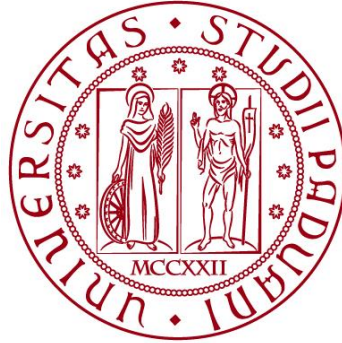


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biologia Molecolare



ELABORATO DI LAUREA

**TRACCE GENETICHE DI SELEZIONE NATURALE INFLUENZATA DA
PATOGENI IN POPOLAZIONI UMANE INDIGENE DELL'AMAZZONIA**

**Tutor: Prof. Luca Pagani
Dipartimento di Biologia**

Laureanda: Camilla Colaci

ANNO ACCADEMICO 2023/2024

INDICE

STATO DELL'ARTE	3
APPROCCIO SPERIMENTALE	5
1. Assemblaggio e controllo della qualità dei dati, fasatura e annotazione del genoma	
2. Scansioni di selezione	
3. Modello demografico e simulazione	
4. Tempo e intensità di selezione	
5. Analisi di sovrarappresentazione e analisi di arricchimento dei set genici	
6. Knockdown del gene candidato in cardiomiociti derivati da hiPSC	
7. Infezione delle hiPSC-CM da parte di <i>Trypanosoma cruzi</i>	
8. Analisi di correlazione	
RISULTATI E DISCUSSIONE	11
1. Geni candidati alla selezione naturale	
2. Caratterizzazione funzionale del gene candidato selezionato: PPP3CA	
3. Frequenza dell'allele rs2659540 selezionato in diverse popolazioni umane	
4. Distribuzione geografica di rs2659540 e incidenza di <i>T. cruzi</i>	
5. Criticità e prospettive	
BIBLIOGRAFIA	18
APPENDICE	19

Abstract

Questo studio analizza i dati genomici di 118 individui appartenenti a 19 popolazioni native allo scopo di identificare tracce di selezione naturale positiva legate alle malattie tropicali che caratterizzano la foresta pluviale amazzonica. Questa regione, infatti, presenta condizioni ecologiche favorevoli alla diffusione di malattie, in particolare quelle trasmesse da vettori, che hanno generato forti pressioni selettive, portando ad un adattamento genetico delle popolazioni native all'ecosistema amazzonico. Dalle analisi genomiche e funzionali, si osserva come un forte segnale di selezione naturale sembra essere associato ad un gruppo di geni che sono coinvolti nell'infezione da *Trypanosoma cruzi*, un protozoo parassita diffuso nella regione che è il patogeno responsabile della malattia di Chagas.

1. STATO DELL'ARTE

La foresta pluviale amazzonica è una delle principali ecoregioni del continente americano, nota per il suo clima tropicale e l'incredibile diversità biologica. Nonostante sia un'ambiente ricco di nutrienti, presenta vari ostacoli per la sopravvivenza umana: la scarsa penetrazione della luce, l'elevata diversità di agenti patogeni, l'instabilità delle risorse alimentari contribuiscono a forti pressioni selettive per la sopravvivenza e la riproduzione umana in questa ecoregione. L'America presenta una vasta gamma di ecoregioni che furono rapidamente esplorate e occupate dai primi esseri umani a raggiungere il continente: questa notevole migrazione umana probabilmente richiese diversi modelli di adattamento genetico e culturale per garantire la sopravvivenza. La regione amazzonica, che comprende la più grande foresta pluviale tropicale e nove paesi sudamericani (Brasile, Perù, Colombia, Venezuela, Ecuador, Bolivia, Guyana, Guyana francese e Suriname), è ineguagliabile per scala, complessità e opportunità. Attualmente è popolata da 1 milione di indigeni, divisi in circa 300 diversi gruppi etnici [1].

Le condizioni ecologiche della regione sono state storicamente favorevoli alla trasmissione di numerose malattie tropicali, soprattutto quelle trasmesse da vettori. Una di queste, considerata in questa tesi, è la malattia di Chagas.

La malattia di Chagas (tripanosomiasi americana) è una malattia trasmessa da vettori, causata dal protozoo *Trypanosoma cruzi*, e viene solitamente trasmessa attraverso diversi insetti triatomici nelle aree endemiche. La malattia colpisce circa 6-8 milioni di persone in tutto il mondo e provoca circa 50.000 decessi all'anno. Altri 65-100 milioni di persone vivono in aree a rischio di infezione in tutto il mondo. Anche se è passato più di un secolo dalla sua scoperta, la malattia di Chagas rimane uno dei principali problemi di salute pubblica per la maggior parte dei paesi dell'America Latina. Negli ultimi decenni, è diventata una preoccupazione anche per luoghi non endemici come Canada, USA, Europa, Australia e Giappone a causa della migrazione costante di individui provenienti da aree endemiche. In queste regioni, la trasmissione avviene principalmente attraverso trasfusioni di sangue, trapianti di organi o trasmissione verticale da madre a figlio.

T. cruzi è un parassita emoflagellato intracellulare che appartiene all'ordine Kinetoplastida, famiglia Trypanosomatidae. È l'agente causale della malattia di Chagas, possedendo la capacità di infettare qualsiasi cellula, principalmente macrofagi, fibroblasti e cellule epiteliali. Durante il suo ciclo di vita, il parassita evolve attraverso tre forme principali: amastigote (forma proliferativa trovata all'interno delle cellule dell'ospite vertebrato), epimastigote (forma proliferativa trovata nell'intestino degli ospiti invertebrati) e tripomastigote (la forma infettiva derivata dagli amastigoti negli ospiti vertebrati e dagli epimastigoti nel tratto digestivo degli ospiti invertebrati). Può danneggiare seriamente il cuore (miocardiopatia chagastica cronica), l'apparato digerente e intestinale e, in misura minore, il sistema nervoso centrale.

La tripanosomiasi americana è una malattia trasmessa da vettori, con il parassita trasmesso da insetti ematofagi (conosciuti anche come "kissing bug") della sottofamiglia Triatominae. È noto che 140 specie di Triatominae sono in grado di

trasmettere *T. cruzi* e sono ampiamente distribuite nelle Americhe. Le specie vettoriali più importanti sono: *Triatoma infestans* in Argentina, Bolivia, Brasile, Cile, Paraguay, Uruguay e Perù; *Rhodnius prolixus* in Colombia, Venezuela e America Centrale; *T. dimidiata* in Ecuador e America Centrale; e *Rhodnius pallescens* a Panama (PAHO). Nella parte meridionale degli USA, l'insetto vettore comune è *Triatoma gerstaeckeri*, seguito da *T. lecticularia* e *T. sanguisuga*. [2]

Ad oggi, la conoscenza riguardo agli adattamenti genetici degli esseri umani all'interno del complesso ecosistema della foresta pluviale amazzonica è in gran parte sconosciuta: esistono numerosi studi sulle epidemie post contatto nelle Americhe (ovvero quelle avvenute in seguito ai fenomeni di migrazione), mentre i dati storici sulle malattie native delle Americhe (pre-contatto) sono insufficienti. Tuttavia, è stato riportato che la tubercolosi (*Mycobacterium tuberculosis*) e la malattia di Chagas (tripanosomiasi americana) erano presenti molto prima dell'arrivo degli europei. La più antica traccia di *T. cruzi* nei resti archeologici umani del Sud America risale a 9000 anni fa in mummie del nord del Cile e del sud del Perù. Resti umani infettati da *T. cruzi* sono stati trovati anche in Brasile circa 7000 anni fa. Tuttavia, non ci sono molti studi sugli adattamenti alla foresta pluviale, incluse le popolazioni amazzoniche. La maggior parte di questi studi è limitata a pochi individui dell'Amazzonia occidentale.

Quanto descritto finora, offre un quadro di difficile sopravvivenza delle popolazioni umane alla regione amazzonica: è stato ipotizzato, quindi, un processo di adattamento genetico a questo ambiente ostile. Lo scopo dello studio effettuato, perciò, è quello di trovare possibili tracce genetiche di un adattamento all'ambiente della foresta pluviale amazzonica, in particolare per identificare segnali di selezione naturale positiva legati alle malattie tropicali, in particolare alla malattia di Chagas.

2. APPROCCIO SPERIMENTALE

2.1 Assemblaggio e controllo della qualità dei dati, fasatura e annotazione del genoma

Il set di dati è stato realizzato combinando i dati relativi a 118 nativi amazzonici con valore di discendenza nativa americana maggiore del 99% (discendenza nativa americana > 0,99), 35 mesoamericani con valore di discendenza nativa americana maggiore del 97% (discendenza nativa americana > 0,97) e 231 asiatici orientali dall'Human Genome Diversity Project. Tramite la piattaforma di genotipizzazione ad alta densità Axiom Human Origin SNP Array (~700K SNPs, Affymetrix/Thermo Fisher Scientific), è stato analizzato un ampio numero di SNPs genomici, circa 700.000, al fine di rilevare la diversità genetica tra le popolazioni umane considerate. Infine, utilizzando il software ADMIXTURE, sulla base dei dati genotipizzati è stata stimata la proporzione di discendenza nativa americana.

Sono state convertite, servendosi del software liftOver, le coordinate genomiche dalla versione GRCh37-hg19 dell'assemblaggio del genoma a quello attuale GRCh38-hg, per aggiornare le posizioni fisiche dei dati genomici ottenuti. È stato necessario effettuare un controllo della qualità dei dati, eliminando gli individui con più del 10% di dati mancanti e gli SNPs con frequenza allelica minore (MAF, minor allele frequency) inferiore al 5% o con più dell'1% di dati mancanti. Per il phasing dei dati è stato utilizzato il software SHAPEIT2, consentendo di prendere in input i genotipi e organizzarli in aplotipi e i file della mappa di ricombinazione del Progetto 1000 Genomi, per determinare in quali posizioni avvengono fenomeni di ricombinazione. In questo processo, non sono stati presi in considerazione pannelli di riferimento a causa della peculiarità della storia evolutiva delle popolazioni native americane, che, quindi, avrebbero potuto essere sottorappresentate e portare a risultati imprecisi.

Infine, sono stati rimossi gli SNPs privi di informazioni riguardanti alleli ancestrali e derivati, consultando il Progetto 1000 Genomi: il set finale di dati a seguito di questi processi risulta contenere 517.984 SNPs.

Il passo successivo è stata l'annotazione genomica, che ha permesso di ricavare informazioni funzionali delle varianti genetiche identificate. Il procedimento è stato effettuato tramite il software ANNOVAR: ogni SNP nel set di dati è stato annotato con il simbolo del gene corrispondente nel genoma umano, utilizzando il database RefGene; è stato usato un criterio di distanza massima di 1 milione di paia di basi (Mb) per associare ogni SNP a un gene e gli identificatori degli SNPs sono stati mappati al riferimento pubblico dell'Axiom Human Origin SNP Array.

2.2 Scansioni di selezione

Sono state effettuate analisi mirate a cercare informazioni sulle varianti genetiche raccolte, in particolare per identificare quelle che sono state soggette a selezione naturale, diventando più comuni in una popolazione, piuttosto che per caso, ad esempio a causa di deriva genetica. Per individuare le tracce della selezione naturale sono stati utilizzati due approcci: metodi basati sull'omozigosi dell'aplotipo esteso (XP-EHH, Cross Population Extended Haplotype Homozygosity), per valutare come la selezione positiva recente abbia influenzato

le frequenze alleliche confrontando l'omozigosi degli aplotipi tra popolazioni; metodi basati sulla differenziazione della popolazione (PBS, Population Branch Statistic), che identificano variazioni importanti di frequenze alleliche in una popolazione rispetto a due popolazioni di riferimento.

Sono state applicate le statistiche XP-EHH, con pacchetto R `rehh`, sul set di dati in fase e polarizzati ponendo come popolazione focale i nativi amazzonici (popolazione principale di interesse) e come popolazioni di confronto i mesoamericani e gli asiatici orientali, per identificare differenze nella selezione. I valori P risultanti sono stati normalizzati per rendere i risultati comparabili, tramite trasformazione negativa \log_{10} .

Prima di applicare le statistiche PBS, sono stati rimossi i siti che risultavano monomorfici in almeno due delle tre popolazioni analizzate, in quanto non fornivano informazioni utili per la differenziazione genetica e permettevano di ridurre il rumore di fondo. SNPs con MAF inferiore al 5% sono stati considerati rari e perciò anch'essi rimossi. Considerando come popolazione focale i nativi amazzonici, come popolazione sorella, ovvero strettamente correlata, i mesoamericani e come gruppo esterno, per confronti più distanti, gli asiatici orientali, è stato applicato un approccio a finestra mobile (con dimensioni di 20 e passi di 5) per ridurre il rumore ed evitare outliers. Per ogni finestra, viene riportato solo lo SNP con il valore PBS più alto insieme alla sua annotazione, in modo da concentrarsi su segnali più forti e significativi.

È stata fatta un'assunzione: a causa del linkage disequilibrium, una variante genica vantaggiosa diventa comune in una popolazione, portando con sé varianti fisicamente vicine; questo è il processo delle selective sweeps, per cui le tracce della selezione positiva sono riconoscibili nelle regioni genomiche adiacenti al gene sotto selezione. Per questo, l'approccio utilizzato considera i risultati derivati dalla media di tutti i valori PBS degli SNP di ciascun gene. Posto come soglia di significatività il 99,9° percentile della distribuzione dei valori PBS, sono stati considerati statisticamente significativi i valori P empirici della coda di destra $\leq 0,001$. Sovrapponendo i valori XP-EHH maggiori di 2 (confronti tra nativi amazzonici e asiatici orientali e tra nativi amazzonici e mesoamericani) e l'analisi PBS superiore al 99,5° della distribuzione sono stati identificati i geni candidati per la selezione positiva e in particolar modo uno SNP candidato.

