NAT2 gene diversity and its evolutionary trajectory in the Americas

ARTICLE *in* THE PHARMACOGENOMICS JOURNAL · NOVEMBER 2015

Impact Factor: 4.23 · DOI: 10.1038/tpj.2015.72

10 AUTHORS, INCLUDING:

Rafael Bisso Machado

University of the Republic, Uruguay

13 PUBLICATIONS **77** CITATIONS

SEE PROFILE

Virginia Ramallo

Centro Nacional Patagonico

30 PUBLICATIONS **75** CITATIONS

SEE PROFILE

Dario A. Demarchi

National Scientific and Technical Research …

55 PUBLICATIONS **343** CITATIONS

SEE PROFILE

Tábita Hünemeier

University of São Paulo

44 PUBLICATIONS **357** CITATIONS

SEE PROFILE

The Pharmacogenomics Journal (2015), 1–7

www.nature.com/tpj

ORIGINAL ARTICLE

NAT2 gene diversity and its evolutionary trajectory in the Americas

R Bisso-Machado[1], V Ramallo[2], VR Paixão-Côrtes[1], V Acuña-Alonzo[3], DA Demarchi[4], JRS Sandoval[5], AAS Granara[5], FM Salzano[1], T Hünemeier[6] and MC Bortolini[1]

N-acetyltransferase 2 (NAT2) is responsible for metabolizing xenobiotics; *NAT2* polymorphisms lead to three phenotypes: rapid, intermediate and slow acetylators. We aimed to investigate *NAT2* diversity in Native Americans. *NAT2* exon 2 was sequenced for 286 individuals from 21 populations (Native American and American Mestizos). Excluding the basal/rapid haplotype *NAT2*4*, the most frequent haplotypes are *NAT2*5B* (35.95%) in hunter-gatherers and *NAT2*7B* (20.61%) and *NAT2*5B* (19.08%) in agriculturalists that were related to the slow phenotype. A new haplotype was identified in two Amerindians. Data from the ~ 44 kb region surrounding *NAT2* in 819 individuals from Africa, East-Asia, Europe and America were used in additional analyses. No significant differences in the acetylator *NAT2* haplotype and phenotype distributions were found between Native American populations practicing farming and/or herding and those practicing hunting and gathering, probably because of the absence or weakness of selection pressures and presence of demographic and random processes preventing detection of any selection signal.

The Pharmacogenomics Journal advance online publication, 27 October 2015; doi:10.1038/tpj.2015.72

INTRODUCTION

The *N-acetyltransferase 2* (*NAT2*) gene is located at 8p22 and has two exons; only the second exon, which is composed of 873 bp without introns, presents variation. *NAT2* encodes a 290-amino acid protein responsible for the metabolism of xenobiotics.[1] NAT2 is a phase II enzyme responsible for catalyzing the *N*- or *O*-acetylation of aromatic and heterocyclic amines and hydrazines present in a wide range of xenobiotics and other substances, including medicines and food.[2–4]

In the 1960s, Evans *et al.*[5,6] demonstrated that the antituberculosis drug isoniazid has a bimodal acetylation distribution that is inherited. Subsequently, it was verified that this phenomenon is related to NAT2 forms, thereby outlining the importance of genetic variation in drug response among individuals and across populations.[7] Today, it is well known that the ability to metabolize isoniazid, as well as many other commonly prescribed drugs such as sulfamethoxazole, that is used to treat AIDS patients suffering from secondary infections is associated with *NAT2* variation.[3,8]

NAT2 polymorphisms lead to distinguishable haplotypes. The presence of these haplotypes, with high or reduced enzymatic activities, has been correlated with fast (ancestral type) and slow (derived) acetylator/metabolizer phenotypes.[3,9,10] An intermediate category, associated with heterozygotes, has also been recognized, indicating a possible co-dominant relationship between the fast and slow alleles. The slow phenotype can lead to complications in xenobiotic metabolism. For example, individuals with tuberculosis and the slow-metabolizer phenotype exhibit collateral effects more frequently than do fast acetylators when treated with isoniazid.[11,12]

Several population studies with *NAT2* have indicated that the frequencies of these metabolizer phenotypes and their corresponding haplotypes are variable worldwide, but some tendencies have been observed. High frequencies of slow, derived haplotypes that lead to reduced activity were found in Middle Eastern and European populations, whereas lower slow-phenotype frequencies were observed in East Asians.[3,13]

The transition from a mode of subsistence based on hunting and gathering/foraging to one based on agriculture and animal domestication, which occurred independently in several continents ~ 6000–10 000 years ago,[14] might have redirected the more recent *NAT2* acetylation in the *Homo sapiens* evolutionary trajectory, probably because this transition led to exposure to new substances and xenobiotics that needed to be metabolized.[3,15] Sabbagh *et al.*[9] have listed some possibilities to explain the human *NAT2* diversity patterns that include drift as well as balancing or positive selection (favoring a more versatile intermediate acetylation metabolism or multiple standing slow-causing haplotypes, respectively). They also suggested that populations practicing farming and herding would have higher prevalence of slow acetylation phenotypes compared with those whose subsistence is mainly based on hunting and gathering, thereby advancing the hypothesis of directional selection for derived variants with reduced enzymatic activity.[9]

