Functional characterization of archaic-specific variants in mitonuclear genes: insights from comparative analysis in S. cerevisiae

Functional Characterization of Archaic-specific Variants in Mitonuclear Genes: Insights from Comparative Analysis in *S. cerevisiae*

Serena Aneli[1°], Camilla Ceccatelli Berti[2°], Alexandru Ionut Gilea[2], Giovanni Birolo[3], Giacomo Mutti[4,5], Angelo Pavesi[2], Enrico Baruffini[2], Paola Goffrini[2**], Cristian Capelli[2,6*]

1. Department of Public Health Sciences and Pediatrics, University of Turin, 10126, Turin, Italy.
2. Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11/a, 43124, Parma, Italy
3. Department of Medical Sciences, University of Turin, 10126, Turin, Italy.
4. Barcelona Supercomputing Centre (BSC-CNS), Plaça Eusebi Güell, 1-3, 08034, Barcelona, Spain
5. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028, Barcelona, Spain
6. Department of Biology, University of Oxford, 11a Mansfield Rd, OX1 3SZ, Oxford, UK

° joint first authors
*Corresponding authors: paola.goffrini@unipr.it; cristian.capelli@unipr.it

Abstract

Neanderthal and Denisovan hybridisation with modern humans has generated a non-random genomic distribution of introgressed regions, the result of drift and selection dynamics. Cross-species genomic incompatibility and more efficient removal of slightly deleterious archaic variants have been proposed as selection-based processes involved in the post-hybridisation purge of archaic introgressed regions. Both scenarios require the presence of functionally different alleles across *Homo* species onto which selection operated differently according to which populations hosted them, but only a few of these variants have been pinpointed so far.

In order to identify functionally divergent archaic variants removed in humans, we focused on mitonuclear genes, which are underrepresented in the genomic landscape of archaic humans. We searched for non-synonymous, fixed, archaic-derived variants present in mitonuclear genes, rare or absent in human populations. We then compared the functional impact of archaic and human variants in the model organism *Saccharomyces cerevisiae*. Notably, a variant within the mitochondrial tyrosyl-tRNA synthetase 2 (YARS2) gene exhibited a significant decrease in respiratory activity and a substantial reduction of Cox2 levels, a proxy for mitochondrial protein biosynthesis, coupled with the accumulation of the YARS2 protein precursor and a lower amount of mature enzyme. Our work suggests that this variant is associated with mitochondrial functionality impairment, thus contributing to the purging of archaic introgression in YARS2. While different molecular mechanisms may have impacted other mitonuclear genes, our approach can be extended to the functional screening of mitonuclear genetic variants present across species and populations.

Introduction

Homo sapiens stands as the sole surviving species within the genus *Homo*. However, for most of its evolutionary timeline, the genus *Homo* has been characterised by the presence of more than one species, several co-existing with anatomically modern humans (AMHs) (1,2). Among these, Denisovans and Neanderthals (often referred to as archaic humans) have been shown to be sister lineages to *Homo sapiens* sharing a common ancestor around 500-600 thousands years ago (kya), the three species involved in a complex scenario of gene-flow and admixture (2–7). The origin of Neanderthals and Denisovans has been related to the spread across Europe and Asia of hominins associated with *H. heidelbergensis*, the taxonomic status and the evolutionary relationships of the latter with other hominins still a matter of debate (8,9). Notably, the diffusion outside Africa had a significant impact on the genetic variation of these groups. On one side, drift-related dynamics resulted in a reduction of genetic diversity, on the other the exposure to different ecological, environmental and climatic backgrounds generated selection-based pressures (10–13). The combination of these dynamics, operating also on the direct ancestors of *H. sapiens* in Africa, has resulted in a number of genetic variants being differentially fixed in AMHs and archaic humans (14,15).

So far the functional significance of derived archaic variants has been almost exclusively investigated by exploring their impact in people bearing regions introgressed from Neanderthals and Denisovans (15–18). However, introgressed archaic genomic regions have been rapidly depleted in humans, by a combination of purifying selection and drift, which means that many archaic derived alleles are either extremely rare or absent in modern day human genomes (11,19–22). Genes that appear to be less introgressed are involved with, among the others, spermatogenesis, keratin and mitochondria, the latter comprising nuclear genes encoding for proteins relevant for mitochondrial functionality (mitonuclear genes) (23). Given the interaction between nuclear and mitochondrial genes and gene products and the absence of Neanderthal/Denisovan mitochondrial lineages in the human gene-pool, the process of selection on mitonuclear genes in humans has been possibly magnified. Different alleles in mitonuclear genes are present in archaic and modern humans, but the functional significance of these differences, if any, is still unknown (23,24).

The surge in sequencing capacity has empowered the identification of genetic variants in individuals affected by mitochondria-related pathologies. However, when novel mutations are identified, it is necessary to confirm that these variants are indeed the cause of the disorder and not just functionally silent polymorphisms. In such cases a model system can be useful to "validate" mutations, and can additionally provide insights into the molecular role played by these variants, the ultimate goal being the development of clinical strategies to rescue their detrimental effects. The budding yeast *Saccharomyces cerevisiae* has been extensively used as a model to validate the functional relevance of variants detected in mitonuclear genes (25–27). This single-celled eukaryote, owing to its good fermentation capacity, can endure the loss of mitochondrial oxidative phosphorylation (OXPHOS) as long as a fermentable carbon source is supplied in the growth medium. Consequently, in the yeast model, variants affecting mitochondrial function result in a respiratory deficient phenotype, thus allowing to easily evaluate the impact of mitonuclear gene variants on mitochondrial activity by comparing oxidative vs. fermentative growth and measuring oxygen consumption rate (OCR) of *ad hoc* yeast strains expressing either the allelic variant or the wild type. Depending on the ability of the human cDNA to complement the yeast mutant deleted in the orthologue of the human

gene, homologous, heterologous or chimeric complementation approaches can be used to evaluate the pathogenicity of genetic variants (27).

Our aim is to delve into the functional relevance of archaic-specific mitonuclear variants and identify differences between humans and Archaics. With this focus, we scanned available Neanderthal and Denisovan genomes to identify archaic nucleotide substitutions potentially altering protein functionality in mitonuclear genes. Then, we conducted a comparative analysis of the functional activity of the human and archaic alleles leveraging the genetic tools offered by yeast. The combination of *in silico* predictions and *in vivo* experiments yielded the first direct evidence for functionally divergent archaic-specific variants within mitonuclear genes.

Results

Identification of archaic-specific mitonuclear variants potentially impacting protein functionality

To explore the genomic landscape of mitonuclear variation among archaic and modern humans, we analysed four high-quality archaic genomes available for three Neanderthals - Altai, Vindija, and Chagyrskaya (28–30) - and one Denisova (31) (Figure 1). The three Neanderthal individuals span a wide geographical and temporal landscape of around 70,000 years across Eurasia (Altai and Vindija are, respectively, the eldest and the youngest, with the former dating back to around 120-130 kya and the latter to 50kya), thus allowing the sampling, to some extent, of Neanderthal chrono-spatial genetic variation (32–34). Conversely, only a single high coverage whole genome Denisovan sample has been generated so far, originating from the same Siberian cave of the Altai Neanderthal genome and dating back to 74-82kya (31) (Figure 1B). Over the past decade, genomic analyses of such specimens have offered an unparalleled perspective on the demographic and adaptive history of the genus *Homo*, providing significant insights into the evolutionary implications of gene-flow between archaic and modern humans (15). We therefore set out to explore the genetic variations harboured in these three species focusing on genes with strong support of mitochondrial localization according to the Human MitoCarta3.0 inventory (35), a complex of genes that has hitherto been marginally explored in the context of Neanderthal introgression (23). In doing so, we found a total of 48,709 variants differing between Neanderthal/Denisovans and AMHs (Suppl. File 1). Of these, 604 were predicted to alter the final protein through aminoacid replacements (missense variants) or by shortening its sequence (i.e., stop gained, splice donor and acceptor). We then focused on derived variants present in homozygosis in either all the archaic genomes or at least in all Neanderthal individuals (28 and 34, respectively). Notably, 47 of these 62 mutations have been predicted by CADD to be phenotypically influential (CADD score above 10) and 26 of these 47 are also rare in modern humans (the alternative and archaic allele shows a frequency below 2% in gnomAD database). The final list included only missense variants, located in 26 different mitonuclear genes (Suppl. File 2).

