PLOS ONE

RESEARCH ARTICLE

NAT2 global landscape: Genetic diversity and acetylation statuses from a systematic review

Jorge E. Gutiérrez-Virgen[1], **Maricela Piña-Pozas**[2], **Esther A. Hernández-Tobías**[3], **Lucia Taja-Chayeb**[4], **Ma. de Lourdes López-González**[1], **Marco A. Meraz-Ríos**[5], **Rocío Gómez**[1]*

1 Departamento de Toxicología, CINVESTAV-IPN, Mexico City, Mexico, 2 Instituto Nacional de Salud Pública, Centro de Información para Decisiones en Salud Pública, Mexico City, Mexico, 3 Facultad de Salud Pública y Nutrición, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, Mexico, 4 División de Investigación Básica, Instituto Nacional de Cancerología, Mexico City, Mexico, 5 Departamento de Biomedicina Molecular, CINVESTAV-IPN, Mexico City, Mexico

* mrgomez@cinvestav.mx

Abstract

Arylamine *N*-acetyltransferase 2 has been related to drug side effects and cancer susceptibility; its protein structure and acetylation capacity results from the polymorphism's arrays on the *NAT*2 gene. Absorption, distribution, metabolism, and excretion, cornerstones of the pharmacological effects, have shown diversity patterns across populations, ethnic groups, and even interethnic variation. Although the 1000 Genomes Project database has portrayed the global diversity of the *NAT*2 polymorphisms, several populations and ethnicities remain underrepresented, limiting the comprehensive picture of its variation. The *NAT*2 clinical entails require a detailed landscape of its striking diversity. This systematic review spans the genetic and acetylation patterns from 164 articles from October 1992 to October 2020. Descriptive studies and controls from observational studies expanded the *NAT*2 diversity landscape. Our study included 243 different populations and 101 ethnic minorities, and, for the first time, we presented the global patterns in the Middle Eastern populations. Europeans, including its derived populations, and East Asians have been the most studied genetic backgrounds. Contrary to the popular perception, Africans, Latinos and Native Americans have been significantly represented in recent years. *NAT*2*4, *5B, and *6A were the most frequent haplotypes globally. Nonetheless, the distribution of *5B and *7B were less and more frequent in Asians, respectively. Regarding the acetylator status, East Asians and Native Americans harboured the highest frequencies of the fast phenotype, followed by South Europeans. Central Asia, the Middle East, and West European populations were the major carriers of the slow acetylator status. The detailed panorama presented herein, expands the knowledge about the diversity patterns to genetic and acetylation levels. These data could help clarify the controversial findings between acetylator states and the susceptibility to diseases and reinforce the utility of *NAT*2 in precision medicine.

Introduction

Arylamine *N*-acetyltransferase 2 (NAT2) is a phase II xenobiotic-metabolising enzyme with medical relevance, responsible for the biotransformation of several therapeutic drugs, environmental and diet compounds [1–5]. The *NAT*2 gene is strikingly diverse; 45 nucleotide variations have been reported hitherto, of which most are single nucleotide polymorphisms (SNPs) and two deletions (Δ859T and Δ3237A) found in South Indian and Japanese populations, respectively [6–9]. The combination of these variants affects the protein structure and the acetylation capacity, thereby producing at least three phenotypes: fast, intermediate, and slow [1]. Such acetylation states modify the efficient detoxification of exogen substances. Thus, the *NAT*2 genetic patterns could influence susceptibility to adverse drug effects and induce genetic damage such as DNA adduct formation [1,2,10]. Although genotype-phenotype associations have remained controversial, lifestyle and the acetylation phenotype have been associated with susceptibility to neoplasia, insulin resistance, and certain cardiometabolic traits [2–4]. On the other hand, absorption, distribution, metabolism, and excretion, cornerstones of the pharmacological effects, have shown relevant differences regarding the ancestral background [11]. *NAT*2 also exhibits allele, haplotype, and phenotype frequency variations across populations and ethnic groups. Demographic events, historical and cultural transmissions of the populations shape the genetic variation. Hence, some authors have pointed out that lifestyle, acetylation state, and genetic background, have delineated the current epidemiological transitions [12].

The 1000 Genomes Project database has portrayed the global diversity of the *NAT*2 polymorphisms; other studies have described its gene variability in specific populations (https://www.internationalgenome.org/). Nonetheless, several populations and ethnicities remain underrepresented. Furthermore, most studies have been limited to populational descriptive data, leaving gaps in the knowledge of the *NAT*2 genetic architecture that observational studies could make accurate.

Despite evidence about the *NAT*2 clinical relevance, the reconstruction of its worldwide diversity remains partial, requiring a comprehensive and detailed landscape. The present systematic review is state of the art, compiling the *NAT*2 genetic and acetylation patterns from 164 articles published from October 1992 to October 2020, representing 80 countries, 243 different populations and 101 ethnic minorities. The articles included descriptive studies and controls from observational studies expanding the diverse landscape of this phase II enzyme. We conducted the diversity analyses from 35,561 genotypes, 51,860 haplotypes and 70,484 phenotypes, providing one of the most complete and detailed panoramas to date. This review expands the knowledge about the diversity patterns of *NAT*2, applicable to drug therapies, pharmacogenetics, and susceptibility to diseases. Our data may even suggest the genetic patterns of unrepresented populations, where their close genetic ties with related populations could constitute the possible scenery of the harboured genetic architecture.

Materials and methods

Eligibility criteria

We conducted a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-analyses 2020 statement (PRISMA) [13]. Inclusion criteria were restricted to articles published in English and Spanish languages, conducted on human populations where at least two individuals were genotyped. These articles included both population genetic papers and observational studies with hospital- and/or population-based control groups evaluating the genetic contributions of NAT2 polymorphisms to allergy, asthma, hypersensitivity, and cancer. Data from observational studies were obtained from control groups to avoid any

skewed diversity related to a health condition. Several databases with an agreement with our institutions to acquire full-text papers (e.g., Embase, Lilacs, PubMed, and Scopus, among others) were used. In those studies where full-text access was denied, the corresponding author(s) was contacted on three occasions via e-mail, requesting the full-text; the article was removed from the analysis database if they failed to respond to our request.

Studies contravening the eligibility criteria in the primary research focus were excluded, as well as those whose genetic variant frequencies were < 1%. Commentaries, newsletters, reviews, overviews and overlapping publications were also removed from the analysis, along with systematic reviews. Articles lacking crucial information in their documentation, such as those lacking information about the number and reference sequence (rs) of SNPs used and those whose authors did not give access to its data, were also omitted.

Information sources and search strategy

Articles indexed in Embase, Lilacs, PubMed, and Scopus databases published from October 1992 to October 2020 were included. The initial date was chosen because, from 1992 onwards, the number of articles, including the terms of the NAT2 gene and polymorphisms, increased exponentially. The search strategy included free-text terms such as "allergy", "asthma", "cancer", "hypersensitivity", "diversity", "ethnic group", "N-acetyltransferase 2" and "NAT2". These headings were combined with the terms "polymorphism, genetic", "polymorphism, single nucleotide", "genetic variation", and "DNA". Relevant articles selected from the reference list of the included items were searched manually to identify additional studies. All studies reviewed and included herein were from published data.

Selection process

Two reviewers independently screened all studies retrieved from the research strategy using the title and/or abstract as eligibility criteria. These two reviewers further participated in the full-text revision of potentially eligible documents and assessed whether the articles met the inclusion criteria for their eligibility. Said reviewers independently carried out the data extraction and quality assessment of all the documents. Three more reviewers worked independently as arbiters to solve inconsistencies and screen disagreements. Disputes among these five reviewers were resolved through group discussion.

Data collection process

Data extraction. Data extraction was conducted following the guidelines for observational studies in epidemiology. First author's name, publication year, country, ethnicity, geographic region, study design (i.e., case-control, cohort, cross-sectional and population-based), sample size, and gender were included in the data extraction [14,15]. Polymorphism was described as a gene variant with at least 1% frequency in the population; the location and "rs" of each single nucleotide polymorphism (SNPs) were also included.

Observed frequencies from alleles and genotypes were collected for each study (S1 Table). In those studies where allele frequency was not reported, it was set up using the haplotype frequencies reported; these data were underlined. Likewise, when the genotype frequencies were not described, these were constructed assuming Hardy-Weinberg equilibrium.

Concerning the haplotypes, observed frequencies, as reported by the authors, were included in the data extraction. Nonetheless, the statistical analyses were made only with those haplotypes representing the consensus nomenclature [https://nat.mbg.duth.gr] and assigned from at least six SNPs. These criteria were also used regarding the acetylator phenotypes frequencies (i.e., fast, intermedia and slow) reported by the authors. The fast phenotype was defined by the

presence of two fast acetylation haplotypes (i.e., *4, *11A, *12A, *12B, *12C, *13A, *18) in agreement with the consensus nomenclature [https://nat.mbg.duth.gr]. The slow phenotype was defined by two slow acetylation haplotypes (i.e., *5A to *5J, *6A to *6E, *7A and *7B, *10, *12D, *14A to *14G, *17 and *19). The intermedia phenotype was defined by the presence of one haplotype fast and one slow. In those studies, reporting very slow phenotypes, such data were added to slow phenotypes. Some authors reported the phenotypes frequencies using at least 6-SNPs and the *tag* SNP rs1495741; in such cases, the first option was solely considered; prior reports have suggested a similar panel to infer accurately the acetylator status [16]. Nonetheless, if the six-SNPs haplotypes included rs1801279, these data were excluded because this SNP was highly conserved amongst the worldwide populations. In those articles where only the *tag* SNP rs1495741 was reported, these data were used to obtain the three phenotype statuses where AA represented the slow phenotype, GG, the fast one and the heterozygous state, the intermedia phenotype. This *tag* has shown similar accuracy to those inferred with seven SNPs panel [16].

Those studies where the authors did not define the specific haplotype (i.e., *NAT2*5A*) and only reported the general haplotype (i.e., *NAT2*5*) were included in the database but excluded from the rest of the analyses to avoid a skewed panorama about the diversity. Likewise, in the studies where the authors only reported the haplotypes without the phenotype statuses, the frequency of these was obtained by adding the number of individuals with fast (to the fast phenotype) or slow haplotypes (to the slow phenotype) and dividing by in the total number of haplotypes reported. In this situation, only fast and slow phenotypes were reported without the intermedia phenotype.

Quality evaluation. The quality, internal validity, risk of bias and comparability were evaluated in each selected study using the Quality of Genetic association studies tool (Q-Genie) [17]. Q-Genie encloses the statements developed both STrengthening the REporting of Genetic Association studies (STREGA) as a means of strengthening the reporting of Genetic RIsk Prediction Studies (GRIPS) [14,15]. STREGA guidelines are built on the STrengthening of the Reporting of OBservational Studies in Epidemiology (STROBE) [18]. From these two lineaments (STREGA and GRIPS), Q-Genie evaluates the quality of genetic studies with eleven items, each one with seven numeric classification answers: one and two suggest poor quality, three and four suggest moderate quality and five to seven suggest high quality [17]. Three reviewers independently assessed the quality of all the articles selected; disagreements were resolved through group discussions with all five reviewers. Only those studies considered good quality were included: for diversity studies, the threshold score was ≥ 40, whereas, for the studies with a control group, the score was ≥ 45.

Effect measures. Given the characteristics of the study, measures of effect were not applied.

Other statements. This review was not registered, and the protocol was not prepared.

Diversity patterns

Allele frequencies were collected for each study (S1 Table) and depicted in global maps obtained from the United States Geological Survey National Map Viewer [https://viewer.nationalmap.gov/viewer/], that is a public domain.

Although multi-ethnic studies were excluded from all analyses and comparisons, they were included in S1 Table.

Statistical analysis

The frequency distribution of the fast and slow phenotypes of the different populations included in the present study was depicted by geographic regions (Africa; AFR, the Americas;

AMR, Asia; ASI, Europe; EUR, the Middle East; MEA, and Oceania; OCE) through violin plots. All continents were subdivided into regions according to the WorldAtlas webpage (http://worldatlas.com), and the frequency data was shown using box plots. Africa was separated into Central (CAf), East (EAf), Nort (NAf), South (SAf), and West (WAf), regions. The Americas were subdivided into Central (CAm), North (NAm) and South (SAm) regions. Its population diversity landscape was also separated into Afrodescendants (from the USA and Brazil), Asian Americans (from the USA), Native Americans, European-derived populations (whites from Canada and the USA represented as non-Hispanic whites, NHW), and Latinos. Asia was separated into Central (CAs), East (EAs), Southeast (SEAs), and South (SAs) regions. Europe was separated into East (EEu), North (NEu), South (SEu), and Western (WEu). Countries belonging to each region appeared in the footnote of each plot.

