

Impact of modular mitochondrial epistatic interactions on the evolution of human subpopulations

Pramod Shinde^{1,a}, Harry J. Whitwell², Rahul Kumar Verma¹, Mikhail Ivanchenko³, Alexey Zaikin^{3,4,5}, Sarika Jalan^{1,3,6,b}

¹*Discipline of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Khandwa road, Simrol, Indore 453552, India.* ²*Chemical Engineering, Imperial College London, London, UK.* ³*Department of Applied Mathematics and Centre of Bioinformatics, Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia.* ⁴*Lobachevsky University, Gagarin avenue 23, Nizhny Novgorod, 603950, Russia.* ⁵*Department of Mathematics and Institute for Women's Health, University College London, London, WC1E 6BT, UK.* ⁶*Complex Systems Lab, Discipline of Physics, Indian Institute of Technology Indore, Khandwa road, Simrol, Indore 453552, India.*
Corresponding authors: ^a*pramodshinde119@gmail.com* , ^b*sarika@iiti.ac.in*

Abstract

Investigation of human mitochondrial (mt) genome variation has been shown to provide insights to the human history and natural selection. By analyzing 24,167 human mt-genome samples, collected for five continents, we have developed a co-mutation network model to investigate characteristic human evolutionary patterns. The analysis highlighted richer co-mutating regions of the mt-genome, suggesting the presence of epistasis. Specifically, a large portion of COX genes was found to co-mutate in Asian and American populations, whereas, in African, European, and Oceanic populations, there was greater co-mutation bias in hypervariable regions. Interestingly, this study demonstrated hierarchical modularity as a crucial agent for these co-mutation networks. More profoundly, our ancestry-based co-mutation module analyses showed that mutations cluster preferentially in known mitochondrial haplogroups. Contemporary human mt-genome nucleotides most closely resembled the ancestral state, and very few of them were found to be ancestral-variants. Overall, these results demonstrated that subpopulation-based biases may

favor mitochondrial gene specific epistasis.

Keywords: Human mitochondria, Genome evolution, Co-mutation network, Epistasis, Hierarchical modularity

1. Introduction

Genetic polymorphism varies among a species as well as within genomes and carries important implications for the evolution and conservation of species. Polymorphism in the mitochondrial (mt) genome is routinely used to trace ancient human migration routes and to obtain absolute dates for genetic prehistory (Chen, et al., 1995). The human mt-genome is very small (16.6 kb), maternally inherited, evolves in both neutral and adaptive fashions, and shows a great deal of variation as a result of divergent evolution. An absence of recombination within mt-genome provides distinct polymorphic loci which have been used to define human genealogy referred to as mt-genome haplogroups (Chen, et al., 1995). These haplogroups are formed as a result of the sequential accumulation of mutations through maternal lineages. Due to population migration, distinct lineages of mt-genome are associated with major global groups (African, American, European, Asian and Oceanic) raising the possibility that mt-genome variation could contribute to the differences in disease prevalence observed among different ethnic groups (Mishmar, et al., 2003; Shriner, and Keita., 2016; Zanelati, et al., 2015).

Conventionally, analyses of mt-genome evolution have focused on individual mutations, particularly in describing haplogroups, and to understand and predict ancestral behavior. However, the evolutionary behavior of mt-genome often involves cooperative changes within and between genes which are difficult to detect using haplogroup analysis. For example, correlated mt-genome mutations were reported among different oxidative phosphorylation subunits, which were found to affect population specific human longevity (Raule, et al., 2014; Fan, et al., 2016; Giuliani, et al., 2018; Conte, et al., 2018). Besides, cooperative activities of both mitochondrial proteins and tRNA genes are