Assortment of archaic variants within genes compatible for evaluation in *S. cerevisiae*

Among the 26 genes identified above we selected those that presented the necessary requirements to perform a functional analysis in yeast. We initially checked which of these mitonuclear genes had an orthologue in *S. cerevisiae* using the information included in the

YeastMine and Biomart databases (Suppl. Table 1). We included all the genes that were present in at least one of the explored databases but only if a single yeast homologous was identified. Of the 11 and 15 mitonuclear genes retrieved with fixed variants in the three Neanderthals and in all four non-human hominins, respectively, five in each group presented a single ortholog in yeast. We manually inspected this list and excluded three additional genes: *DHOD*, as the suggested yeast ortholog *URA1* is not localised in the mitochondria, *PDSS2*/*COQ1* and *SLC25A23*/*SAL1* as the two yeast genes were indicated as orthologs also for other human genes (*PDSS1* and *SLC25A25*, respectively) (37,38). Finally, to facilitate the functional evaluation of the impact of the selected variants on mitochondrial function, we further focused on genes whose deletion was previously shown to generate OXPHOS-related phenotypes by consulting the Saccharomyces Genome Database (SGD; (https://www.yeastgenome.org/). In particular, we focused on genes affecting growth on oxidative carbon sources, respiratory activity and/or mtDNA stability. Of the remaining seven orthologous genes, only four were retained, all characterised by having an impact on oxidative growth and/or on respiratory activity (Suppl. Table 1). The final set comprised four mitonuclear genes each harbouring one fixed derived archaic variant: *FDXR* and *LYRM7,* which presented a variant derived in all the four archaic genomes, and *COQ2* and *YARS2,* with variants derived only in all of the three Neanderthal genomes (Figure 2, Suppl. Table 1, Suppl. File 2). Two of these genes are involved in the synthesis or assembly of the mitochondrial respiratory chain complexes: *LYRM7* (LYR motif-containing protein 7) encodes a complex III (CIII) assembly factor involved in the UQCRFS1 insertion step that promotes the formation of the mature and functional CIII complex (39), while *COQ2* encodes a polyprenyl transferase involved in the synthesis of CoQ (ubiquinone) that serves as a redox carrier in the mitochondrial respiratory chain (40). The other two genes play a more general role within the mitochondria: *FDXR* encodes the unique human mitochondrial ferredoxin reductase involved in the biosynthesis of iron-sulphur (Fe-S) clusters and heme formation (41–43), while *YARS2* encodes the mitochondrial tyrosyl-tRNA synthetase, an essential enzyme for the mitochondrial translation, that catalyses the covalent binding of tyrosine to its cognate tRNA (44). YARS2 and LYRM7 are localised in the mitochondrial matrix, while COQ2 and FDXR are associated with the mitochondrial inner membrane (39,45–48).

Protein sequence conservation and *in silico* stability prediction

The positions of the four aminoacidic changes are indicated in Figure 2, which also includes the location of motifs and functional sites, whose coordinates were obtained from the literature (49–52) or directly inferred using MITOFATES (53). We further explored the degree of nucleotide and aminoacid conservation across primates focusing on the regions harbouring the archaic-specific variants (Figure 2; Supplementary Figure 1).

The archaic variant in the protein FDXR is located at position 470 within the final FAD-binding domain and encodes the aminoacid Serine instead of Alanine (Ala470Ser, Figure 2 and Supplementary Figure 2). All the aligned protein sequences present an Alanine in this position, except the Strepsirrhine *Microcebus murinus* which harboured the same aminoacid present in the archaic FDXR. Alanine and Serine differ from each other for the presence of a hydroxyl group, which changes the non-polar Alanine to the polar Serine. The residue in the archaic protein LYRM7 involves the substitution at position 34 of an Isoleucine with another aliphatic, apolar aminoacid (Leucine). All primates present an Isoleucine at this position, except the gibbon *Hylobates moloch* which bears a Valine, another aliphatic, apolar aminoacid (Figure 2; Supplementary Figure 3).

Position 35 of YARS2, which differs between humans and Neanderthals, is characterised in all tested primates by the presence of hydrophobic, non polar residues (Alanine, Proline and Valine), except for Neanderthals, where a hydrophilic, negatively charged residue is present (Aspartic acid; Figure 2; Supplementary Figure 4).
The substitution at position 192 of the protein COQ2 found in Neanderthals involves the non-polar aminoacid Alanine instead of the polar, uncharged Threonine. Only the common marmoset (*Callithrix jacchus*) shares the same residue found in Neanderthals, while all the other primates carry the aminoacid found in humans (Figure 2; Supplementary Figure 5).
Overall, the archaic residues are either shared with one or more primate species or equivalent in their chemical properties to variants present in other primates, except for the Neanderthal YARS2 residue, unique across primates and chemically different from all the others.

We further investigated how archaic-specific missense variants potentially affected protein stability by using the softwares DDGun Seq (54) and ACDC-NN Seq (55). Both methods are designed to predict the unfolding free energy difference between the wild type and mutant protein from the protein sequence. While CADD has predicted the four genetic variants to rank in the top 10% of substitutions with a functional impact, their inferred stability change falls within the neutrality range [-0.5:0.5]. This implies that the protein residue variations in the archaic versions of these proteins do not significantly alter the free energy of the final structure (Suppl. Table 2).

Population distribution and phenotypic associations of archaic mitonuclear variants

We investigated the distribution of the four selected archaic mitonuclear variants within human populations by relying on different genomic resources, such as the 1000 genomes project and gnomAD (Suppl. File 3). Given the rarity of such variants (the allele frequency below 2% in modern humans was one of the filtering criteria), gnomAD turned out to be the most informative repository for occurrence in humans. For the same reason, it proved difficult to link each variant to a specific phenotypic effect. Indeed, our scan for phenotypic associations recovered data only for the archaic variants in *LYRM7* and *YARS2* (Suppl. File 4). rs200336982 in *LYRM7* was found mildly associated with Neutrophil percentage of white cells and different conditions related to the musculoskeletal system and connective tissue (even though none of the signals was strong enough to reach the GWAS threshold; Suppl. File 4). Several associations were recovered for the archaic variant rs149447502 in *YARS2*, including plasma measurements (mean platelet volume), musculoskeletal and heart diseases, mental health (anxiety and depression), as well as other conditions related to the nervous system (e.g., the strongest signal was “Extrapyramidal and movement disorders”) and other biological processes (e.g., “Duration of moderate activity”; Suppl. File 4).

Functional evaluation of archaic mitonuclear variants in yeast

We proceeded in our investigation of the functional impact of archaic variants by transforming yeast strains deleted for the human orthologs of the four selected genes. To make the functional analysis of the variants more meaningful and informative, the yeast models were constructed using the heterologous complementation approach in which the variant is directly introduced into the human cDNA and then its effects are evaluated in the corresponding null yeast mutant. The success of this approach relies on the ability of human cDNA to

complement, at least partially, the lack of function of the yeast null mutant. Such information was available only for two of the four genes selected above. Whereas the human cDNAs encoding FDXR and COQ2 proteins are in fact known to be able to complement the corresponding null yeast mutants *arh1*Δ (56,57) and *coq2*Δ (52), such information was not available for *LYRM7* and *YARS2* cDNAs. For these two genes, we initially verified this capacity and then proceeded with the functional testing as for the other two genes, as described below (Supplementary Information, Supplementary Figure 6).