Violin and box plots were made with R software using GGplot2 [19]. The median differences among and within continents were performed using the chi-square test (χ^2) with MedCalc® Statistical Software v20.118 [20]. P-values ≤ 0.05 were considered significant. Bar plots with proportions, area charts with the allele frequencies and doughnut charts were made using the Numbers app v12.2 (Apple Inc., 2022).

Haplotype diversity (h) and mean pairwise differences (MPD) were conducted only in the most frequent haplotypes to have a comparison panorama. These two calculous were made with Arlequin v3.5 using 1000 permutations [21]. MPD statistical differences among the different geographic regions were made using the Wilcoxon's test with MedCalc® Statistical Software v20.11 [20]. P-values ≤ 0.05 were considered as significative.

Comparison with other populations. Data were compared with several populations bearing similar ancestral and geographic backgrounds from the 1000 Genomes Project database (1KGP; https://www.internationalgenome.org/). These comparisons were only made in Africa, the Americas, Asia, and Europe.

Results

A total of 1090 publications, including 31 additional articles and 61 records identified by citation searching, were obtained from the first screening. Of these, 926 articles were excluded for various reasons (Fig 1). Two hundred and forty-three potential full-text articles were thoroughly assessed, of which 164 were included in the present study (S1 Table).

The generalities of the studies included

The selected articles represented 80 countries, 243 different populations and 101 ethnic minorities. Of the total of studies, ~ 30% were from the Americas, followed by Asia (24%), Africa (21%), Europe (19%), the Middle East (5.263%), and Oceania (0.330%). From each geographic region, several sub-regions were analysed regarding the number of studies (Fig 2). The most studied area in Africa and Asia was the East (28% and 70%, respectively), whereas in the Americas and Europe was the South region (~ 54% and 35%, respectively).

The countries contributing to the significant number of populations studied in Africa were Cameroon, with ten populations, and Nigeria and Tanzania, with six populations (S1 Table). In the Americas were the USA and Brazil (26 and 18 populations, respectively), excluding the multi-ethnic studies reported in the USA, whereas in Asia were China and Japan with 18 and 12 populations, respectively. Germany (seven populations) and Spain (nine populations) were the European countries with the most population studied. However, the Russian Federation has contributed to 12 populations both in Europe and Asia regions. The Middle East has studied 15 populations.
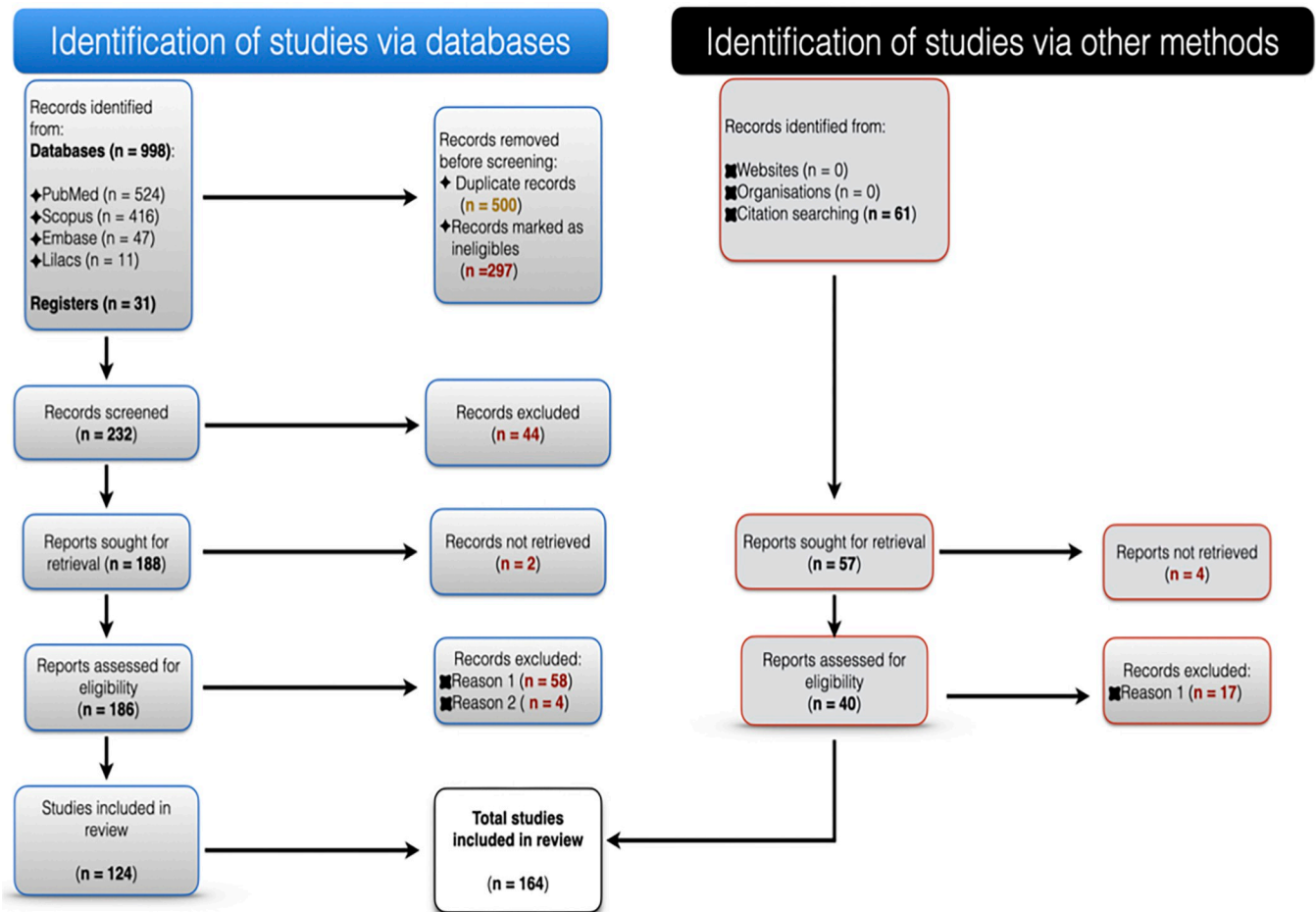
Fig 1. Selection process used in the systematic review following the PRISMA 2020 statement. Note: Reason 1, low quality. Reason 2, duplicated data.

https://doi.org/10.1371/journal.pone.0283726.g001

Regarding the data source, 60% of the studies came from descriptive population studies; the European ones contributed to the most observational studies.

Based on ethnicity, 96% of the articles were from well-established geographic regions; the remaining studies involved multi-ethnic (more than three ethnicities) origins, which were excluded from all analyses. Present-day, European and European-derived populations have been the most studied.

Allele and genotype diversity

The most studied polymorphisms within *NAT2* were rs1801279, rs1041983, rs1801280, rs1799929, rs1799930, rs1208, rs1799931, and rs149574 (S1–S5 Figs). Of these, rs1801279 and rs1799931 depicted a conserved distribution of ancestral alleles being the most prominent. In the case of rs1801279, the African populations and the United Arab Emirates presented the major frequencies of the derivative allele. By contrast, Asians and Latinos showed the highest frequencies of the derivative allele regarding the SNP rs179931. Of note is the high frequency of this allele in Swedish (0.364) and Emiratis (0.244). Another SNP with similar distribution worldwide was rs1041983, whose ancestral allele frequencies presented a range from 0.591 (in Asians) to 0.710 (in Europeans). SNPs such as rs1801280, rs1799929, and rs1208 exhibited
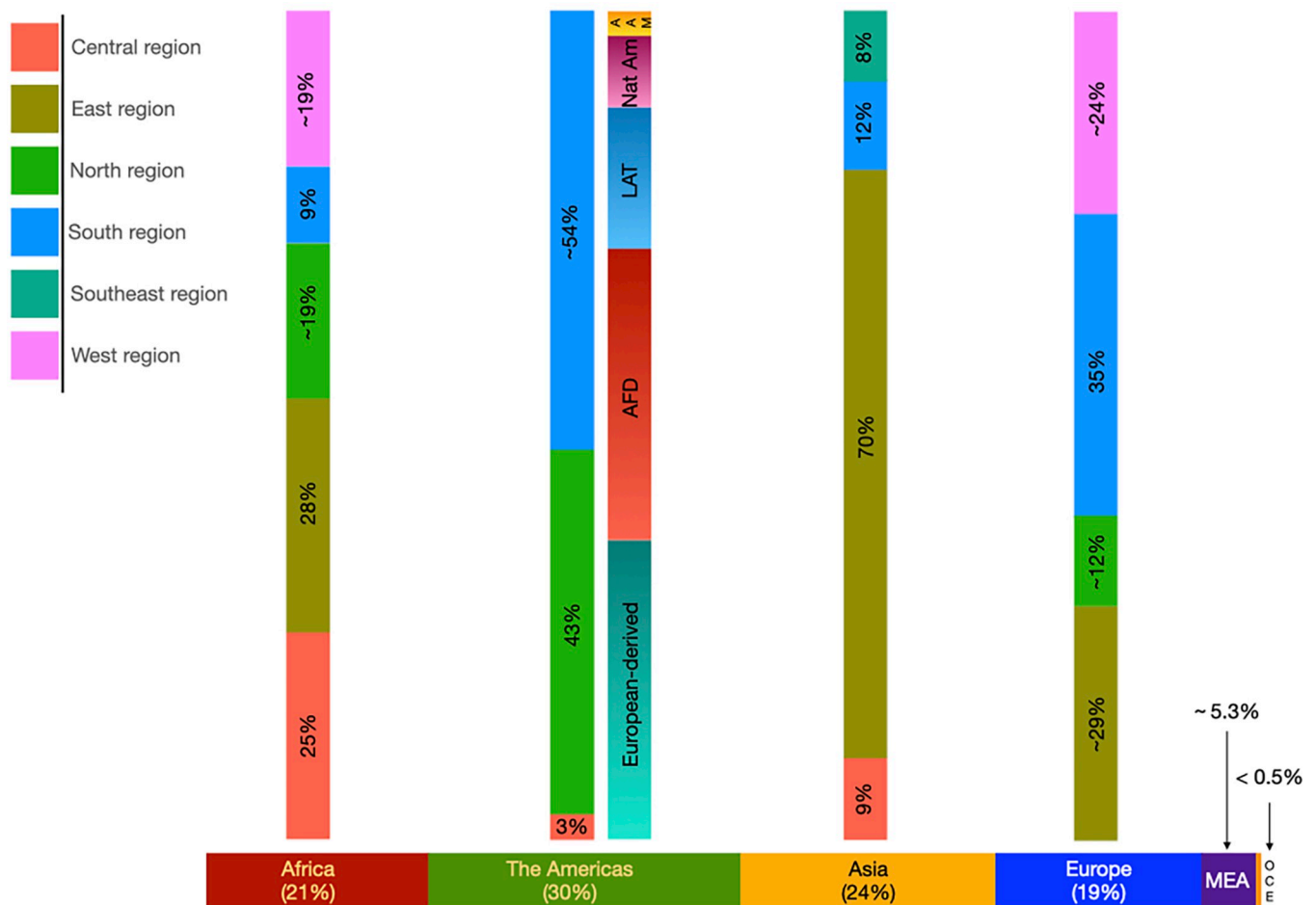
Fig 2. Percentage of articles by region included in the systematic review considering geographic regions and subregions. Note: AAM, Asian Americans; AFD, Afro-descendants; LAT, Latinos; MEA, Middle East, Nat Am, Native Americans; OCE, Oceania.

several distribution patterns with remarkably high frequencies of the ancestral allele in the Asian populations.

Regarding the derivative alleles of these three SNPs, they presented the highest frequencies within European populations. About rs1799930 polymorphism, the highest frequency of the ancestral allele was shown in Latinos (range: 0.775 in Brazilians to 0.999 in Ecuadorians) and Papuans, this last population with a low portrayal (n = 2). The derivative allele was uniformly distributed worldwide (except in Latinos), although Swedish exhibited a remarkable frequency (0.663).

Concerning the *tag* SNP rs1495741 (Fig 3), it has been the least studied with high frequencies of the derivative allele representing the slow phenotype, which was more frequent in South Asia (*f* = 0.779) and Europe (*f* = 0.756). Worthy of note are the distributions of the derivative allele in the Mali population and the opposite pattern in Brazil.

