Доклади на Българската академия на науките
Comptes rendus de l'Académie bulgare des Sciences
Tome 76, No 4, 2023

BIOLOGY
Molecular genetics

PREVALENCE OF RARE VARIANTS ASSOCIATED WITH MONOGENIC DISEASES IN PRE-CONTACT CARIBBEAN COMMUNITIES

Draga Toncheva[#], Maya Atanasoska[*,], Maria Marinova[***], Plamenka Borovska[****], Dimitar Serbezov**

Received on March 5, 2021
Accepted on March 30, 2021

Abstract

Genomic technologies in archaeology have been used to study migration patterns, origin of and genetic similarity between ancient populations. Researchers are, however, just beginning to employ these technologies to evaluate the genomic health of ancient communities. In this study we screen publicly available whole-genome sequencing data of pre-contact Caribbean individuals for the presence of pathogenic mutations causing monogenic diseases. Our results show that Caribbean communities from the Ceramic Age (3100–400 BP) had an unusually high frequency of pathogenic variants associated with three monogenic conditions, Classical Phenylketonuria, Hypohidrotic Ectodermal Dysplasia and *LRRK2*-Related Parkinson's disease. The estimated frequency of these variants from the analyzed ancient samples is severalfold higher than that in contemporary populations. Three conditions classified today as rare diseases thus might have had much higher prevalence in the past. Studying the history of monogenic diseases from ancient communities will expand the list of rare diseases with larger prevalence in the past.

Key words: ancient DNA, rare diseases, Caribbean

Introduction. Mitochondrial DNA (mtDNA), isolated from teeth and bone remains, has been used in recent studies to evaluate migration patterns, origin

[#]Corresponding author.
DOI:10.7546/CRABS.2023.04.04

of and genetic similarity between ancient populations [1]. The intensive studies performed with mtDNA have led to new understanding of past events in human history. With the advent of whole-genome sequencing technologies it has also become possible to detect genetic disturbances in ancient populations. The sequencing of the Neanderthal genome was the first major accomplishment in the field of molecular archeology [2].

Research dealing with ancient pathogenic DNA mutations is currently limited to analyses on target regions in a handful of genes. Polymorphic sites have been identified in the β-Globin gene, analyzed with DNA isolated from bone remains dated 10 000 years BCE [3]. A study on Minoan DNA samples from early-to mid-Bronze Age (2000 BCE) aimed to find thalassemia mutations in a small intron target region, the IVS 1 region of the β-Globin gene and its splicing junctions but none were found [4]. A Japanese research team had sequenced loci in the ABO blood group gene from ancient DNA from different periods (120 000–400 years BCE) [5]. Berens et al. [6] estimate the genetic disease risk for 3180 loci in 147 ancient genomes, and find it to be similar to that of modern day humans. Analyses of individual genomes, however, indicate that the overall genomic health of the Altai Neanderthal is worse than 97% of present day humans and that Ötzi the Tyrolean Iceman had a genetic predisposition to gastrointestinal and cardiovascular diseases [6].

Whole-genome sequencing of 25 Thracian mtDNA samples (dated 3000–2000 BCE) detected 608 mtDNA variants [7]. The variant m.15326A>G (rs2853508) was established, designated as probably pathogenic for being associated with familial breast cancer [8]. Six mitochondrial variants designated as being "confirmed pathogenic" in modern patients were established in the ancient Eurasian mtDNA genomes from different periods (n=1443) from the publicly accessible AmtDB database [9]. The oldest m.7510T>C in the MT-TS1 gene, was detected in a sample from the Neolithic period, dated 5800–5400 BCE [10].

Rare monogenic diseases (Orphan diseases) are considerable health burden in contemporary populations. The majority of these are marked by severe clinical manifestations, lead to disabilities, their diagnosis is costly, treatment options are limited, and often have lethal outcome. Less than one in 2000 individuals in Europe and less than one in 1250 in USA must be affected by a condition in order for it to be classified as a rare disease. So far, 5856 have been described, the majority being genetically heterogeneous [11]. The number of genes with variants associated with rare diseases has considerably increased with the advent of NGS technologies and is currently around 4000 [12].