NAT2 variability in Native American populations has not been studied as extensively as that in other human continental autochthonous groups. Fuselli *et al.*[2] investigated nine Amerindian/high Amerindian ancestry populations from North, Central and South America, and suggested that, as a whole,

[1]Laboratório de Evolução Humana e Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [2]Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Puerto Madryn, Argentina; [3]Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia, Mexico City, Mexico; [4]Instituto de Antropología de Córdoba, CONICET, Universidad Nacional de Córdoba, Córdoba, Argentina; [5]Facultad de Medicina Humana, Universidad de San Martin de Porres, Lima, Peru and [6]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. Correspondence: Dr MC Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, Porto Alegre 91501-970, RS, Brazil.
E-mail: maria.bortolini@ufrgs.br
Received 7 May 2015; revised 21 August 2015; accepted 8 September 2015

these populations could be considered homogeneous for the distribution of acetylator phenotypes.

The present study described *NAT2* diversity in Native American populations. Some of these populations have retained their hunter-gathering/foraging habits, whereas others adopted agriculture centuries before the arrival of European colonizers.[16] Therefore, these modes of subsistence were incorporated in our analyses. In addition, data from other native continental groups, Mestizo populations with high Amerindian ancestry as well as a Paleoindian whose genome had a coverage of 14.4x (named 'Anzick', an individual belonging to the classical North American hunter-gathering Clovis culture who lived ~13 000 calendar years ago)[17] were also considered in our analyses.

MATERIALS AND METHODS

Samples and laboratory procedures

In total, 286 individuals from the following 21 populations were sequenced for exon 2, in which *NAT2* variation has been detected: (1) Native Americans: Apalaí, Arara, Gavião, Jamamadi, Suruí, Wai-wai, Xavánte and Zoró (named the Amazonian/Brazilian Central Plateau; $n=83$); Guarani (named Southern; $n=34$); Lengua (named Chaco; $n=22$); and Totonaco (named Mesoamerican; $n=8$); (2) Andean natives and other communities with high Amerindian ancestry ($n=67$): Uro, Amantani, Anapia, Andoas, Cabanaconde, Chivay, Taquile and Yanque; (3) Other: urban Mexican Mestizo ($n=34$); and (4) Native Asians: Siberian Eskimo ($n=38$). The geographic localizations of these populations are listed in Supplementary Table S2. Ethical approval for the present study was provided by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98) for the Amerindian and Siberian samples. Approval was also obtained from the following ethics committees: (1) Universidad Nacional Autónoma de México, Ciudad de México, México (Mexican samples) and (2) Universidad San Martín de Porres, Lima, Perú (Andean samples). Individual and/or tribal informed oral consents were obtained from illiterate participants. All sampling procedures were obtained according to the Helsinki Declaration. The ethics committees approved the oral consent procedure as well as the use of these samples in population and evolutionary studies.

Data collection

DNA from the 286 volunteers from 21 populations was extracted from plasma, glycerolized red blood cells as well as total blood stored in our laboratories in previous studies (review in Salzano[18]) by using the QIAamp DNA MiniKit (Qiagen, Hilden, Germany). The following primers were used to sequence *NAT2* exon 2: *NAT2* forward : 5′-TGAGCACCAGATCCGGGCTGTT-3′; *NAT2* reverse (5′ to 3′): 5′-CCATCACCAGGTTTGGGCACGA-3′. The PCR conditions were as follows: first step (initial denaturation): 95 °C for 15 min; second step (35 cycles): 94 °C for 30 s, 70 °C for 30 s and 72 °C for 40 s; third step (final extension): 72 °C for 15 min. Amplification products were purified and sequenced on both DNA strands (sequencing service provided by Macrogen, Seoul, Korea). Sequences obtained were deposited in GenBank (KT756895-KT757128).

Sequences were aligned and their qualities, as well as the assessment of the accuracy of the resulting data, were ascertained using the CodonCode Aligner 4.1 software (http://www.codoncode.com/aligner/). Sequencing of *NAT2* exon 2 (873 bp) revealed seven known single-nucleotide polymorphisms (SNPs; 191G>A, rs1801279; 282C>T, rs1041983; 341T>C, rs1801280; 481C>T, rs1799929; 590G>A, rs1799930; 803A>G, rs1208; 857G>A and rs1799931; Supplementary Figure S1). Genotypic frequencies were obtained by direct counting.

Additional data from SNP rs1208 and a genomic region covering 44 kb surrounding *NAT2* with 20 known SNPs were also used to extend this investigation. The data were obtained from 819 individuals from 73 populations of four continents (Africa, Asia, Europe and America), and were compiled from Reich *et al.*[19] and the HGDP-CEPH Human Genome Diversity Cell Line Panel (http://www.cephb.fr/en/hgdp_panel.php; Supplementary Table S1). Note that only one SNP (rs1208) located at *NAT2* exon 2 is present in the above-cited genetic data banks. Thus, 21 SNPs were used in the analyses by using this second set of data.