Haplotype diversity by geographic region

The present systematic review depicted the distribution of 97 different haplotypes (S1 and S2 Tables); 68 were determined using at least 6-SNPs. Such 6-SNPs haplotypes were obtained from 19,301 individuals (38,601 haplotypes). Of these, 34 singletons were found. Overall, the
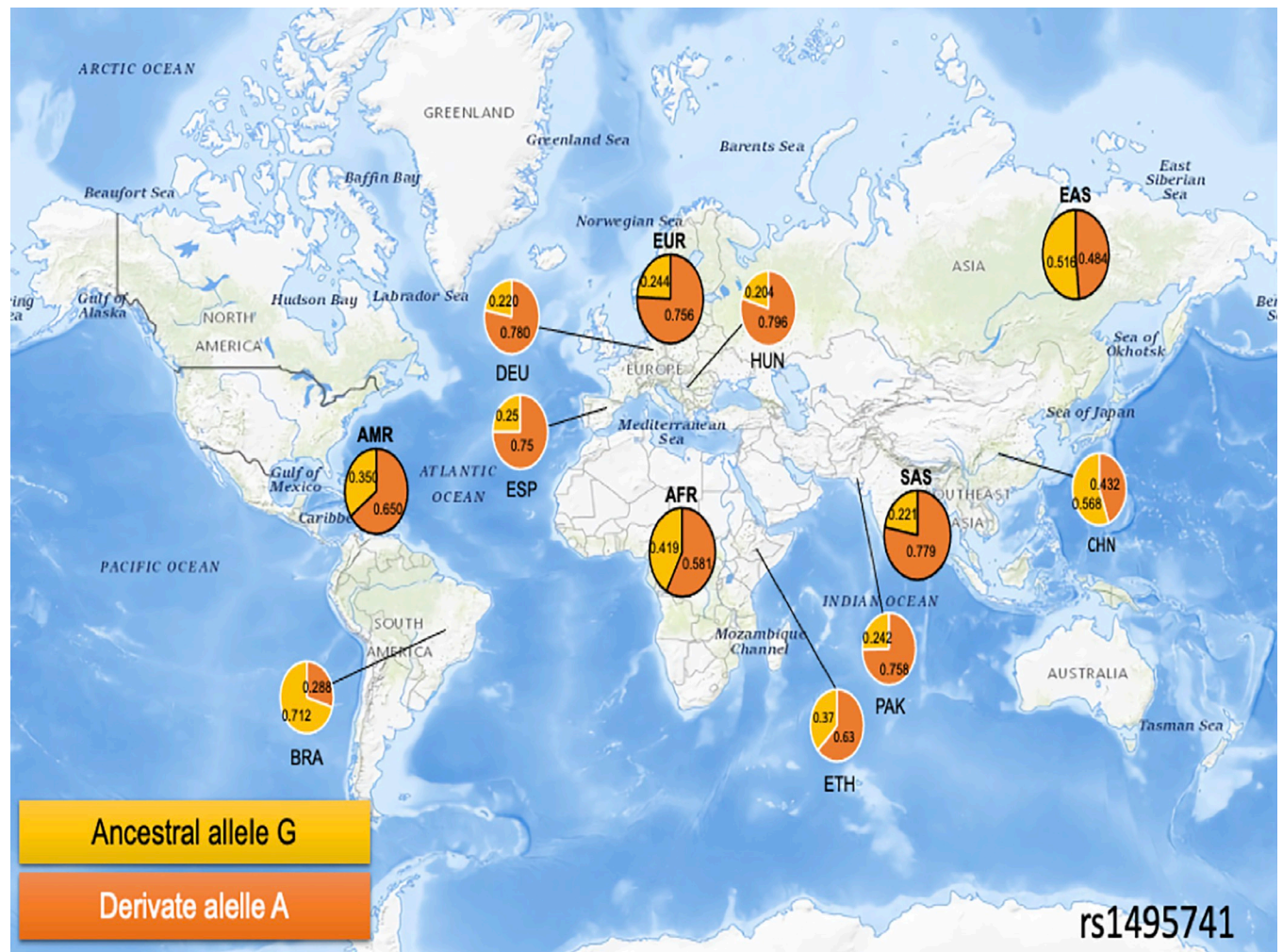
Fig 3. Frequency of the ancestral and derivative allele of rs1495741. Note: AFR, Africa; AMR, The Americas; BRA, Brazil; CHN, China; DEU, Germany; EAS, East Asia; ESP, Spain; ETH, Ethiopia; EUR, Europe; HUN, Hungary; MLI, Mali; PAK, Pakistan; SAS, South Asia. The map was obtained from the United States Geological Survey National Map Viewer [https://viewer.nationalmap.gov/viewer/], that is a public domain.

https://doi.org/10.1371/journal.pone.0283726.g003

haplotype *NAT2*4* (wild type) was the most common globally, followed by *5B, *6A, and *7B (S6 Fig). Other haplotypes with critical frequencies were *12A (in African, European and the Americas populations), *5A, *5C, *7A (Asian and the Americas populations) and *14B (in Africa and the Middle East).

Regarding the different haplotypes found by region, Africa presented 41 haplotypes, of which twenty have been described, as yet, only in this region (singletons). Such singletons have been characterised, mainly within *6 and *12 haplotype clusters. The Middle East was represented by 32 haplotypes and nine singletons within the *5 haplotype cluster. The Americas exhibited 31 haplotypes and six singletons within *6 and *7 haplotype clusters, whereas Europe presented 26 haplotypes and three singletons. Asia was the least diverse region, with seventeen haplotypes and one singleton. Oceania was only represented by two haplotypes from two individuals. Such patterns may present bias because they depend on the resolution power of each study, the sample size and the dates on which they were made; the new technologies have the advantage of the resolution within haplotypes. Hence, the haplotype diversity and mean pairwise differences were conducted with the data from the eight most represented haplotypes

Table 1. Diversity patterns for *NAT*2 haplotypes by continental regions.

Region	N	h	Haplotype diversity	MPD
Africa	1,955	15	0.728 ± 0.007	2.508 ± 1.352
America	5,362	20	0.689 ± 0.004	2.648 ± 1.414
Asia	3,206	13	0.648 ± 0.006	2.002 ± 1.129
Europe	13,176	21	0.583 ± 0.003	2.444 ± 1.324
Middle East	2,957	16	0.715 ± 0.005	2.736 ± 1.452

Note: *N*, number of samples; *h*, Number of different haplotypes found; MPD, mean pairwise differences.

excepting *NAT*2*4 (i.e., *12A, *5A, *5B, *5C, *6A, *7A, *7B, and *14B). These results showed comparable patterns to those described as a whole without any significant difference amongst the different regions (Table 1). In Africa, the western and central countries contributed to the greatest diversity, principally in those haplotypes within *6 and *12. The countries within the southern region were the least diverse but also the minor studied. In Europe, the most diverse region was the southern one, followed by the Easter and Western.

No significant values (p-values ≥ 0.05) were found among MPD by geographic regions by the Wilcoxon's test.

Acetylation capacities

Slow phenotype. The substantial charge of the slow phenotype worldwide was a consequence of the high frequency of the slow haplotypes. This status was more frequent in the Middle East, where most data were distributed around the median (0.782) with remarkable frequencies in the Ashkenazi Jews, Emiratis and Pakistanis (S1 Table). Significant differences ($p \leq 0.0001$) were found between MEA and Asia (~1.4 times lower) and the Americas (~1.3 times lower) when comparing the median values (Fig 4). Africa and Europe also presented high frequencies of this phenotype (median values: 0.758 and 0.751, respectively, without significant differences). About Africa, the north region presented the highest median value (0.823), showing marked differences with CAf, SAf, and WAf ($p \leq 0.001$) (S7 Fig). The prominent frequencies were presented in Cameroon (CAf) within the Fulani ethnicity and Tanzania (EAf) in Burunges, Hazdas and Maasais. Of note, the contrasting frequencies are even within the same country (i.e., Cameroon). Conversely, SAf presented the lowest median value (0.391) with significant differences ($p \leq 0.0001$) with all regions.

Europe regions presented similar slow phenotype distributions with median values in a rank of 0.736 (SEu) to 0.795 (NEu). Thus, any significant difference ($p \geq 0.05$) among the different regions was identified (S8 Fig). The highest frequencies were seen in NEu (Sweden) and WEu (France), whereas the lowest ones were found in SEu (Serbia) and France. The Americas (0.612) and Asia (0.565) presented the lowest median values (S9 and S10 Figs). The lowest frequencies were observed in Japan (EAs) and within the Native Americans. Inside the Americas, the lowest median values (0.147) were seen in CAm, which was ~ 4.5 and 3.8 times lower in comparison with NAm and SAm, respectively ($p \leq 0.0001$); this interpretation should be taken with caution. The patterns of this phenotype were dissimilar depending on the ancestral background. Latinos presented the most significant frequencies of the slow phenotype: Brazil (0.480) and Mexico (0.560). High frequencies were also observed within the Afro-descendants from Brazil (0.290) and the USA (0.260) and in the European-derived populations from Canada and the USA (0.310) (Fig 5).

Regarding the Asian region, the highest values of the slow phenotype were presented in SAs (0.839), represented by two Indian studies. SEAs (0.640), CAs (0.615) also presented similar
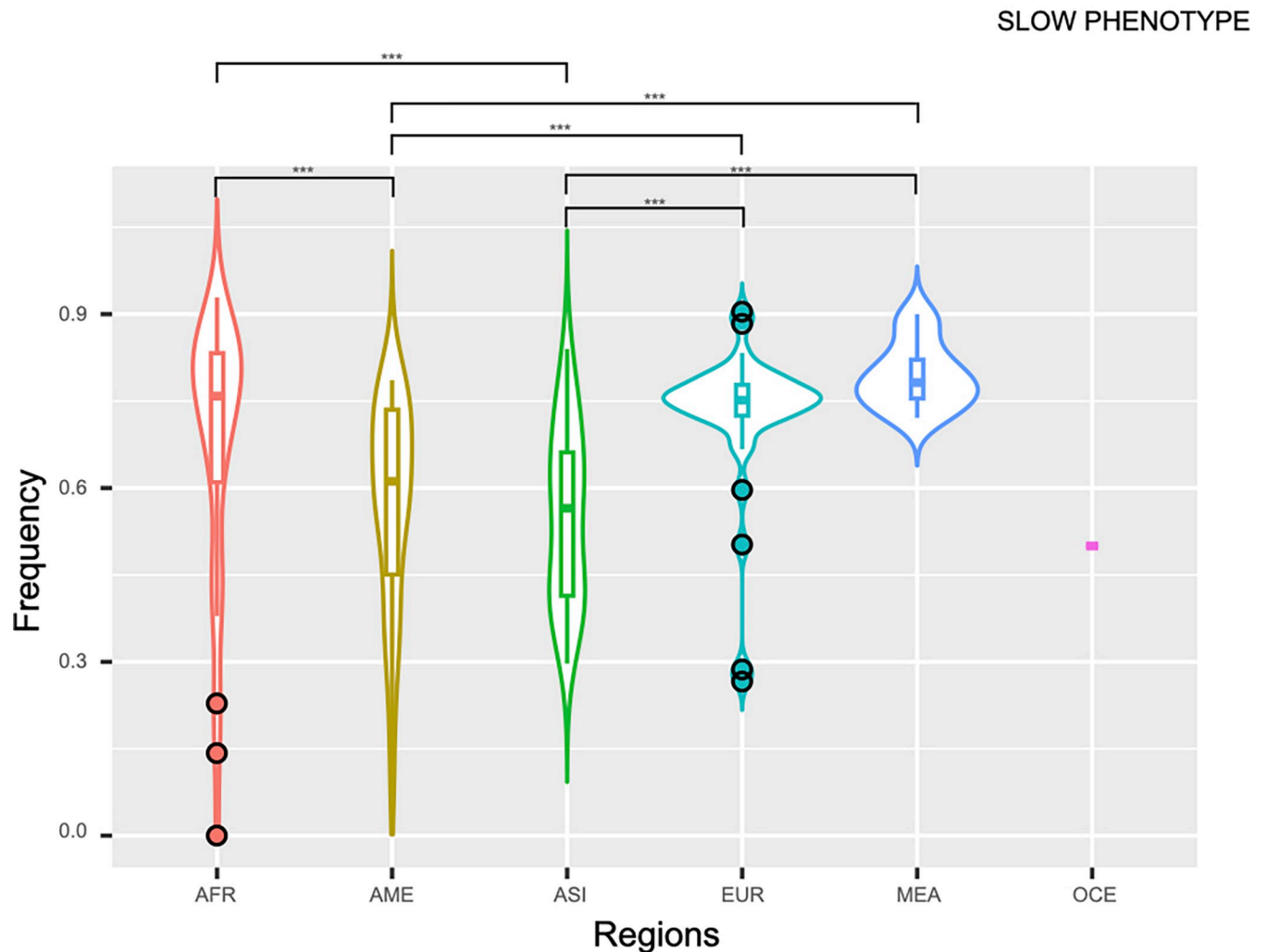
Fig 4. Violin plots of the slow phenotype distribution by geographic region. Note: AFR, Africa; AME, The Americas; ASI, Asia; EUR, Europe; MEA, Middle East, OCE, Oceania.

frequencies (S10 Fig). By contrast, EAs (0.429) exhibited the lowest values, particularly within Japanese populations, except in the Eskimos and Yakuts from Siberia. Yet, the pattern of this region should not be generalised, given the paucity number of studies.