2.3 Modello demografico e simulazione

Prima di proseguire con l'analisi, è stato costruito un modello demografico assumendo una stasi in Beringia, area che collegava l'Asia e l'America del Nord, a circa 26.000 anni fa, la migrazione e l'insediamento verso l'America a circa 15.000 anni fa e la divergenza tra le popolazioni dell'America centrale e del Sud America circa 13.000 anni fa. Tramite i software ASCEND e IBDNe e basandosi sul dataset di Castro e Silva e colleghi, è stata determinata la dimensione effettiva della popolazione e i suoi cambiamenti nel tempo. È stato assunto come tasso di mutazione 10^{-8} , numero di mutazioni per sito per generazione e un tempo di generazione (tempo medio tra la nascita dei genitori e la nascita dei figli) di 25 anni. Stabiliti questi parametri, sono state effettuate simulazioni demografiche per singoli SNP tramite il software `ms`, per simulare la storia demografica delle

popolazioni e generare dati genetici simulati. In seguito, 10.000 simulazioni sono state utilizzate per generare una distribuzione nulla dei valori PBS, cioè una distribuzione teorica che rappresenta che ci si aspetta sotto uno scenario senza selezione. Questa distribuzione nulla è stata quindi confrontata con i valori PBS osservati (paragrafo 2.3), determinando se questi ultimi sono statisticamente significativi, ovvero se sono significativamente diversi dalla distribuzione nulla e quindi indicano segnali di selezione positiva.

2.4 Tempo e intensità di selezione

Se il set di dati genetici non rappresenta accuratamente la variabilità genetica della popolazione, a causa ad esempio di possibili errori di campionamento, è necessario affrontare questo bias per una migliore rappresentatività dei dati. Sono stati imputati gli SNPs nel cromosoma, che dalle analisi precedenti risultano avere varianti statisticamente interessanti allo scopo di questo studio, utilizzando il software IMPUTE4 e i dati di sequenziamento dell'intero genoma del progetto di diversità genetica umana (HGDP whole genome sequence) come pannello di riferimento per l'imputazione. Successivamente, sono stati selezionati gli SNP situati entro 500 kb su entrambi i lati dello SNP candidato per un totale di regione genomica di 1Mb. Infine, sulla base dei SNP osservati, è stata calcolata la frequenza degli alleli derivati (DAF, derived allele frequency) e l'integrated haplotype score (iHS), la misura dell'estensione dell'omozigosi degli aplotipi attorno al SNP candidato. Alti valori suggeriscono che la variante allelica candidata è stata selezionata positivamente.

Utilizzando il software msms si è valutato quando e con quale intensità è avvenuto questo evento di selezione per lo SNP candidato, simulando l'evoluzione di alleli sotto scenari di selezione naturale. Sono state simulate 10.000 sequenze di 1 Mb ciascuna, prendendo in considerazione il modello demografico descritto nel paragrafo 2.4 e assumendo che: l'allele fosse già segregato nelle popolazioni con una frequenza di 0,25 (valore basato sui dati osservati nella popolazione degli asiatici orientali); la distribuzione a priori sia uniforme per i periodi di tempo da 1000 a 12.500 anni e per il coefficiente di selezione da 0 a 0.2, ovvero che ogni valore all'interno di questi intervalli sia considerato ugualmente probabile. Da queste sequenze simulate sono state selezionati solo gli SNP che corrispondono alle posizioni osservate nei dati reali. Si è ottenuto un sottoinsieme di SNPs, che hanno permesso di calcolare le statistiche DAF e iHS.

È stato utilizzato un approccio ABC (Approximated Bayesian Computation) implementato nel pacchetto R abc, per stimare se i parametri che hanno prodotto i dati simulati possono essere accettati o rifiutati in base alla distanza euclidea (ϵ) tra le statistiche di sintesi osservate e simulate. Le simulazioni con una distanza euclidea inferiore a ϵ vengono considerate accettabili e utilizzate per stimare la distribuzione a posteriori dei parametri di interesse: per ogni modello solo l'1% delle simulazioni con la distanza euclidea più vicina a zero rispetto alle statistiche di sintesi osservate viene mantenuto e utilizzato per stimare la distribuzione posteriore dei parametri.

2.5 Analisi di sovra rappresentazione e analisi di arricchimento dei set genici

I risultati delle scansioni di selezione sono stati utilizzati per condurre analisi che hanno potuto identificare associazioni significative tra varianti geniche e caratteristiche fenotipiche, ruoli biologici e pathway biologici. A questo scopo è stata eseguita un'analisi di sovra rappresentazione (ORA, Overrepresentation Analysis) e un'analisi di arricchimento di set di geni (GSEA, Gene Set Enrichment Analysis). L'analisi ORA ha permesso di identificare se funzioni biologiche sono sovrarappresentate nei geni selezionati rispetto a quanto ci si aspetterebbe per caso; è stata condotta utilizzando gli strumenti FUMA GWAS ed Enrichr che identificano associazioni tra varianti genetiche e caratteristiche fenotipiche riportate nel GWAS Catalog (database che contiene informazioni su varianti geniche associate a tratti e malattie specifiche). Inoltre, tramite Enrichr, utilizzato assieme a GOATOOLS, è stato analizzato il database Gene Ontology (GO) per l'identificazione di categorie funzionali di geni sovrarappresentati nei dati delle scansioni di selezione. L'analisi GSEA ha permesso di valutare in quali processi biologici cellulari e funzioni molecolari l'insieme dei geni selezionati è arricchito; è stato utilizzato WebGestalt come strumento per eseguire l'analisi sul database KEGG Pathway permettendo di identificare pathways biologici significativamente arricchiti nei risultati delle scansioni di selezione. Inoltre, WebGestalt è stato eseguito anche su GO per l'individuazione di categorie funzionali dei geni arricchiti.

Sono stati fissati i seguenti criteri di selezione: i valori soglia sono stati impostati al minimo necessario per ottenere risultati robusti, per assicurarsi che i risultati fossero statisticamente significativi e biologicamente rilevanti; per l'analisi ORA sono stati selezionati geni mappati da SNP con valori di scansione superiori ai percentili 99.99°, 99.95°, 99°, mentre per l'analisi GSEA sono stati utilizzati tutti i geni con punteggi medi.

A questo punto, è stato necessario verificare la correttezza dei risultati trovati. Questa operazione è stata condotta utilizzando il pacchetto R GOfuncR, uno strumento bioinformatico che, ha permesso di calcolare il tasso di errore familiare (FWER, probabilità di avere almeno un falso positivo quando si effettuano confronti multipli. Il tasso di errore familiare utilizza insiemi casuali di geni per determinare la significatività statistica delle categorie funzionali osservate; è stata definita una soglia di FWER < 0.1, è stata accettata, perciò, una probabilità inferiore al 10% di ottenere almeno un falso positivo. Tramite GOfuncR è stato eseguito un test ipergeometrico, un metodo statistico utilizzato per determinare se la rappresentazione del set di geni in una categoria è maggiore di quanto ci si aspetterebbe per caso, tenendo conto della lunghezza dei geni (geni più lunghi possono avere più SNP e quindi maggiore probabilità di essere associati ad una funzione biologica per caso) e il loro raggruppamento spaziale (geni vicini fisicamente sul cromosoma possono avere una maggiore probabilità di essere coinvolte in funzioni correlate; per questo è stata considerata una distanza fisica di 2 Mb). È stata, quindi, confermata la significatività delle categorie funzionali identificate. Come ultimo passo, l'analisi è stata raffinata, sono stati rimossi, cioè, i geni significativi dall'analisi e i dati sono stati rianalizzati per verificare se la categoria funzionale rimanesse

significativa. In questo modo, è stata controllato che l'associazione osservata sia robusta e non dipenda da pochi geni specifici.

2.6 Knockdown del gene candidato in cardiomiociti derivati da hiPSC

Cellule staminali pluripotenti indotte umane (hiPSC) sono state coltivate e differenziate in cardiomiociti (hiPPSC-CM) seguendo il protocollo descritto da Sharma e collaboratori. Al giorno 21 del processo di differenziazione, le cellule hiPSC-CM sono state ripiastrate in due tipi di piastre, a seconda dell'analisi prevista: in una piastra a 12 pozzetti, per analisi proteiche, con 2×10^6 di cellule per pozzetto; in una piastra a 96 pozzetti per analisi delle immagini, con 9×10^4 cellule per pozzetto. Al giorno 28 le hiPSC-CM sono state trattate con particelle lentivirali contenenti shRNA (prodotto specifico MISSION shRNA Lentiviral Transduction Particles #SHCLNV-NM_000944, fornito da Sigma-Aldrich, con i cloni RCN0000352645, TRCN0000342619, TRCN0000352706, TRCN0000002750 e TRCN0000002751), corrispondenti a cinque diverse regioni della stessa proteina codificata dal gene candidato. Il rapporto tra ogni clone di lentivirus e le cellule bersaglio (MOI, multiplicity of infection) è di 200, quindi, è stata garantita un'alta probabilità di infezione. Come controllo negativo è stato utilizzato uno shRNA scramble (acquisito da Addgene, plasmide #1864), trasdotto alla stessa MOI, che non ha come bersaglio nessun gene specifico ed è essenziale per determinare se gli effetti osservati sono specifici per il knockdown del gene candidato. Tramite Western Blot è stata fatta l'analisi proteica: l'efficacia del knockdown è indicata dalla percentuale di riduzione dell'espressione della proteina target; le proteine sono state estratte con buffer per saggio di radioimmunoprecipitazione e caricate su un gel di poliaccrilamide a gradiente di bis-tris al 4-20%. Le proteine sono separate a seconda delle loro dimensioni e successivamente trasferite su una membrana di polivinilidene difluoride e bloccate con latte scremato al 5% in tampone salino fosfato-tween® (PBST) al 0,5%, per prevenire il legame aspecifico di anticorpi. Le proteine sono state, infine, sondate usando anticorpi per la proteina espressa dal gene candidato e per la deidrogenasi della gliceraldeide 3-fosfato (GAPDH), che essendo una proteina generalmente costante nelle cellule, permette di normalizzare i dati e assicurarsi che le differenze osservate non siano dovute a variazioni nella quantità di proteina caricata.

2.7 Infezione delle hiPSC-CM da parte di *Trypanosoma cruzi*

Le cellule hiPSC-CM, dopo 48 ore dal trattamento con shRNA per la riduzione dell'espressione del gene candidato, sono state fatte infettare dalla forma tripomastigote di *T. cruzi*. Il parassita, *T. cruzi* ceppo Y con tag GFP, modificato per esprimere la proteina fluorescente verde (GFP), è stato inoculato dal surnatante LCMK2, nei cardiomiociti derivati con MOI di 5. Dopo 24 ore dall'inoculazione, si è condotta l'analisi dell'infezione tramite analisi delle immagini. Le cellule sono state fissate con paraformaldeide al 4% dopo lavaggi con tampone fosfato salino e colorate con 4,6-diamidino-2-fenilindolo (DAPI) per la marcatura dei nuclei e Alexa Fluor 555-falloidina per la marcatura del citoscheletro. Le immagini sono state catturate utilizzando il sistema

automatizzato di screening ad alto contenuto Molecular Devices IXM-C che consente di acquisire un grande numero di immagini, con ingrandimento di x20. Per ogni pozzetto sono state acquisite 12 immagini, ogni differenziazione delle cellule hiPS-CM è stata replicata in cinque pozzetti, aumentando la robustezza dei dati e sono state replicate due differenziazioni indipendenti per infezione assicurandosi che i risultati siano riproducibili. I nuclei delle cellule e i parassiti presenti nelle 12 immagini sono stati contati automaticamente con MATLAB. Dopo 24 ore dall'infezione, sono stati contati i parassiti all'interno delle cellule per ogni pozzetto: sono stati analizzati 10 pozzetti con il controllo negativo di shRNA scrambled e 10 pozzetti con shRNA specifico. Infine, è stato effettuato un test t per confrontare i due gruppi di cellule, controllo negativo e knockdown, per determinare se ci sia una differenza significativa nel numero di parassiti che infettano le cellule trattate con shRNA specifici e quelle trattate con shRNA scrambled.

2.8 Analisi di correlazione

Sono state eseguite due ultime analisi per esplorare la correlazione tra la distribuzione dell'allele presuntivamente selezionato e la malattia di Chagas, utilizzando il linguaggio R, con il pacchetto stats. Nella prima analisi, sono stati utilizzati i dati dei modelli di distribuzione geografica del numero di vettori per la malattia di Chagas, ovvero le specie di triatomine, descritti da [3] e sono state considerate le coordinate geografiche di ciascuna popolazione analizzata in questo studio. Sono stati correlati i dati sui vettori della malattia con la frequenza del presunto allele selezionato in diverse regioni geografiche. Lo scopo è stato quello di determinare se le aree con una maggiore presenza di vettori mostrano una frequenza più alta dell'allele selezionato. Nella seconda analisi, invece, sono stati correlati i dati sull'incidenza della malattia di Chagas (numero di casi per 100.000 abitanti) in diverse regioni del Brasile (ottenuti dal Sistema di Informazione di Agravos de Notificação SINAN, del Ministero della Salute) e del Messico, tra gli anni 2003 e 2013, e la frequenza dell'allele candidato. Si è voluto valutare se l'incidenza della malattia nelle diverse regioni è correlata con la frequenza dell'allele.

3. RISULTATI E DISCUSSIONE

3.1 Geni candidati alla selezione naturale

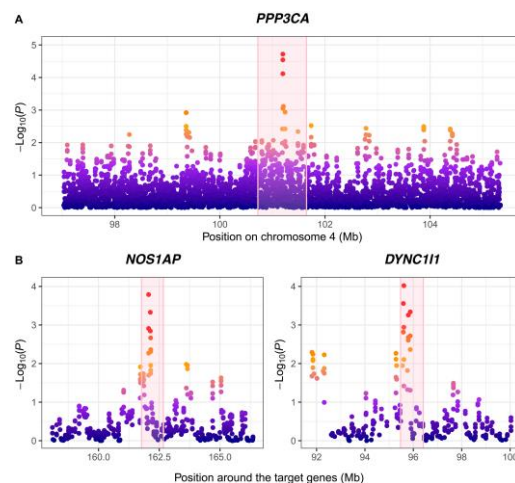
Dalla combinazione delle analisi statistiche degli indici di selezione positiva XP-EHH e PBS, descritte nel paragrafo 2.2, è stata trovata la correlazione della distribuzione risultante con un gruppo di geni, che quindi sono stati ritenuti candidati ad essere sotto selezione. In particolare, i geni individuati sono correlati a tratti metabolici e cardiovascolari (MTRR, DNAJA4, KCMA1 e MTPN), a tratti legati al sistema immunitario (KCMA1 e GCA) e infezioni da patogeni (PPP3CA e DYNC1/1 coinvolti nella sieropositività a *T. cruzi*), come riportato nella *Tabella 1*.