Studies on ancient populations analyzing the diversity and frequency of monogenetic nuclear DNA mutations are currently lacking. Population sizes and migration rates were small in ancient compared to modern populations. The small effective population sizes and the associated increased levels of homozygosity might result in accumulation of pathogenic mutations. It can therefore be speculated

that the occurrence of rare diseases varied substantially among different ancient communities as well as among different time periods. It can also be presumed that some of what today are classified as rare diseases were common diseases thousands of years ago. New insights about monogenic mutations can be gained by studying their occurrence and frequency in ancient populations. Knowledge about the genetic history of orphan diseases might change how researchers deal with these pathologies.

The aim of this study is to analyze the prevalence of known disease causing rare mutations in pre-contact Caribbean communities that had inhabited the Bahamas, Haiti and the Dominican Republic (collectively, Hispaniola), Puerto Rico, Curacao and Venezuela [13]. Ceramic age individuals (between ca. 3100 and 400 calibrated years BP) are particularly interesting from medico-genetic perspective as they are largely a genetically homogeneous group. They have migrated from north-eastern areas of South America, and their effective population size estimates range from 500 to 1500 for single populations and from 1500 to 8000 across islands.

Materials and methods. Of the publicly available WGS data of 171 pre-contact Caribbean individuals from the Archaic and Ceramic Ages the following locations were screened for the presence of pathogenic mutations:

Ceramic Age Samples

- **Greater Antilles:** Bahamas ($n = 24$). Eastern Greater Antilles ($n = 43$), Southeast coast of Dominican Republic ($n = 86$), Haiti ($n = 2$)
- **Lesser Antilles:** Curacao ($n = 5$), Venezuela ($n = 8$)

Archaic Age Samples

- **Greater Antilles:** Cueva del Perico, Cuba ($n = 1$), Cueva Roja, Dominican Republic ($n = 2$)

We searched the data for variants that the publicly available DisGeNet database designates as disease associated [14]. Of the variants found, we then selected those with very low contemporary population frequency (< 0.001 by the genome database gnomeAD [15]). We assume that variants with no estimates are of very low contemporary population frequencies. We excluded upstream/downstream, synonymous, intergenic and intronic variants as their role in human disease etiology is currently unclear. The association with disease variants was additionally reviewed in the online platforms ClinVar [16] and Varsome [17]. These tools are routinely used to verify the pathogenicity and disease association of variants established in patients. Ancient DNA molecules are, however, often retrieved in low copy number, highly fragmented and possibly contaminated, and precaution should be taken when making diagnostic inferences.

Results. The pre-contact Caribbean samples were screened for the presence of 32 644 variants associated with disease as reported by DisGeNet. Forty of

Table 1

Pathogenic variants in pre-contact Caribbean samples

dbSNP	Gene	Alleles	Consequence	Disease
rs5030846	*PAH*	G/A	stop/gain	Classical PKU
rs5030853	*PAH*	C/A	missense	Classical PKU
rs5030858	*PAH*	G/A	missense	Classical PKU
rs5030851	*PAH*	G/A	missense	Classical PKU
rs76296470	*PAH*	G/A	stop/gain	Classical PKU
rs79931499	*PAH*	C/T	missense	Classical PKU
rs121908450	*EDAR*	C/T	missense	Hypohidrotic Ectodermal Dysplasia
rs121908453	*EDAR*	C/T	missense	Hypohidrotic Ectodermal Dysplasia
rs34637584	*LRRK2*	G/A	missense	Parkinson's disease

these variants are with contemporary population frequency below 0.001. Nine of these pathogenic variants are established to be in homozygous state (Table 1). Six variants in 16 samples are associated with classical phenylketonuria (PKU), two variants in eight samples – with Hypohidrotic Ectodermal Dysplasia, and one variant in a single sample – with Parkinson's disease (Table 1 and Table 2).

High prevalence (16.1%) of pathogenic homozygous genotypes was established in Ceramic Age samples from Eastern Greater Antilles, SE coast of Dominican Republic, Haiti and Bahamas (Table 2).

The geographical distribution of the established monogenic disease cases is presented in Fig. 1.