critical for mt-genome evolution. The importance of co-mutational interactions has been well documented in the genomics field (Lane, et al., 2012; Chen, et al., 2013; Haddad, et al., 2018). Increasing evidence suggests that interactions among polymorphic sites may confer a cumulative association of multiple mutations with many diseases (Chen, et al., 2013). Interactions among polymorphic sites have also effectively been used to infer ancestry and functional convergence in the human populations (Ioannidis, et al., 2001). Commonly used methods include tree ensembles, functional nodal mutations, and single nucleotide polymorphism (SNP) based enrichment (Lunetta, et al., 2004). Important information about mt-genome evolutionary behavior, which is contained in the correlated changes between nucleotide positions both within and between genes, is not captured by these techniques. Despite strong evidence that mt-genome variation plays a role in the development and progression of complex human diseases, mitochondrial genetic variation has been largely ignored in the context of co-mutations and particularly the mechanisms by which these co-mutations occur (Boles, et al., 1998; Goodman, et al., 2006). Investigation of co-mutation effects can, therefore, improve the explanatory ability of genetics twofold. Firstly, the interaction between two informative genomic positions to explain a part of the trait heritability. Secondly, finding significant statistical links between mutations could provide strong indications of molecular-level interactions that differ between distant populations (Hartwig, et al., 2013).

Complex network science revolves around the hypothesis that the behavior of complex systems can be elucidated in terms of structural and functional relationships between their constituents employing a graph representation (Albert, and Barabási., 2002; Shinde, et al., 2015; Shinde, and Jalan., 2015; Whitwell, et al., 2018; Rai, et al., 2018; Ho, et al., 2014). The basis of the current study is that genome positions can impact each other and co-mutate within genomes (Shinde, et al., 2018; Du, et al., 2008; Sun, et al., 2014). The interaction between two or more genetic loci is referred to here as the

57 co-mutation of nucleotide positions. There are previous studies which have
 58 used genomic co-mutations as a basis of the evolution of human H3N2 and
 59 Ebola viruses (Du, et al., 2008; Deng, et al., 2015). These viral genome models
 60 have identified the co-mutating nucleotide clusters, apparently underpinning
 61 the dynamics of virus evolution since these clusters were antigenic regions
 62 of the viral capsid proteins (Du, et al., 2008; Deng, et al., 2015). In another
 63 study, Shinde et al. (Shinde, et al., 2018) demonstrated the impact of
 64 codon position bias while forming co-mutations using human mt-genomes.
 65 These studies have considered perfect co-mutation as causing factor for co-
 66 mutations. However, the role of the co-mutation frequency in these studies
 67 remains unclear. Here, we thoroughly examined a set of networks associated
 68 with a range of co-mutation frequencies and chose a particular co-mutation
 69 frequency for further network construction. Whilst pair-wise co-mutations
 70 can be straightforwardly perceived, the identification of larger sized functional
 71 units is not straightforward. Here, we used community detection algorithms
 72 to enumerate lists of modules formed within networks and described the
 73 functional relationships among nucleotide positions forming these modules.

74 We set out to develop a comprehensive approach to understand mitochon-
 75 drial diversity using mitochondrial co-mutations. To this end, we conducted
 76 a comparative analysis of 24,167 sequenced mt-genomes. The paper is orga-
 77 nized as follows. In the first section, we briefly described the level of diversity
 78 observed among underlying subpopulations concerning polymorphic site vari-
 79 ations in human mitochondrial genomes. In the second section, we described
 80 the framework to investigate co-mutations, which are critical in underlying
 81 complex mitochondrial evolution. For this, we constructed co-mutation net-
 82 works which were used to identify modules of co-mutations and also compared
 83 these results with those of the corresponding random networks. In the third
 84 and fourth sections, we identified local topological phenomena, which were
 85 crucial agents for co-mutation networks make-up. We listed down modules
 86 comprised of co-mutations and demonstrated that the identified modules

indeed correspond to ancestry based associations. Overall, revealing the importance of co-mutational biases among different human subpopulations, our analysis identified local preferences, which were key agents in forming mt-genome epistatic interactions.