Chromosome	Position	Gene	SNP ID*	Alleles (ancestral> derived)	Association
4	101198533	PPP3CA	rs2659540	G>A	<i>T. cruzi</i> infection
5	7897170	MTRR	rs12347	G>A	Metabolic syndrome and heart disease
1	162060864	NOS1AP	rs10918593	T>C	Mosquito bite size
10	77465512	KCNMA1	rs607483	G>A	Obesity
6	154400423	CNKSR3	rs6917661	T>C	Diabetes type 2
2	162342778	GCA	rs2033300	T>G	Relate to Immune innate system
7	95872168	DYNC111	rs12535988	C>T	<i>T. cruzi</i> infection
15	78266167	DNAJA4	rs2037347	A>G	Metabolic tract - cholesterol
7	135973421	MTPN	rs7801598	G>A	Cardiovascular disease
3	10260685	TATDN2	rs451952	C>T	Metabolic tract - Ghrelin interaction
15	66135032	MEGF11	rs55648467	A>G	Body Index Mass

Tabella 1: Elenco dei geni candidati per la selezione naturale. L'elenco dei geni è stato stabilito sui geni sovrapposti con xpEHH > 2 in entrambe le analisi (amazzonici x asiatici orientali e amazzonici x mesoamericani) e nel 99° percentile nell'analisi PBS.

Sono stati presi in considerazione tre geni, utili per l'analisi, quali PPPCA3 e DYNC1/1 che sono coinvolti nella risposta immunitaria e associati alla sieropositività per *T. cruzi* e NOS1AP, correlato alla reazione alla puntura di zanzara (*Figura 1*). Considerando i cromosomi in cui sono posizionati, questi geni presentano valori PBS più alti.

Figura 1. Log₁₀ dei p-value, dell'analisi PBS, intorno ai geni PPP3CA (A), NOS1AP e DYNC1/1 (B): i grafici mostrano fino a 4 Mb di distanza dalla posizione iniziale e finale del gene e le aree in rosa corrispondono a 300 kb dei limiti descrittivi del gene



Il segnale di selezione più forte trovato nelle popolazioni amazzoniche è quello relativo alla regione del gene PPPCA3 (*Figura 1* e *Figura 2*)

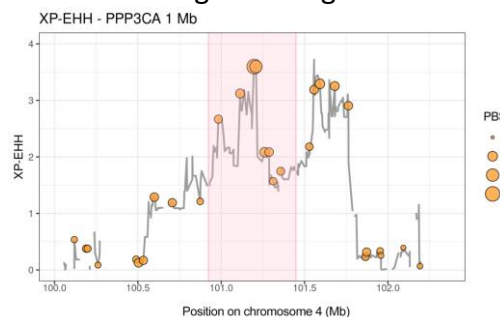


Figura 2: valori XP-EHH e PBS nella regione del gene PPP3CA. I cerchi indicano i valori PBS per SNP e l'asse Y indica i valori XP-EHH. L'asse X indica la posizione genomica in Mb. L'area ombreggiata in rosa corrisponde a 1 Mb dei limiti del gene PPP3CA.

Le simulazioni di coalescenza neutra, che hanno permesso di ricavare la distribuzione nulla di valori PBS e XP-EHH, confrontata con i valori PBS e XP-EHH osservati, indicano che le deviazioni nella frequenza allelica, in particolare della variante allelica rs2659540 G>A, sono statisticamente significative, con p-value $P < 0,0075$ (Figura 3). Questo è coerente con l'azione della selezione positiva, a differenza dell'effetto della deriva genetica nelle popolazioni amazzoniche.

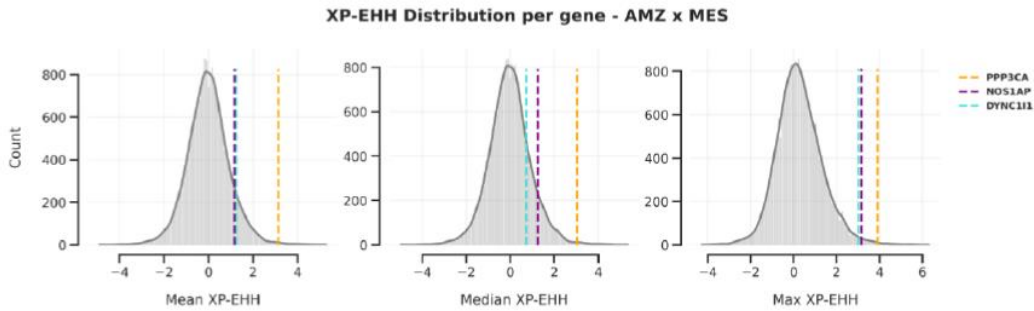


Figura 3: distribuzione dei valori XP-EHH per Amazzonici e Mesoamericani. I grafici indicano i valori medi (pannello di sinistra), mediani (pannello centrale) e massimi (pannello di destra) della distribuzione XP-EHH. Ogni regione del gene candidato è indicata da linee tratteggiate di colori diversi: PPP3CA in giallo, NOS1AP in viola e DYNC11L in blu.

È stato stimato un inizio della selezione a 7500 fa, con intervallo di confidenza CI = 1560 a 12.035, e un coefficiente di selezione di 0.05, CI = 0.015 a 0.20. Questo risultato suggerisce che la selezione ha iniziato ad agire su questo gene dopo la separazione tra le popolazioni amazzoniche, della costa del Pacifico e delle Ande. Uno studio precedente su tessuti antichi di 283 individui, datati da 9000 a 450 anni fa, dell'era costiera del sud del Perù e del nord del Cile ha mostrato un leggero aumento del tasso di infezione nel tempo [4]. Questi risultati supportano l'ipotesi di un segnale di selezione che è probabilmente esclusivo delle popolazioni della foresta pluviale amazzonica.

Dalle analisi di arricchimento Gene Ontology (approccio utilizzato per identificare categorie funzionali di geni che sono sovrarappresentate in un insieme di geni di interesse) basata sia sull'analisi XP-EHH, sia su quella PBS, si è osservato un arricchimento significativo dei geni coinvolti nella risposta alla chemiotassi dei protozoi e dei granulociti eosinofili (Tabella 2), che hanno un ruolo noto nelle infezioni parassitarie.

Method	Association	P value FDR	FWER	Refine P value*
XP-EHH	Defense response to protozoan	0.01	0.08	3.25×10^{-5}
PBS	Eosinophil chemotaxis	0.008	0.03	1.61×10^{-4}
	Eosinophil migration	0.008	0.04	8.11×10^{-4}

Tabella 2: analisi di arricchimento Gene Ontology. Il 95° percentile estremo della distribuzione XP-EHH e PBS ha mostrato un arricchimento statisticamente significativo di geni [tasso di scoperta falsa (FDR) $< 0,05$, FWER $< 0,1$ e valore P affinato], dovuto alla risposta ai protozoi e alla selezione delle cellule T.

*Rimuovere i geni con FWER $< 0,1$ e ripetere il test di arricchimento per verificare se un'associazione significativa rimarrebbe significativa.

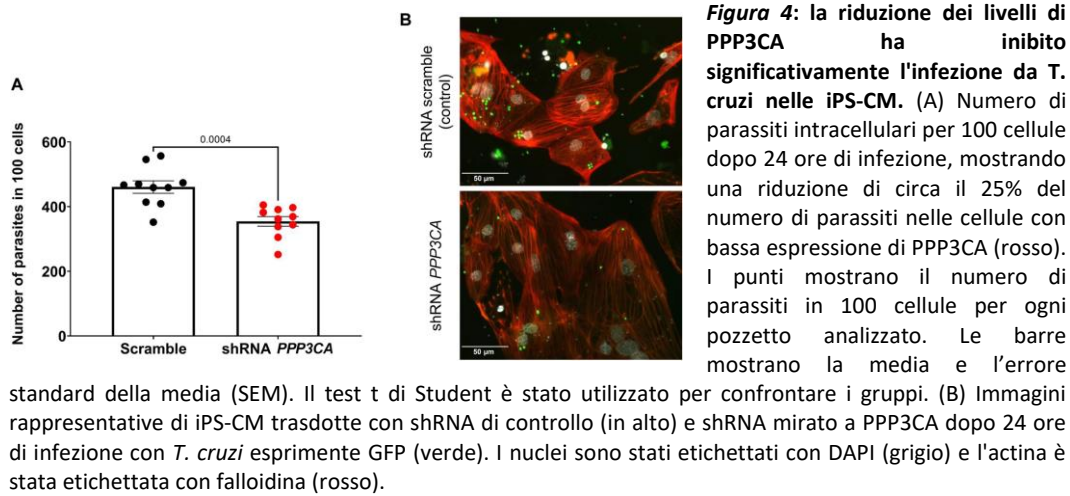
Applicando FUMA GWAS ed Enrichr, sono state individuate associazioni tra le varianti genetiche e tratti fenotipici, come il comportamento di ricerca di novità, tratti metabolici e pressione sistolica/diastolica. Studi su popolazioni indigene brasiliane, come i Xavante, hanno rilevato alti tassi di malattie metaboliche (obesità) e cardiache: il 66% della popolazione soffre di obesità, diabete o malattie coronarie. Uno studio ha mostrato un forte segnale di selezione convergente nei geni correlati al cuore, studiando cacciatori/raccoglitori asiatici e africani come adattamenti compensatori alla loro bassa statura; questo potrebbe essere anche il caso degli indigeni americani. L'ipotesi che è stata formulata suggerisce che la selezione naturale abbia portato a adattamenti genetici per comportamenti di ricerca di novità, come il consumo di caffeina e nicotina, che potrebbero essere stati vantaggiosi per lo stile di vita di cacciatori/raccoglitori del passato, aiutandoli ad esplorare nuovi territori e risorse.

3.2 Caratterizzazione funzionale del gene candidato selezionato: PPP3CA

PPP3CA codifica la subunità catalitica alfa della proteina fosfatasi 3, nota come calcineurina. La calcineurina è una proteina fosfatasi serina/treonina dipendente dal calcio e dalla calmodulina che svolge un ruolo in una vasta gamma di processi biologici, incluso essere un regolatore chiave del riciclaggio delle vescicole sinaptiche nei terminali nervosi. Non a caso, è stato dimostrato che mutazioni di questo gene provocano gravi malattie neurologiche con convulsioni [5]. La calcineurina è coinvolta nella via di segnalazione G- $\beta\gamma$: il complesso G- $\beta\gamma$ è un elemento essenziale nella cascata di segnalazione dei recettori accoppiati a proteine G, che svolge un ruolo importante nell'attivazione delle cellule immunitarie. PPP3CA è coinvolto anche nella segnalazione della calcineurina, un canale chiave nella risposta immunitaria innata e nell'internalizzazione di *T. cruzi*. La calcineurina può attivare fattori di trascrizione come NFAT (Nuclear Factor of Activated T cells) che entrano nel nucleo e regolano l'espressione genica di diversi geni coinvolti nella risposta immunitaria. Il polimorfismo rs2659540, la cui frequenza è risultata statisticamente significativa, è la variante regolatoria attiva nelle cellule immunitarie e nel tessuto cardiaco.

Si è posta l'attenzione sull'associazione del gene con la risposta immunologica ai protozoi, eseguendo uno studio funzionale di PPP3CA in cardiomiociti derivati da cellule staminali pluripotenti indotte umane (hiPSC) infettati con *T. cruzi*. L'analisi ha evidenziato il ruolo attivo del gene PPP3CA nell'infezione intracellulare da parte del parassita e nel risultato della malattia associata: utilizzando shRNA contro PPP3CA si è scoperto che nelle cellule con ridotta espressione genica, pari a circa il 65% rispetto al shRNA controllo, *T. cruzi* ha significativamente diminuito la capacità infettiva di circa il 25% rispetto al shRNA scrambled, come si può osservare nella *Figura 4* (risultato statisticamente significativo, $P < 0,0004$). Inoltre, un'analisi in silico ha mostrato che l'allele derivato rs2659540 ha portato a una significativa riduzione dell'espressione di FLJ20021, un lncRNA nella regione 3'-UTR del promotore di PPP3CA, nell'appendice atriale del cuore. È importante notare che la regione atriale del cuore è principalmente colpita dalla malattia di Chagas, quindi la riduzione dell'espressione genica in questa regione potrebbe avere implicazioni significative per lo sviluppo e l'andamento della

malattia e la riduzione dell'espressione di FLJ20021 potrebbe svolgere un ruolo protettivo contro l'infezione da *T. cruzi* portando ad un fenotipo di malattia più lieve.



Si è ipotizzato che la variante del lncRNA regoli l'RNA messaggero, puntando direttamente alla regione 3'-UTR di PPP3CA; questo suggerisce che FLJ20021 potrebbe agire come regolatore trascrizionale, influenzando l'espressione di PPP3CA attraverso interazioni specifiche con la sua regione 3'-UTR. Secondo un'altra ipotesi, invece, il polimorfismo in considerazione è in linkage disequilibrium con altre varianti del gene PPP3CA (che influenzano direttamente o indirettamente la funzione o l'espressione del gene) nelle popolazioni peruviane con alta ascendenza nativa americana. Nelle popolazioni africane, questo pattern di linkage disequilibrium non è stato trovato, sebbene queste popolazioni mostrino anche un'alta frequenza dell'allele derivato. Questo suggerisce che l'allele derivato rs2659540 potrebbe essere associato a una minore attività di PPP3CA e che l'associazione tra rs2659540 e varianti funzionali di PPP3CA è specifica delle popolazioni native americane.