The hcDNA of the four selected genes (*FDXR*, *COQ2, LYRM7*, *YARS2*), cloned in appropriate plasmids, were mutagenized to introduce the archaic variants of interest. The plasmids containing the resulting acDNAs were then each introduced in yeast strains deleted for the orthologous gene and compared for their growth ability on oxidative carbon sources and oxygen consumption rate (OCR) with deleted strains transformed with the hcDNA and with empty vector (Figure 3A, B; Supplementary Information). The complementation of the yeast deletion with human cDNAs did not fully rescue the OCR phenotype (40-75% vs the wild-type strain; Figure 3B). This is not surprising, since human cDNAs often manage to complement the function of their yeast ortholog only in part (58). Human and archaic versions of *FDXR, COQ2* and *LYRM7* did not show major differences in oxidative growth. The archaic variant in *LYRM7* generated a 20% OCR increase, acting as a moderate hypermorphic allele whereas ac*COQ2* is associated with a 30% reduction of OCR. On the contrary, the strain expressing the archaic variant in the *YARS2* gene appeared to be severely affected in both oxidative growth and OCR (Figure 3A, B).

Given the role of FDXR in assembling Fe-S clusters, we tested the activity of the enzyme aconitase, a Fe-S cluster containing enzyme, as a proxy for the overall impact on similar enzymes. Notably, the aconitase enzymatic activity was reduced by 15% in the *arh1*Δ/*FDXR*^{A470S} strain when compared to the strain expressing *FDXR* hcDNA suggesting a slight impairment in Fe-S cluster biogenesis in the “archaic” strain (Figure 3C). Yeast strains with defective Arh1 have been shown to present dysregulated iron metabolism while iron overload has been observed in fibroblasts of affected FDXR individuals (56). Considering that increased iron accumulation in yeast led to an increased sensitivity to this ion (27,59,60), we evaluated the inhibition of cellular growth in the *arh1*Δ/*FDXR*^{A470S} strain by increasing iron concentrations in the medium. The *arh1*Δ strain expressing archaic protein displays a slight growth defect after exposure to iron, suggesting that the intracellular iron concentration could be altered in *FDXR*^{A470S} strains (Supplementary Figure 7).

To evaluate if the increased oxygen consumption observed in the *mzm1*Δ/*LYRM7*^{I34L} archaic strain was associated or not with an increase in ATP production, OCR was measured in the presence of the mitochondrial uncoupler carbonyl cyanide 3-chlorophenylhydrazone (CCCP). In both strains *mzm1*Δ/*LYRM7* and *mzm1*Δ/*LYRM7*^{I34L}, the presence of CCCP leads to a doubling of oxygen consumption indicating that also the increased percentage of OCR observed in the strain expressing the archaic variant is coupled to ATP production (data not shown).

The strain *msy1*Δ/*YARS2*^{A35D} showed a severe OXPHOS growth defect, indicating that the A35D variant significantly impacts mitochondrial functionality (Figure 3A). Consistent with this result, the oxygen consumption rate was reduced by about 50% in *msy1*Δ/*YARS2*^{A35D} when compared to the deleted strain transformed with the hc*YARS2* (Figure 3B). As *YARS2* is involved in the mitochondrial protein synthesis, we measured the steady state level of Cox2 protein, a subunit of CIV (cytochrome *c* oxidase), to investigate if the observed OXPHOS defective phenotype was due to a reduction in the synthesis of proteins encoded by the mitochondrial genome. As expected, no Cox2 signal was detected in *msy1*Δ strain while in

*msy1*Δ/*YARS2*^{A35D} Cox2 level, normalised for Por1, was decreased by 50% when compared to *msy1*Δ strain expressing *YARS2* hcDNA (Figure 3D, Supplementary Figure 8). Altogether these results support the hypothesis that the A35D archaic variant of the mt tyrosyl-tRNA synthetase affects mitochondrial protein synthesis resulting in a severe dysfunction of mitochondria. Similar phenotypes are usually scored as pathological when mitonuclear variants discovered in patients are tested in yeast (27,61,62).

Molecular stability and processing of *YARS2*^{A35D}

Of the four archaic variants tested, *YARS2*^{A35D} was the one associated with the most substantial alteration of mitochondrial functionality when compared to the human allele (Figure 3). Eubacterial and mitochondrial tRNATyr have a G1-C72 pair, which is involved in the binding by their cognate tRNA synthetase, contrary to cytoplasmic tRNATyr that has a C1-G72 pair recognized by the corresponding cytoplasmic tyrosyl-tRNA synthetase (63). Human YARS2 can aminoacylate tRNATyr with either G1-C72 or C1-G72, indicating that the specific kind of pair is dispensable for the aminoacylation (64). The binding and catalytic activity involves two helices of 14 amino acids, known as clusters 1 and 2, which constitute the active site and require the presence of A73 to properly function (49,65). Both C1-G72 and A73 are present in the yeast mt-tRNATyr and clusters 1 and 2 have approximately 80% similarity between the yeast and human mitochondrial tyrosyl-tRNA synthetases (Supplementary Figure 9). To date, specific information on the binding and catalysis of mt-tRNATyr aminoacylation in yeast is missing. However, the structure of human YARS2 is highly similar throughout its whole sequence to that of *Neurospora crassa*, the only published structure of a fungal mitochondrial tyrosyl-tRNA synthetase and which shares 60% similarity with the *S. cerevisiae* one (49,66).

In YARS2, Ala35 is located in the very beginning of the catalytic domain of the protein. The predicted structure places the residue on the outer surface of the protein, far away from the catalytic site containing the tyrosyl adenylase (Figure 2B and Figure 4A). YARS2 operates as a dimer, as other aminoacyl transferases (45), but the Ala35 does not lay close to the dimerisation site (Figure 4A). Overall these observations suggest that the archaic residue might not alter the catalytic activity of the enzyme. We therefore evaluated if the Neanderthal aminoacidic variant might be responsible for the observed phenotypes by altering either protein stability and/or maturation. Steady-state protein levels of the YARS2 archaic and human variants were assessed by Western Blot. Immunoblot analysis of YARS2 performed in the strain expressing the hcYARS2 revealed a single band compatible with the size of the mature enzyme, whereas in that expressing acYARS2 two bands were observed (Figure 4B). Of the two bands, one had the same molecular weight as the one present in the yeast strain expressing hcYARS2, the other displayed instead a higher molecular weight, compatible with the unprocessed YARS2 containing the MTS. The ratio between the two forms of YARS2 was about 2:1, which suggests that the archaic variant partially affected the processing of the YARS2 protein (Figure 4B). The overall amount of YARS2 (precursor and mature) in the acYARS2 strain is comparable to the amount present in the strain hcYARS2, suggesting that protein stability is similar between the two forms (Figure 4C). On the contrary, the observed accumulation of the YARS2 precursor is unique to the strain expressing the archaic variant A35D, indicating the presence of Asp35 may affect the removal of the MTS. A reduced amount of the mature, fully formed YARS2 enzyme could possibly be related to the mitochondrial phenotypes reported for the acYARS2 yeast strain.

Discussion

The hybridisation between humans, Neanderthals and Denisovans has left signatures in the genomes of modern humans. However, the non-random distribution of archaic introgressed regions across the genome has pointed to quick selection-related processes shaping the chromosomal introgression landscape across populations (67). The removal of specific regions has been linked to genomic incompatibility between archaic regions and modern human genomic background and/or more effective selection operating on archaic regions once introduced in the larger human population (10–13). Both scenarios imply the presence of functional differences between human and archaic alleles, resulting in differential fitness and selection (14,15,68). Phenotypic differences between human and archaic alleles have been reported and positively selected introgressed archaic regions have been identified (69–72). However, the focus has been so far on the characterisation of the phenotypic impact of either human-specific changes or positively selected archaic introgressed variants. Less is known about the functional impact of poorly introgressed archaic-specific genetic variants and their characterisation is essential to fully appreciate their role in shaping the observed patterns of introgression (24). Yet, the segregation at low allele frequencies of negatively selected variants hinders the possibility of confidently associating such variants to phenotypic effects using standard association studies.