Table 2

Incidence of homozygous genotypes of monogenic disorders in Ceramic Age pre-contact Greater Antilles communities

Disease/ Gene	Variant/ Genotype	Eastern Greater Antilles ($n = 43$)	SE coast Dominican Republic ($n = 86$)	Haiti ($n = 2$)	Bahamas ($n = 24$)	Total ($n = 155$)
PKU (*PAH*)	rs5030846 (AA)	1	1			2
	rs5030853 (AA)		1			1
	rs5030858 (AA)		1	1		2
	rs5030851 (AA)				1	1
	rs76296470 (AA)	1	3			4
	rs79931499 (TT)	2	2		2	6
Ectodermal Dysplasia (*EDAR*)	rs121908450 (TT)	3	2		1	6
	rs121908453 (TT)	1	1			2
Parkinson's disease (*LRRK2*)	rs34637584 (AA)		1			1
Total		8	12	1	4	25

Fig. 1. Geographical distribution of monogenic diseases in Ceramic Age Caribbean communities on Greater Antilles. • Phenylketonuria, ▲ Hypohidrotic Ectodermal Dysplasia, ♢ Parkinson's disease

DNA samples from the Lesser Antilles: Ceramic Age Curacao ($n = 5$) and Venezuela ($n = 8$) and from Archaic Greater Antilles, Cuban Cueva Del Perico ($n = 1$) and Dominican Cueva Roja ($n = 2$), did not contain any rare mutations associated with monogenic diseases.

Discussion. Our study reveals for the first time the range and prevalence of monogenic diseases in pre-contact Caribbean communities from the Ceramic Age (3100 to 400 years BP). The results suggest that classical phenylketonuria was the most prevalent monogenetic disease, manifesting as a result of homozygous carriership of 6 different *PAH* gene pathogenic variants.

Phenylketonuria (PKU) is a metabolic disorder caused by deficit of phenylalanine hydroxylase and the ensuing disruption of phenylalanine metabolism. PKU is autosomal recessive disorder. With no treatment increased phenylalanine levels in the blood and the brain have toxic effect and lead to intellectual disability, light skin and hair, seizures, developmental delays, behavioural problems, and psychiatric disorders.

The estimated frequency of PKU in Ceramic Age Greater Antilles communities is 9.8%, a thousandfold higher than the worldwide contemporary population frequency (ca. 1 in 10 000). Considering that the condition seems to be equally common across the sampled Greater Antilles islands, PKU incidence in this age and region classifies it as common disease. In contemporary populations, however,

it is a rare disease, the highest frequency being established in isolated populations practicing monoethnic marriages, e.g. one in 332 newborns among the nation of Karachays in the Karachay-Cherkess Republic, Russia [18].

Another condition with high prevalence in pre-contact communities of the Greater Antilles is Hypohidrotic Ectodermal Dysplasia (HED). We establish homozygous pathogenic HED genotypes in 8 pre-contact Caribbean samples. In parallel with PKU, the estimated prevalence (4.9%) is a thousandfold higher than that in modern populations (1 in 20 000 newborns worldwide). HED is a genetic skin disease that can be inherited in an autosomal dominant, autosomal recessive, or in X-linked recessive manner. HED is characterized by sparseness of scalp and body hair (hypotrichosis), reduced ability to sweat (hypohidrosis), and congenital absence of teeth (hypodontia). The basic features of classic HED appear during childhood. The scalp hair is thin, lightly pigmented and slow-growing. Sweating is markedly deficient, leading to episodes of hyperthermia. Only a few abnormally formed teeth grow belatedly. Different types of HEDs are caused by specific mutations in at least four genes, and can be inherited in a variety of ways. EDAR gene mutations account for a smaller percentage of cases and are inherited in an autosomal dominant or recessive manner.

In addition, we also established a single sample homozygous for the A allele of the rs34637584 in the *LRRK2* gene. It is germ line dominant mutation established as pathogenic/risk factor associated with Parkinson's disease. LRRK2 Parkinson's disease (PD) is characterized by features consistent with idiopathic PD: initial motor features of slowly progressive asymmetric tremor at rest and/or bradykinesia, cogwheel muscle rigidity, postural instability, and gait abnormalities. Certain non-motor symptoms in LRRK2-PD, especially REM sleep behaviour disorder and cognitive decline, may occur at similar or slightly reduced frequency compared to typical idiopathic PD. Onset is generally after age 50, although early age onset and late age onset manifestations have been described.

The estimated frequency of all nine disease associated variants is 100 to 1000-fold higher than that in contemporary populations (Table 3). The *PAH* gene missense mutation rs79931499 is the most common pathogenic variant detected in the analyzed samples, with estimated frequency varying from 0.023 to 0.061 among samples. Founder effect and low genetic differentiation across islands might explain the existence of variants with high and similar estimated frequencies across islands, e.g. the *PAH* gene variants rs79931499 and rs76296470.