3.3 Frequenza dell'allele rs2659540 selezionato in diverse popolazioni umane

L'allele selezionato è presente in diverse frequenze in varie regioni geografiche, variando dal 10% in Europa al 59% in Africa e alcuni studi hanno sottolineato la sua importanza nella risposta adattiva ai cambiamenti ambientali. Le foreste tropicali dell'Africa subsahariana e dell'Amazzonia hanno somiglianze ecologiche e patologiche, inclusi i tripanosomi (*T. brucei* in Africa e *T. cruzi* in Amazzonia). Una ipotesi plausibile, quindi, è che PPP3CA sia associato a una risposta all'infezione da tripanosomiasi in Africa e in America, che esercita quindi pressioni selettive simili sul gene. Una volta che l'allele è già segregante nella popolazione ancestrale, la selezione naturale diventa efficace, anche con un coefficiente selettivo di 0,05; nonostante ciò, è essenziale notare che i test XP-EHH non rilevano un segnale di selezione positiva nelle popolazioni africane in questo locus. Si ipotizza che la bassa frequenza dell'allele in Europa (circa del 10%), Asia meridionale (circa del 15%) e Asia orientale (circa del 25%) potrebbe essere dovuta a processi evolutivi come la deriva genetica e a differenze ecologiche che comporterebbero una minore pressione selettiva da parte di

questi patogeni o anche da altri fattori selettivi, come la selezione negativa (l'ipotesi della selezione negativa nelle regioni in cui la frequenza dell'allele rs2659540 è bassa considera il gene PPP3CA associato a casi di interesse clinico). Questo studio non si è focalizzato sulle regioni geografiche sopra citate, per questo si è parlato di ipotesi; per meglio elucidare questa questione, si dovrebbero condurre altri studi che utilizzino sequenze dell'intero genoma (whole genome sequencing) per evitare il bias della genotipizzazione basata sugli SNP.

3.4 Distribuzione geografica di rs2659540 e incidenza di *T. cruzi*

Una delle domande più importanti che sono state poste, ma che rimane senza risposta è perché solo il 30% degli individui infetti da malattia di Chagas sviluppa forme croniche di cardiomiopatia terminale. Non ci sono informazioni, inoltre, su chi tra gli affetti svilupperà malattie cardiache più avanti nella vita o sull'impatto delle varianti genetiche e l'esito della malattia. Quello che è noto è che la malattia di Chagas colpisce circa 6 milioni di persone solo in America latina ed è una delle principali cause di morte della regione; è una malattia emergente negli Stati Uniti e in Europa, dove si stima che circa 450.000 individui siano affetti. Studi che coinvolgono la popolazione attuale dell'Amazzonia non hanno ancora riportato il patogeno ad alta frequenza, nonostante l'ampia distribuzione di vettori tra mammiferi selvatici e triatomine in questa regione, sebbene la malattia fosse già presente nel periodo pre-contatto. In popolazioni come Alto Xingu, Asuriní del Pará, Karitiana, Suruí della Rondônia e Xavante del Mato Grosso, l'analisi dei sierotipi ha rivelato una prevalenza molto bassa o l'assenza della malattia di Chagas. Ciò suggerisce che la malattia è poco diffusa o addirittura assente in queste popolazioni. Contrariamente, nelle popolazioni andine, la malattia di Chagas è considerata endemica e la sua frequenza è alta in tutte le regioni studiate. Questo porta alla conclusione che la malattia di Chagas in Amazzonia è principalmente enzootica e non endemica, la trasmissione della malattia, cioè, avviene principalmente tra animali non umani (enzootica), mentre i casi umani sono sporadici e limitati a specifici cluster geografici.

Dalle analisi descritte nel paragrafo 2.8 e come si può notare nella *Figura 5 e Figura 6*, la distribuzione geografica dell'allele trovato sotto selezione naturale coincide con le aree a bassa infezione da *T. cruzi* in Sud America. Inoltre, è stata trovata una correlazione positiva tra la diversità delle specie vettoriali



Figura 5: distribuzione di rs2659540. L'area gialla sulla mappa rappresenta la regione endemica della malattia di Chagas. L'area arancione rappresenta la distribuzione dell'allele potenzialmente selezionato (rs2659540 G>A) nelle popolazioni studiate

(Tiatomine) e la frequenza dell'allele derivato per il polimorfismo rs2659540 (Figura 7). Questo risultato rafforza l'ipotesi di selezione naturale che porta a un fenotipo protettivo nei nativi americani in questo ecosistema, poiché suggerisce un potenziale di infezione più elevato (dato dalla maggiore diversità di specie vettoriali) nella regione in cui la variante è ad alta frequenza.

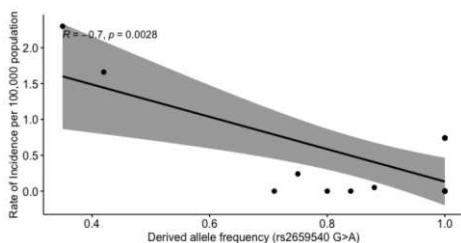


Figura 6: correlazione tra il tasso di incidenza della malattia di Chagas e la frequenza dell'allele candidato (rs2659540 G>A). Tasso medio di incidenza della malattia di Chagas per 100.000 abitanti tra gli anni 2003-2013 nelle regioni geografiche del Messico e del Brasile, sovrapposte alla posizione delle popolazioni di Nativi Americani valutate nel presente studio.

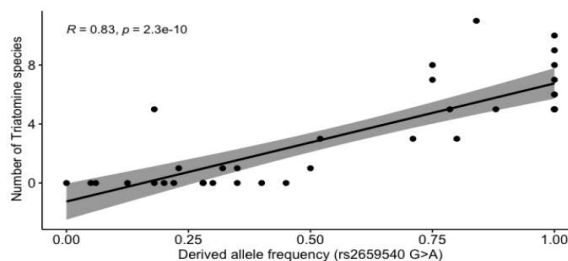


Figura 7: correlazione tra il numero di specie vettore della malattia di Chagas (triatomine) e la frequenza dell'allele candidato (rs2659540 G>A). Ogni punto nero corrisponde alle popolazioni dell'Amazzonia, Mesoamerica e Asia orientale analizzate nel presente studio.

Con circa 2 milioni di persone infette da *T. cruzi*, il Brasile è uno dei paesi più colpiti dalla malattia di Chagas. Gli individui brasiliani con ascendenze africane e nativo-americane sono stati correlati a livelli più elevati di infezione da *T. cruzi*. Tuttavia, è importante notare che queste due componenti dell'ascendenza sono anche associate a uno stato socioeconomico più basso. Anche se la componente di ascendenza nativo-americana nei brasiliani misti è relativamente bassa (5,4%, di cui 7,1% infetti e 4,8% non infetti), è stata osservata un'associazione con un'alta prevalenza di malattia di Chagas. Questo suggerisce che anche una piccola proporzione di ascendenza nativo-americana potrebbe essere un fattore di rischio significativo per l'infezione da *T. cruzi* o un proxy per lo stato socioeconomico. Studi di mappatura della mescolanza nelle popolazioni miste colombiane hanno trovato un'associazione protettiva con la malattia di Chagas all'interno della regione del complesso maggiore di istocompatibilità; tuttavia, non è stata osservata alcuna associazione con la variante PPP3CA. Questa discrepanza può essere spiegata da diverse ragioni: la regione in studio potrebbe avere una composizione di ascendenza diversa da altre regioni, con una prevalenza maggiore di ascendenza andina, perciò, poiché l'allele sottoposto a selezione ha una frequenza inferiore in questa popolazione potrebbe non emergere come significativo nei risultati delle analisi; se la selezione si è verificata in un periodo diverso da quello considerato nell'analisi o se è stata meno intensa rispetto ad altre regioni, potrebbe essere difficile rilevarla nei dati di mappatura della mescolanza; infine, se la selezione è stata localizzata o non è stata diffusa uniformemente nella popolazione, potrebbe essere più difficile identificarla nei dati aggregati utilizzati per l'analisi.

Nelle popolazioni del Messico meridionale, sono stati identificati diversi geni putativi legati alle risposte immunitarie e infiammatorie che sono sottoposti a

selezione naturale, tra cui PPP3CA. Questa regione è caratterizzata dalla presenza di parassiti come *T. cruzi* e *Leishmania mexicana*, che sono endemici nelle foreste locali. Nonostante la presenza di parassiti endemici, i nativi messicani del sud che vivono in queste aree sembrano mostrare una progressione più lieve delle principali cardiomiopatie associate alla malattia di Chagas rispetto ai soggetti provenienti da altre regioni del Sud America. Questo suggerisce che potrebbero esistere alleli sottoposti a selezione naturale che potrebbero influenzare positivamente la progressione della malattia in queste popolazioni.

3.5 Criticità e prospettive

Una criticità metodologica riscontrata è la possibile introduzione di bias data dalla genotipizzazione basata sugli SNPs: l'utilizzo di array di SNP potrebbe introdurre bias nella rilevazione delle varianti genetiche. Poiché, le variazioni genetiche presenti in un individuo non vengono tutte analizzate, questi array potrebbero non coprire tutte le varianti presenti nelle popolazioni. Una soluzione per evitare questo bias potrebbe essere quello di utilizzare il sequenziamento dell'intero genoma che, quindi, permette di identificare tutte le variazioni genetiche. Un'altra criticità riguarda, l'impossibilità, in alcune situazioni, di dissociare l'effetto dell'ascendenza genetica da altri fattori come lo status socioeconomico, che possono influenzare la prevalenza di infezione da parte di *T. cruzi*; ad esempio, se individui con ascendenza nativa-americana vivono spesso in condizioni socioeconomiche svantaggiate, potrebbe essere difficile determinare se l'alto tasso di infezione da *T. cruzi* è dovuto alla genetica o alle condizioni di vita. Lo studio descritto in questa tesi fornisce indicazioni sull'adattamento della popolazione alla foresta amazzonica.

È stato dimostrato che la variante rs2659540 del gene PPP3CA ha probabilmente avuto un effetto protettivo sull'esito dell'infezione della malattia di Chagas nelle popolazioni amazzoniche in passato, probabilmente a causa della riduzione del tasso di infezione dei portatori. L'identificazione di un ruolo di PPP3CA nell'infezione da *T. cruzi* contribuisce alla descrizione dei meccanismi molecolari dell'infezione del patogeno e, quindi, potenzialmente allo sviluppo di terapie per la malattia. La riduzione mirata dell'espressione di PPP3CA potrebbe essere esplorata come strategia terapeutica per ridurre l'infettività di *T. cruzi* nelle cellule umane, tramite ad esempio la tecnologia CRISPR-Cas9, utile nel modificare specificatamente la variante allelica identificata. La creazione successivamente, di modelli animali geneticamente modificati potrebbe essere utile per studiare l'effetto a livello di organismo, ad esempio utilizzando topi transgenici con rs2659540. Uno studio successivo potrebbe, invece, coinvolgere altri geni risultati rilevanti in questo studio, come NOS1AP e DYNC1/1 per comprenderne meglio il ruolo nelle risposte immunitarie legate a *T. cruzi*. Anche se alcuni bersagli molecolari sono stati identificati sulla superficie del parassita [6], manca una comprensione approfondita della loro interazione e partecipazione all'adesione e invasione del parassita; tale conoscenza amplierebbe le possibilità di scoprire nuovi agenti terapeutici per combattere la malattia di Chagas.

BIBLIOGRAFIA

- [1] F. M. Salzano, *The Amerindian Microcosm: Anthropology, Comparative History, Ecology, Genetics and Evolution* (Cambridge Scholars Publishing, 2019).
- [2] Lidani KCF, Andrade FA, Bavia L, Damasceno FS, Beltrame MH, Messias-Reason IJ, Sandri TL. Chagas Disease: From Discovery to a Worldwide Health Problem. *Front Public Health*. 2019
- [3] F. E. Eberhard, S. Cunze, J. Kochmann, S. Klimpel, Modelling the climatic suitability of Chagas disease vectors on a global scale. *eLife* **9**, e52072 (2020)
- [4] A. C. Aufderheide, W. Salo, M. Madden, J. Streitz, J. Buikstra, F. Guhl, B. Arriaza, C. Renier, L. E. Wittmers Jr., G. Fornaciari, M. Allison, A 9,000-year record of Chagas' disease. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2034–2039 (2004)
- [5] Myers CT, Stong N, Mountier EI, Helbig KL, Freytag S, Sullivan JE, Ben Zeev B, Nissenkorn A, Tzadok M, Heimer G, Shinde DN, Rezazadeh A, Regan BM, Oliver KL, Ernst ME, Lippa NC, Mulhern MS, Ren Z, Poduri A, Andrade DM, Bird LM, Bahlo M, Berkovic SF, Lowenstein DH, Scheffer IE, Sadleir LG, Goldstein DB, Mefford HC, Heinzen EL. De Novo Mutations in PPP3CA Cause Severe Neurodevelopmental Disease with Seizures. *Am J Hum Genet*. 2017
- [6] Rodríguez-Bejarano OH, Avendaño C, Patarroyo MA. Mechanisms Associated with *Trypanosoma cruzi* Host Target Cell Adhesion, Recognition and Internalization. *Life* (Basel). 2021



GENETICS

Indigenous people from Amazon show genetic signatures of pathogen-driven selection

Cainã M. Couto-Silva^{1†}, Kelly Nunes^{1†}, Gabriela Venturini^{2,3}, Marcos Araújo Castro e Silva^{1,4}, Lygia V. Pereira¹, David Comas⁴, Alexandre Pereira^{2,3}, Tábita Hünemeier^{1,5*}

Ecological conditions in the Amazon rainforests are historically favorable for the transmission of numerous tropical diseases, especially vector-borne diseases. The high diversity of pathogens likely contributes to the strong selective pressures for human survival and reproduction in this region. However, the genetic basis of human adaptation to this complex ecosystem remains unclear. This study investigates the possible footprints of genetic adaptation to the Amazon rainforest environment by analyzing the genomic data of 19 native populations. The results based on genomic and functional analysis showed an intense signal of natural selection in a set of genes related to *Trypanosoma cruzi* infection, which is the pathogen responsible for Chagas disease, a neglected tropical parasitic disease native to the Americas that is currently spreading worldwide.