The bioinformatic scan here performed pinpointed a set of potential functionally divergent fixed archaic-derived variants present in mitonuclear genes, a group of genes previously shown to be less introgressed in modern humans (23). The extent of this underrepresentation was found significant but small for Neanderthals and not significant for Denisovans, the latter possibly the result of a more conservative identification of Denisovan introgressed regions. These results suggested that human-archaic mitonuclear differences might have had a minimal, but not negligible, impact on affecting the Archaic introgression landscape. However, the authors stressed that "...the lack of strong genome-wide signatures of mitonuclear incompatibilities does not preclude the possibility of strong effects on individual loci", a prediction that is in agreement with our results. In fact, our functional comparison of archaic and human mitonuclear variants in the model organism *Saccharomyces cerevisiae* highlighted differences in all the tested genes, showing variation in the extent of their phenotypic impact. Three, *FDXR*, *LYRM7* and *COQ2* generated mild phenotypes possibly with limited impact on individual fitness: an imbalance of iron homeostasis, an increased oxygen consumption and a slight decrease in OCR. The evolutionary significance of the observed phenotypic differences, if any, remains to be fully evaluated. Among these, the hypermorphic *LYRM7* archaic allele is particularly interesting given the uneven expression profile of this gene across human tissues, particularly high in the brain (73). It is tempting to speculate that the higher oxygen consumption associated with the LYRM7 archaic variant might have contributed to drive the removal of the Neanderthal allele from the human genome. However further investigations are necessary to test this hypothesis, in particular by evaluating the impact of the archaic LYRM7 allele in a more human-like genomic context and its overall impact on the whole organism.

Of the four tested variants, the one in YARS2 generated a mitochondrial functionality impairment of such a degree that if identified in a patient would be classified as pathogenic when similarly tested in yeast (62,74). The presence of such a variant in the Neanderthal genome could be explained by considering this variant as mildly deleterious and assuming a less efficient action of selection given the smaller population size of Neanderthals (10,15,75).

The human and archaic versions of YARS2 do not show any significant difference in terms of stability. However, a substantial fraction of the protein was not processed in the mature form, as about one third of the total amount of the protein was present in its precursor form in the yeast strain expressing the archaic allele of *YARS2*. It cannot be excluded that the detrimental effect observed in yeast could be not so severe in humans. In fact, the import and processing machinery is not fully conserved between the two species (76) and therefore the mature form of the Neanderthal YARS2 might be generated more efficiently on a human than a yeast genomic background. Future investigations in human cell lines and/or other higher eukaryotes will be able to assess the functional impact and the phenotypic relevance of the *YARS2*^{A35D} variant in that genetic context.

As our goal was the identification of mitonuclear variants differing in functionality across humans and Archaics, we did not engage in the additional characterization of the molecular basis of the incomplete maturation of the archaic YARS2 protein. However, it is possible to speculate that in addition to, or in alternative of, a less efficient selective pressure operating in Neanderthals, a compensatory mutation (or more) might be present in *H. neanderthalensis* on any of the proteins operating in the mitochondrial processing and import of YARS2. Of the 26 proteins identified in our original search for variants to be tested, DNAJC4 is the only one potentially involved in the matrix import and homeostasis of mitochondrial proteins (35) (Suppl. File 2). DNAJC4 is in fact a) bound to the internal mitochondrial membrane (77), b) indicated as associated with mitochondrial protein import, sorting and homeostasis (35) and c) with an in-silico prediction of a role in Response To Unfolded Protein (78). In addition, among the molecular interactors of DNAJC4 listed in the Biogrid database are Mortalin/HSPA9 and GRPEL1, two members of the PAM complex localised in the matrix and directly involved in the translocation of the imported proteins into the matrix (79,80). The co-localisation and the interaction with these proteins, as well as the sharing of the DNAj motif with other molecules known to be involved with the PAM complex (DNAJC15 and DNAJC19) (81), suggest the involvement of DNAJC4 in the protein import in the mitochondrial matrix, potentially of YARS2 too. The expression profile of DNACJ4 across human tissues identifies a preferential expression in the testis and suggests its potential relevance for spermatogenesis, a biological process particularly targeted by purifying selection operating on Neanderthal introgressed regions (19,82). The future characterization of the mitochondrial role played by DNAJC4 will enable a more direct evaluation of the functional relevance of the associated Neanderthal variant and its potential involvement with the mitochondrial import of YARS2. Additional scenarios might explain the persistence of the *YARS2*^{A35D} variant in Neanderthals: mild deleterious variants can survive in small populations as the result of drift dominated dynamics, something that has been suggested as possibly relevant in explaining the genomic landscape of Neanderthal introgression in humans (11,83). Last but not least, a genomic background different from yeast might provide the context for a mitigated phenotypic effect, less extreme in its impact on the fitness of individuals and therefore less exposed to purifying selection. It is worth stressing here that yeast served as an efficient model to sieve across candidates and identify genes and variants of interest. However, predicting the full impact of the YARS2 archaic variant when placed onto a full human genomic background poses challenges. In this context, our findings establish a foundation for further characterising the molecular and phenotypic impact of this variant in humans. Given the rarity of *YARS2*^{A35D} in our species, genetic engineering of cell lines is necessary to introduce the archaic variant into the genome and investigate related molecular and functional changes.

Finally, the approach employed, combining bioinformatic analysis for identifying potential variants and *in vivo* evaluation using yeast to assess their impact on mitochondrial

functionality, proves to be a resource-efficient approach to screen for functional variants differing across evolutionary close species and populations. However, despite its many advantages, it is worth emphasising that the expression of mammalian genes in yeast has limitations. First, being *S. cerevisiae* a single cell organism, the effect of heterologous expression cannot be analysed at the scale of tissues, organs or complex multicellular organisms. Secondly, and more generally, it is noteworthy that, when heterologous expression is carried out in a divergent model system such as yeast, complementation is partial in most cases. Epistatic, additive or synergistic interactions can be present between different interactors, shaping phenotypes. Such interactors might be present in humans but not in yeast or, if present, they may have a different sequence or structure. Consequently, when a human protein is expressed in yeast, it is possible that amino acids that are critical for the interaction with other human macromolecules have a minor role in the interaction with the yeast orthologs, and vice versa. Despite these obvious limitations, the screening strategy presented here can help filtering mitonuclear variants to identify those worth further exploring in an appropriate genomic background and could be extended beyond humans for comparisons between other primates and more broadly across mammals.

Materials and Methods

Identification of mitonuclear archaic-specific variants

In order to identify archaic-specific variants present in mitonuclear genes, we leveraged the whole genome sequences available for three Neanderthals - Altai, Vindija, and Chagyrskaya (28–30) - and one Denisovan individual (31). In particular, we focused on variants located within the 910 mitonuclear genes with reported evidence of being mitochondrial proteins (i.e., Tmito) from the Human MitoCarta3.0 (35). We also retrieved a high-quality catalogue of genes essential for the human oxidative phosphorylation pathway (OXPHOS) from the literature (84). Using the softwares SnpEff and vcfanno, we annotated single nucleotide substitutions on the canonical transcript of each mitonuclear gene and added clinical significance information (e.g., CADD v1.4, ClinVar v20180805, Spidex), and population frequencies from gnomAD database (version r2.1) (85–88). In order to access archaic specific variants functionally interesting for mitochondria, we performed a series of filtering steps (Suppl. File 1). We initially selected the variants in mitonuclear genes different between humans and Neanderthals/Denisovans. Among these, we then selected the variants with a potential functional impact, by focusing on protein altering variants (i.e., missense, stop gained, splice donor and acceptor) and then filtered these by selecting positions that were derived in the archaic genomes (the derived allele information was retrieved from dbNSFP database v2.9.3 (89)). Finally, we retained only derived variants present in homozygosis in all the archaic genomes, or just in all of the three Neanderthal genomes.