Fast phenotype. Regarding the fast phenotype, it was more frequent in Asia (median = 0.435) and the Americas (median = 0.388). Thus, significant differences ($p \leq 0.0001$) were found to compare with Africa (median = 0.242), Europe (median = 0.249) and MEA (median = 0.218), where it was almost twice less frequent (Fig 6). Some African populations, such as Baka and Bakola Pygmies from Cameroon, Biaka Pygmies (Central African Republic), and San (Namibia), also presented high frequencies of this phenotype (S1 Table). A similar pattern was found in Serbia and France within the European region.

Inside Asia, the Eastern region presented the higher median frequencies (0.571); Chinese, Han Chinese, Koreans and Western Siberians exhibited the highest values. Thus, significant differences ($p \leq 0.0001$) were found to compare it with the other regions (CAs = 0.385; SEAs = 0.360; and SAs = 0.161) (S11 Fig). About the Americas, overall, the Native American populations exhibited the major frequencies; Emberas and Ngawbes from Panama presented
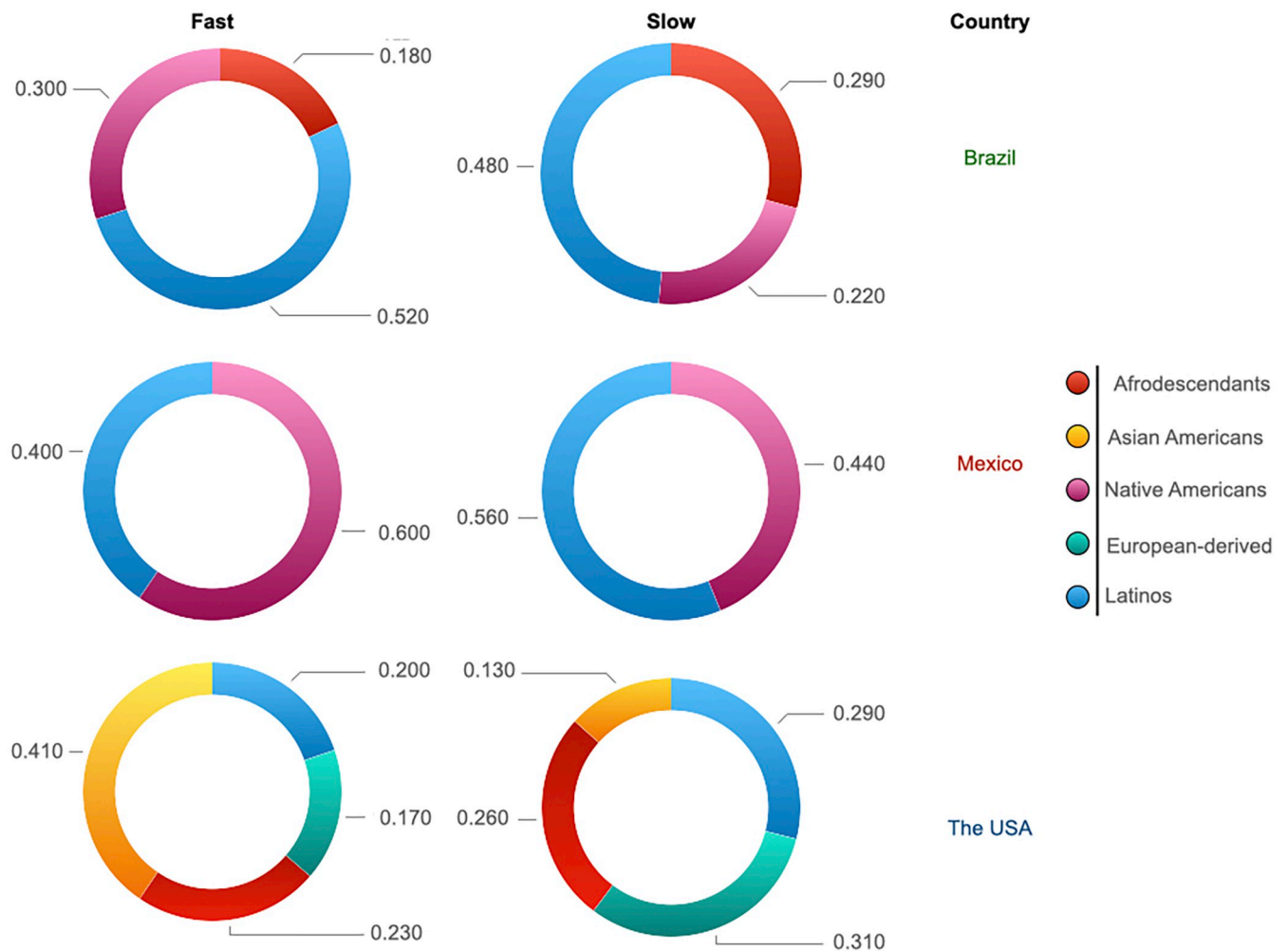
Fig 5. Doughnut charts representing the fast and slow phenotypes distributions by ethnicity in Brazil, Mexico and the United States of the America.

the highest frequencies (0.853 and 0.924, respectively) (S12 Fig). Latinos also presented high values of this phenotype (Fig 5).

About Africa, the highest frequencies of the fast phenotype were found within SAf populations (median = 0.609), followed by the Central region where Bakola Pygmies (Cameroon) presented the greatest values, accompanied by Namibia (SAf, 0.857) and the Wolaitas from Ethiopia (EAf, 0.621) (S13 Fig). Significant values were found amongst the different regions (range $p \leq 0.05$—$p \leq 0.0001$). Among the different countries belonging to the Middle East, the highest frequencies were presented in Jordan (0.279), and Druze from Israel (0.273); the Ashkenazi Jews (0.100) and the Emiratis (0.119) presented the lowest frequencies (S1 Table).

Europe depicted similar median distributions among the regions (range:0.205 in Neu to 0.264 in SEu) without any significative differences (S14 Fig). Of note is the high frequencies of the fast phenotypes in Serbia (0.730) and France (0.711).

Intermedia phenotype. Although this phenotype has not been fully reported, those studies that included it have illustrated high frequencies in East and South African populations as well as African Americans. By contrast, the lowest values were found among Europeans, its descendant populations, and in the Middle East (Fig 7).
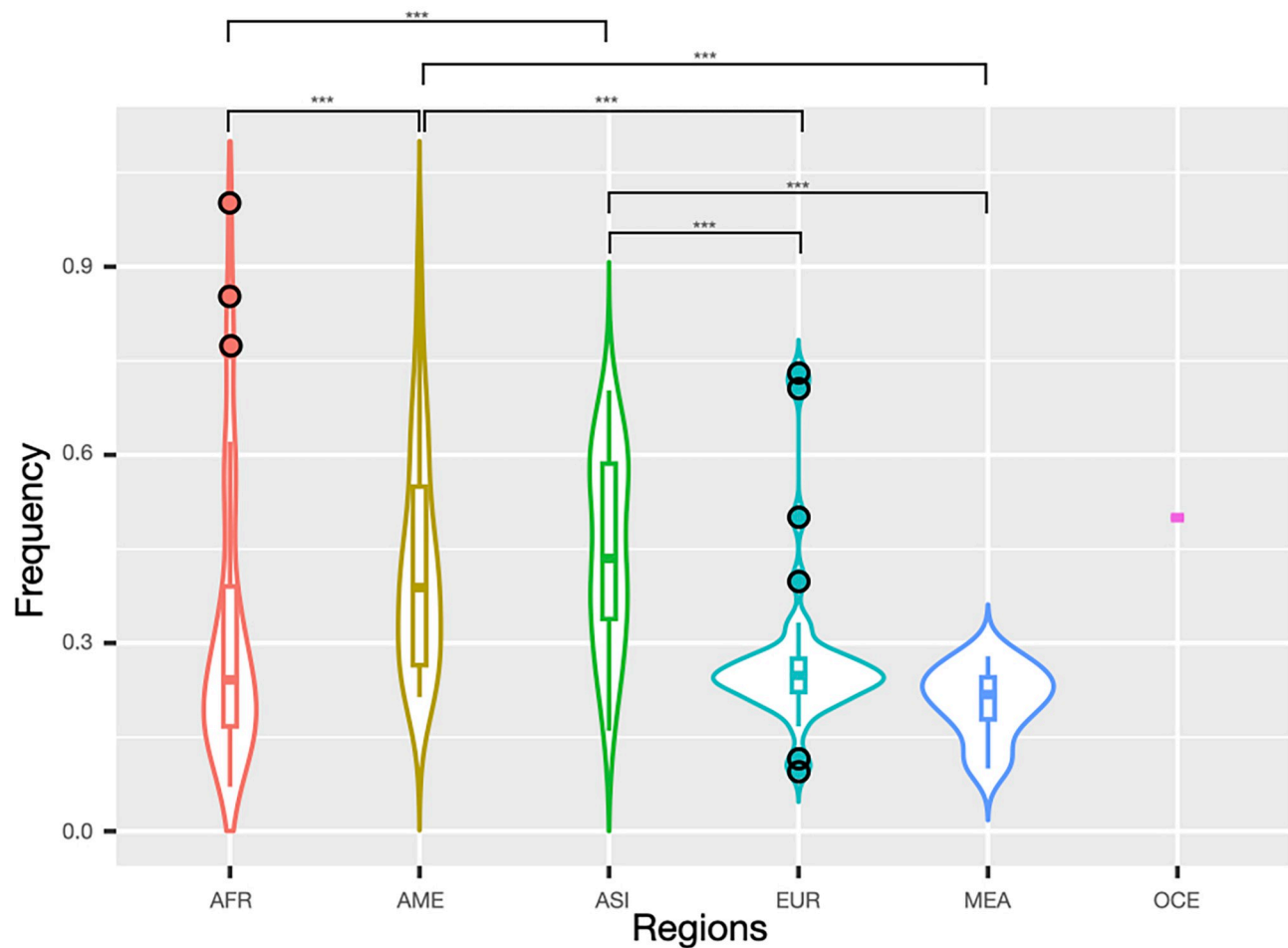
Fig 6. Violin plots of the fast phenotype distribution by geographic region. Note: AFR, Africa; AME, The Americas; ASI, Asia; EUR, Europe; MEA, Middle East, OCE, Oceania.

Comparative with 1000 genomes project data

Overall, the allele frequency distributions of the eight most studied *NAT2* polymorphisms were congruent with those reported in 1KGP. However, the panorama of populations analysed herein portrayed a more accurate frequency distribution and included more populations and individuals. Our contribution to the scenery of genetic diversity in Africa comprised 22 populations. The Americas had several populations (Natives, Afro-descendants, Asian Americans, and European-derived populations) from Canada, the USA, Mexico, Nicaragua, Panama, Colombia, Argentina, Paraguay, Brazil, Peru and Ecuador. Inside Asia, Central and Southeast regions were included herein, enlarging the diverse representation landscape of East Asian and South populations. About Europe, our study represented populations from the East, West, North and South of the continent.