INTRODUCTION

America presents a wide range of ecoregions that were quickly explored and occupied by the first humans to reach the continent (1). This remarkable human migration likely required different genetic and cultural adaptation patterns to ensure successful subsistence (2–5). The Amazon rainforest is one of the main ecoregions of the American continent, which is well known for its tropical climate and extraordinary biological diversity. Despite being a nutrient-rich environment, it is also hostile, presenting different obstacles to long-term human survival. Several challenges, including the instability of food resources (6, 7), low light penetration (8), and high diversity of pathogens (9), probably contribute to strong selective pressures for human survival and reproduction in this ecoregion.

The Amazon region, which encompasses the single largest tropical rainforest and nine South American countries, is virtually unrivaled in scale, complexity, and opportunity. It is currently populated by 1 million indigenous people, divided into approximately 300 different ethnic groups (10). Ecological conditions in the region have historically been favorable for transmitting numerous tropical diseases, especially vector-borne diseases (11–13). Although there are numerous studies on postcontact epidemics (14) in the Americas, historical data on precontact diseases (i.e., diseases native to the Americas) are inadequate. However, it has been reported that tuberculosis (*Mycobacterium tuberculosis*) (15, 16) and Chagas disease (i.e., American trypanosomiasis) (17) were present long before the Europeans arrived. Chagas disease is a vector-borne disease caused by the protozoan *Trypanosoma cruzi*, and it is usually transmitted through different triatomine bugs in endemic areas. The oldest record of *T. cruzi* in South American human archaeological remains dates back to 9000 years in mummies from

northern Chile and southern Peru (18). Human remains infected with *T. cruzi* were also found in Brazil about 7000 years before the present (yr B.P.) (19). There are, however, not many studies on the adaptations to the rainforest including Amazonian populations. Most of these are limited to a few individuals from the western Amazonia (20, 21). To date, knowledge regarding genetic adaptations in humans within this complex ecosystem is largely unknown. Motivated by this lack of knowledge, we searched for possible footprints of genetic adaptation to the Amazon rainforest environment by analyzing the genomic data of 118 nonadmixed individuals belonging to 19 native populations (table S1). We were specifically interested in identifying signals for positive natural selection related to tropical diseases.

To search for signals of positive selection, we applied two distinct approaches: (i) Population Branch Statistics (PBS), which identify alleles that have experienced strong changes in frequency in one population relative to two reference populations (19), and (ii) Cross-Population Extended Homozygosity Haplotype (XP-EHH) statistics, which contrast the extended haplotype homozygosity within and between populations (22). We then explored these results through gene pathway enrichment analyses [the Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) (23), Gene Ontology (GO) (24), Kyoto Encyclopedia of Genes and Genomes (KEGG) (25), and Reactome (26) and the Gene Set Analysis Toolkit (GESTALT) (27)]. To formally test whether natural selection underlies the cases of extreme differentiation, rejecting genetic drift as a cause, PBS values were compared against those obtained with neutral coalescent simulations generated according to a plausible demographic scenario for the peopling of South America. Last, we performed a functional follow-up analysis to characterize the role of the putative selected gene.

RESULTS AND DISCUSSION

Genetic adaptation to the Amazon rainforest

The upper distribution of the combined positive selection indices (PBS + XP-EHH) is highly correlated to cardiovascular and metabolic traits (*MTRR*, *DNAJA4*, *KCMA1*, and *MTPN*), immune-related traits (*KCMA1* and *GCA*), and pathogen infection (table

¹Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, 05508090, Brazil. ²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ³Laboratório de Genética e Cardiologia Molecular, Instituto do Coração, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil. ⁴Institut de Biologia Evolutiva, Departament de Medicina i Ciències de la Vida, Universitat Pompeu Fabra, Barcelona 08003, Spain. ⁵Institut de Biologia Evolutiva (CSIC/Universitat Pompeu Fabra), Barcelona 08003, Spain.

†These authors contributed equally to this work.

*Corresponding author. Email: hunemeier@gmail.com

S2). Among the three genes that were highlighted in our analysis, *PPP3CA* and *DYNC11I* were suggestively associated with *T. cruzi* seropositivity and immune response (Fig. 1 and table S2) (28), while *NOS1AP* was related to the mosquito bite reaction (Fig. 1) (29). The strongest selection signature found in Amazonian populations was around the *PPP3CA* gene region (Fig. 1 and figs. S1 and S4). Neutral coalescent simulations indicate that these deviations in allele frequency (rs2659540 G>A) are statistically significant ($P < 0.0075$; fig. S3), which is consistent with the action of ositive selection as opposed to a genetic drift effect in the Amazonian populations. We estimated a selection starting time of 7500 yr B.P. [confidence interval (CI) = 1560 to 12,035] and a selection coefficient of 0.05 (CI = 0.015 to 0.20) (fig. S6). This result suggests that selection started acting on this gene after the split between Amazonian, Coastal Pacific, and Andes populations (30). A previous study on ancient tissues of 283 individuals, dating from 9000 to 450 yr B.P., from South America's coastal area of southern Peru and northern Chile showed a slight increase in the infection rate over time (18). These findings support our inference of a selection signal, which is likely exclusive to Amazonian rainforest populations.

The GO enrichment analysis based on both the XP-EHH (95th percentile) and PBS (95th percentile) analysis showed statistically significant enrichment of genes involved in the response to protozoan and eosinophil chemotaxis, which have a known role in parasitic infections (Table 1 and table S3). When FUMA GWAS and Enrichr were applied, associations with other traits were detected, including novelty-seeking behavior, metabolic traits, and systolic/diastolic pressure (tables S4 and S5). Several studies, including Brazilian indigenous populations, have shown high rates of obesity and cardiopathies (31, 32). For instance, in the Xavante population, 66% of individuals suffer from obesity, diabetes, or coronary heart disease (33). Bergey *et al.* (34) showed a strong signal of convergent selection in heart-related networks in studying Asian and African hunter-gatherers as compensatory adaptations to their short stature, which may possibly also be the case in Native Americans. Furthermore, we also identify associations that could raise a hypothesis for increased novelty-seeking behavior, currently recognized through caffeine and nicotine consumption (tables S4 and S5) (35, 36). Novelty-seeking behaviors may have been important in

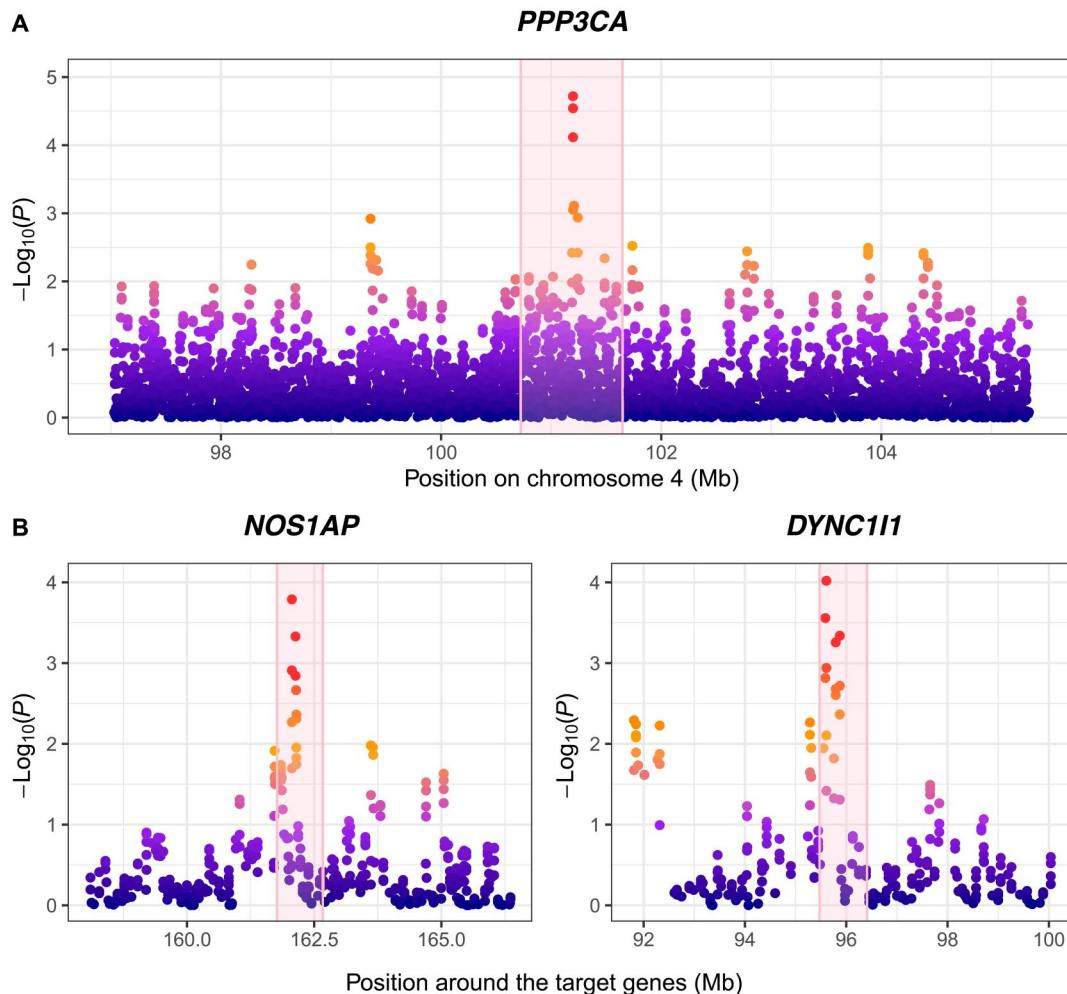


Fig. 1. PBS-windowed scores for target genes. Log_{10} of PBS P values around the (A) *PPP3CA* and (B) *NOS1AP* and *DYNC11I* genes, chromosomes 1 and 4, respectively. Graphs show up to 4 Mb away from the initial and final gene position. Pink shadow areas correspond to 300 kb of the gene limit description. All these genes have the highest PBS values for their respective chromosomes.

Table 1. Gene Ontology enrichment analysis. The extreme 95th percentile of XP-EHH and PBS distribution showed a statistically significant [false discovery rate (FDR) < 0.05, FWER < 0.1, and refine *P* value] enrichment of genes, owing to protozoan response and T cell selection.

Method	Association	<i>P</i> value FDR	FWER	Refine <i>P</i> value*
XP-EHH	Defense response to protozoan	0.01	0.08	3.25×10^{-5}
PBS	Eosinophil chemotaxis	0.008	0.03	1.61×10^{-4}
	Eosinophil migration	0.008	0.04	8.11×10^{-4}

*Remove genes FWER < 0.1 and repeat the enrichment test to check whether a significant association would still be significant.

the past to the hunter-gatherer lifestyle, helping to explore new territories and search for resources (35, 37, 38).

Functional characterization of the putative selected gene

PPP3CA encodes the protein phosphatase 3 catalytic subunit α and is involved in the G- $\beta\gamma$ signaling pathway. The G- $\beta\gamma$ complex is an essential element in the G protein-coupled receptor signaling cascade, which plays an important role in the activation of immune cells. *PPP3CA* is also involved in calcineurin signaling, a key channel in the innate immune response and *T. cruzi* internalization (39–41). The rs2659540 polymorphism is experimentally described as a regulatory variant in the Ensembl database, being active in immune cells and heart tissue (<http://ensembl.org/> on 10 May 2022).

Focusing on the association with the immunological response to protozoa, we performed a functional study of the *PPP3CA* gene in human induced pluripotent stem cell (hiPSC)-derived cardiomyocytes infected with *T. cruzi*. Using short hairpin RNAs (shRNAs) against *PPP3CA*, we found that in cells exhibiting reduced gene expression (65% reduced compared to control shRNA), *T. cruzi* significantly decreased the infectivity capacity by approximately 25% compared to control shRNA ($P < 0.0004$; Fig. 2), indicating that *PPP3CA* plays an active role in parasite intracellular infection and the disease outcome. Further, in silico analysis showed that rs2659540-derived allele (A) led to a significant reduction in the expression of *FLJ20021*, a long noncoding RNA (lncRNA) in the 3' untranslated region (3'UTR) region of the *PPP3CA* promoter, on atrial heart appendage (<https://gtexportal.org/home/snp/rs2659540> on 7 December 2022). The atrial region of the heart is primarily affected by Chagas disease (42, 43). In this sense, this gene expression reduction in the atrium could lead to a milder disease phenotype by decreasing *T. cruzi* infection. One hypothesis is that the lncRNA variant regulates differentially expressed mRNA by directly targeting *PPP3CA* 3'UTR region. However, more functional studies directly targeting the polymorphism found under natural selection are necessary to corroborate this hypothesis. Another possibility would be that this polymorphism is in linkage disequilibrium with functional variants of *PPP3CA* only in Native Americans, leading to differential expression of this gene in individuals carrying the haplotype containing the derived allele. The rs2659540 polymorphism is in strong linkage disequilibrium with

several other variants of the *PPP3CA* gene in Peruvian populations with high Native American ancestry. In African populations, this pattern of linkage disequilibrium is not found, although these populations also showed a high frequency of the derived allele (<http://ensembl.org> on 7 December 2022).

Putatively selected allele frequency in different human populations

The putatively selected allele (A) of the rs2659540 polymorphism also segregates in other geographic regions (ranging from 10% in Europe and 59% in Africa). Studies have reiterated the importance of natural selection in standing variation in local adaptation outside Africa (44). In particular, alleles with intermediate frequency and under balancing selection in Africa have the greatest potential for rapid adaptive response to environmental changes and new selective pressures in the non-African population (45). The sub-Saharan African and Amazon tropical forests have ecological similarities, including trypanosomes (*T. brucei* in Africa and *T. cruzi* in the Amazon). Therefore, a plausible hypothesis is that *PPP3CA* is associated with a response to trypanosomiasis infection in Africa and America. Once the allele is already segregating in the ancestral population, natural selection becomes effective, even with a selective coefficient of 0.05. However, it is essential to note that extended haplotype homozygosity tests do not detect a positive selection signal in the African populations in this locus (<https://pgb.ibe.upf.edu> on 7 December 2022). The lower frequency of the derived allele in Europe, South Asia, and East Asia (~10, ~15, and ~25%, respectively) may result from different evolutionary processes such as genetic drift. In addition, ecological differences between regions may trigger less selective pressure by these pathogens or even by other selective factors (e.g., negative selection—because the *PPP3CA* gene has been associated with other characteristics of clinical interest). Studies should focus on these geographic regions to better elucidate this question, preferably using whole genome sequences to avoid single-nucleotide polymorphism (SNP)-array genotyping bias.