We further focused on variants with a putative functional impact, by selecting those with a phred-scaled CADD score higher than 10, meaning that they are predicted to be within the 10% substitutions which are most likely to have a functional effect. Finally, we selected the variants with a frequency below 2% in the gnomAD database (considering the “controls” subsection of both the exome and the genome datasets). We chose this frequency filter in order to focus on the variants that arose along the archaic lineage while considering also the admixture events that occurred between AMHs and Neanderthals which led to ~ 2% of archaic introgression within modern Eurasian individuals (3). Indeed, in the presence of purifying

selection, we expect that less than 2% of modern humans would have inherited such fixed and archaic-specific variants, on average. The final list of variants (26) is reported in Suppl. File 2.

Yeast functional comparison of human and archaic mitonuclear variants

The mitonuclear variants identified as described above were then screened for being compatible with their functional evaluation in *Saccharomyces cerevisiae.* The first selection was made by checking which of these mitonuclear genes had an orthologue in yeast, a prerequisite for assessing the impact of variants in the model considered, by searching in the YeastMine and Biomart databases (90,91). We then retained only genes whose deletion or missense mutations in yeast have been previously shown to be associated with OXPHOS-related phenotypes (https://www.yeastgenome.org/) such as a growth defect on oxidative carbon sources and/or a decrease in respiratory activity and/or an increase in mutability/stability of mitochondrial DNA. These phenotypes are relatively simple to test when performing functional analyses in yeast (27). The final set of genes (4) is reported in Suppl. Table 1.

S. cerevisiae strains deleted in the orthologous human genes identified above were either already available or were constructed in this work (see Supplementary Information). The coding sequences of the human cDNAs (hcDNAs) encoding for *LYRM7*, *YARS2* and *COQ2* were PCR-amplified and cloned in appropriate expression vectors as indicated in Supplementary Information. A vector containing *FDXR* hcDNA was already available (Supplementary Information; (56)). Archaic missense variants were inserted in the corresponding hcDNA by PCR QuikChange™(Agilent) using KOD Hot Start DNA Polymerase (Merck, US) and appropriate primers (Suppl. Table 3). All archaic-containing cDNAs (acDNAs) constructs were verified by Sanger sequencing. The vectors containing hcDNAs or acDNAs were transformed into the appropriate yeast strain, i.e., a yeast strain deleted in the yeast gene orthologous of the human one, using the lithium acetate, single-stranded DNA, polyethylene glycol method (92). When required for strain viability (*arh1*Δ) or mitochondrial DNA maintenance (*msy1*Δ *and coq2*Δ), the plasmid containing the wild type gene was lost through plasmid-shuffling on 5-FOA medium (93).

The ability of the different cDNAs to complement the respiratory deficient phenotype of the yeast null mutants was tested by spot assay analysis, evaluating the ability of growth in media containing oxidative carbon sources such ethanol or glycerol as detailed in Supplementary Information. Oxygen consumption rate was measured on whole cells at 30°C using a Clark-type oxygen electrode (Oxygraph System Hansatech Instruments England) (94,95). Aconitase activity was tested by the aconitase-isocitrate dehydrogenase-coupled assay (60). Cox2 and YARS2 steady state levels were measured relative to Por1 through Western blot after incubation of whole protein extracts with appropriate secondary antibodies, as reported in Supplementary Information. Differences across strains were tested via one-way analysis of variance (ANOVA) followed by Bonferroni's post hoc test. P-values below 0.05 were considered statistically significant.

Multispecies sequence comparisons (genes and proteins)

We used the Ensembl section “Phylogenetic Context" in order to perform multiple alignment analyses over 21-long genomic regions harbouring the target variants. The sequence alignments were performed among the 10 primates considered within the alignment called “10

primates EPO". Then, we used the online tool WebLogo 3 to graphically represent the DNA sequence alignments (Supplementary Figure 1). Given that this function has been disabled in Ensembl for the GRCh37 genomic assembly, we reported in Supplementary Figure 1 the DNA sequence alignments using the GRCh38 genomic build.

Then, we investigated the multiple sequence alignment on a larger set of primates at the protein level using Clustal Omega (96). We represented protein alignments encompassing the variants using WebLogo 3 in Figure 2B and MView in Supplementary Figure 2, Supplementary Figure 3, Supplementary Figure 4 and Supplementary Figure 5.

Protein stability predictions, protein domains and YARS2 3D structure

We further investigated the protein stability for the selected genes using the softwares DDGun Seq (54) and ACDC-NN Seq (55), which predict the free energy changes upon specific protein residue variations.

Protein domains were either retrieved from the literature (49–52) or predicted using MITOFATES (53). Finally, they were represented using Prosite My Domains (97). The structure of human YARS2 in complex with a tyrosyladenylate analogue has been resolved (PDB number: 2PID), but the structure lacks the first aminoacids of the mature enzyme, including Ala35 (49). Starting from this structure, the structure of whole YARS2 was predicted with MiSynPat at misynpat.org (98). The first 30 amino acids, corresponding to the mitochondrial targeting signal (MTS) removed during the processing of YARS2 in mitochondria, were removed, and the structure was superimposed with the resolved structure using the Magic fit tool of Swiss PDB viewer. The final structure was visualised with Rasmol.

Population distribution of putatively fixed archaic mitonuclear variants

We retrieved the frequencies of the selected archaic specific variants in modern human populations from the repositories 1000 Genomes Project (99) and gnomAD v2.1.1 (85,99). The allele counts and allele frequency of archaic-specific alleles from these repositories are reported in Suppl. File 3 reports (note that in each case the archaic allele matches the alternative alleles by definition).

Phenotypic associations of archaic mitonuclear variants in humans

We scanned the literature for known phenotypic associations using public catalogues such as Gene ATLAS (100), Open Targets Genetics (101) and the AstraZeneca PheWAS Portal (102). We also checked all clinical submissions and information in ClinVar and Varsome related to our variants of interest. The complete list of associations is reported in Suppl. File 4.

Acknowledgements

This work benefited from the equipment and framework of the COMP-HUB and COMP-R Initiatives, funded by the 'Departments of Excellence' program of the Italian Ministry for University and Research (MIUR, 2018-2022 and MUR, 2023-2027). This work was supported by Programma Nazionale della Ricerca PNR 2021-2027 e PON "Ricerca e Innovazione" 2014-2020 – progetti di ricerca su tematiche "Innovazione" e "Green" [to S.A.]; and the Italian Ministry of Health, [RF-2016-02361241 to C.C.B. and P.G.]; the Ph.D. fellowship of A.I.G. was co-funded by the University of Parma and Italian Telethon Foundation [GGP19287A to A.I.G.].

This research additionally benefits from the HPC (High-Performance Computing) facility of the University of Parma, Italy; the Ph.D. program in Biotecnologie e Bioscienze (University of Parma). The authors would like to thank Cristina Dallabona for comments and suggestions, Antonietta Cirasolo for technical assistance.

Conflict of Interest Statement

Authors declare no conflict of interest.

Data availability

We downloaded the VCF of each individual from: http://cdna.eva.mpg.de/neandertal/Vindija/VCF/Altai/ (Altai Neanderthal), http://cdna.eva.mpg.de/neandertal/Vindija/VCF/Vindija33.19/ (Vindija Neanderthal), http://ftp.eva.mpg.de/neandertal/Chagyrskaya/VCF/ (Chagyrskaya Neanderthal); http://cdna.eva.mpg.de/neandertal/Vindija/VCF/Denisova/ (Denisova). High-quality positions of each sequence have been downloaded from http://cdna.eva.mpg.de/neandertal/Vindija/FilterBed/ and http://ftp.eva.mpg.de/neandertal/Chagyrskaya/FilterBed/.