Of note is the behaviour of the distribution of rs1495741 in the Americas and African populations, which presented different patterns (e.g., Brazil and Mali, respectively) to those described by 1KGP. In addition, the Middle East populations were represented for the first time.
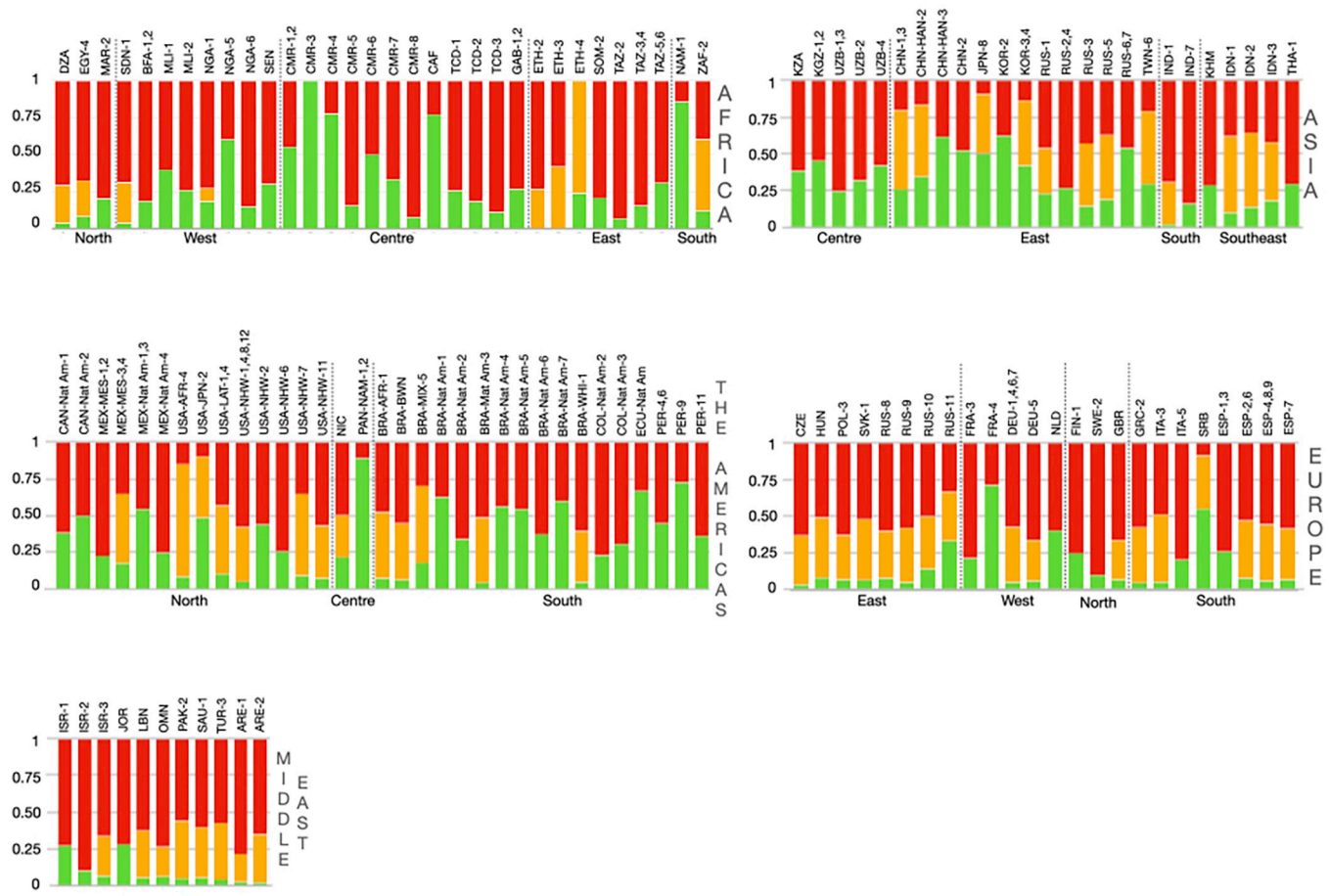
Fig 7. Bar plots with the distributions of the slow, fast and intermedia phenotypes by geographic region. Note: ARE, United Arab Emirates; BFA, Burkina Faso; BRA, Brazil; CAF, Central African Republic; CAN, Canada; CHN, China; CMR, Cameroon; COL, Colombia; Czech Republic; DZA, Algeria; DEU, Germany; EGY, Egypt; ESP, Spain; ETH, Ethiopia; FIN, Finland; FRA, France; GAB, Gabon; GBR; United Kingdom; GRC, Greece; HUN, Hungary; IDN, Indonesia; IND, India; ISR, Israel; ITA, Italy; JOR, Jordan; JPN, Japan; KGZ, Kyrgyzstan; KHM, Cambodia; KOR, Korea; KZA, Kazakhstan; LBN, Lebanon; MAR, Morocco; MEX, Mexico; MLI, Mali; NAM, Namibia; NGA, Nigeria; NIC, Nicaragua; NLD, Netherlands; OMN, Oman; PAK, Pakistan; PAN, Panama; PER, Peru; POL, Poland; RUS, Russian Federation; SAU, Saudi Arabia; SEN, Senegal; SDN, Sudan; SOM, Somalia; SRB, Serbia; SVK, Slovakia; SWE, Sweden; TCD, Chad; THA, Thailand; TUR, Turkey; TWN, Taiwan; TZA, Tanzania; USA, The United State of America; UZB, Uzbekistan; ZAF, South Africa. LAT, Latinos; Nat Am, Native Americans NHW, mom-Hispanic Whites representing the European-derived populations from Canada and the United States of America.

https://doi.org/10.1371/journal.pone.0283726.g007

Discussion

*NAT*2 polymorphisms have shown variations in the allelic distributions across populations and at inter-ethnic and inter-individual levels. Although, several studies have described the gene variability of *NAT*2, these have been limited to populational descriptive data, leaving gaps in the knowledge of its genetic architecture. The present systematic review explored, in detail, the global *NAT*2 diversity patterns from 304 populations, including 164 articles from populational-descriptive and observational studies.

Particularities of the studies included

Akin to other documents related to several pathologies and pharmacogenetic studies, Europeans (including European-derived populations from Canada and the USA) have been the most characterised, mainly to avoid spurious results given its genetic homogeneity [22–26]. Such

homogeneity was demonstrated in the distribution of both the slow and fast phenotypes in the different regions of Europe without significant differences among them (S8 and S14 Figs).

East Asia, particularly the Han Chinese populations was the second most studied region. This ethnic group has also been considered homogeneous, given its age (traced back to the Neolithic) and the gene-flow with surrounding populations [27,28]. The remarkable number of studies in East Asia could be associated with the earliest stages of sedentism and plant cultivation found in northern China, the second oldest domestication centre in Eurasia [29]. In this region, several agricultural systems (e.g., millets and rice) emerged independently and gradually increased in the Hexi Corridor and the Yellow River Basin [29,30]. Several human groups interacted during this period, and with the human expansions, the cultures spread southward and northwards (almost simultaneously) as well as central China [29,31,32]. The genetic background of the Han Chinese population has been associated with different subsistence strategies such as hunter-gatherers (Mongolia and Amur River Basin), farmers (from Yellow River Basin) and pastoralists (from western Mongolia), all around 3000 years Before Christ (BC) [33]. Thus, East Asians may derive from different mixture proportions, which makes their study remarkably interesting regarding the acetylator statuses, but also particularly complex. Han Chinese populations depicted a genetic cline where farmers from the Upper and Middle Yellow River share a gene pool with the north Han group [33]. The Yellow River Valley connected to China and Southeast Asia, and in turn, Han Chinese also shared a gene pool with Southeast Asians, probably, through southern Chinese agriculturalists [33]. These particularities could explain several phenotype patterns described in Han populations.

Other East Asian populations (e.g., Japan, Korea and Taiwan) and its counterparts in Western countries (i.e., Asian-Americans) were also represented. Nevertheless, the findings in these last populations might lead to flagrant under- or overestimation of their diversity, given the remarkable inbreeding rates regarding the outbred populations from which they arise [34,35]. Similar findings have been described in the equivalent of South Asian populations (i.e., India and Pakistan) [35,36]. In turn, the reported data regarding Asian Americans should be interpreted with prudence because their patterns could be closely related to the genetic architecture of each population, limiting its applicability.

Historically ethnicities such as Africans, Latinos and Native Americans' populations have been understudied. Nonetheless, our findings depicted an increase in the number of studies involving them in recent years [37–39]. Particularly, the subsistence mode and the diversity in climatic zones and biomes have been proposed as keystones in the evolution of NAT2, exerting a positive selection [9]. Thus, studies on African ethnic groups have been interested in elucidating the adaptation signals related to diet and lifestyle practices [40–43]. African populations are ethnically and genetically diverse, forging a cornerstone for answering such questions [42]. Thus, it was not wholly surprising that the most extraordinary NAT2 diversity has been found in Africa.

Latinos are a cradle of diversity; their genetic background was shaped by the fusion amongst European and East Asian migratory waves peopling the New World and African populations [44,45]. Thus, the Americas are a melting pot of diversity that emerged 500 years ago (ya). Both European and African migrants came from several regions. The European migrants that colonised the Americas came from England, France, Portugal and Spain. Regarding the Iberian peninsula's geographic position, it favoured the trade interchange with circum-Mediterranean cultures (i.e., Greeks, Phoenicians and Carthaginians) besides the colonisation by the Romans and the Muslim domination [46]. English, French and the Netherlands migrations, as pirates and corsair, also invaded the Caribbean region, which also concentrated the enslaved Africans from Angola, Congo, Gambia, Ghana, Guineas, Mozambique, and Senegal, amongst others [46–49].

Furthermore, the Jewish and Muslim Diaspora in Latin America as *Conversos* lead to genetic diversification [46,50]. Native Americans traced their origins back to a host of regions from Eurasia, explaining their numerous diversities in languages and lineages [49,51]. A remarkable intrapopulation diversity has been observed in the Native Americans, supporting the results found in this review [37,52]. Likewise, the genetic wealth of contemporary Native Americans, in conjunction with the several admixture degrees in their non-Native American populations, makes the Americas an excellent candidate for pharmacogenetic studies [39,49,53,54].

Hence, Africans and Americans were the most diverse populations regarding *NAT*2 diversity, reinforcing the relevance of including ethnic minorities in studies on diversity. The presence of new variants in these two ethnic backgrounds depicted their diversities and possibly a recent growth within their demographic history. Nonetheless, such variants could mirror the ability of modern technologies (e.g., sequence of the whole gene and array-based genotyping platforms) compared to genotyping with a limited number of SNPs. The findings from prior studies do not rule out the diversity of the populations, which could be skewed by the number of SNPs selected (i.e., not-informative markers). In turn, the number of singletons described and the diversity within haplotypes should be taken with caution.

Concerning those similar patterns described in Native Americans and East Asians, there seem to be related to the Americas peopling, and the bottlenecks and genetic drive underwent for the first settler populations [49,55]. Nonetheless, the Asian and Native American populations included herein did not represent the full diversity because most belonged to ethnic and religious groups besides the small sample sizes described. Similar arguments also might explain the diverse patterns found in Asia, where the least number of haplotypes were described. As mentioned earlier, the diversity depends on sample sizes and the genetic drift effect in these small and endogamic populations, the distinct ethnic backgrounds and the resolution power of the technologies employed [9].

Despite the reduced number of studies, the remarkable diversity of the Middle East was remarkable. Akin to Africa, this feature could be related to lifestyle practices. The transition from hunter-gatherers to an agricultural lifestyle has been associated with the domestication of wild cereals and plants [56]. Wheat, barley, among other cereals were domesticated in the Fertile Crescent during the Neolithic era (circa ~8,000–10,000 ya) [57–60]. This region spans the current countries of Iraq, Syria, Lebanon, Israel, Palestine, Jordan, Kuwait (northern region), Iran (western region), and Turkey (southern region). Early evidence of cultivation and domestication have been reported in Syria (11,150–10,450 BC), Jordan valley (9,700–8800 BC), South-eastern Turkey, the Upper Euphrates valley (Abu Hureyra, Syria), and Jericho (Israel) [59,60]. Besides, the Levant was one of the earliest regions where agriculture and animal domestication emerged [61]. Archaeological and genetic studies have reinforced that the north of Iraq (8,000–11,000 years Before Present) was the core of sheep's initial domestication [62]. Both plant cultivation and animal domestication are cornerstones of human societies' modification because these were learned from one region to another, favouring the genetic exchange, which could explain the diversity patterns [59].

SNP diversity patterns

As in other studies, SNPs such as rs1801279 and rs1799931 depicted particular distributions, especially in Africa and the Americas, respectively [9]. The other SNPs' distribution patterns were similar to those previously reported by other research groups [9]. Yet, the peculiar pattern of rs1495741 distribution in Brazilian populations might be a reflection of the demographic events in the peopling of this young population (~ 500 years) and the subsequent bottleneck events [46].

Slow acetylators explanations

NAT genes (*NAT*1 and *NAT*2), as others (i.e., *FADS*1), have been targets for natural selection [63]. Several studies have explained the prevalence of the slow acetylator status through selective pressures exerted by environment and lifestyle. Briefly, the transition from the fast acetylators to slow ones has been associated with the emergence of agriculture and pastoralism, replacing the hunting-gathering subsistence mode [40,63,64]. The shift in lifestyles involved the introduction of new foods to the diet with different nutrients and fats, as well as exposures to new pathogen [63]. These changes, in turn, could be implicated in the participation of metabolic pathway genes and those polymorphisms related to slow haplotypes, conferring certain advantages [63,65,66]. Finally, these genetic variants were fixed in the populations, increasing their frequency to favour adaptation. Hence, the slow acetylator phenotype has been more frequent within food-producing populations from Central and Southern Asia, North and Central Africa, Europe, and the Middle East [65–67]. By contrast, in hunter-gatherer populations, the rapid acetylator phenotype has been the most frequent [41,65]. The heterogeneous phenotypes in Africa and the Americas, which could be explained by their intrinsic diversity and the remarkable differences among and within populations [9].