T. cruzi incidence and the geographic distribution of the selected allele

Chagas disease affects approximately 6 million people in Latin America alone and is still a leading cause of death in the region. It is also an emerging infectious disease in the United States and Europe (46), where approximately 450,000 individuals are estimated to have Chagas disease. One of the most important and unanswered questions about this disease is why only 30% of infected individuals develop chronic forms of end-stage cardiomyopathy. There is also no information about who among the infected individuals will develop heart disease later in their lives, or about the impact of genetic variants and ancestry on the disease outcome. Although the disease was already present in Amazonians in the pre-contact period, studies involving the current population have not yet reported the pathogen at a high frequency, despite the wide distribution of vectors among wild mammals and triatomines in this region (47). Serotype analysis in different Amazonian populations, such as Alto Xingu and Asuriní from Pará, Karitiana and Suruí from Rondônia, and Xavante from Mato Grosso, indicates a very low prevalence or absence of the disease. Contrarily, in Andean populations, Chagas disease is considered endemic and its frequency is high in all studied regions (48). This leads to the conclusion that

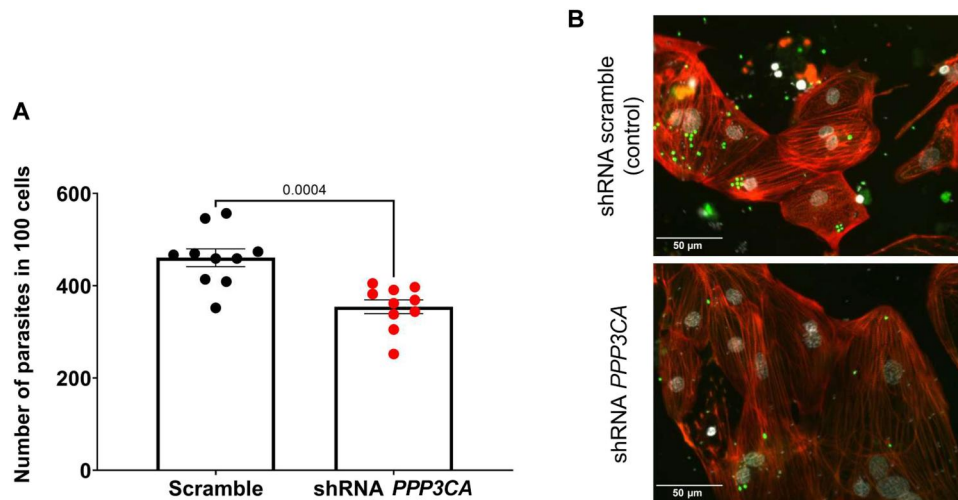


Fig. 2. Reduction in *PPP3CA* levels significantly inhibited iPS-CM *T. cruzi* infection. (A) Number of intracellular parasites per 100 cells after 24 hours of infection showing a reduction of ~25% in parasite numbers in cells with lower expression of *PPP3CA* (red). Dots show the number of parasites in 100 cells for each well analyzed. The bars show the mean and SEM. Student's *t* test was used to compare groups. (B) Representative images of iPS-CM transduced with scramble shRNA (top) and shRNA targeting *PPP3CA* after 24 hours of infection with *T. cruzi* expressing GFP (green). Nuclei were labeled with DAPI (gray), and actin was labeled with phalloidin (red).

Chagas disease in Amazonia is mostly enzootic and not endemic, with only sporadic cases occurring in some geographic clusters. Noticeably, the geographic distribution of the allele found under natural selection coincides with the areas of low *T. cruzi* infection in South America (Fig. 3 and fig. S7). In addition, we found a positive correlation between the vector species diversity (Tiatomine) and the frequency of the allele derived (A) for the rs2659540 polymorphism (fig. S8). This result reinforces our hypothesis of natural selection leading to a protective phenotype in Native Americans in this ecosystem, as it suggests a higher potential for infection in the region where the variant is in high frequency.

Brazil alone has approximately 2 million people infected with *T. cruzi* (49). Urban southeastern admixed Brazilians exhibiting African and Native American ancestries were correlated with higher levels of *T. cruzi* infection. However, these two ancestry components were also associated with poor socioeconomic status, making it difficult to dissociate these two factors (50). Nevertheless, the Native American ancestry component of that admixed Brazilians sample has a median contribution of only 5.4% (7.1% infected and 4.8% noninfected); this low ancestry proportion could be another possible confounding factor related to the association between Native American ancestry and high prevalence of Chagas disease. Admixture mapping studies in Colombian admixed populations found a protective association with Chagas disease within the major histocompatibility complex region (51); however, no association was observed with the *PPP3CA* variant. This can be explained by the differences in the ancestry composition of the region that predominantly exhibited an Andean ancestry (52), and wherein the allele found under selection exhibits a lower frequency (0.277). Furthermore, selection did not necessarily occur in the postcontact period; if it did occur, it may not have been strong or widely distributed enough to be detected in admixture mapping analyses. In Southern Mexican populations, several putative genes related to immune and inflammatory responses were identified under natural selection, including *PPP3CA* (53). The forests in this region are abundant with parasites such as *T. cruzi*

and *Leishmania mexicana*, which are endemic. Nevertheless, Southern Native Mexicans are chronically affected by Chagas disease. Individuals affected by this disease show milder progression of the main Chagas-related cardiomyopathies when compared to subjects from other regions of South America. This suggests that there could be alleles under natural selection that could likely be involved in improving disease progression.

Our results provide insights into the population's adaptation to the Amazonian Rainforest. Combining temporal epidemiological data, genomic analyses of selection, the geographic distribution of the putatively selected allele, and functional analysis of the candidate gene, we showed that a variant of the *PPP3CA* gene likely had a protective effect on the infection outcome of Chagas Disease in Amazonian populations in the past, probably due to the reduction in the infection rate of carriers. Moreover, the identification of a role for *PPP3CA* in *T. cruzi* infection contributes to the dissection of the molecular mechanisms of pathogen infection and, thus, potentially to the development of therapies for the disease.

MATERIALS AND METHODS

Dataset assembly

The dataset consists of combined data from 118 non-admixed (Native American ancestry > 0.99) native Amazonians (table S1 and fig. S9), 35 non-admixed Mesoamericans (Native American ancestry > 0.97), and 231 East Asians from the Human Genome Diversity Project (HGDP; dataset 11, <http://cephb.fr/hgdp/>). Samples were genotyped with the high-density SNP array Axiom Human Origins (~700K SNPs, Affymetrix/Thermo Fisher Scientific). Native American ancestry proportions were estimated using ADMIXTURE (54).

Data quality control and phasing

First, we used the liftOver software (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) to update the physical positions from GRCh37-



Fig. 3. High-frequency distribution of the putatively selected *PPP3CA* allele. The yellow area on the map represents the endemic region of Chagas disease. The orange area represents the distribution of the putatively selected allele (rs2659540 G>A) in the studied populations.

hg19 to the current genome assembly, GRCh38-hg38. Individuals with more than 10% of missing data and those SNPs with minor allele frequency (MAF) lower than 5% or with more than 1% of missing data were removed. We further computationally phased the data using SHAPEIT2 (55) and the recombination map files from the 1000 Genomes Project (56) to phase the haplotypes. No reference panels were used because of the peculiarities of evolutionary history and the lack of representation of Native American populations in these panels. We used the ancestral and derived allele information annotated in the 1000 Genomes Project from EPO pipeline (56), according to Ensembl FTP site (http://ftp.ensembl.org/pub/release-71/fasta/ancestral_alleles/homo_sapiens_ancestor_GRCh37_e71.tar.bz2), and removed SNPs without such information so that our final phased and polarized datasets contain 517,984 SNPs.

Genome annotation

We used ANNOVAR (57) and in-house scripts for the genome annotation. We annotated gene symbols to each SNP from both

datasets using the RefGene database from ANNOVAR itself, taking 1 Mb as a maximum distance threshold to generate the gene annotation. The SNP IDs were mapped to the publicly available reference from the Axiom Human Origins SNP Array (<https://thermofisher.com/br/en/home/life-science/microarray-analysis/microarray-data-analysis/genechip-array-annotation-files.html>).

Selection scans

To identify the signature of positive selection, we applied methods relying on extended haplotype homozygosity (XP-EHH) and population differentiation (PBS) as described below. We used the phased and polarized dataset to apply Cross Population Extended Haplotype Homozygosity (XP-EHH) statistics, using the rehh R package (22, 58). *P* value normalization was performed using a negative \log_{10} transformation. We structured the XP-EHH analysis using the Native Amazonians as the focal population and the Mesoamericans or East Asians in the comparative analysis.

We also applied the PBS as published by Yi *et al.* (59). Before calculating the PBS values, we removed all monomorphic sites in

at least two of the three analyzed populations, as well as the SNPs with MAF lower than 5% when considering the three populations altogether. We structured our analysis by placing the Native Amazonians as the focal population, Mesoamericans as the sister population, and East Asians as the outgroup. We then applied a rolling-mean approach (window size = 20, steps = 5) to reduce the noise and avoid spurious outliers. For each window, only the SNP with the highest PBS value and its related annotation are reported.

In addition, assuming that selective sweeps leave signatures in adjacent genomic regions, we took a gene approach where we summarized the results by taking the mean of all the SNP PBS values of each gene. We considered as significant those SNPs or genes located in the extreme 99.9th percentile of the distribution (empirical P value from right tail ≤ 0.001). Candidate genes for positive selection were accessed by overlapping XP-EHH greater than 2 (Amazonians \times East Asians and Amazonians \times Mesoamericans) and PBS analysis upper percentile of distribution (99.5th percentile).

Demographic model and simulation

The demographic model (fig. S3A) was built assuming a standstill in Beringia at 26,000 yr B.P., the settlement of America at 15,000 yr B.P., and divergence between Central and South American populations at 13,000 yr B.P. The effective population size and its changes over time were based on the Castro e Silva *et al.* (60) dataset inferred through ASCEND (61) and IBDNe (62) softwares. We assume a mutation rate of 1×10^{-8} and a generation time of 25 years. Demographic simulations for single SNPs were performed using the ms software (63). We used 10,000 simulations according to this scenario to generate a null distribution of the PBS values to which observed PBS values were compared. This procedure allowed us to evaluate the significance of the differences and get the outlier PBS values.

Selection time and intensity

First, to deal with the ascertainment bias of the array data, we imputed [IMPUTE4 (64)] SNPs from chromosome 4 using the HGDP Whole Genome Sequence data as the reference panel (65). Next, 500 kb flanking SNPs adjacent to each side of the candidate SNP (rs2659540) were selected, totaling a 1-Mb genomic region. On the basis of the SNPs observed in this region, DAF (derived allele frequency) and iHS (integrated haplotype score) were inferred (58).

To assess when and the intensity of the selective event in the candidate SNP, we used the msms software (66), with the demographic model described above, assuming that the allele already segregated in populations with a frequency of 0.25 (basal frequency observed in East Asian) and assuming uniform prior distribution for time periods (1000 to 12,500 years) and selection coefficient (0 to 0.2) to simulate 10,000 sequences of 1 Mb each. From the simulated sequences, we selected only the markers in positions also observed in the real data. With this subset of SNPs, the summary statistics DAF and iHS were inferred.

Using the ABC approach, implemented in the abc R package [v.2.2.1 (67)], the parameters that produced the simulated data are accepted or rejected according to the Euclidean distance (ϵ) between observed and simulated summary statistics. As a threshold for the rejection algorithm, for each simulated model, we retained 1% of the top simulations with Euclidean distance between observed and simulated summary statistics closer to zero. The posterior distribution was estimated on the basis of these retained simulations.

Overrepresentation and gene set enrichment analyses

We performed both overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA) on our selection scans' results using the GWAS Catalog (<https://ebi.ac.uk/gwas/>), KEGG Pathway (<https://genome.jp/kegg/pathway.html>), and GO (<http://geneontology.org/docs/go-enrichment-analysis/>) databases. We used WebGestalt (<http://webgestalt.org/>) to apply GSEA, and FUMA GWAS (<https://fuma.ctglab.nl/>), Enrichr (<https://maayanlab.cloud/Enrichr/>), and GO for ORA. The GWAS Catalog database was analyzed using both FUMA GWAS and Enrichr, because FUMA GWAS requires a custom gene background and Enrichr does not. We also used Enrichr alongside GOATOOLS to analyze ORA in the GO database. Last, we used WebGestalt to address GSEA in KEGG. For ORA, we selected genes mapped to SNPs with scan values above the 99.99th, 99.95th, and 99th percentiles, while we used all genes with mean scores for GSEA. Threshold values were set at the minimum necessary to return robust results in the enrichment analysis.

To verify the results found, we used the R package GOfuncR (68), which uses random sets to compute the family-wise error rate (FWER; probability of having at least one false positive when multiple comparisons are being tested), performing a hypergeometric test, taking the gene length and spatial clustering of genes (2-Mb physical distance) in account. We then used a threshold FWER of <0.1 and refined the analyses, removing significant genes and reanalyzing the data to verify that the associated category remains significant.

PPP3CA knockdown hiPSC-derived cardiomyocytes

hiPSCs were differentiated into cardiomyocytes (hiPSC-CM) according to Sharma and collaborators (69). On day 21 of differentiation, hiPSC-CM cells were replated in a 12-well plate (2×10^6 cells per well) for protein analysis or in a 96-well plate (9×10^4 cells per well) for image analysis. On day 28, hiPSC-CM cells were transduced with shRNA lentivirus (MISSION shRNA Lentiviral Transduction Particles, Sigma-Aldrich, #SHCLNV-NM_000944, clones RCN0000352645, TRCN0000342619, TRCN0000352706, TRCN0000002750, and TRCN0000002751) corresponding to five different regions of the protein at a multiplicity of infection (MOI) of 200 for each clone. Scramble shRNA was used as a negative control and was acquired from Addgene (plasmid #1864) gifted by D. Sabatini (70) and transduced at the same MOI. Knockdown was evaluated by the reduction percent measured by Western blot. Then, a radioimmunoprecipitation assay buffer was used to extract the proteins, which were then loaded onto a 4 to 20% gradient bis-tris gel (Novex–Thermo Fisher Scientific). Proteins were transferred to a polyvinylidene difluoride membrane, blocked with 5% fat-free milk in 0.5% phosphate buffered saline tween® (PBST), and probed using PPP3CA (Abcam, #ab3673) or glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Santa Cruz Biotechnology, sc-32233) antibodies.