References

1. Higham, T., Douka, K., Wood, R., Ramsey, C.B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruíz, C., *et al.* (2014) The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, **512**, 306–309.
2. Villanea, F.A. and Schraiber, J.G. (2019) Multiple episodes of interbreeding between Neanderthal and modern humans. *Nat Ecol Evol*, **3**, 39–44.
3. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710–722.
4. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., *et al.* (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, **468**, 1053–1060.
5. Jacobs, G.S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C.C., Lawson, D.J., Mondal, M., Pagani, L., Ricaut, F.-X., Stoneking, M., *et al.* (2019) Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*, **177**, 1010–1021.e32.
6. Dannemann, M. and Racimo, F. (2018) Something old, something borrowed: admixture and adaptation in human evolution. *Curr. Opin. Genet. Dev.*, **53**, 1–8.
7. Kuhlwilm, M., Gronau, I., Hubisz, M.J., de Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H.A., Lalueza-Fox, C., de la Rasilla, M., *et al.* (2016) Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature*, **530**, 429–433.
8. Sánchez-Quinto, F. and Lalueza-Fox, C. (2015) Almost 20 years of Neanderthal palaeogenetics: adaptation, admixture, diversity, demography and extinction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **370**, 20130374.
9. Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., *et al.* (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, **531**, 504–507.
10. Castellano, S., Parra, G., Sánchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., *et al.* (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 6666–6671.
11. Juric, I., Aeschbacher, S. and Coop, G. (2016) The Strength of Selection against Neanderthal Introgression. *PLoS Genet.*, **12**, e1006340.
12. Yair, S., Lee, K.M. and Coop, G. (2021) The timing of human adaptation from Neanderthal introgression. *Genetics*, **218**.
13. Findley, A.S., Zhang, X., Boye, C., Lin, Y.L., Kalita, C.A., Barreiro, L., Lohmueller, K.E., Pique-Regi, R. and Luca, F. (2021) A signature of Neanderthal introgression on molecular mechanisms of environmental responses. *PLoS Genet.*, **17**, e1009493.
14. Dannemann, M. and Kelso, J. (2017) The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. *Am. J. Hum. Genet.*, **101**, 578–589.
15. Reilly, P.F., Tjahjadi, A., Miller, S.L., Akey, J.M. and Tucci, S. (2022) The contribution of Neanderthal introgression to modern human traits. *Curr. Biol.*, **32**, R970–R983.
16. McArthur, E., Rinker, D.C. and Capra, J.A. (2021) Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nat. Commun.*, **12**, 4481.
17. Brand, C.M., Colbran, L.L. and Capra, J.A. (2022) Predicting Archaic Hominin Phenotypes from Genomic Data. *Annu. Rev. Genomics Hum. Genet.*, **23**, 591–612.
18. Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., Crosslin, D.R., Hebbring, S.J., Jarvik, G.P., Kullo, I.J., *et al.* (2016) The phenotypic legacy of admixture between modern humans and Neandertals. *Science*, **351**, 737–741.
19. Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N. and Reich, D. (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, **507**, 354–357.
20. Vernot, B. and Akey, J.M. (2014) Resurrecting surviving Neandertal lineages from modern human genomes. *Science*, **343**, 1017–1021.

21. Vernot, B., Tucci, S., Kelso, J., Schraiber, J.G., Wolf, A.B., Gittelman, R.M., Dannemann, M., Grote, S., McCoy, R.C., Norton, H., *et al.* (2016) Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science*, **352**, 235–239.
22. Jégou, B., Sankararaman, S., Rolland, A.D., Reich, D. and Chalmel, F. (2017) Meiotic Genes Are Enriched in Regions of Reduced Archaic Ancestry. *Mol. Biol. Evol.*, **34**, 1974–1980.
23. Sharbrough, J., Havird, J.C., Noe, G.R., Warren, J.M. and Sloan, D.B. (2017) The Mitonuclear Dimension of Neanderthal and Denisovan Ancestry in Modern Human Genomes. *Genome Biol. Evol.*, **9**, 1567–1581.
24. Trujillo, C.A., Rice, E.S., Schaefer, N.K., Chaim, I.A., Wheeler, E.C., Madrigal, A.A., Buchanan, J., Preissl, S., Wang, A., Negraes, P.D., *et al.* (2021) Reintroduction of the archaic variant of in cortical organoids alters neurodevelopment. *Science*, **371**.
25. Barrientos, A. (2003) Yeast Models of Human Mitochondrial Diseases. (2003) Yeast Models of Human Mitochondrial Diseases. *IUBMB Life* , ***55***, 89–95.
26. Baile, M.G. and Claypool, S.M. (2013) The power of yeast to model diseases of the powerhouse of the cell. *Front. Biosci.* , **18**, 241–278.
27. Ceccatelli Berti, C., di Punzio, G., Dallabona, C., Baruffini, E., Goffrini, P., Lodi, T. and Donnini, C. (2021) The Power of Yeast in Modelling Human Nuclear Mutations Associated with Mitochondrial Diseases. *Genes* , **12**.
28. Mafessoni, F., Grote, S., de Filippo, C., Slon, V., Kolobova, K.A., Viola, B., Markin, S.V., Chintalapati, M., Peyrégne, S., Skov, L., *et al.* (2020) A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 15132–15136.
29. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–49.
30. Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., *et al.* (2017) A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, **358**, 655–658.
31. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., *et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science*, **338**, 222–226.
32. Peyrégne, S., Slon, V., Mafessoni, F., de Filippo, C., Hajdinjak, M., Nagel, S., Nickel, B., Essel, E., Le Cabec, A., Wehrberger, K., *et al.* (2019) Nuclear DNA from two early Neandertals reveals 80,000 years of genetic continuity in Europe. *Sci Adv*, **5**, eaaw5873.
33. Hajdinjak, M., Fu, Q., Hübner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., Narasimham, V., Rougier, H., Crevecoeur, I., *et al.* (2018) Reconstructing the genetic history of late Neanderthals. *Nature*, **555**, 652–656.
34. Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., *et al.* (2018) The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*, **561**, 113–116.
35. Calvo, S.E., Clauser, K.R. and Mootha, V.K. (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Research* **(2016)** , *44*, D1251–D1257.
36. Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N. and Reich, D. (2023) The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. *bioRxiv*.
37. Mollet, J., Giurgea, I., Schlemmer, D., Dallner, G., Chretien, D., Delahodde, A., Bacq, D., de Lonlay, P., Munnich, A. and Rötig, A. (2007) Prenyldiphosphate synthase, subunit 1 (PDSS1) and OH-benzoate polyprenyltransferase (COQ2) mutations in ubiquinone deficiency and oxidative phosphorylation disorders. *J. Clin. Invest.*, **117**, 765–772.
38. Alliance of Genome Resources https://www.alliancegenome.org/gene/SGD:S000005027 (accessed Jul 18, 2023).