Isolation, adaptation and migration are evolutionary processes that shape the genetic structure of populations. Modern human populations are markedly differentiated and researchers have amassed large data on how genetic variants associate with diverse population structures [19]. Genetic differentiation is largely shaped by frequency differences of common genetic variants, but rare genetic variants might also play a role in shaping population differentiation patterns [20]. A recent study finds that different pathogenic *PAH* gene variants are differentially

Table 3

Minor allele frequency (MAF) of pathogenic variants associated with monogenic disorders in pre-contact Caribbean samples

Gene	Pathogenic allele	Eastern Greater Antilles	SE coast Dominican Republic	Haiti	Bahamas	Total	Contemporary population (GnomAD exomes)
PAH	rs79931499 (T) missense	0.047	0.023		0.061	0.037	0.00006
	rs76296470 (A) stop/gain	0.023	0.035			0.024	0.00003
	rs5030858 (A) missense		0.012	0.500		0.023	0.0008
	rs5030846 (A) stop/gain	0.023	0.012			0.012	0.00004
	rs5030853 (A) missense		0.012			0.006	0.0006
	rs5030851 (A) missense				0.030	0.006	0.0001
EDAR	rs121908450 (T) missense	0.070	0.023		0.030	0.037	0.000004
	rs121908453 (T) missense	0.023	0.012			0.012	NA
LRRK2	rs34637584 (A) missense		0.012			0.006	0.0005

enriched in different geographical areas, likely as a result of a combination of evolutionary and adaptive events in populations with distinct ancestries [21].

Conclusion. Studies like ours lay the foundation of a novel scientific field, the field of archeological genomic medicine. The employment of genomic technologies in archaeology allows ancient genomes to be screened for disease associated mutations. Our results indicate that Ceramic Age Caribbean communities that have lived 3100–400 BP had an unusually high frequency of pathogenic variants associated with three monogenic conditions, Classical Phenylketonuria, Hypohidrotic Ectodermal Dysplasia and *LRRK2*-Related Parkinson's disease. The estimated frequency of these variants from the analyzed ancient samples is severalfold higher than that in contemporary populations. Such high frequencies of disadvantageous mutations might have caused the effects by genetic drift bolstered by small effective population sizes. Enhanced effects of natural selection resulting from increased effective population sizes, coupled with medical advances could have reduced the frequency of these mutations. Three conditions classified today as rare diseases thus might have been more common in the past. Studying the history of monogenic diseases from ancient communities will expand the list of

rare diseases with larger prevalence in the past. Data on the historical prevalence might alter our perspective on rare diseases and could encourage the introduction of more extensive screening programmes and development of new technologies for non-invasive prenatal prevention of rare diseases.

REFERENCES

[1] Fu Q., P. Rudan, S. Pääbo, J. Krause (2012) Complete mitochondrial genomes reveal neolithic expansion into Europe, PloS one, **7**(3), e32473.

[2] Green R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel et al. (2010) A draft sequence of the Neandertal genome, Science, **328**(5979), 710–722.

[3] Béraud-Colomb E., R. Roubin, J. Martin, N. Maroc, A. Gardeisen et al. (1995) Human beta-globin gene polymorphisms characterized in DNA extracted from ancient bones 12,000 years old, Am. J. Hum. Genet., **57**(6), 1267–1274.

[4] Hughey J. R., M. Du, Q. Li, M. Michalodimitrakis, G. Stamatoyannopoulos (2012) A search for β thalassemia mutations in 4000 year old ancient DNAs of Minoan Cretans, Blood Cells Mol. Dis., **48**(1), 7–10.

[5] Sato T., H. Kazuta, T. Amano, H. Ono, H. Ishida et al. (2010) Polymorphisms and allele frequencies of the ABO blood group gene among the Jomon, Epi-Jomon and Okhotsk people in Hokkaido, northern Japan, revealed by ancient DNA analysis, J. Hum. Genet., **55**(10), 691–696.

[6] Berens A. J., T. L. Cooper, J. Lachance (2017) The genomic health of ancient hominins, Human Biol., **89**(1), 7–20.

