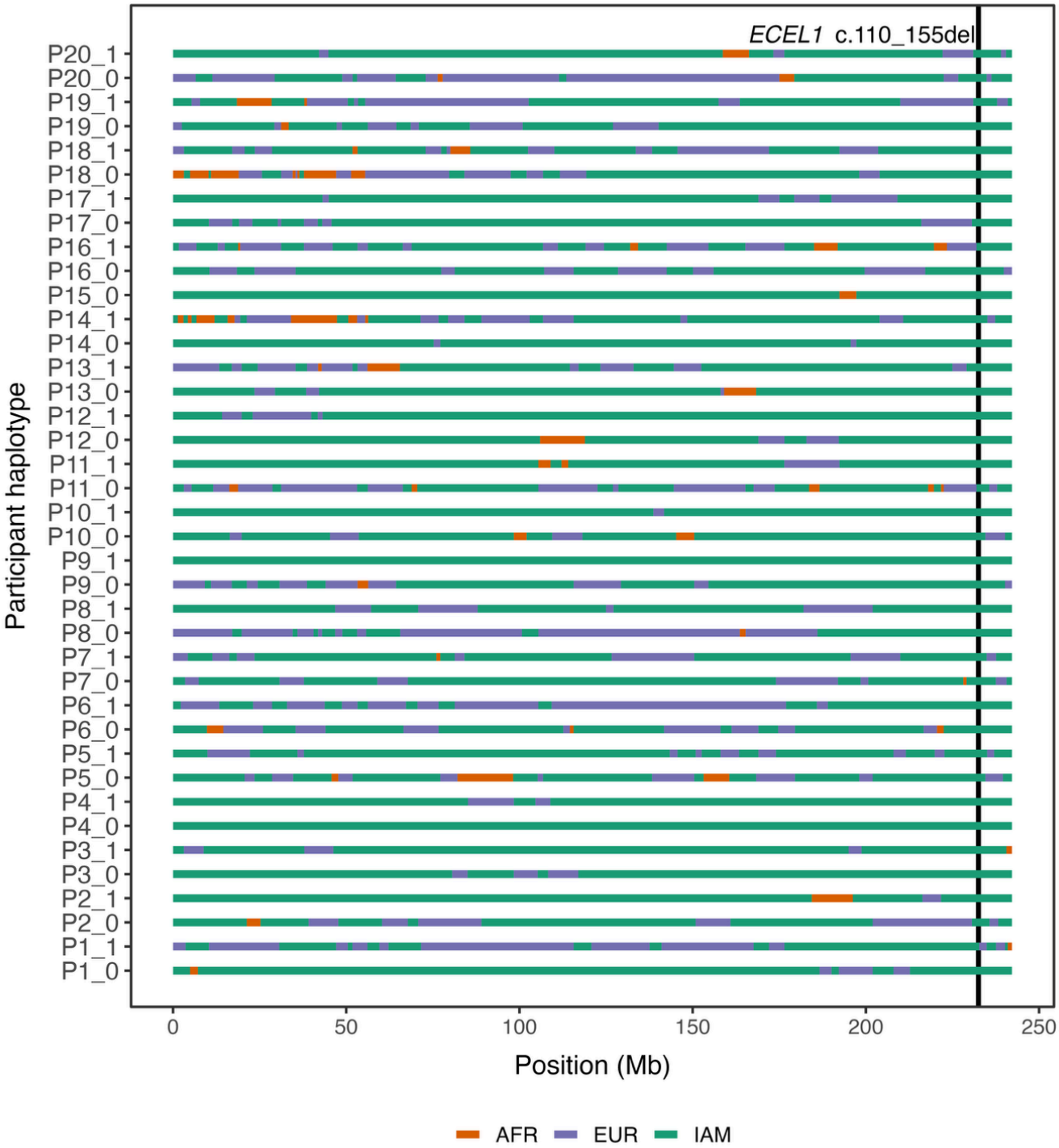


Figure 4

Local ancestry inference of the shared *ECEL1*-associated haplotype on chromosome 2. Local ancestry across chromosome 2 was inferred using RFMix v2.03 on haplotypes phased with Eagle v2.4.1 for 20 unrelated individuals homozygous for the ECEL1 c.110_155del variant (40 total haplotypes). Ancestry inference was performed using a Mexican (Los Angeles; MXL) recombination map from the 1000 Genomes Project Phase 3 and reference panels phased with SHAPEIT5. Each row represents a participant haplotype, and genomic position (Mb) is shown along the x-axis. Local ancestry assignments

are color-coded as Indigenous American (IAM; green), European (EUR; purple), and African (AFR; orange). The black vertical line indicates the position of the *ECEL1* c.110_155del variant (hg38 chr2:232,486,498). Across all 40 participant haplotypes, the genomic interval harboring the shared disease-associated haplotype is uniformly assigned Indigenous American ancestry, supporting an Indigenous Mexican founder origin of the extended haplotype carrying the *ECEL1* c.110_155del variant in the Mexican AMC cohort.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.pdf