In summary, population analyses were performed using two sets of data: (1) original data for 7 SNPs, 286 individuals and 21 populations and (2) data available from the literature and public banks considering 21 SNPs, 819 individuals and 73 native continental populations.

Finally, data from the Paleoamerican 'Anzick' genome[17] were also compiled, and the *NAT2* haplotype and corresponding phenotype were determined.

Data analysis

Haplotypes based on *NAT2* exon 2 SNPs (data set 1) as well as haplotypes constructed on the basis of 21 SNPs (data set 2; data from HGDP-CEPH plus Reich *et al.*[19]) were estimated using PHASE 2.1 [refs 20,21] (command line: /PHASE < filename.inp > < filename.out10000110000). The haplotypes were then named and their respective acetylation status defined according to the arylamine NAT official nomenclature (http://nat.mbg.duth.gr/).

Because we did not have the metabolic phenotypes for data set 2, we inferred this information by using the 1000 Genomes data and the following three steps. (1) Complete data for the 27 SNPs were obtained for Mexicans living in Los Angeles, USA (MXL population from 1000 Genomes/Ensembl), whose Amerindian ancestry is well known. It is also known that this population has some degree of admixture; however, from among the populations that comprise the 1000 Genomes data, this population can be considered the closest to the Amerindian populations. (2) Linkage disequilibrium data based on these 27 SNPs were obtained for this Mexican population. (3) Metabolic phenotypes of data set 2 were finally inferred using these linkage disequilibrium data; that is, well-known phenotypes resulting from the presence of the haplotypes defined by the seven SNPs of *NAT2* exon 2 were used to infer the phenotypes of the other SNPs.

Tests for heterogeneity across mode-of-subsistence categories were performed with the nonparametric Kruskal–Wallis test by using the IBM SPSS 20.0 software (http://www-01.ibm.com/software/analytics/spss/products/statistics/downloads.html).

Median-joining haplotype networks[22] were constructed using the Network software (version 4.6.1.2; http://www.fluxus-engineering.com). Possible reticulations were resolved by MP calculation.[23]

Analysis of molecular variance[24–26] was performed using Arlequin 3.5.1.2.[27] A *P*-value of ≤ 0.01 was considered statistically significant for these tests.

Analyses of population structure were performed by dividing a selected set of Amerindian/high Amerindian ancestry populations into the following three categories, according to their more traditional modes of subsistence and geographical location: (1) those with a relatively recent history of hunter-gathering/foraging (Amazonian/Brazilian Central Plateau populations-BHG: Apalaí, Arara, Gavião, Jamamadi, Suruí, Wai-Wai, Xavánte and Zoró); (2) Ancient Andean agriculturalists-AA (Amantani, Anapia, Andoas, Cabanaconde, Chivay, Taquile, and Yanque); and (3) Ancient Mesoamerican agriculturalists-MA (Totonaco). The Uro were not included in the population structure analyses because they live in the Andes but are not recognized as farmers. Four other populations (Guarani, Lengua, Siberian Eskimo and Mexican Mestizo) were also excluded from the analyses because they could not be included in any of the three classifications adopted here.

Ewens–Watterson tests of neutrality were performed using Arlequin 3.5.1.2.[27] This test can potentially reveal whether demographic events (for example, rapid population growth) and directional (observed homozygosity higher than expected homozygosity) and/or balancing (observed homozygosity lesser than expected homozygosity) selection are operating on a particular locus across populations. Here, a *P*-value of <0.05 was used to detect a signal of balancing selection (or recent bottleneck), and a *P*-value of >0.95 was used to detect a signal of positive selection (or population growth), according to an earlier suggestion.[28]

RESULTS

Table 1 presents the haplotype frequencies and their probable acetylator status. Haplotype *NAT2*4* was the most frequent haplotype in agriculturalists (44%) and the second most frequent haplotype (31%) in hunter-gatherers/foragers from the Amazonian/Brazilian Central Plateau. This result was expected, as *NAT2*4* is the basal rapid haplotype.[2] The most frequent haplotype in hunter-gatherers/foragers is the slow acetylator *NAT2*5B* (36%), whereas in agriculturalists, *NAT2*7B* (21%) and *NAT2*5B* (19%), both of which are related to the slow phenotype, are also relatively common. The rare *7Ee haplotype, described earlier in a Brazilian tuberculosis patient[29](www.louisville.edu/medschool/pharmacology/NAT2.html), was found in one Mexican Mestizo.