Fast acetylators explanations

Although slow acetylator status is the most frequent worldwide, fast acetylators have remained mainly in EAS and Native American populations. These results support the findings of former studies [41,68]. The hunter-gatherer lifestyles in ancient populations have been documented in Eastern Asians [63]. Regarding the Native Americans, it is likely that they have maintained the lifestyles of their ancestors. Cultural diffusion (i.e., subsistence practices) also shapes gene frequency patterns [69]. Studies have reinforced the ancestral connection between Native Americans and the East Asian populations [49,70–72]. Given the fossil evidence described in this region (i.e., bison, horse, and mammoth, among others), ancient North Americans' diet has been associated with the hunter-gatherer subsistence mode [73–75]. Other processes could explain the genetic architecture. Demographic events such as bottlenecks sustained prior to or during the colonisation of the Americas could also involve the frequency of fast acetylators distribution in these populations [76]. Instead, the effect of genetic drift is most substantial in small populations. At the same time, the number of migrants during the peopling of the Americas is controversial; the effective population size could have been small [49,77]. These migratory waves could be carriers of heritable traits "fixed" amongst populations or sub-populations under selective pressures, transmitting the "modified" haplotypes to other geographic regions [78]. Notably, the folate-rich diet and green leafy vegetables have also been associated with this phenotype [41,66,79,80]. Fish and soy, both folate-rich sources, are key ingredients of East Asian cuisine [81–83].

These adaptations could also have arisen independently in other geographic regions. Thus, one possibility that explains the subtle difference patterns in the frequency of the fast phenotype in Africa could be that the north region has mainly been occupied by hunter-gatherer populations of at least 5,000 ya [84]. Amongst these, the ethnic groups from Cameroon, Gabon, and Namibia depicted a significant frequency of fast acetylators. A comparable argument might explain Europe's highest proportions of the fast phenotype in Europe. Hunter-gatherer ancient populations have been established in Central, North, and South European regions [63]. By contrast, the East and West European regions have been associated chiefly with agricultural practices and, in turn, with carriers of slow phenotypes [63]. Nonetheless, the frequency patterns in this geographic region did not necessarily correspond to the agriculture diffusion, suggesting that the different gene flow levels could have influenced the frequency

distributions in Eurasian populations [85]. As mentioned before, demographic events play a critical role in gene frequency variations. The out-of-Africa and the Arabian aridification by climate change are two bottlenecks that could have impacted the diversity patterns of the Middle East and Eurasian populations [86,87].

Because not all natural xenobiotics were related to Nat2, other xenobiotic biotransformation genes could have been affected by the selective pressures exerted by environment and lifestyle [88,89]. Such selection effect could bring about behavioural adjustments in physiological and biochemical pathways as well as in the gut microbiome [63,90–92]. Complex biological pathways regulate metabolisms; in turn, hundreds of genes are likely involved in this physiological process. Nat2 is expressed in the intestines and liver; thus, possible coevolution would entail. Consequently, it is unlikely that the adaptation proceeding acted on single genes but rather, was a polygenic selection process that could also shape the *NAT*2 frequency patterns. Genes related to diet and metabolisms have been persistent in the models of polygenic selection [63,75].

A similar argument might explain the dissimilarities between the frequencies of *NAT*2*5B in worldwide populations regarding some areas of Asia, where its haplotype is less frequent, replicating the findings of prior studies [41]. On the one hand, the rs1801280-C allele, encoding for the altered slow phenotype, is more frequent in Central and Western Eurasians (range 0.287 to 0.500) than in East Asians (range 0.037 to 0.269). Such differences could result from a selection process, given that the rs1801280-C allele has increased its frequency significantly in Eurasians [65,89]. Again, the emergence of agriculture could be the selective pressure to the shift from the ancestral state rs1801280-T to the derivative one [65]. By contrast, the low frequency of the *NAT*2*5 haplotype in East Asians could be related to its liking for the hunter-gatherer lifestyle and other aspects mentioned before [63,93–96]. The *NAT*2*5 haplotype has shown an association with *NAT*1*4 in western and central Eurasians by 80% [65]. This association is twice and four times more than those found in Eastern Eurasians and sub-Saharan populations, respectively [65,89]. The effective metabolisms of environmental xenobiotics should require the collective action of phase I and II enzymes. Since these two genes are located on the same chromosome coevolution, it should not be unlikely [89]. Hence, the *NAT*2*5B selective advantage could affect the evolution of *NAT*1 as other genes [65]. Other studies have suggested that *NAT*1 and *NAT*2 could evolve under distinct selective regimes. These two genes have a physical distance fairly close to 200 kilobases, exhibiting linkage equilibrium among them [9].

In addition to 341 C > T, another three sets of polymorphisms (191 G>A, 590 G>A, and 857 G>A) are encoded for the slow acetylator state. Of these, 857G>A have depicted more frequency in Asians than Europeans [89]. The hunter-gatherer subsistence mode could also explain these differences, which acquires food from their surrounding environment. However, it did not discard the effect of demographic events. Contemporary populations maintaining this subsistence mode have shown a correlation between population density and local primary production [97]. While contemporary populations are not analogues to the ancient ones, the climate conditions to which ancient populations were exposed could use up the aliments from the environment with the subsequent bottleneck. Similar to the out-of-Africa model, the famine could have favoured migrations like those peopling the Americas.

Nevertheless, in addition to the diet, the diversity patterns could also reflect the environmental xenobiotic insults, the epigenetic regulation, the history and specific pressures of the populations, and climate features, among others [9].

Among the strengths of the present systematic review is the detailed landscape depicting the diversity of *NAT*2 from 35,561 genotypes, 51,860 haplotypes, and 70,484 phenotypes. These data portrayed the eight most studied SNPs, and for the first time, the *NAT*2 diversity of the Middle East populations, which has not been reported in any former studies. Likewise, the present diversity panorama discriminates between the most prominent ancestries in the

Americas: Latinos, Native Americans, and non-Hispanic whites. These features reflect the diversity among populations and individuals and could be a cornerstone for having a possible scenario regarding other ethnicities [11]. Diversity within *NAT*2 has been related to the developing drug side effects such as hepatotoxicity, peripheral neuropathy, lupus, and susceptibility to some kinds of cancer [1,98,99]. It is also necessary to highlight some of the study's shortcomings, including that not all studies included information on the diversity patterns of the eight polymorphisms and data from the specific haplotypes.

Conclusion

The global diversity that occurred in ancestry and demographic events begs an understanding of the variation within genes of tremendous importance, such as *NAT*2. The present study provided the most up-to-date overview of the *NAT*2 diversity to allele, genotype, haplotype, and acetylator status with implications in pharmacogenetics and certain complex disease susceptibility. The study of this set of approaches could further illuminate its value and usefulness in personalised and precision medicine. Nonetheless, further studies are needed to unravel such diversity in ethnic minorities besides correlating the worldwide population diversity with pharmacodynamics and pharmacokinetics strategies.

Supporting information

S1 Fig. Frequency of the ancestral and derivative allele of rs1801279, rs1801280, rs1799929, rs1799930, rs1799931, rs1041983, and rs1208 in the African populations. Note: A, Adenine; C, Cytosine, G, Guanine; T, Thymine.
(TIFF)

S2 Fig. Frequency of the ancestral and derivative allele of rs1801279, rs1801280, rs1799929, rs1799930, rs1799931, rs1041983, and rs1208 in the Americas populations. Note: A, Adenine; C, Cytosine, G, Guanine; T, Thymine.
(TIFF)

S3 Fig. Frequency of the ancestral and derivative allele of rs1801279, rs1801280, rs1799929, rs1799930, rs1799931, rs1041983, and rs1208 in the Asian populations. Note: A, Adenine; C, Cytosine, G, Guanine; T, Thymine.
(TIFF)

S4 Fig. Frequency of the ancestral and derivative allele of rs1801279, rs1801280, rs1799929, rs1799930, rs1799931, rs1041983, and rs1208 in the European populations. Note: A, Adenine; C, Cytosine, G, Guanine; T, Thymine.
(TIFF)

S5 Fig. Frequency of the ancestral and derivative allele of rs1801279, rs1801280, rs1799929, rs1799930, rs1799931, rs1041983, and rs1208 in the Middle East populations. Note: A, Adenine; C, Cytosine, G, Guanine; T, Thymine.
(TIFF)

S6 Fig. Frequency of the most common haplotypes among the geographic regions.
(TIF)

S7 Fig. Box plots of the slow phenotype distribution among the Africa regions. Note: CAf, Central Africa; EAf, East Africa; NAf, North Africa; SAf, South Africa; Waf, West Africa. CMR, Cameroon; ETH, Ethiopia; NAM, Namibia; NGA, Nigeria; TZA, Tanzania; ZAF, South

Africa.
(TIF)

S8 Fig. Box plots of the slow phenotype distribution among the European regions. Note: EEu, East Europe; Neu, North Europe; SEu, South Europe; WEu, West Europe. CZE, Czech Republic; FIN, Finland; FRA, France; GRC, Greece; ITA, Italy; NLD, the Netherlands; RUS, the Russian Federation; SRB, Serbia; SWE, Sweden.
(TIF)

S9 Fig. Box plots of the slow phenotype distribution among the Americas regions. Note: CAf, Central America; Nam, North America; Sam, South America. BRA, Brazil; COL, Colombia; ECU, Ecuador; MEX, Mexico; NIC, Nicaragua; PAN, Panama; USA, the United States of America.
(TIF)

S10 Fig. Box plots of the slow phenotype distribution among the Asia regions. Note: Cas. Central Asia; EAs, East Asia; SEAs, Southeast Asia; SAs, South Asia. JPN, Japan; KGZ, Kirghizstan; RUS, the Russian Federation; UZB, Uzbekistan.
(TIF)

S11 Fig. Box plots of the fast phenotype distribution among the Asia regions. Note: Cas. Central Asia; EAs, East Asia; SEAs, Southeast Asia; SAs, South Asia. JPN, Japan; KGZ, Kirghizstan; RUS, the Russian Federation; UZB, Uzbekistan.
(TIF)

S12 Fig. Box plots of the fast phenotype distribution among the Americas regions. Note: CAf, Central America; Nam, North America; Sam, South America. BRA, Brazil; COL, Colombia; ECU, Ecuador; NIC, Nicaragua; PAN, Panama; USA, the United States of America.
(TIF)

S13 Fig. Box plots of the fast phenotype distribution among the Africa regions. Note: CAf, Central Africa; EAf, East Africa; NAf, North Africa; SAf, South Africa; Waf, West Africa. CMR, Cameroon; ETH, Ethiopia; NAM, Namibia; NGA, Nigeria; TZA, Tanzania; ZAF, South Africa.
(TIF)

S14 Fig. Box plots of the fast phenotype distribution among the European regions. Note: EEu, East Europe; Neu, North Europe; SEu, South Europe; WEu, West Europe. CZE, Czech Republic; DEU, Germany; FRA, France; ITA, Italy; NLD, the Netherlands; RUS, the Russian Federation; SRB, Serbia.
(TIF)

S1 Table. Data extraction of allele, genotype, haplotypes, and acetylator status from all articles included in this systematic review.
(XLSX)

S2 Table. Single nucleotide polymorphisms with nucleotide changes and phenotype of all haplotypes found in the present systematic review.
(XLSX)

Acknowledgments

disinterested help in the proofreading. The authors would like to thank the anonymous reviewers; their suggestions remarkably increased the quality of our work.

Author Contributions

Conceptualization: Jorge E. Gutiérrez-Virgen, Maricela Piña-Pozas, Rocío Gómez.

Data curation: Jorge E. Gutiérrez-Virgen, Esther A. Hernández-Tobías, Marco A. Meraz-Ríos, Rocío Gómez.

Formal analysis: Jorge E. Gutiérrez-Virgen, Esther A. Hernández-Tobías, Rocío Gómez.

Investigation: Jorge E. Gutiérrez-Virgen, Maricela Piña-Pozas, Lucia Taja-Chayeb, Ma. de Lourdes López-González, Marco A. Meraz-Ríos, Rocío Gómez.

Methodology: Jorge E. Gutiérrez-Virgen, Maricela Piña-Pozas, Esther A. Hernández-Tobías, Lucia Taja-Chayeb, Ma. de Lourdes López-González, Marco A. Meraz-Ríos, Rocío Gómez.

Project administration: Rocío Gómez.

Resources: Maricela Piña-Pozas, Rocío Gómez.

Supervision: Lucia Taja-Chayeb, Marco A. Meraz-Ríos, Rocío Gómez.

Validation: Rocío Gómez.

Visualization: Rocío Gómez.

Writing – original draft: Jorge E. Gutiérrez-Virgen, Rocío Gómez.

Writing – review & editing: Rocío Gómez.