2. Methods and Material

2.1. Acquisition of genomic data

Analysis of mt-genome variations in continental populations has revealed the most ancient of all human continent-specific haplogroups in Africa and their subsequent migration and settlement in other continents (Chen, et al., 1995). Therefore, continents are constituents to defining different global-ancestral lineages beyond being just landmasses. The global lineages among each continent have been shown to explain a variety of signatures including demographic history, climate and environmental changes, local-admixture patterns (Conte, et al., 2018; Fonseca, et al., 2008; Derenko, et al., 2001; Hudson, et al., 2013). Each continent has its own signatures as well as shared signatures, as human migrations are known to happen differently among different continents (Mishmar, et al., 2003).

Having this notion, we prepared an extensive collection of mitochondrial genomes of geographically diverse *Homo sapiens* populations (Fig 1) from the Human Mitochondrial Database (Hmtdb) (Rubino, et al., 2012). All downloaded genome sequences were in FASTA format. In total, the dataset comprised of 24,167 mitochondrial genome sequences from the five world continents (genome groups), including 3426 African (AF), 2650 American (AM), 8483 Asian (AS), 8060 European (EU) and 1548 Oceanic (OC) genomes. Antarctica was excluded from the present analysis since no data was available. It should be noted that these genome groups are multiethnic cohorts representing a range of populations across the continent (Fig 1A). A brief description of all the genomes and their origin is provided in S1 File.