Table 1. Haplotype observed in the populations evaluated in the present study (in the Americas) compared with those from the populations in the other three continents

		NAT2 inferred phenotypes[a]															*Mode of subsistence*	
		Rap	*Slow*	*Slow*	*Slow*	*Slow*	*Slow*	*Slow*	*Slow*	*ND*	*Rap*	*Rap*	*Rap*	*Rap*	*Slow*			*Reference*
		NAT2 haplotypes																
SNP	*(n)*	**4*	**5A*	**5B*	**5C*	**5D*	**6A*	**6C*	**7B*	**7Ee*	**12A*	**12B*	**12C*	**13A*	*New*	*Others*		
191G>A (rs1801279)		.	.	.	.	.	.	.	.	.	.	.	.	.	.			
282 C>T (rs1041983)		.	.	.	.	.	T	T	T	T	.	T	T	T	.			
341T>C (rs1801280)		.	C	C	C	C	.	.	.	.	.	.	.	.	C			
481C>T (rs1799929)		.	T	T	.	.	.	.	.	T	.	.	T	.	T			
590 G>A (rs1799930)		.	.	.	.	.	A	A	.	.	.	.	.	.	A			
803A>G (rs1208)		.	.	G	G	.	.	G	.	.	G	G	G	.	G			
857 G>A (rs1799931)		.	.	.	.	.	.	.	A	A	.	.	.	.	.			
Africa	(2100)	0.097	0.005	0.296	0.035	0.001	0.224	0.001	0.023		0.157	0.013	0.001	0.066		0.081		Sabbagh *et al.*[9]
Europe	(9848)	0.224	0.027	0.402	0.026	0.001	0.271	0.000	0.021		0.007	0.001	0.002	0.006		0.012		Sabbagh *et al.*[9]
Asia	(8500)	0.484	0.009	0.115	0.012	0.002	0.233	0.004	0.098		0.005	0.001	0.003	0.010		0.024		Sabbagh *et al.*[9]
Native American																		
Mesoamerican	(16)	0.250	.	0.250	0.062	.	.	.	0.437	.	.	.	.	.	.		Agriculturalists	Present study
Amazonian and Brazilian Central Plateau	(166)	0.307	.	0.362	.	.	0.048	.	0.247	.	0.006	0.012	0.006	.	0.012		Hunter-gatherers	Present study
Southern	(68)	0.441	.	0.103	0.015	.	0.059	.	0.382	.	.	.	.	.	.		Agriculturalists	Present study
Chaco	(44)	0.772	.	0.136	.	.	.	.	0.091	.	.	.	.	.	.		Agriculturalists	Present study
Mestizo/High Amerindian ancestry																		
Mexican	(68)	0.250	.	0.103	0.015	.	0.162	.	0.456	0.015	.	.	.	.	.		Urban	Present study
Andean	(134)	0.358	0.007	0.246	.	0.015	0.157	0.015	0.127	.	0.007	0.007	0.015	0.045	.		Agriculturalists	Present study
Siberian																		
Eskimo	(76)	0.263	.	0.355		.	0.250	.	0.118	.	.	.	0.013	.	.		Hunter-gatherers	Present study

Abbreviations: *NAT2*, *N-acetyltransferase 2*; ND, not determined; Rap, rapid: high enzyme activity; Slow, reduced enzyme activity. Observation: the haplotypes were named according to the official nomenclature (http://nat.mbg.duth.gr/). [a]Number of chromosomes are in parentheses.