39. Sánchez, E., Lobo, T., Fox, J.L., Zeviani, M., Winge, D.R. and Fernández-Vizarra, E. (2013) LYRM7/MZM1L is a UQCRFS1 chaperone involved in the last steps of mitochondrial Complex III assembly in human cells. *Biochim. Biophys. Acta*, **1827**, 285–293.
40. Turunen, M., Olsson, J. and Dallner, G. (2004) Metabolism and function of coenzyme Q. Metabolism and function of coenzyme Q. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **(2004)** , *1660*, 171–199.
41. Jung, Y.S., Gao-Sheridan, H.S., Christiansen, J., Dean, D.R. and Burgess, B.K. (1999) Purification and biophysical characterization of a new [2Fe-2S] ferredoxin from Azotobacter vinelandii, a putative [Fe-S] cluster assembly/repair protein. *J. Biol. Chem.*, **274**, 32402–32410.
42. Ewen, K.M., Kleser, M. and Bernhardt, R. (2011) Adrenodoxin: the archetype of vertebrate-type [2Fe-2S] cluster ferredoxins. *Biochim. Biophys. Acta*, **1814**, 111–125.
43. Stehling, O., Wilbrecht, C. and Lill, R. (2014) Mitochondrial iron–sulfur protein biogenesis and human disease. *Biochimie*, **100**, 61–77.
44. Ardissone, A., Lamantea, E., Quartararo, J., Dallabona, C., Carrara, F., Moroni, I., Donnini, C., Garavaglia, B., Zeviani, M. and Uziel, G. (2015) A Novel Homozygous YARS2 Mutation in Two Italian Siblings and a Review of Literature. *JIMD Rep.*, **20**, 95–101.
45. Bonnefond, L., Fender, A., Rudinger-Thirion, J., Giegé, R., Florentz, C. and Sissler, M. (2005) Toward the full set of human mitochondrial aminoacyl-tRNA synthetases: characterization of AspRS and TyrRS. *Biochemistry*, **44**, 4805–4816.
46. Desbats, M.A., Morbidoni, V., Silic-Benussi, M., Doimo, M., Ciminale, V., Cassina, M., Sacconi, S., Hirano, M., Basso, G., Pierrel, F., *et al.* (2016) The COQ2 genotype predicts the severity of coenzyme Q10 deficiency. *Hum. Mol. Genet.*, **25**, 4256–4265.
47. Imamichi, Y., Mizutani, T., Ju, Y., Matsumura, T., Kawabe, S., Kanno, M., Yazawa, T. and Miyamoto, K. (2014) Transcriptional regulation of human ferredoxin reductase through an intronic enhancer in steroidogenic cells. *Biochim. Biophys. Acta*, **1839**, 33–42.
48. Yi, S., Zheng, Y., Yi, Z., Wang, Y., Jiang, Y., Ouyang, J., Li, S., Xiao, X., Sun, W., Wang, P., *et al.* (2023) -Associated Oculopathy: Congenital Amaurosis and Early-Onset Severe Retinal Dystrophy as Common Presenting Features in a Chinese Population. *Genes* , **14**.
49. Bonnefond, L., Frugier, M., Touzé, E., Lorber, B., Florentz, C., Giegé, R., Sauter, C. and Rudinger-Thirion, J. (2007) Crystal structure of human mitochondrial tyrosyl-tRNA synthetase reveals common and idiosyncratic features. *Structure*, **15**, 1505–1516.
50. Paukstelis, P.J., Chari, N., Lambowitz, A.M. and Hoffman, D. (2011) NMR Structure of the C-terminal domain of a tyrosyl-tRNA synthetase that functions in group I intron splicing. *Biochemistry*, **50**, 3816–3826.
51. Atkinson, A., Smith, P., Fox, J.L., Cui, T.-Z., Khalimonchuk, O. and Winge, D.R. (2011) The LYR protein Mzm1 functions in the insertion of the Rieske Fe/S protein in yeast mitochondria. *Mol. Cell. Biol.*, **31**, 3988–3996.
52. Forsgren, M., Attersand, A., Lake, S., Grünler, J., Swiezewska, E., Dallner, G. and Climent, I. (2004) Isolation and functional expression of human COQ2, a gene encoding a polyprenyl transferase involved in the synthesis of CoQ. *Biochem. J*, **382**, 519–526.
53. Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P. and Imai, K. (2015) MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteomics*, **14**, 1113–1126.
54. Montanucci, L., Capriotti, E., Birolo, G., Benevenuta, S., Pancotti, C., Lal, D. and Fariselli, P. (2022) DDGun: an untrained predictor of protein stability changes upon amino acid variants. *Nucleic Acids Res.*, **50**, W222–7.
55. Pancotti, C., Benevenuta, S., Repetto, V., Birolo, G., Capriotti, E., Sanavia, T. and Fariselli, P. (2021) A Deep-Learning Sequence-Based Method to Predict Protein Stability Changes Upon Genetic Variations. *Genes* , **12**.
56. Paul, A., Drecourt, A., Petit, F., Deguine, D.D., Vasnier, C., Oufadem, M., Masson, C.,

Bonnet, C., Masmoudi, S., Mosnier, I., *et al.* (2017) FDXR Mutations Cause Sensorial Neuropathies and Expand the Spectrum of Mitochondrial Fe-S-Synthesis Diseases. *Am. J. Hum. Genet.*, **101**, 630–637.
57. Stenton, S.L. and Prokisch, H. (2020) Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. *EBioMedicine*, **56**, 102784.
58. Dunham, M.J. and Fowler, D.M. (2013) Contemporary, yeast-based approaches to understanding human genetic variation. *Curr. Opin. Genet. Dev.*, **23**, 658–664.
59. Foury, F. and Cazzalini, O. (1997) Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria. *FEBS Lett.*, **411**, 373–377.
60. Patil, V.A., Fox, J.L., Gohil, V.M., Winge, D.R. and Greenberg, M.L. (2013) Loss of cardiolipin leads to perturbation of mitochondrial and cellular iron homeostasis. *J. Biol. Chem.*, **288**, 1696–1705.
61. Gilea, A.I., Ceccatelli Berti, C., Magistrati, M., di Punzio, G., Goffrini, P., Baruffini, E. and Dallabona, C. (2021) Saccharomyces cerevisiae as a Tool for Studying Mutations in Nuclear Genes Involved in Diseases Caused by Mitochondrial DNA Instability. *Genes*, **12**.
62. Figuccia, S., Degiorgi, A., Ceccatelli Berti, C., Baruffini, E., Dallabona, C. and Goffrini, P. (2021) Mitochondrial Aminoacyl-tRNA Synthetase and Disease: The Yeast Contribution for Functional Analysis of Novel Variants. *Int. J. Mol. Sci.*, **22**.
63. Quinn, C.L., Tao, N. and Schimmel, P. (1995) Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair. *Biochemistry*, **34**, 12489–12495.
64. Bonnefond, L., Frugier, M., Giegé, R. and Rudinger-Thirion, J. (2005) Human mitochondrial TyrRS disobeys the tyrosine identity rules. *RNA*, **11**, 558–562.
65. Nair, S., Ribas de Pouplana, L., Houman, F., Avruch, A., Shen, X. and Schimmel, P. (1997) Species-specific tRNA recognition in relation to tRNA synthetase contact residues. *J. Mol. Biol.*, **269**, 1–9.
66. Paukstelis, P.J., Coon, R., Madabusi, L., Nowakowski, J., Monzingo, A., Robertus, J. and Lambowitz, A.M. (2005) A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface. *Mol. Cell*, **17**, 417–428.
67. Petr, M., Pääbo, S., Kelso, J. and Vernot, B. (2019) Limits of long-term selection against Neandertal introgression. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 1639–1644.
68. Wei, X., Robles, C.R., Pazokitoroudi, A., Ganna, A., Gusev, A., Durvasula, A., Gazal, S., Loh, P.-R., Reich, D. and Sankararaman, S. (2023) The lingering effects of Neanderthal introgression on human complex traits. *Elife*, **12**.
69. Zeberg, H. and Pääbo, S. (2021) A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proc. Natl. Acad. Sci. U. S. A.*, **118**.
70. Jagoda, E., Marnetto, D., Senevirathne, G., Gonzalez, V., Baid, K., Montinaro, F., Richard, D., Falzarano, D., LeBlanc, E.V., Colpitts, C.C., *et al.* (2023) Regulatory dissection of the severe COVID-19 risk locus introgressed by Neanderthals. *Elife*, **12**.
71. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**, 75–78.
72. Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, **512**, 194–197.
73. Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics*, **13**, 397–406.
74. Sommerville, E.W., Ng, Y.S., Alston, C.L., Dallabona, C., Gilberti, M., He, L., Knowles, C., Chin, S.L., Schaefer, A.M., Falkous, G., *et al.* (2017) Clinical Features, Molecular Heterogeneity, and Prognostic Implications in YARS2-Related Mitochondrial Myopathy. *JAMA Neurol.*, **74**, 686–694.