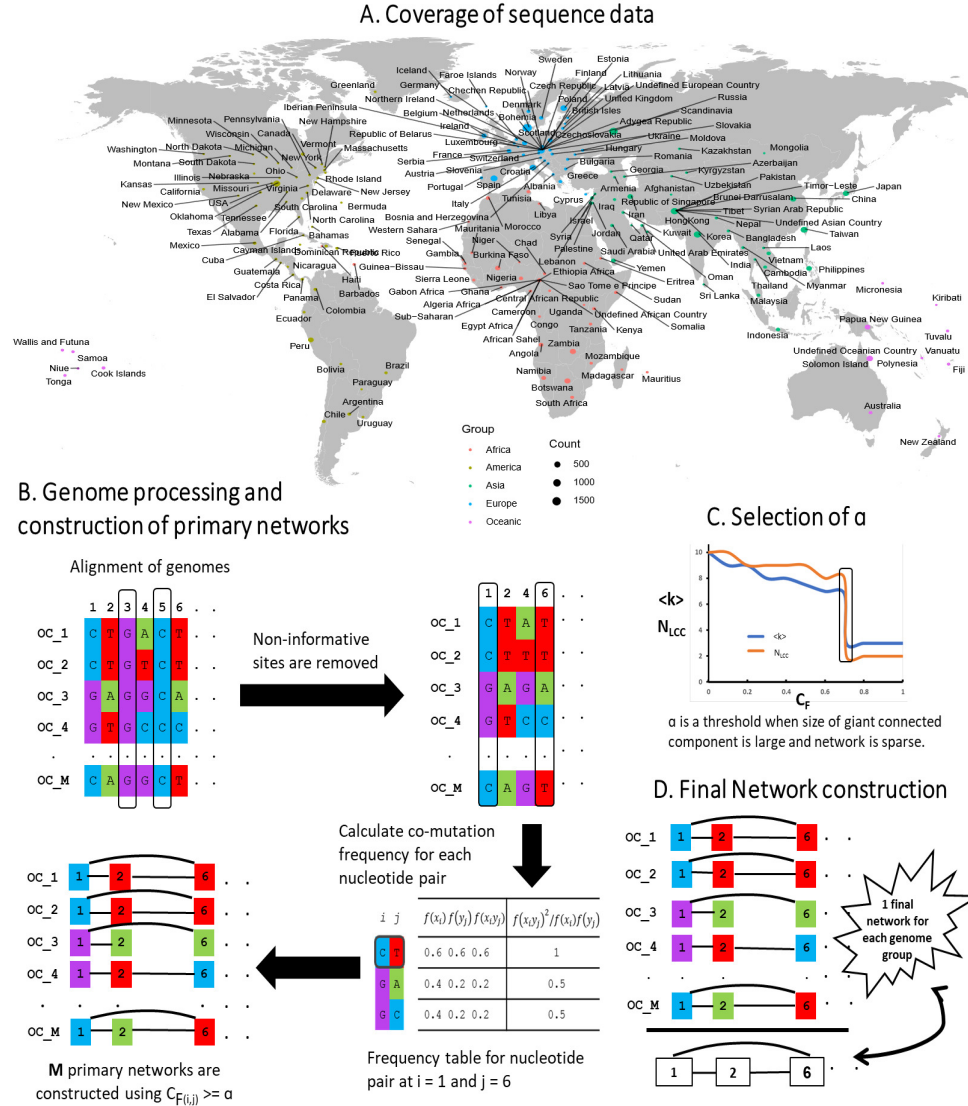


Figure 1: Schematic representation of mtDNA co-mutation network construction and analysis. (A) World map shows sequence data taken for the current study covered a good distribution across the entire globe. (B) A schematic diagram is drawn for a genome group with 5 sample sequences. The schematic diagram depicts (1) Alignment of genomes. All mitochondrial genomes in a genome group were end to end aligned, and therefore all aligned sequences had the same length. (2) Removal of non-informative sites. A genome position consist of a single nucleotide among all samples was removed from the analysis. (3) Calculation of co-mutation frequency (C_F) for each nucleotide pair. (C) Selection of network efficiency score (α). α was a threshold when the average degree ($\langle k \rangle$) of a network is small, and the size of the largest connected component (N_{LCC}) is high. For each genome group, α was found to be different. (D) Each genome group has M genomes *i.e.* M networks. A unique list of edges was picked up from M networks from a genome group to construct a final weighted network for M networks in a genome group. Likewise, five networks were constructed for five genome groups.

115 2.2. Construction and preliminary analysis of co-mutation networks

116 Co-mutation calculations were carried out on each genome group distinctly.
117 Co-mutation network construction is broadly divided into two parts, con-
118 struction of primary networks for each genome sequence followed by the
119 construction of final networks for each genome groups. Each primary co-
120 mutation network represents an individual sequence, and thus for each genome
121 group, M primary co-mutation networks were generated where M is the num-
122 ber of sequences in the genome group. In a co-mutation network (for both
123 primary and final), nodes represent genome positions, and edges between
124 nodes represent genomic co-mutations. We constructed five co-mutation
125 networks for each genome group using their primary co-mutation networks.
126 Final co-mutation networks comprise of qualitative information of interactions
127 between genome positions. The methodology for constructing primary and
128 final co-mutation networks is schematically represented in Fig 1 and described
129 as follows:

130 2.2.1. Primary co-mutation network

131 (1) Genome sequence data was end to end aligned. (2) All non-variable
132 genome positions within samples of a genome group were removed, leaving only
133 polymorphic genome positions. The number of polymorphic sites (N_P) is given
134 in Table 2. (3) Using only polymorphic nucleotide positions, we calculated
135 the frequency of occurrence of all the nucleotide pairs $f(x_i y_j) = N(x_i y_j)/M$
136 where, $N(x_i y_j)$ denoted the number of co-mutation pairs $(x_i y_j)$ at position
137 (i, j) . We then calculated the frequency of occurrence of single nucleotides
138 $f(x_i) = N(x_i)/M$ and $f(y_j) = N(y_j)/M$ where, $N(x_i)$ and $N(y_j)$ denoted
139 the number of single nucleotides at their respective positions i and j (Du,
140 et al., 2008). (4) Co-mutation of two nucleotides (C_F) at position (i, j) was
141 denoted as,

$$C_{Fi,j} = \frac{f(x_i y_j)^2}{f(x_i) f(y_j)} \quad (1)$$

142 For a particular co-mutation frequency threshold, here, termed as network

143 efficiency score (see Section 2.3), we constructed primary co-mutation networks.
 144 A network can be represented mathematically by an adjacency matrix (A)
 145 with binary entries.

$$A_{ij} = \begin{cases} 1 & \text{if } C_{Fi,j} \geq \alpha \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

146 As each genome sequence has its own information of co-mutating genome
 147 positions, a total M primary co-mutation networks were generated for each
 148 genome group.