References

1. Hein DW, Millner LM. Arylamine N-acetyltransferase acetylation polymorphisms: paradigm for pharmacogenomic-guided therapy- a focused review. Expert Opin Drug Metab Toxicol. 2021; 17(1):9–21. https://doi.org/10.1080/17425255.2021.1840551 PMID: 33094670

2. Brian ZH, Songren W, David B, Lynne RW, Lang W, William JB, et al. Red meat consumption, cooking mutagens, NAT1/2 genotypes and pancreatic cancer risk in two ethnically diverse prospective cohorts. International journal of cancer. 2021. https://doi.org/10.1002/ijc.33598

3. Berrandou T, Mulot C, Cordina-Duverger E, Arveux P, Laurent-Puig P, Truong T, et al. Association of breast cancer risk with polymorphisms in genes involved in the metabolism of xenobiotics and interaction with tobacco smoking: A gene-set analysis. Int J Cancer. 2019; 144(8):1896–908. https://doi.org/10.1002/ijc.31917 PMID: 30303517

4. Knowles JW, Xie W, Zhang Z, Chennamsetty I, Assimes TL, Paananen J, et al. Identification and validation of N-acetyltransferase 2 as an insulin sensitivity gene. J Clin Invest. 2016; 126(1):403. https://doi.org/10.1172/JCI85921 PMID: 26727231

5. Salazar-Gonzalez RA, Turijan-Espinoza E, Hein DW, Milan-Segovia RC, Uresti-Rivera EE, Portales-Perez DP. Expression and genotype-dependent catalytic activity of N-acetyltransferase 2 (NAT2) in human peripheral blood mononuclear cells and its modulation by Sirtuin 1. Biochem Pharmacol. 2018; 156:340–7. https://doi.org/10.1016/j.bcp.2018.08.034 PMID: 30149019

6. Laurieri NAS, Edith. Arylamine N-Acetyltransferases in Health and Disease: World Scientific; 2018.

7. Anitha A, Banerjee M. Arylamine N-acetyltransferase 2 polymorphism in the ethnic populations of South India. Int J Mol Med. 2003; 11(1):125–31. PMID: 12469231

8. Sekine A, Saito S, Iida A, Mitsunobu Y, Higuchi S, Harigae S, et al. Identification of single-nucleotide polymorphisms (SNPs) of human N-acetyltransferase genes NAT1, NAT2, AANAT, ARD1 and L1CAM in the Japanese population. J Hum Genet. 2001; 46(6):314–9. https://doi.org/10.1007/s100380170065 PMID: 11393533

9. Sabbagh A, Darlu P, Vangenot C, Poloni ES. Arylamine N-Acetyltransferases in Anthropology. Arylamine N-Acetyltransferases in Health and Disease: World Scientific; 2018. p. 165–93.

10. Trumble BC, Finch CE. The Exposome in Human Evolution: From Dust to Diesel. Q Rev Biol. 2019; 94 (4):333–94. https://doi.org/10.1086/706768 PMID: 32269391

11. Yang HC, Chen CW, Lin YT, Chu SK. Genetic ancestry plays a central role in population pharmacogenomics. Commun Biol. 2021; 4(1):171. https://doi.org/10.1038/s42003-021-01681-6 PMID: 33547344

12. Johnston FH, Melody S, Bowman DM. The pyrohealth transition: how combustion emissions have shaped health through human history. Philos Trans R Soc Lond B Biol Sci. 2016; 371(1696). https://doi.org/10.1098/rstb.2015.0173 PMID: 27216506

13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. J Clin Epidemiol. 2021; 134:178–89. https://doi.org/10.1016/j.jclinepi.2021.03.001 PMID: 33789819

14. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ, Group G. Strengthening the reporting of Genetic RIsk Prediction Studies: the GRIPS Statement. PLoS Med. 2011; 8(3):e1000420. https://doi.org/10.1371/journal.pmed.1000420 PMID: 21423587

15. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med. 2009; 6 (2):e22. https://doi.org/10.1371/journal.pmed.1000022 PMID: 19192942

16. García-Closas M, Hein DW, Silverman D, Malats N, Yeager M, Jacobs K, et al. A single nucleotide polymorphism tags variation in the arylamine N-acetyltransferase 2 phenotype in populations of European background. Pharmacogenet Genomics. 2011; 21(4):231–6. https://doi.org/10.1097/FPC.0b013e32833e1b54 PMID: 20739907

17. Sohani ZN, Meyre D, de Souza RJ, Joseph PG, Gandhi M, Dennis BB, et al. Assessing the quality of published genetic association studies in meta-analyses: the quality of genetic studies (Q-Genie) tool. BMC Genet. 2015; 16:50. https://doi.org/10.1186/s12863-015-0211-2 PMID: 25975208

18. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet. 2007; 370(9596):1453–7. https://doi.org/10.1016/s0140-6736(07)61602-x

19. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 1 ed: Springer International Publishing; 2016. 213 p.

20. MedCalc Statistical Software version 20.11 (MedCalc Software Ltd, Ostend, Belgium https://www.medcalc.org; 2022).

21. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010; 10(3):564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x PMID: 21565059

22. Ramos E, Doumatey A, Elkahloun AG, Shriner D, Huang H, Chen G, et al. Pharmacogenomics, ancestry and clinical decision making for global populations. Pharmacogenomics J. 2014; 14(3):217–22. https://doi.org/10.1038/tpj.2013.24 PMID: 23835662

23. Guerrero S, Lopez-Cortes A, Indacochea A, Garcia-Cardenas JM, Zambrano AK, Cabrera-Andrade A, et al. Analysis of Racial/Ethnic Representation in Select Basic and Applied Cancer Research Studies. Sci Rep. 2018; 8(1):13978. https://doi.org/10.1038/s41598-018-32264-x PMID: 30228363

24. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature. 2021; 600(7890):675–9. https://doi.org/10.1038/s41586-021-04064-3 PMID: 34887591

25. Hindorff LA, Bonham VL, Brody LC, Ginoza MEC, Hutter CM, Manolio TA, et al. Prioritizing diversity in human genomics research. Nat Rev Genet. 2018; 19(3):175–85. https://doi.org/10.1038/nrg.2017.89 PMID: 29151588

26. Claudio-Campos K, Duconge J, Cadilla CL, Ruano G. Pharmacogenetics of drug-metabolizing enzymes in US Hispanics. Drug Metab Pers Ther. 2015; 30(2):87–105. https://doi.org/10.1515/dmdi-2014-0023 PMID: 25431893

27. Yu X, Li H. Origin of ethnic groups, linguistic families, and civilizations in China viewed from the Y chromosome. Mol Genet Genomics. 2021; 296(4):783–97. https://doi.org/10.1007/s00438-021-01794-x PMID: 34037863

28. Yang X, Wang XX, He G, Guo J, Zhao J, Sun J, et al. Genomic insight into the population history of Central Han Chinese. Ann Hum Biol. 2021; 48(1):49–55. https://doi.org/10.1080/03014460.2020.1851396 PMID: 33191788

29. Shelach-Lavi G, Teng M, Goldsmith Y, Wachtel I, Stevens CJ, Marder O, et al. Sedentism and plant cultivation in northeast China emerged during affluent conditions. PLoS One. 2019; 14(7):e0218751. https://doi.org/10.1371/journal.pone.0218751 PMID: 31318871

30. Zhou X, Yu J, Spengler RN, Shen H, Zhao K, Ge J, et al. 5,200-year-old cereal grains from the eastern Altai Mountains redate the trans-Eurasian crop exchange. Nat Plants. 2020; 6(2):78–87. https://doi.org/10.1038/s41477-019-0581-y PMID: 32055044

31. Yi B, Liu X, Yan X, Zhou Z, Chen J, Yuan H, et al. Dietary shifts and diversities of individual life histories reveal cultural dynamics and interplay of millets and rice in the Chengdu Plain, China during the Late Neolithic (2500–2000 cal. BC). Am J Phys Anthropol. 2021; 175(4):762–76. https://doi.org/10.1002/ajpa.24259

32. Long T, Leipe C, Jin G, Wagner M, Guo R, Schröder O, et al. The early history of wheat in China from (14)C dating and Bayesian chronological modelling. Nat Plants. 2018; 4(5):272–9. https://doi.org/10.1038/s41477-018-0141-x PMID: 29725102

33. Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, et al. Genomic insights into the formation of human populations in East Asia. Nature. 2021; 591(7850):413–9. https://doi.org/10.1038/s41586-021-03336-2 PMID: 33618348

34. Sengupta D, Choudhury A, Basu A, Ramsay M. Population Stratification and Underrepresentation of Indian Subcontinent Genetic Diversity in the 1000 Genomes Project Dataset. Genome Biol Evol. 2016; 8(11):3460–70. https://doi.org/10.1093/gbe/evw244

35. Narasimhan VM, Hunt KA, Mason D, Baker CL, Karczewski KJ, Barnes MR, et al. Health and population effects of rare gene knockouts in adult humans with related parents. Science. 2016; 352 (6284):474–7. https://doi.org/10.1126/science.aac8624 PMID: 26940866

36. Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, Khetarpal SA, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature. 2017; 544(7649):235–9. https://doi.org/10.1038/nature22034 PMID: 28406212

37. Fuselli S, Gilman RH, Chanock SJ, Bonatto SL, De Stefano G, Evans CA, et al. Analysis of nucleotide diversity of NAT2 coding region reveals homogeneity across Native American populations and high intra-population diversity. Pharmacogenomics J. 2007; 7(2):144–52. https://doi.org/10.1038/sj.tpj.6500407 PMID: 16847467

38. Rodrigues-Soares F, Penas-Lledo EM, Tarazona-Santos E, Sosa-Macias M, Teran E, Lopez-Lopez M, et al. Genomic Ancestry, CYP2D6, CYP2C9, and CYP2C19 Among Latin Americans. Clin Pharmacol Ther. 2020; 107(1):257–68. https://doi.org/10.1002/cpt.1598 PMID: 31376146

39. Gonzalez-Covarrubias V, Morales-Franco M, Cruz-Correa OF, Martinez-Hernandez A, Garcia-Ortiz H, Barajas-Olmos F, et al. Variation in Actionable Pharmacogenetic Markers in Natives and Mestizos From Mexico. Front Pharmacol. 2019; 10:1169. https://doi.org/10.3389/fphar.2019.01169 PMID: 31649539

40. Luca F, Perry GH, Di Rienzo A. Evolutionary adaptations to dietary changes. Annu Rev Nutr. 2010; 30:291–314. https://doi.org/10.1146/annurev-nutr-080508-141048 PMID: 20420525

41. Sabbagh A, Darlu P, Crouau-Roy B, Poloni ES. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. PLoS One. 2011; 6(4):e18507. https://doi.org/10.1371/journal.pone.0018507 PMID: 21494681

42. Mortensen HM, Froment A, Lema G, Bodo JM, Ibrahim M, Nyambo TB, et al. Characterization of genetic variation and natural selection at the arylamine N-acetyltransferase genes in global human populations. Pharmacogenomics. 2011; 12(11):1545–58. https://doi.org/10.2217/pgs.11.88 PMID: 21995608

43. Valente C, Alvarez L, Marks SJ, Lopez-Parra AM, Parson W, Oosthuizen O, et al. Exploring the relationship between lifestyles, diets and genetic adaptations in humans. BMC Genet. 2015; 16:55. https://doi.org/10.1186/s12863-015-0212-1 PMID: 26018448

44. Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, et al. The Great Migration and African-American Genomic Diversity. PLoS Genet. 2016; 12(5):e1006059. https://doi.org/10.1371/journal.pgen.1006059 PMID: 27232753

45. Micheletti SJ, Bryc K, Ancona Esselmann SG, Freyman WA, Moreno ME, Poznik GD, et al. Genetic Consequences of the Transatlantic Slave Trade in the Americas. Am J Hum Genet. 2020; 107(2):265–77. https://doi.org/10.1016/j.ajhg.2020.06.012 PMID: 32707084

46. Gómez R, Schurr T, Meraz-Rios M. Diversity of Mexican Paternal Lineages Reflects Evidence of Migration and 500 Years of Admixture. Human Migration: Biocultural Perspectives Oxford Academic; 2021. p. 139–52.