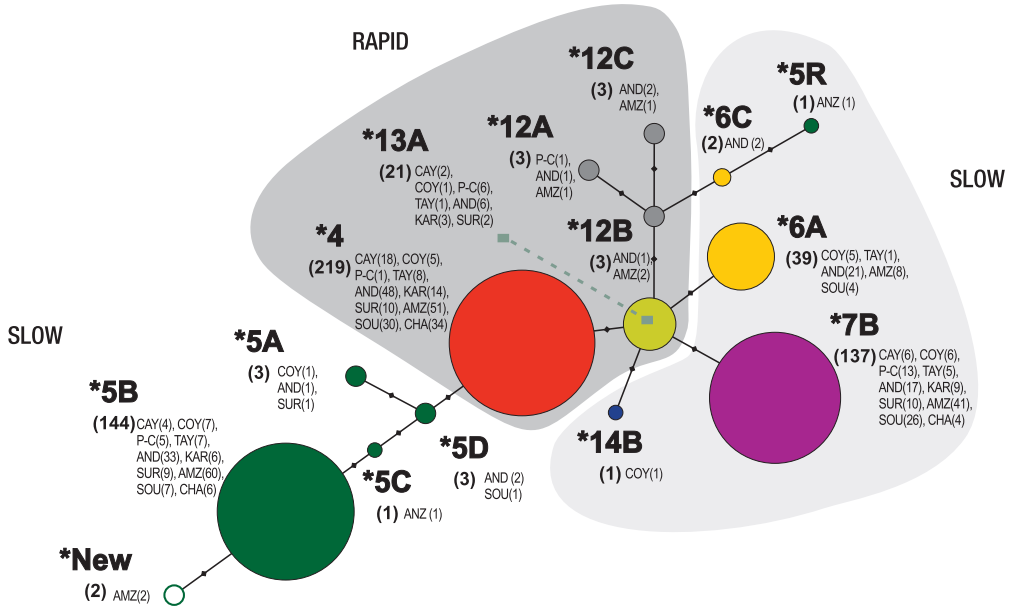

Figure 1. Network of *N-acetyltransferase 2* (*NAT2*) South American Native haplotypes plus Paleoamerindian data (Anzick). Data obtained in this study: AMZ, Amazonian/Brazilian Central Plateau; AND, Andean; CHA, Chaco; SOU, Southern. Data compiled from the literature: ANZ, Anzick (Rasmussen *et al.*[17]); CAY, Cayapo; COY, Coyaima; KAR, Karitiana; P-C, Piapoco-Curripaco; SUR, Suruí (data from the HGDP-CEPH Human Genome Diversity Cell Line Panel; http://www.cephb.fr/en/hgdp/diversity.php/); TAY, Tayacaja (Fuselli *et al.*[2]). Numbers and letters associated with asterisks represent the name of each haplotype. Each group of haplotypes is represented by a different color. Numbers in parentheses indicate numbers of individuals with the corresponding haplotype. The size of the spheres is proportional to the number of individuals belonging to that haplotype who were sampled. Transverse bars represent mutational steps between haplotypes.

A new haplotype was identified in Brazilian natives (Table 1). This new haplotype differed from *NAT2*5B* by a single mutational step, suggesting that the new haplotype was probably derived from the common *NAT2*5B* haplotype (Figure 1; haplotype network constructed using only data from South American natives plus Anzick). This new haplotype was found in one Suruí and one Xavánte who speak different languages belonging to the Tupi and Je major branches, respectively. As the new haplotype was derived from a slow acetylator, it probably has the same phenotype. The split between these two major South American linguistic branches occurred ~5000–6000 years ago,[30] indicating that this haplotype could have a more ancient (if present in the ancestral population) or a more recent origin, where it originated in one Je speaker population and was introduced into Tupi speakers (or vice versa) because of gene flow.

In America, the proportion of all slow acetylator haplotypes ranges from ~23% (Chaco) to ~75% (Andean agriculturalists and urban Mexican Mestizos). An intermediate number is found when hunter-gatherers/foragers from Amazonia and Brazilian Central Plateau are considered (67%). Haplotype *6A has been considered a marker of post-Columbian admixture.[2] However, its presence in the hunter-gatherers/forager Brazilian natives (~5%), in which admixture is virtually absent, indicates that it could have arrived with the first migrants who colonized America.

Our analyses also revealed that Anzick was a heterozygote for the *NAT2* locus, as evidenced by the presence of the *NAT2*5C* and *NAT2*5R* haplotypes. The *NAT2*5C* haplotype (slow) is found in low frequencies in modern human populations, including Native Americans, whereas *NAT2*5R* (probable slow haplotype) has only been reported in the arylamine *N*-acetyltransferases official database (http://nat.mbg.duth.gr/) without any indication of the population in which this haplotype was observed. This finding indicates that the population to which Anzick belonged could have a slightly different *NAT2* genetic background compared with modern Native Americans. However, note that the acetylation phenotype (Figure 1) is that expected for a hunter-gatherer.[9]

The among-population and between-group components of variance (F_{ST} and F_{CT}, respectively) for *NAT2* haplotypes and phenotypes (set of data 1; see Materials and methods) were calculated just for the Amazonian/Brazilian Central Plateau hunter-gatherer/forager and Andean agriculturalist groups, because of the larger number of populations investigated for each group. For the haplotype data, the hunter-gatherer/forager group showed an F_{ST} value of 0.11 ($P < 0.001$), whereas a lower but still significant structure was detected for Andean agriculturalists ($F_{ST} = 0.05$; $P = 0.008$). These values are not consistent with those obtained when the acetylator phenotypes were considered, where no differences were detected (hunter-gatherer/forager group: $F_{ST} = 0.027$, $P = 0.248$; Andean agriculturalists: $F_{ST} = 0.021$, $P = 0.312$). These differences can be attributed to allelic heterogeneity, as different alleles can cause fast or slow phenotypes.

On the other hand, for F_{CT}, no significant differences, using both *NAT2* haplotype and NAT2 phenotype data, were detected. Similar to the F_{CT} results, the Kruskal–Wallis test revealed that the distributions of the slow and fast acetylator haplotypes/phenotypes are not significantly different between the two modes-of-subsistence categories.