[7] Modi A., D. Nesheva, S. Sarno, S. Vai, S. Karachanak-Yankova et al. (2019) Ancient human mitochondrial genomes from Bronze Age Bulgaria: new insights into the genetic history of Thracians, Sci. Rep., **9**(1), 5412.

[8] Landrum M. J., J. M. Lee, G. R. Riley, W. Jang, W. S. Rubinstein et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype, Nucleic Acids Res., **42**(Database issue), D980-D5.

[9] Ehler E., J. Novotný, A. Juras, M. Chyleński, O. Moravčík, J. Pačes (2018) AmtDB: a database of ancient human mitochondrial genomes, Nucleic Acids Res., **47**(D1), D29-D32.

[10] Toncheva D., D. Serbezov, S. Karachanak-Yankova, D. Nesheva (2020) Ancient mitochondrial DNA pathogenic variants putatively associated with mitochondrial disease, PloS one, **15**(9), e0233666.

[11] Schieppati A., J. I. Henter, E. Daina, A. Aperia (2008) Why rare diseases are an important medical and social issue, Lancet (London, England), **371**(9629), 2039–2041.

[12] Lee C. E., K. S. Singleton, M. Wallin, V. Faundez (2020) Rare Genetic Diseases: Nature's Experiments on Human Development, iScience, **23**(5), 101123.

[13] Fernandes D. M., K. A. Sirak, H. Ringbauer, J. Sedig, N. Rohland et al. (2021) A genetic history of the pre-contact Caribbean, Nature, **590**(7844), 103–110.

[14] Piñero J., J. M. Ramírez-Anguita, J. Saüch-Pitarch, F. Ronzano, E. Centeno et al. (2020) The DisGeNET knowledge platform for disease genomics: 2019 update, Nucleic Acids Res., **48**(D1), D845-D55.

[15] Karczewski K. J., L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi et al. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans, Nature, **581**(7809), 434–443.
[16] Landrum M. J., J. M. Lee, M. Benson, G. R. Brown, C. Chao et al. (2018) ClinVar: improving access to variant interpretations and supporting evidence, Nucleic Acids Res., **46**(D1), D1062-d7.
[17] Kopanos C., V. Tsiolkas, A. Kouris, C. E. Chapple, M. Albarca Aguilera et al. (2019) VarSome: the human genomic variant search engine, Bioinformatics, **35**(11), 1978–1980.
[18] Gundorova P., R. A. Zinchenko, I. A. Kuznetsova, E. A. Bliznetz, A. A. Stepanova et al. (2018) Molecular-genetic causes for the high frequency of phenylketonuria in the population from the North Caucasus, PloS one, **13**(8), e0201489.
[19] Mallick S., H. Li, M. Lipson, I. Mathieson, M. Gymrek et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations, Nature, **538**(7624), 201–206.
[20] Persyn E., R. Redon, L. Bellanger, C. Dina (2018) The impact of a fine-scale population stratification on rare variant association test results, PloS one, **13**(12), e0207677.
[21] Oussalah A., E. Jeannesson-Thivisol, C. Chéry, P. Perrin, P. Rouyer et al. (2020) Population and evolutionary genetics of the PAH locus to uncover overdominance and adaptive mechanisms in phenylketonuria: Results from a multiethnic study, EBioMedicine, **51**, 102623.

Department of Medical Genetics
Medical Faculty
Medical University-Sofia
2 Zdrave St
1431 Sofia, Bulgaria
e-mails: dragatoncheva@gmail.com
dimitase@gmail.com

**Genetic Laboratory*
Gynecology and Assisted Reproduction
Hospital "Malinov MD"
46 Gotse Delchev Blvd
1680 Sofia, Bulgaria
e-mail: majaatanasoska@gmail.com

***Department of Genetics*
Faculty of Biology
Sofia University "St. Kliment Ohridski"
8 Dragan Tsankov Blvd
1164 Sofia, Bulgaria

****Department of Computer Systems and Technologies*
Faculty of Electronics and Automation
Technical University – Sofia, Branch Plovdiv
25 Tsanko Dyustabanov St
4000 Plovdiv, Bulgaria
e-mail: m_marinova@tu-plovdiv.bg

*****Department of Informatics*
Faculty of Applied Mathematics and Informatics
Technical University of Sofia
8 Kliment Ohridski Blvd
1756 Sofia, Bulgaria
e-mail: pborovska@tu-sofia.bg