47. Lasso M. Race War and Nation in Caribbean Gran Colombia, Cartagena, 1810–1832. The American Historical Review. 2006; 111(2):336–61. https://doi.org/10.1086/ahr.111.2.336

48. Eltis D. The Rise of African Slavery in the Americas. Cambridge: Cambridge University Press; 1999.

49. Gomez R, Vilar MG, Meraz-Rios MA, Veliz D, Zuniga G, Hernandez-Tobias EA, et al. Y chromosome diversity in Aztlan descendants and its implications for the history of Central Mexico. iScience. 2021; 24 (5):102487. https://doi.org/10.1016/j.isci.2021.102487 PMID: 34036249

50. Velez C, Palamara PF, Guevara-Aguirre J, Hao L, Karafet T, Guevara-Aguirre M, et al. The impact of Converso Jews on the genomes of modern Latin Americans. Hum Genet. 2012; 131(2):251–63. https://doi.org/10.1007/s00439-011-1072-z PMID: 21789512

51. Sandoval K, Moreno-Estrada A, Mendizabal I, Underhill PA, Lopez-Valenzuela M, Peñaloza-Espinosa R, et al. Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas. Am J Phys Anthropol. 2012; 148(3):395–405. https://doi.org/10.1002/ajpa.22062 PMID: 22576278

52. Bisso-Machado R, Ramallo V, Paixao-Cortes VR, Acuna-Alonzo V, Demarchi DA, Sandoval JR, et al. NAT2 gene diversity and its evolutionary trajectory in the Americas. Pharmacogenomics J. 2016; 16 (6):559–65. https://doi.org/10.1038/tpj.2015.72 PMID: 26503810

53. Kengne AP, Nakamura K, Barzi F, Lam TH, Huxley R, Gu D, et al. Smoking, diabetes and cardiovascular diseases in men in the Asia Pacific region. J Diabetes. 2009; 1(3):173–81. https://doi.org/10.1111/j.1753-0407.2009.00028.x PMID: 20923536

54. Martinez-Magana JJ, Genis-Mendoza AD, Villatoro Velazquez JA, Camarena B, Martin Del Campo Sanchez R, Fleiz Bautista C, et al. The Identification of Admixture Patterns Could Refine Pharmacogenetic Counseling: Analysis of a Population-Based Sample in Mexico. Front Pharmacol. 2020; 11:324. https://doi.org/10.3389/fphar.2020.00324 PMID: 32390825

55. Fagundes NJR, Tagliani-Ribeiro A, Rubicz R, Tarskaia L, Crawford MH, Salzano FM, et al. How strong was the bottleneck associated to the peopling of the Americas? New insights from multilocus sequence data. Genet Mol Biol. 2018; 41(1 suppl 1):206–14. https://doi.org/10.1590/1678-4685-GMB-2017-0087 PMID: 29668018

56. Jones H, Leigh FJ, Mackay I, Bower MA, Smith LM, Charles MP, et al. Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. Mol Biol Evol. 2008; 25(10):2211–9. https://doi.org/10.1093/molbev/msn167 PMID: 18669581

57. Shaaf S, Sharma R, Baloch FS, Badaeva ED, Knüpffer H, Kilian B, et al. The grain Hardness locus characterized in a diverse wheat panel (Triticum aestivum L.) adapted to the central part of the Fertile Crescent: genetic diversity, haplotype structure, and phylogeny. Mol Genet Genomics. 2016; 291(3):1259–75. https://doi.org/10.1007/s00438-016-1180-5 PMID: 26898967

58. Balfourier F, Bouchet S, Robert S, De Oliveira R, Rimbert H, Kitt J, et al. Worldwide phylogeography and history of wheat genetic diversity. Sci Adv. 2019; 5(5):eaav0536. https://doi.org/10.1126/sciadv.aav0536 PMID: 31149630

59. Fuller DQ, Willcox G, Allaby RG. Early agricultural pathways: moving outside the 'core area' hypothesis in Southwest Asia. J Exp Bot. 2012; 63(2):617–33. https://doi.org/10.1093/jxb/err307 PMID: 22058404

60. Morrell PL, Clegg MT. Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. Proc Natl Acad Sci U S A. 2007; 104(9):3289–94. https://doi.org/10.1073/pnas.0611377104 PMID: 17360640

61. Eshed V, Gopher A, Pinhasi R, Hershkovitz I. Paleopathology and the origin of agriculture in the Levant. Am J Phys Anthropol. 2010; 143(1):121–33. https://doi.org/10.1002/ajpa.21301 PMID: 20564538

62. Mustafa SI, Schwarzacher T, Heslop-Harrison JS. Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors. Mitochondrial DNA A DNA Mapp Seq Anal. 2018; 29(8):1180–93. https://doi.org/10.1080/24701394.2018.1431226 PMID: 29385875

63. Colbran LL, Johnson MR, Mathieson I, Capra JA. Tracing the Evolution of Human Gene Regulation and Its Association with Shifts in Environment. Genome Biol Evol. 2021; 13(11). https://doi.org/10.1093/gbe/evab237 PMID: 34718543

64. Gowlett JA. The discovery of fire by humans: a long and convoluted process. Philos Trans R Soc Lond B Biol Sci. 2016; 371(1696). https://doi.org/10.1098/rstb.2015.0164 PMID: 27216521

65. Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, et al. Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. Am J Hum Genet. 2006; 78(3):423–36. https://doi.org/10.1086/500614 PMID: 16416399

66. Luca F, Bubba G, Basile M, Brdicka R, Michalodimitrakis E, Rickards O, et al. Multiple advantageous amino acid variants in the NAT2 gene in human populations. PLoS One. 2008; 3(9):e3136. https://doi.org/10.1371/journal.pone.0003136 PMID: 18773084

67. Magalon H, Patin E, Austerlitz F, Hegay T, Aldashev A, Quintana-Murci L, et al. Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. Eur J Hum Genet. 2008; 16(2):243–51. https://doi.org/10.1038/sj.ejhg.5201963 PMID: 18043717

68. Sabbagh A, Langaney A, Darlu P, Gérard N, Krishnamoorthy R, Poloni ES. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. BMC Genet. 2008; 9:21. https://doi.org/10.1186/1471-2156-9-21 PMID: 18304320

69. MacDonald K, Scherjon F, van Veen E, Vaesen K, Roebroeks W. Middle Pleistocene fire use: The first signal of widespread cultural diffusion in human evolution. Proc Natl Acad Sci U S A. 2021; 118(31). https://doi.org/10.1073/pnas.2101108118 PMID: 34301807

70. Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, et al. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. Am J Hum Genet. 2012; 90(2):229–46. https://doi.org/10.1016/j.ajhg.2011.12.014 PMID: 22281367

71. Raghavan M, Steinrucken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al. POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science. 2015; 349(6250):aab3884. https://doi.org/10.1126/science.aab3884 PMID: 26198033

72. Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 2014; 506(7487):225–9. https://doi.org/10.1038/nature13025 PMID: 24522598

73. Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, et al. Postglacial viability and colonization in North America's ice-free corridor. Nature. 2016; 537(7618):45–9. https://doi.org/10.1038/nature19085 PMID: 27509852

74. Mulligan CJ, Szathmary EJ. The peopling of the Americas and the origin of the Beringian occupation model. Am J Phys Anthropol. 2017; 162(3):403–8. https://doi.org/10.1002/ajpa.23152 PMID: 28101962

75. Hsieh P, Hallmark B, Watkins J, Karafet TM, Osipova LP, Gutenkunst RN, et al. Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations. Mol Biol Evol. 2017; 34(11):2913–26. https://doi.org/10.1093/molbev/msx226 PMID: 28962010

76. Lesnek AJ, Briner JP, Lindqvist C, Baichtal JF, Heaton TH. Deglaciation of the Pacific coastal corridor directly preceded the human colonization of the Americas. Sci Adv. 2018; 4(5):eaar5040. https://doi.org/10.1126/sciadv.aar5040 PMID: 29854947

77. John FH, Scott AE, Dennis HOR, Scott GR, Nancy HB. Beringia and the global dispersal of modern humans. Evolutionary anthropology. 25:64–78. https://doi.org/10.1002/evan.21478 PMID: 27061035

78. Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, et al. Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci U S A. 2010; 107 Suppl 2:8924–30. https://doi.org/10.1073/pnas.0914625107 PMID: 20445095

79. Podgorná E, Diallo I, Vangenot C, Sanchez-Mazas A, Sabbagh A, Černý V, et al. Variation in NAT2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in Africa. BMC Evol Biol. 2015; 15:263. https://doi.org/10.1186/s12862-015-0543-6 PMID: 26620671

80. Aklillu E, Carrillo JA, Makonnen E, Bertilsson L, Djordjevic N. N-Acetyltransferase-2 (NAT2) phenotype is influenced by genotype-environment interaction in Ethiopians. Eur J Clin Pharmacol. 2018; 74(7):903–11. https://doi.org/10.1007/s00228-018-2448-y PMID: 29589062

81. Sonoda T, Nagata Y, Mori M, Miyanaga N, Takashima N, Okumura K, et al. A case-control study of diet and prostate cancer in Japan: possible protective effect of traditional Japanese diet. Cancer Sci. 2004; 95(3):238–42. https://doi.org/10.1111/j.1349-7006.2004.tb02209.x PMID: 15016323

82. Choi MK, Jun YS. Analysis of boron content in frequently consumed foods in Korea. Biol Trace Elem Res. 2008; 126(1–3):13–26. https://doi.org/10.1007/s12011-008-8179-7 PMID: 18665334

83. Mo H, Kariluoto S, Piironen V, Zhu Y, Sanders MG, Vincken JP, et al. Effect of soybean processing on content and bioaccessibility of folate, vitamin B12 and isoflavones in tofu and tempe. Food Chem. 2013; 141(3):2418–25. https://doi.org/10.1016/j.foodchem.2013.05.017 PMID: 23870976

84. Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, et al. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A. 2011; 108(13):5154–62. https://doi.org/10.1073/pnas.1017511108 PMID: 21383195

85. Gopalan S, Berl REW, Myrick JW, Garfield ZH, Reynolds AW, Bafens BK, et al. Hunter-gatherer genomes reveal diverse demographic trajectories during the rise of farming in Eastern Africa. Curr Biol. 2022; 32(8):1852–60 e5. https://doi.org/10.1016/j.cub.2022.02.050 PMID: 35271793

86. Ashraf B, Lawson DJ. Genetic drift from the out-of-Africa bottleneck leads to biased estimation of genetic architecture and selection. Eur J Hum Genet. 2021; 29(10):1549–56. https://doi.org/10.1038/s41431-021-00873-2 PMID: 33846580

87. Almarri MA, Haber M, Lootah RA, Hallast P, Al Turki S, Martin HC, et al. The genomic history of the Middle East. Cell. 2021; 184(18):4612–25 e14. https://doi.org/10.1016/j.cell.2021.07.013 PMID: 34352227

88. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. PLoS Genet. 2014; 10(8): e1004412. https://doi.org/10.1371/journal.pgen.1004412 PMID: 25102153

89. Tiis RP, Osipova LP, Lichman DV, Voronina EN, Filipenko ML. Studying polymorphic variants of the NAT2 gene (NAT2*5 and NAT2*7) in Nenets populations of Northern Siberia. BMC Genet. 2020; 21 (Suppl 1):115. https://doi.org/10.1186/s12863-020-00909-4 PMID: 33092525

90. Harris EE, Meyer D. The molecular signature of selection underlying human adaptations. Am J Phys Anthropol. 2006;Suppl 43:89–130. https://doi.org/10.1002/ajpa.20518 PMID: 17103426

91. Amato KR, Jeyakumar T, Poinar H, Gros P. Shifting Climates, Foods, and Diseases: The Human Microbiome through Evolution. Bioessays. 2019; 41(10):e1900034. https://doi.org/10.1002/bies.201900034 PMID: 31524305

92. Zierer J, Jackson MA, Kastenmuller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet. 2018; 50(6):790–5. https://doi.org/10.1038/s41588-018-0135-7 PMID: 29808030

93. Jiang HE, Li X, Zhao YX, Ferguson DK, Hueber F, Bera S, et al. A new insight into Cannabis sativa (Cannabaceae) utilization from 2500-year-old Yanghai Tombs, Xinjiang, China. J Ethnopharmacol. 2006; 108(3):414–22. https://doi.org/10.1016/j.jep.2006.05.034 PMID: 16879937

94. Kilpatrick K, Pajak A, Hagel JM, Sumarah MW, Lewinsohn E, Facchini PJ, et al. Characterization of aromatic aminotransferases from Ephedra sinica Stapf. Amino Acids. 2016; 48(5):1209–20. https://doi.org/10.1007/s00726-015-2156-1 PMID: 26832171

95. Struthers R, Hodge FS. Sacred tobacco use in Ojibwe communities. J Holist Nurs. 2004; 22(3):209–25. https://doi.org/10.1177/0898010104266735 PMID: 15296576

96. Araujo AM, Carvalho F, Bastos Mde L, Guedes de Pinho P, Carvalho M. The hallucinogenic world of tryptamines: an updated review. Arch Toxicol. 2015; 89(8):1151–73. https://doi.org/10.1007/s00204-015-1513-x PMID: 25877327

97. Zhu D, Galbraith ED, Reyes-Garcia V, Ciais P. Global hunter-gatherer population densities constrained by influence of seasonality on diet composition. Nat Ecol Evol. 2021; 5(11):1536–45. https://doi.org/10.1038/s41559-021-01548-3 PMID: 34504317

98. Mitchell SC. N-acetyltransferase: the practical consequences of polymorphic activity in man. Xenobiotica. 2020; 50(1):77–91. https://doi.org/10.1080/00498254.2019.1618511 PMID: 31092097

99. Agundez JA. Polymorphisms of human N-acetyltransferases and cancer risk. Curr Drug Metab. 2008; 9 (6):520–31. https://doi.org/10.2174/138920008784892083 PMID: 18680472