Infection of hiPSC-CM by *T. cruzi*

T. cruzi [Y strain–green fluorescent protein (GFP)–tag (71)] trypanomastigote form from LCMK2 supernatant was inoculated in human induced pluripotent stem cell-derived cardiomyocytes (iPS-CM) 48 hours after the shRNA transfection at MOI of 5. For infection analysis [24 hours post infection (hpi)], cells were fixed with 4% paraformaldehyde after phosphate-buffered saline washes. For image

analysis, cells were stained with 4',6-diamidino-2-phenylindole (DAPI) for nuclei labeling and Alexa Fluor 555–phalloidin (Thermo Fisher Scientific, #A34055) for cytoskeleton labeling. Images were captured at a magnification of $\times 20$ using Molecular Devices IXM-C high-content screening.

We acquired 12 images per well, five wells per differentiation, and two independent differentiations per infection. Nuclei and parasites were counted using MATLAB and are the sum of 12 images of each well. After 24 hours of infection, we counted the number of parasites inside the cells in each well. We analyzed 10 wells with negative control shRNA (scrambled) and 10 wells with shRNA targeting *PPP3CA*, and each dot in Fig. 2 is the measure of one well. We performed a *t* test to compare the number of parasites that infected cells in each group.

Correlation analysis

To investigate the correlation of the geographic distribution of the putative-selected allele (rs2659540 G>A) and Chagas disease, we performed two analyses. In the first, from the geographic coordinates of each population analyzed on the present study, we used data from geographic distribution models of the number of Chagas disease vectors (Triatomine species), described by (13) and (72) and correlated with the frequency of the putatively selected rs2659540 G>A in different geographic regions. In the second analysis, we correlated data on the incidence rate (cases per 100,000 population) of Chagas disease in different regions of Brazil (Sistema de Informação de Agravos de Notificação SINAN, Brazilian Ministry of Health) and Mexico (73), between the years 2003 and 2013, and the frequency of the candidate allele rs2659540 G>A. Correlation analysis was performed in R language with the stats package.

Supplementary Materials

This PDF file includes:

Figs. S1 to S9

Legends for tables S1 to S5

References

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S5

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

- B. Llamas, L. Fehren-Schmitz, G. Valverde, J. Soubrier, S. Mallick, N. Rohland, S. Nordenfeldt, C. Valdiosera, S. M. Richards, A. Rohrlach, M. I. B. Romero, I. F. Espinoza, E. T. Cagigao, L. W. Jiménez, K. Makowski, I. S. L. Reyna, J. M. Lory, J. A. B. Torrez, M. A. Rivera, R. L. Burger, M. C. Ceruti, J. Reinhard, R. S. Wells, G. Politis, C. M. Santoro, V. G. Standen, C. Smith, D. Reich, S. Y. W. Ho, A. Cooper, W. Haak, Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* **2**, e1501385 (2016).
- V. Acuña-Alonzo, T. Flores-Dorantes, J. K. Kruit, T. Villarreal-Molina, O. Arellano-Campos, T. Hünemeier, A. Moreno-Estrada, M. G. Ortiz-López, H. Villamil-Ramírez, P. León-Mimila, M. Villalobos-Comparan, L. Jacobo-Albavera, S. Ramírez-Jiménez, M. Sikora, L.-H. Zhang, T. D. Pape, M. d. A. Granados-Silvestre, I. Montufar-Robles, A. M. Tito-Alvarez, C. Zurita-Salinas, J. Bustos-Arriaga, L. Cedillo-Barrón, C. Gómez-Trejo, R. Barquera-Lozano, J. P. Vieira-Filho, J. Granados, S. Romero-Hidalgo, A. Huertas-Vázquez, A. González-Martín, A. Gorostiza, S. L. Bonatto, M. Rodríguez-Cruz, L. Wang, T. Tusié-Luna, C. A. Aguilar-Salinas, R. Lisker, R. S. Moises, M. Menjivar, F. M. Salzano, W. C. Knowler, M. C. Bortolini, M. R. Hayden, L. J. Baier, S. Canizales-Quinteros, A functional *ABCA1* gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Hum. Mol. Genet.* **19**, 2877–2885 (2010).
- T. Hünemeier, C. E. G. Amorim, S. Azevedo, V. Contini, V. Acuña-Alonzo, F. Rothhammer, J.-M. Dugoujon, S. Mazières, R. Barrantes, M. T. Villarreal-Molina, V. R. Paixão-Côrtes, F. M. Salzano, S. Canizales-Quinteros, A. Ruiz-Linares, M. C. Bortolini, Evolutionary responses to a constructed niche: Ancient Mesoamericans as a model of gene-culture co-evolution. *PLOS ONE* **7**, e38862 (2012).
- C. E. G. Amorim, K. Nunes, D. Meyer, D. Comas, M. C. Bortolini, F. M. Salzano, T. Hünemeier, Genetic signature of natural selection in first Americans. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2195–2199 (2017).
- J. Lindo, R. Haas, C. Hofman, M. Apatá, M. Moraga, R. A. Verdugo, J. T. Watson, C. Viviano Llave, D. Witonsky, C. Beall, C. Warinner, J. Novembre, M. Aldenderfer, A. Di Rienzo, The genetic prehistory of the Andean highlands 7000 years BP through European contact. *Sci. Adv.* **4**, eaau4921 (2018).
- T. B. Hart, J. A. Hart, The ecological basis of hunter-gatherer subsistence in African rain forests: The Mbuti of Eastern Zaire, in *Case Studies in Human Ecology*, D. G. Bates, S. H. Lees, Eds. (Springer, 1996), pp. 55–83.
- R. C. Bailey, G. Head, M. Jenike, B. Owen, R. Rechtman, E. Zechenter, Hunting and gathering in tropical rain forest: Is it possible? *Am. Anthropol.* **91**, 59–82 (1989).
- J. Ratnam, W. J. Bond, R. J. Fensham, W. A. Hoffmann, S. Archibald, C. E. R. Lehmann, M. T. Anderson, S. I. Higgins, M. Sankaran, When is a “forest” a savanna, and why does it matter? *Glob. Ecol. Biogeogr.* **20**, 653–660 (2011).
- V. Guernier, M. E. Hochberg, J.-F. Guégan, Ecology drives the worldwide distribution of human diseases. *PLOS Biol.* **2**, e141 (2004).
- F. M. Salzano, *The Amerindian Microcosm: Anthropology, Comparative History, Ecology, Genetics and Evolution* (Cambridge Scholars Publishing, 2019).
- P. L. Tautil, The status of infectious disease in the Amazon region. *Emerg. Infect. Dis.* **15**, 625 (2009).
- U. E. C. Confalonieri, C. Margonari, A. F. Quintão, Environmental change and the dynamics of parasitic diseases in the Amazon. *Acta Trop.* **129**, 33–41 (2014).
- F. E. Eberhard, S. Cunze, J. Kochmann, S. Klimpel, Modelling the climatic suitability of Chagas disease vectors on a global scale. *eLife* **9**, e52072 (2020).
- R. Walker, L. Sattenspiel, K. Hill, Mortality from contact-related epidemics among indigenous populations in Greater Amazonia. *Sci. Rep.* **5**, 14032 (2015).
- M. J. Allison, D. Mendoza, A. Pezzia, Documentation of a case of tuberculosis in Pre-Columbian America. *Am. Rev. Respir. Dis.* **107**, 985–991 (1973).
- J. E. Buikstra, S. Williams, Tuberculosis in the Americas: Current perspectives, in *Human Paleopathology, Current Syntheses and Future Options*, D. Ortner, A. Aufderheide, Eds. (Smithsonian Institution Press, Washington DC, ed. 1, 1991), pp. 161–172.
- F. Guhl, C. Jaramillo, R. Yockteng, G. A. Vallejo, F. Cárdenas-Arroyo, Trypanosoma cruzi DNA in human mummies. *Lancet* **349**, 1370 (1997).
- A. C. Aufderheide, W. Salo, M. Madden, J. Streitz, J. Buikstra, F. Guhl, B. Arriaza, C. Renier, L. E. Wittmers Jr., G. Fornaciari, M. Allison, A 9,000-year record of Chagas' disease. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2034–2039 (2004).
- V. S. Lima, A. M. Iniguez, K. Otsuki, L. Fernando Ferreira, A. Araújo, A. C. P. Vicente, A. M. Jansen, Chagas disease in ancient hunter-gatherer population, Brazil. *Emerg. Infect. Dis.* **14**, 1001–1002 (2008).
- C. E. G. Amorim, J. T. Daub, F. M. Salzano, M. Foll, L. Excoffier, Detection of convergent genome-wide signals of adaptation to tropical forests in humans. *PLOS ONE* **10**, e0121557 (2015).
- V. Borda, I. Alvim, M. Mendes, C. Silva-Carvalho, G. B. Soares-Souza, T. P. Leal, V. Furlan, M. O. Scliar, R. Zamudio, C. Zolini, G. S. Araújo, M. R. Luizon, C. Padilla, O. Cáceres, K. Levano, C. Sánchez, O. Trujillo, P. O. Flores-Villanueva, M. Dean, S. Fuselli, M. Machado, P. E. Romero, F. Tassi, M. Yeager, T. D. O'Connor, R. H. Gilman, E. Tarazona-Santos, H. Guio, The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32557–32565 (2020).
- M. Gautier, R. Vitalis, reh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
- K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
- H. Mi, A. Muruganujan, D. Ebert, X. Huang, P. D. Thomas, PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
- H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
- J. Griss, G. Viteri, K. Sidiropoulos, V. Nguyen, A. Fabregat, H. Hermjakob, ReactomeGSA-efficient multi-omics comparative pathway analysis. *Mol. Cell. Proteomics* **19**, 2115–2125 (2020).
- B. Zhang, S. Kirov, J. Snoddy, WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741–W748 (2005).