We extended our investigation using an additional genomic data set (set of data 2; see Materials and methods), covering ~44 kb surrounding *NAT2* in 819 individuals from the 4 continents. Inclusion of this data set expanded the number of investigated populations to 73 (HGDP: http://www.hagsc.org/hgdp/; and Reich *et al.*[19]). The 21 SNPs found revealed the presence of 19 haplotypes in Native Americans (Table 2 and Supplementary Table S1). Five of these haplotypes are found exclusively in Native Americans (h5, h7, h13, h14 and h17), but their frequencies are low. Moreover, the analysis revealed that Amerindians have high identity with Asians. For instance, the rapid haplotype 1 has virtually the same frequency in Native Americans and Asians

Table 2. Worldwide distribution of the 19 haplotypes found in Native Americans[a]

Haplotype	*Predicted phenotype*[b]	*America (536)*	*Europe (314)*	*Africa (352)*	*East-Asia (436)*
h1	Rapid	0.42500	0.10200	0.04370	0.42600
h2	Slow	0.31000	0.01910	0.01640	0.14900
h3	Slow	0.04970	0.22000	0.10400	0.17700
h4	Slow	0.14600	0.32500	0.19900	0.04910
h5	Slow	0.00171	0.00000	0.00000	0.00000
h6	Rapid	0.00856	0.06690	0.01910	0.02640
h7	Rapid	0.00342	0.00000	0.00000	0.00000
h8	Slow	0.00685	0.00000	0.00820	0.00189
h9	Rapid	0.02230	0.00318	0.00273	0.00000
h10	Slow	0.00171	0.00637	0.00000	0.00000
h11	Rapid	0.00171	0.00000	0.00000	0.00377
h12	Unknown	0.00342	0.00000	0.00000	0.00000
h13	Rapid	0.00514	0.00000	0.00000	0.00000
h14	Slow	0.00171	0.00000	0.00000	0.00000
h15	Slow	0.00171	0.02230	0.04640	0.00000
h16	Rapid	0.00342	0.03180	0.00273	0.00377
h17	Slow	0.00514	0.00000	0.00000	0.00000
h18	Rapid	0.00171	0.00000	0.00000	0.00189
h19	Slow	0.00171	0.01590	0.01090	0.00377

[a]*America:* Parakanã, Pima, Quechua, Surui, Wayuu, Waunana, Zapotec, Zenu, Palikur, Aymara, Karitiana, Jamamadi, Mapuche, Mixe, Mixtec, Ticuna, Arhuaco, Bribri, Cabecar, Chilote, Chipewyan, Chono, Colombian, Cree, Diaguita, Embera, Guarani, Guaymi, Hullice, Ingano, Kaingang, Kaqchikel, Kogi, Maleku, Maya, Ojibwa, Teribe, Uro and Yagane. *Europe:* Tuscan, Basque, Adygei, French, Italian, Orcadian, Russian and Sardinian. *Africa:* Yoruba, Bantu Kenya, Bantu South Africa, Bedouin, Pygmy Biaka, Mandenka, Pygmy Mbuti, Mozabite and San. *East-Asia:* Cambodian, Dai, Daur, Han, Hezhen. Japanese, Lahu, Miao, Mongol, Naxi, Oroqen, She, Tu, Tujia, Uygur, Xibo and Yi. [b]Based on linkage disequilibrium (see Materials and Methods). The number of chromosomes studied is indicated in parentheses.

(42%), whereas the values for the other continents/regions are lower. This result probably represents the Native American and Asian modal fast haplotype *4 (Figure 1 and Table 1 of the present study; Fuselli *et al.*[2]). On the basis of variants with known acetylator status and our linkage disequilibrium analysis, we estimated that 53% of the 19 haplotypes correspond to slow phenotypes in America. Among the exclusive haplotypes, three are slow (5, 14 and 17) and two are rapid (7 and 13; Table 2).

The analysis of molecular variance with this larger genetic and population data set allowed examination of the following three groups of interest considering geography and modes of subsistence: (1) hunter-gatherer/foragers (Jamamadi, Guarani, Karitiana, Zenu, Kogi, Ticuna, Embera, Wayuu, Palikur, Yaganes, Waunana, Kaingang, Surui and Parakanã); (2) Andean agriculturalists (Mapuche, Aymara, Quechua, Chilote, Hullice, Ingano, Arhuaco and Chonos); and (3) Mesoamerican agriculturalists (Mixe, Mixtec, Zapotec, Maya, Kaqchikel, Cabecar and Guaymi). The F_{ST} value for hunter-gatherer/foragers (0.11; $P = 0.002$) indicates high level of differentiation, whereas the populations from the Andean area are homogeneous ($F_{ST} = 0.008$, $P = 0.330$). This pattern was expected, considering the putative neutral markers in these two populations. The difference may be attributed to the distinguishing patterns of gene flow and historical effective sizes, cultural differences as well as paleoclimatic and environmental changes to which these indigenous populations[31] have been subjected. The Mesoamerican agriculturalist group presents an intermediate level, but the value is not significant ($F_{ST} = 0.039$, $P = 0.016$). No difference was detected in the pairwise between-group comparisons ($F_{CT} =$ 0.01–0.03; $P > 0.01$). When the acetylator phenotypes were considered, the same tendency was observed (hunter-gatherer/forager group: $F_{ST} = 0.13$, $P = 0.008$; Andean agriculturalists: $F_{ST} = 0.01$, $P = 0.361$; and Mesoamerican agriculturalists: $F_{ST} = 0.04$, $p = 0.039$). No significant differences in the distribution of the NAT2 phenotypes between the three groups were detected.