75. Mafessoni, F. and Prüfer, K. (2017) Better support for a small effective population size of Neandertals and a long shared history of Neandertals and Denisovans. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E10256–E10257.
76. Schneider, A. (2022) Evolution and diversification of mitochondrial protein import systems. *Curr. Opin. Cell Biol.*, **75**, 102077.
77. Piette, B.L., Alerasool, N., Lin, Z.-Y., Lacoste, J., Lam, M.H.Y., Qian, W.W., Tran, S., Larsen, B., Campos, E., Peng, J., *et al.* (2021) Comprehensive interactome profiling of the human Hsp70 network highlights functional differentiation of J domains. *Mol. Cell*, **81**, 2549–2565.e8.
78. Alliance of Genome Resources Consortium (2022) Harmonizing model organism data in the Alliance of Genome Resources. *Genetics*, **220**.
79. Herrmann, J.M. and Bykov, Y. (2023) Protein translocation in mitochondria: Sorting out the Toms, Tims, Pams, Sams and Mia. *FEBS Lett.*, **597**, 1553–1554.
80. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., *et al.* (2021) The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.*, **30**, 187–200.
81. Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., *et al.* (2021) MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.*, **49**, D1541–D1547.
82. McCoy, R.C., Wakefield, J. and Akey, J.M. (2017) Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. *Cell*, **168**, 916–927.e12.
83. Harris, K. and Nielsen, R. (2016) The Genetic Cost of Neanderthal Introgression. *Genetics*, **203**, 881–891.
84. Arroyo, J.D., Jourdain, A.A., Calvo, S.E., Ballarano, C.A., Doench, J.G., Root, D.E. and Mootha, V.K. (2016) A Genome-wide CRISPR Death Screen Identifies Genes Essential for Oxidative Phosphorylation. *Cell Metab.*, **24**, 875–885.
85. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
86. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. and Kircher, M. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.*, **47**, D886–D894.
87. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
88. Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K.C., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., *et al.* (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, **347**, 1254806.
89. Liu, X., Jian, X. and Boerwinkle, E. (2013) dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *Human Mutation* **(2013)** , *34*, E2393–E2402.
90. Balakrishnan, R., Park, J., Karra, K., Hitz, B.C., Binkley, G., Hong, E.L., Sullivan, J., Micklem, G. and Cherry, J.M. (2012) YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database* , **2012**, bar062.
91. Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., *et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* , **2011**, bar030.
92. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
93. Baruffini, E., Ferrero, I. and Foury, F. (2010) In vivo analysis of mtDNA replication defects in yeast. *Methods*, **51**, 426–436.

94. Goffrini, P., Ercolino, T., Panizza, E., Giache, V., Cavone, L., Chiarugi, A., Dima, V., Ferrero, I. and Mannelli, M. (2009) Functional study in a yeast model of a novel succinate dehydrogenase subunit B gene germline missense mutation (C191Y) diagnosed in a patient affected by a glomus tumor. *Human Molecular Genetics* **(2009)** , *18*, 1860–1868.
95. Panizza, E., Ercolino, T., Mori, L., Rapizzi, E., Castellano, M., Opocher, G., Ferrero, I., Neumann, H.P.H., Mannelli, M. and Goffrini, P. (2013) Yeast model for evaluating the pathogenic significance of SDHB, SDHC and SDHD mutations in PHEO-PGL syndrome. *Hum. Mol. Genet.*, **22**, 804–815.
96. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
97. Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L. and Xenarios, I. (2012) New and continuing developments at PROSITE. *Nucleic Acids Res.*, **41**, D344–D347.
98. Moulinier, L., Ripp, R., Castillo, G., Poch, O. and Sissler, M. (2017) MiSynPat: An integrated knowledge base linking clinical, genetic, and structural data for disease-causing mutations in human mitochondrial aminoacyl-tRNA synthetases. *Hum. Mutat.*, **38**, 1316–1324.
99. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
100. Canela-Xandri, O., Rawlik, K. and Tenesa, A. (2018) An atlas of genetic associations in UK Biobank. *Nature Genetics*, 1593–1599.
101. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., *et al.* (2021) Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.*, **49**, D1311–D1320.
102. Wang, Q., Dhindsa, R.S., Carss, K., Harper, A.R., Nag, A., Tachmazidou, I., Vitsios, D., Deevi, S.V.V., Mackay, A., Muthas, D., *et al.* (2021) Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature*, **597**, 527–532.

Figure legends

Figure 1. (***A)*** *Geographical localization of the archaic individuals considered in this study. (**B)** Age (BCE, before common era) of archaic specimens (line ranges refer to 95.4% CI calibrated radiocarbon age, when present, or to the archaeological context range according to (36)). (**C)** Schematic representation of nuclear phylogeny of modern humans, Neanderthals and Denisovans. Stars refer to Neanderthal-Denisovan shared (dark grey) and Neanderthal-specific (light grey) lineage specific mutations.*
ALT text: This image represents the geographical localization and timeline of the archaic individuals considered in this study, along with a simplified modern human, Neanderthal and Denisovan phylogeny.

Figure 2. *Localization of archaic residues along the protein sequence and multiple alignments on primates. The locations of the variants are marked with pins. The domain organisation has been retrieved from the literature (see Material and Methods).*
ALT text: Localization of archaic residues on protein sequence and primate alignments.

Figure 3. *Functional comparison of human and archaic variants in S. cerevisiae. (A) Oxidative growth phenotype. The strains arh1Δ, mzm1Δ, msy1Δ and coq2Δ were transformed with plasmids carrying their respective human or archaic variants and with the empty vector (EV) when indicated. Equal amounts of serial dilutions of cells (10^5, 10^4, 10^3, 10^2 cells) were spotted onto plates supplemented with 2% glucose or 2% glycerol and 2% ethanol. The growth was scored after 3 days of incubation at 28°C. The experiment was performed on at least three independent clones for each strain. (**B**) Oxygen consumption rates. Respiration was measured in cells grown in medium supplemented with 0.6% glucose at 28°C. For each strain, the value of three biological replicates and their means ± SD were reported and expressed as nmol O_2/(min*mg of protein). (**C**) Aconitase activity in FDXR and FDXR^{A470S} expressing strains was recorded in whole-cell extracts and assayed by the aconitase-isocitrate dehydrogenase-coupled assay. For each strain, the aconitase activity of three biological replicates and their means ± SD were reported and expressed as mU/mg of protein. **(D)** Protein level of Cox2 in YARS2 and YARS2^{A35D} strains. Normalized signals plot quantifying the intensities of the protein bands was performed using the Image Lab software. The signals were normalized according to the control signal (Por1). Proteins extraction was performed as indicated in Supplementary Information in three biological replicates as shown in the scatter plot. *: p<0.05; **: p<0.01; ***: p<0.001 using ANOVA followed by Bonferroni's post hoc test. See Supplementary Information for additional details on the location of residues on cloned isoforms.*

ALT text: (A) Oxidative growth phenotype of the strains transformed with plasmids carrying their respective human or archaic variants at serial cells dilutions. On the left (B), the oxygen consumption rates for each strain. At the bottom, the aconitase activity in FDXR and FDXR^{A470S} (C) and the protein level of Cox2 in YARS2 and YARS2^{A35D} strains (D).

Figure 4. ***(A)*** *Structure of mature YARS2 predicted as described in Materials and Methods. Ala35 is reported in red as "spacefilling spheres", whereas the tyrosyladenilate analog is reported as "ball and stick"* ***(B)*** *Expression levels of YARS2 and YARS2^{A35D} proteins. Representative image of a Western blot analysis, using anti-YARS2 polyclonal antibody, on extracts obtained from the msy1Δ strains transformed with the hcYARS2 or the acYARS2^{A35D} (**C**) Quantification of the expression levels of YARS2 and mature and precursor YARS2^{A35D} proteins using the Image Lab software (Bio-Rad). The signals were normalized to the control signal (Por1); protein quantification was performed in three biological replicates as shown in the scatter plot.*

ALT text: (A) Predicted structure of mature YARS2 with Ala35 highlighted in red. (B) Western blot showing YARS2 and YARS2^{A35D} expression levels. (C) Quantification of protein expression using Image Lab software, normalized to Por1 control signal.