28. R. B. Kennedy, I. G. Ovsyannikova, V. S. Pankratz, I. H. Haralambieva, R. A. Vierkant, G. A. Poland, Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. *Hum. Genet.* **131**, 1403–1421 (2012).
29. A. V. Jones, M. Tilley, A. Gutteridge, C. Hyde, M. Nagle, D. Ziemek, D. Gorman, E. B. Fauman, X. Chen, M. R. Miller, GWAS of self-reported mosquito bite size, itch intensity and attractiveness to mosquitoes implicates immune-related predisposition loci. *Hum. Mol. Genet.* **26**, 1391–1406 (2017).
30. D. N. Harris, W. Song, A. C. Shetty, K. S. Levano, O. Cáceres, C. Padilla, V. Borda, D. Tarazona, O. Trujillo, C. Sanchez, M. D. Kessler, M. Galarza, S. Capristano, H. Montejo, P. O. Flores-Villanueva, E. Tarazona-Santos, T. D. O'Connor, H. Guio, Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6526–E6535 (2018).
31. A. E. P. Lourenço, R. V. Santos, J. D. Y. Orellana, C. E. A. Coimbra, Nutrition transition in Amazonia: Obesity and socioeconomic change in the Suruí Indians from Brazil. *Am. J. Hum. Biol.* **20**, 564–571 (2008).
32. S. G. A. Gimeno, D. Rodrigues, E. N. Canó, E. E. S. Lima, M. Schaper, H. Pagliaro, M. M. Lafer, R. G. Baruzzi, Cardiovascular risk factors among Brazilian Karib indigenous peoples: Upper Xingu, Central Brazil, 2000–3. *J. Epidemiol. Community Health* **63**, 299–304 (2009).
33. L. P. Soares, A. L. D. Fabbro, A. S. Silva, D. S. Sartorelli, L. F. Franco, P. C. Kuhn, R. S. Moises, J. P. B. Vieira-Filho, L. J. Franco, Prevalence of metabolic syndrome in the Brazilian Xavante indigenous population. *Diabetol. Metab. Syndr.* **7**, 105 (2015).
34. C. M. Bergey, M. Lopez, G. F. Harrison, E. Patin, J. A. Cohen, L. Quintana-Murci, L. B. Barreiro, G. H. Perry, Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11256–E11263 (2018).
35. R. Redolat, A. Perez-Martinez, M. Carrasco, P. Mesa, Individual differences in novelty-seeking and behavioral responses to nicotine: A review of animal studies. *Curr. Drug Abuse Rev.* **2**, 230–242 (2009).
36. T. Wingo, T. Nesil, J.-S. Choi, M. D. Li, Novelty seeking and drug addiction in humans and animals: From behavior to molecules. *J. Neuroimmune Pharmacol.* **11**, 456–470 (2016).
37. L. Tovo-Rodrigues, S. M. Callegari-Jacques, M. L. Petzl-Erler, L. Tsuneto, F. M. Salzano, M. H. Hutz, Dopamine receptor D4 allele distribution in Amerindians: A reflection of past behavioral differences? *Am. J. Phys. Anthropol.* **143**, 458–464 (2010).
38. A. de Rubira, L. Georges, L. Fehren-Schmitz, Ancient DNA reveals that the variability of the DRD4-521 C/T SNP associated with novelty seeking behavior is influenced by selection in Western South American populations. *Adapt. Hum. Behav. Physiol.* **2**, 77–91 (2016).
39. S. Feske, H. Okamura, P. G. Hogan, A. Rao, Ca²⁺/calcineurin signalling in cells of the immune system. *Biochem. Biophys. Res. Commun.* **311**, 1117–1132 (2003).
40. P. R. Orrego, H. Olivares, E. M. Cordero, A. Bressan, M. Cortez, H. Sagua, I. Neira, J. González, J. F. da Silveira, N. Yoshida, J. E. Araya, A cytoplasmic new catalytic subunit of calcineurin in *Trypanosoma cruzi* and its molecular and functional characterization. *PLOS Negl. Trop. Dis.* **8**, e2676 (2014).
41. T. Mizuguchi, M. Nakashima, M. Kato, N. Okamoto, H. Kurahashi, N. Ekhilevitch, M. Shiina, G. Nishimura, T. Shibata, M. Matsuo, Loss-of-function and gain-of-function mutations in PPP3CA cause two distinct disorders. *Hum. Mol. Genet.* **27**, 1421–1433 (2018).
42. M. C. P. Nunes, A. Beaton, H. Acquatella, C. Bern, A. F. Bolger, L. E. Echeverría, W. O. Dutra, J. Gascon, C. A. Morillo, J. Oliveira-Filho, A. L. P. Ribeiro, J. A. Marin-Neto: American Heart Association Rheumatic Fever; Endocarditis and Kawasaki Disease Committee of the Council on Cardiovascular Disease in the Young; Council on Cardiovascular and Stroke Nursing; and Stroke Council, Chagas cardiomyopathy: An update of current clinical knowledge and management: A scientific statement from the American Heart Association. *Circulation* **138**, e169–e209 (2018).
43. V. D. Aiello, P. F. P. de Campos, Chronic Chagas cardiomyopathy. *Autops. Case Rep.* **5**, 7–9 (2015).
44. J. S. Rees, S. Castellano, A. M. Andrés, The genomics of human local adaptation. *Trends Genet.* **36**, 415–428 (2020).
45. C. de Filippo, F. M. Key, S. Ghirrotto, A. Benazzo, J. R. Meneu, A. Weihmann; NISC Comparative Sequence Program, G. Parra, E. D. Green, A. M. Andrés, Recent selection changes in human genes under long-term balancing selection. *Mol. Biol. Evol.* **33**, 1435–1447 (2016).
46. K. C. F. Lidani, F. A. Andrade, L. Bavia, F. S. Damasceno, M. H. Beltrame, I. J. Messias-Reason, T. L. Sandri, Chagas disease: From discovery to a worldwide health problem. *Front. Public Health* **7**, 166 (2019).
47. J. Pèneau, A. Nguyen, A. Flores-Ferrer, D. Blanchet, S. Gourbière, Amazonian triatomine biodiversity and the transmission of chagas disease in french Guiana: In Medio Stat Sanitas. *PLOS Negl. Trop. Dis.* **10**, e0004427 (2016).
48. R. V. Santos, C. amp;Epsilon; A. Coimbra Jr, *Saúde e Povos Indígenas* (Editora FIOCRUZ, 1994).
49. A. R. de Arias, C. Monroy, F. Guhl, S. Sosa-Estani, W. S. Santos, F. Abad-Franch, Chagas disease control-surveillance in the Americas: The multinational initiatives and the practical impossibility of interrupting vector-borne *Trypanosoma cruzi* transmission. *Mem. Inst. Oswaldo Cruz* **117**, e2011030 (2022).
50. M. F. Lima-Costa, J. Macinko, J. V. d. M. Mambrini, S. V. Peixoto, A. C. Pereira, E. Tarazona-Santos, A. L. P. Ribeiro, Genomic African and Native American ancestry and chagas disease: The Bambuí (Brazil) epigen cohort study of aging. *PLOS Negl. Trop. Dis.* **10**, e0004724 (2016).
51. D. Casares-Marfil, B. Guillen-Guio, J. M. Lorenzo-Salazar, H. Rodríguez-Pérez, M. Kerick, M. A. Jaimes-Campos, M. L. Díaz, E. Estupiñán, L. E. Echeverría, C. I. González, J. Martín, C. Flores, M. Acosta-Herrera, Admixture mapping analysis reveals differential genetic ancestry associated with Chagas disease susceptibility in the Colombian population. *Hum. Mol. Genet.* **30**, 2503–2512 (2021).
52. H. Ossa, J. Aquino, R. Pereira, A. Ibarra, R. H. Ossa, L. A. Pérez, J. D. Granda, M. C. Lattig, H. Groot, E. F. de Carvalho, L. Gusmão, Outlining the ancestry landscape of Colombian admixed populations. *PLOS ONE* **11**, e0164414 (2016).
53. C. Ojeda-Granados, P. Abondio, A. Setti, S. Sarno, G. A. Gnecci-Ruscione, E. González-Orozco, S. De Fanti, A. Jiménez-Kaufmann, H. Rangel-Villalobos, A. Moreno-Estrada, M. Sazzini, Dietary, cultural, and pathogens-related selective pressures shaped differential adaptive evolution among native Mexican populations. *Mol. Biol. Evol.* **39**, msab290 (2022).
54. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
55. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
56. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
57. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
58. M. Gautier, A. Klassmann, R. Vitis, rehh 2.0: A reimplement of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17**, 78–90 (2017).
59. X. Yi, Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, Huasang, J. Luosang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J. Wang, J. Wang, Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
60. M. A. C. e. Silva, T. Ferraz, C. M. Couto-Silva, R. B. Lemes, K. Nunes, D. Comas, T. Hünemeier, Population histories and genomic diversity of south American natives. *Mol. Biol. Evol.* **39**, msab339 (2022).
61. R. Tournebize, G. Chu, P. Moorjani, Reconstructing the history of founder events using genome-wide patterns of allele sharing across individuals. *PLOS Genet.* **18**, e1010243 (2022).
62. S. R. Browning, B. L. Browning, Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am. J. Hum. Genet.* **97**, 404–418 (2015).
63. R. R. Hudson, Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
64. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
65. A. Bergstrom, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanche, J. F. Deleuze, H. Cann, S. Mallick, D. Reich, M. S. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
66. G. Ewing, J. Hermisson, MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**, 2064–2065 (2010).
67. K. Csilléry, O. François, M. G. B. Blum, abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **3**, 475–479 (2012).
68. S. Grote, GOfuncR: Gene Ontology Enrichment Using FUNC, R package version 1.18.0; <https://bioconductor.org/packages/release/bioc/html/GOfuncR.html>.
69. A. Sharma, C. N. Toepfer, M. Schmid, A. C. Garfinkel, C. E. Seidman, Differentiation and contractile analysis of GFP-sarcomere reporter hiPSC-cardiomyocytes. *Curr. Protoc. Hum. Genet.* **96**, 21.12.1–21.12.12 (2018).

70. D. D. Sarbassov, D. A. Guertin, S. M. Ali, D. M. Sabatini, Phosphorylation and regulation of Akt/PKB by the Rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).
71. M. I. Ramirez, L. M. Yamauchi, L. H. de Freitas Jr, H. Uemura, S. Schenkman, The use of the green fluorescent protein to monitor and improve transfection in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **111**, 235–240 (2000).
72. F. A. Monteiro, C. Weirauch, M. Felix, C. Lazoski, F. Abad-Franch, Evolution, systematics, and biogeography of the triatominae, vectors of Chagas disease. *Adv. Parasitol.* **99**, 265–344 (2018).
73. E. M. Shelly, R. Acuna-Soto, K. C. Ernst, C. R. Sterling, H. E. Brown, A critical assessment of officially reported Chagas disease surveillance data in Mexico. *Public Health Rep.* **131**, 59–66 (2016).
74. P. Skoglund, S. Mallick, M. C. Bortolini, N. Chennagiri, T. Hünemeier, M. L. Petzl-Erler, F. M. Salzano, N. Patterson, D. Reich, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
75. M. Araújo Castro E. Silva, T. Ferraz, M. C. Bortolini, D. Comas, T. Hünemeier, Deep genetic affinity between coastal Pacific and Amazonian natives evidenced by Australasian ancestry. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025739118 (2021).
76. M. A. Castro E. Silva, K. Nunes, R. B. Lemes, A. Mas-Sandoval, C. E. G. Amorim, J. E. Krieger, J. G. Mill, F. M. Salzano, M. C. Bortolini, A. D. C. Pereira, D. Comas, T. Hünemeier, Genomic insight into the origins and dispersal of the Brazilian coastal natives. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2372–2377 (2020).
77. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
78. X. Deng, E. C. Sabino, E. Cunha-Neto, A. L. Ribeiro, B. Ianni, C. Mady, M. P. Busch, M. Seielstad, Genome wide association study (GWAS) of Chagas cardiomyopathy in *Trypanosoma cruzi* seropositive subjects. *PLOS ONE* **8**, e79629 (2013).
79. A. Xu, W. Wang, X. Jiang, The roles of *MTRR* and *MTHFR* gene polymorphisms in congenital heart diseases: A meta-analysis. *Biosci. Rep.* **38**, BSR20181160 (2018).
80. Y. R. Kim, S.-H. Hong, Associations of *MTRR* and *TSER* polymorphisms related to folate metabolism with susceptibility to metabolic syndrome. *Genes Genomics* **41**, 983–991 (2019).
81. A. P. Sowton, N. Padmanabhan, S. J. Tunster, B. D. McNally, A. Murgja, A. Yusuf, J. L. Griffin, A. J. Murray, E. D. Watson, *Mtrr* hypomorphic mutation alters liver morphology, metabolism and fuel storage in mice. *Mol. Genet. Metab. Rep.* **23**, 100580 (2020).
82. H. Jiao, P. Arner, J. Hoffstedt, D. Brodin, B. Dubern, S. Czernichow, F. van't Hooft, T. Axelsson, O. Pedersen, T. Hansen, T. I. A. Sorensen, J. Hebebrand, J. Kere, K. Dahlman-Wright, A. Hamsten, K. Clement, I. Dahlman, Genome wide association study identifies *KCNMA1* contributing to human obesity. *BMC Med. Genomics* **4**, 51 (2011).
83. S. K. Iyengar, J. R. Sedor, B. I. Freedman, W. H. L. Kao, M. Kretzler, B. J. Keller, H. E. Abboud, S. G. Adler, L. G. Best, D. W. Bowden, A. Burlock, Y.-D. I. Chen, S. A. Cole, M. E. Comeau, J. M. Curtis, J. Divers, C. Drechsler, R. Duggirala, R. C. Elston, X. Guo, H. Huang, M. M. Hoffmann, B. V. Howard, E. Ipp, P. L. Kimmel, M. J. Klag, W. C. Knowler, O. F. Kohn, T. S. Leak, D. J. Leehey, M. Li, A. Malhotra, W. März, V. Nair, R. G. Nelson, S. B. Nicholas, S. J. O'Brien, M. V. Pahl, R. S. Parekh, M. G. Pezzolesi, R. S. Rasooly, C. N. Rotimi, J. I. Rotter, J. R. Schelling, M. F. Seldin, V. O. Shah, A. M. Smiles, M. W. Smith, K. D. Taylor, F. Thameem, D. P. Thornley-Brown, B. J. Truitt, C. Wanner, E. J. Weil, C. A. Winkler, P. G. Zager, R. P. Igo Jr, R. L. Hanson, C. D. Langefeld, Family investigation of nephropathy and diabetes (find), genome-wide association and trans-ethnic meta-analysis for advanced diabetic kidney disease: Family investigation of nephropathy and diabetes (FIND). *PLOS Genet.* **11**, e1005352 (2015).
84. E. Dahlström, N. Sandholm, Progress in defining the genetic basis of diabetic complications. *Curr. Diab. Rep.* **17**, 80 (2017).
85. F. Liu, L. I. U. Fengzhi, H. Shinomiya, T. Kirikae, H. Hirata, Y. Asano, Characterization of murine grancalcin specifically expressed in leukocytes and its possible role in host defense against bacterial infection. *Biosci. Biotechnol. Biochem.* **68**, 894–902 (2004).
86. P. Xu, J. Roes, A. W. Segal, M. Radulovic, The role of grancalcin in adhesion of neutrophils. *Cell. Immunol.* **240**, 116–121 (2006).
87. T. W. Kim, S. Hong, A. H. Talukder, V. Pascual, Y.-J. Liu, Grancalcin (GCA) modulates Toll-like receptor 9 (TLR9) mediated signaling through its direct interaction with TLR9. *Eur. J. Immunol.* **46**, 712–724 (2016).
88. C. Robichon, M. Varret, X. Le Liepvre, F. Lasnier, E. Hajduch, P. Ferré, I. Dugail, DnaJA4 is a SREBP-regulated chaperone involved in the cholesterol biosynthesis pathway. *Biochim. Biophys. Acta* **1761**, 1107–1113 (2006).
89. R. J. O'Brien, I. Loke, J. E. Davies, I. B. Squire, L. L. Ng, Myotrophin in human heart failure. *J. Am. Coll. Cardiol.* **42**, 719–725 (2003).
90. I. Seim, S. L. Carter, A. C. Herington, L. K. Chopin, Complex organisation and structure of the ghrelin antisense strand gene *GHRLOS*, a candidate non-coding RNA gene. *BMC Mol. Biol.* **9**, 95 (2008).
91. W.-H. Wei, G. Hemani, A. Gyenesei, V. Vitart, P. Navarro, C. Hayward, C. P. Cabrera, J. E. Huffman, S. A. Knott, A. A. Hicks, I. Rudan, P. P. Pramstaller, S. H. Wild, J. F. Wilson, H. Campbell, N. D. Hastie, A. F. Wright, C. S. Haley, Genome-wide analysis of epistasis in body mass index using multiple human populations. *Eur. J. Hum. Genet.* **20**, 857–862 (2012).

Acknowledgments: We thank C. E. G. Amorim and M. C. Bortolini for comments and suggestions, O. Lao for his help in the analysis, and T. Ferraz for graphic assistance. We also thank all the native communities who participated in the study. **Funding:** This work was supported by São Paulo Research Foundation (FAPESP) grants 17/14916-8 (to C.M.C.-S.), 15/26875-9 (T.H.), and 19/11821-1 (to G.V.), NHLBI R01HL133165 (to A.P.), and NIH grants R01GM075091 (to K.N.) and SU19AI098461 (to A.P.). **Author contributions:** Conceptualization: T.H. Methodology: C.M.C.-S., K.N., L.V.P., and A.P. Investigation: C.M.C.-S., K.N., G.V., and T.H. Funding acquisition: D.C., A.P., and T.H. Project administration: T.H. Supervision: T.H. Writing—original draft: K.N., M.A.C.e.S., L.V.P., A.P., and T.H. Writing—review and editing: T.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The in house scripts for gene adaptation can be found at https://zenodo.org/record/7620975#_Y_Don3bm12x.

Submitted 7 January 2022
Accepted 8 February 2023
Published 8 March 2023
10.1126/sciadv.abo0234