On the other hand, these set of data exhibited deviation from the standard neutral model when Mesoamerican agriculturalists were considered (Ewens–Watterson test, $P = 0.988$), indicating an excess of homozygote carriers of the rapid and most common h1 haplotype, an opposite tendency than would be expected in agriculturalist environments. Thus, we have adopted a more conservative position, attributing this finding to a relatively rapid population growth experienced by the Mesoamerican populations, even in pre-Columbian times. The values for Andean agriculturalists and South American hunter-gatherers were not significant ($P = 0.77$ and 0.64, respectively).

It should be mentioned that small sample sizes can lead to low statistical power. However, our sample had an 80% statistical power to detect significant values if the difference between slow and rapid haplotype frequencies was > 0.15 (calculated using data from Brazilian hunter-gatherer/forager and Andean agriculturalist groups, and StatMate software 2.0, $\alpha = 0.05$; Table 1 and http://www.graphpad.com/scientific-software/statmate/, respectively).

DISCUSSION

The common origin of Native Americans and their biological and cultural autochthonous evolution make the American continent an excellent model for studies on gene–culture coevolution. In terms of social development, in places such as Mesoamerica and the Andes, hunter-gatherer/forager societies gave rise to sedentary agriculturalist communities, and magnificent civilizations developed centuries before the arrival of Christopher Columbus.[16] Elsewhere, such as in Amazonia and the Brazilian Central Plateau, small and relatively isolated populations remained, basically surviving because of hunting and gathering until recently, or even to the present day (http://www.funai.gov.br/index.php/nossas-acoes/povos-indigenas-isolados-e-de-recente-contato; http://www.survivalinternational.org/tribes/uncontacted-brazil). Based on this and other facts, it is not easy to obtain research material from these Native American populations. Probably for this reason they are underrepresented in the scientific literature, and our study was conducted to fill at least part of this gap. Thus, the main objective of the present study was to expand our knowledge on *NAT2* gene pattern diversity and to understand why some trends in haplotype

distribution and their corresponding rapid, intermediate and slow metabolizer phenotypes occur. The haplotype distribution found in this study (Table 1) is in accordance with that reported in previous studies.[2,3] The exception is the high frequency of the ancestral rapid haplotype *4 in the agriculturalist Lengua from Chaco, Paraguay. On the other hand, some common slow haplotypes (*5B and *7B) are widespread among Native Americans, Siberian Eskimo as well as in other populations with high Amerindian ancestry. In contrast, we detected a potentially novel reduced-activity haplotype (341T > C; 481 C > T; 590 G > A; 803A > G) that occurred only among the Amazonian Suruí and Xavánte of the Brazilian Central Plateau (one individual in each population). To the best of our knowledge, this is the first time that a potentially exclusive Native American haplotype has been reported. It should be noted that these two tribes are geographically located far apart, and their members speak languages that belong to different major linguistic branches (Tupi and Je, respectively). In addition, probable Native American exclusive haplotypes were also detected with our larger data set, reinforcing the idea of unique *NAT2* haplotypes in Native Americans.

The evolutionary history of the South American hunter-gatherer/forager populations has been marked by dramatic genetic drift; this can also be inferred from the following aspects of our results: (1) the description of probable exclusive Native American haplotypes; (2) the high level of differentiation (F_{ST}) among hunter-gatherer/forager populations; and (3) the presence of the probable slow haplotype *NAT2*5R* in Anzick, a Paleoindian belonging to the hunter-gathering Clovis culture that is apparently absent in contemporary Amerindian populations.

It is reasonable to assume that distinct diet and lifestyles influence the genetic background of human populations, including Native Americans. For instance, earlier studies with the *ABCA1* metabolism gene, which encodes a transmembrane protein that promotes cholesterol efflux, have suggested that the Native American exclusive ABCA1 230Cys allele favors intracellular cholesterol and energy storage during unstable periods of food production, as experienced by the Mesoamericans during the transition to a sedentary lifestyle based mainly on maize cultivation.[32,33] Because NAT2 plays a role in the detoxification of exogenous substances, it has long been considered as a likely target of population-specific selective pressures, including new diet habits.[3,9] Thus, we hypothesized that, similar to what was described for *ABCA1*, the diversity of *NAT2* in Native Americans could be associated with their markedly different modes of subsistence. However, we were unable to detect significant differences in the *NAT2* haplotype and acetylator phenotype distributions when comparing Native American populations practicing farming and/or herding with those practicing hunting and gathering. In addition, we cannot exclude demography as the cause for the large number of h1 homozygotes found in the Mesoamerican group.

CONCLUSION

The modes of subsistence considered here do not seem to be associated with the *NAT2* genetic pattern variability, but demographic and random processes can mask possible and perhaps subtle adaptive signals. Small sample sizes and the absence of data on possible physiological variations between phenotypes are also factors that may have contributed to the difficulty in finding a clear evidence of a connection between the genetic variability and modes of subsistence.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are grateful to the individuals who donated the samples analyzed here and to the Fundação Nacional do Índio (FUNAI) for logistic support in the Brazilian sample collections. We thank Sandro L Bonatto for donating the Siberian Eskimo samples. We are also grateful to Michele Aramburu Serafini for technical assistance, Luciana Tovo-Rodrigues for information about the software used and Sidia M Callegari-Jacques and Diego Rovaris for help with the statistical analysis. Financial support was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, Programa de Apoio a Núcleos de Excelência (FAPERGS/PRONEX). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1 Blum M, Grant DM, McBride W, Heim M, Meyer UA. Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA Cell Biol* 1990; **9**: 193–203.
2 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**: 144–152.
3 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**: e21.
4 Mahasneh A, Jubaili A, El Bateiha A, Al-Ghazo M, Matalka I, Malkawi M. Polymorphisms of arylamine N-acetyltransferase2 and risk of lung and colorectal cancer. *Genet Mol Biol* 2012; **35**: 725–733.
5 Evans DA, Manley KA, Mc KV. Genetic control of isoniazid metabolism in man. *Br Med J* 1960a; **2**: 485–491.
6 Evans DA, Storey PB, Wittstadt FB, Manley KA. The determination of the isoniazid inactivator phenotype. *Am Rev Respir Dis* 1960b; **82**: 853–861.
7 Weber WW. *The Acetylator Genes and Drug Response*. Oxford University Press: New York, 1987.
8 Grimwade K, Gilks C. Cotrimoxazole prophylaxis in adults infected with HIV in low-income countries. *Curr Opin Infect Dis* 2001; **14**: 507–512.
9 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**: e18507.
10 Ruiz JD, Martínez C, Anderson K, Gross M, Lang NP, García-Martín E *et al*. The differential effect of NAT2 variant alleles permits refinement in phenotype inference and identifies a very slow acetylation genotype. *PLoS One* 2012; **7**: e44629.
11 Huang YS, Chern HD, Su WJ, Wu JC, Lai SL, Yang SY *et al*. Polymorphism of the N-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug-induced hepatitis. *Hepatology* 2002; **35**: 883–889.
12 Santos NP, Callegari-Jacques SM, Ribeiro Dos Santos AK, Silva CA, Vallinoto AC, Fernandes DC *et al*. N-acetyl transferase 2 and cytochrome P450 2E1 genes and isoniazid-induced hepatotoxicity in Brazilian patients. *Int J Tuberc Lung Dis* 2013; **17**: 499–504.
13 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**: 423–436.
14 Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature* 2002; **418**: 700–707.
15 Cordain L, Eaton SB, Sebastian A, Mann N, Lindeberg S, Watkins BA *et al*. Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* 2005; **81**: 341–354.
16 Salzano FM, Bortolini MC. *Evolution and Genetics of Latin American Populations*. Cambridge University Press: Cambridge, 2002.
17 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**: 225–229.
18 Salzano FM. Molecular variability in Amerindians: widespread but uneven information. *An Acad Bras Cienc* 2002; **74**: 223–263.
19 Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N *et al*. Reconstructing Native American population history. *Nature* 2012; **488**: 370–374.
20 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.
21 Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *Am J Hum Genet* 2005; **76**: 449–462.
22 Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.

23 Polzin T, Daneschmand SV. On Steiner trees and minimum spanning trees in hypergraphs. *Operat Res Lett* 2003; **31**: 12–20.
24 Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984; **38**: 1358–1370.
25 Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
26 Weir BS. The second National Research Council report on forensic DNA evidence. *Am J Hum Genet* 1996; **59**: 497–500.
27 Excoffier L, Lischer HEL. Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 2010; **10**: 564–567.
28 McClelland EK, Ming TJ, Tabata A, Kaukinen KH, Beacham TD, Withler RE *et al.* Patterns of selection and allele diversity of class I and class II major histocompatibility loci across the species range of sockeye salmon (*Oncorhynchusnerka*). *Mol Ecol* 2013; **22**: 4783–4800.
29 Teixeira RL, Morato RG, Cabello PH. Genetic polymorphisms of NAT2, CYP2E1 and GST enzymes and the occurrence of antituberculosis drug-induced hepatitis in Brazilian TB patients. *Mem Inst Oswaldo Cruz* 2011; **106**: 716–724.
30 Ramallo V, Bisso-Machado R, Bravi C, Coble MD, Salzano FM, Hünemeier T *et al.* Demographic expansions in South America: enlightening a complex scenario with genetic and linguistic data. *Am J Phys Anthropol* 2013; **150**: 453–463.
31 Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM *et al.* Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 2001; **68**: 1485–1496.
32 Acuña-Alonzo V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hünemeier T *et al.* A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Hum Mol Genet* 2010; **19**: 2877–2885.
33 Hünemeier T, Amorim CE, Azevedo S, Contini V, Acuña-Alonzo V, Rothhammer F *et al.* Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One* 2012; **7**: e38862.

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)