149 2.2.2. Final co-mutation network

150 Unique edges from all primary co-mutation networks of a genome group
 151 were used to construct a final co-mutation network (Fig 1D). These five
 152 final co-mutation networks were used for network analysis and community
 153 detection.

154 We extracted hierarchical modules from final co-mutation networks and
 155 compared these networks with random networks, hypothesizing that hierar-
 156 chical modularity is the underlying phenomena of co-mutation networks and
 157 is not a mere outcome of the random evolutionary process. Furthermore, we
 158 characterized the identified module structures using ancestral markers.

159 2.2.3. Preliminary analysis of co-mutation networks

160 The degree of a node (k_i) is defined as a number of edges connected to
 161 the node such as $k_i = \sum_{j=1}^N A_{ij}$ where N denoted the number of nodes in a
 162 network. The average degree connectivity $\langle k \rangle$ is the average nearest neighbor
 163 degree of nodes with degree k . The clustering coefficient (C) is a measure
 164 of the extent to which nodes in a network tend to cluster together. An
 165 average clustering coefficient of a network can be written as $\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$.
 166 Another property of the network which turns out to be crucial in distinguishing
 167 the individual networks was the assortative coefficient (r), which measures
 168 the tendency of nodes with the similar numbers of edges to connect. The

assortative coefficient, r , was defined as the Pearson correlation coefficient of degree between pairs of linked nodes (Newman, et al., 2003). The value of r being zero corresponds to a random network, whereas the negative (positive) values correspond to dis (assortative) networks.

2.3. Selection of network efficiency score (α)

Network efficiency score (α) was used to filter edges required for network construction. Selecting an α value for each network should require iterating through a range of C_F values. To consider a network with C_F values of least 10^{-4} precision would require the construction of $24,167 * 10^4$ networks in total, which would be a very computationally intensive process. Therefore, we performed statistical sampling on each genome group interdependently by selection analysis of m samples from each population. The sample size was determined by Cochran's sample size formula (Cochran,, 1997) with critical value ($z = 1.96$). As the population was finite, the sample size was corrected by Cochran's adjustment (Cochran,, 1997).

A zero α value would result in co-mutation between each mutation and all others, whereas α equal to one would give only those pairs of mutations which have co-mutated perfectly in a genome group. In other words, zero α value would result in the globally connected network (Fig 2B) and $\alpha = 1$ would result networks with many globally connected small sub-graphs (Fig 2E). Even when the α value was as high as 0.99, networks remained very densely connected (Fig 2C). Therefore, it was reasonable to propose a criterion to select an α value, otherwise generated networks would be saturated structures holding no information about co-mutations. In order to tackle this, we plotted $\langle k \rangle$ and the size of the largest connected component (N_{LCC}) against all the α values. We observed surprising network phenomena where at a particular α value, $\langle k \rangle$ is small whilst N_{LCC} is large. At this point, networks are sparser as compared to previous α values (Fig 2D). By a sparse network, we would mean that the majority of elements of the adjacency matrix are zeroes. After exceeding this α value, the network breaks into several disconnected

199 components.

200 With this criterion, we chose a particular α value for each genome group
201 and constructed primary co-mutation networks. Although the α value applied
202 to each genome group was very high (close to 1), this value was sufficient
203 to capture more than 50% of the polymorphic sites in each genome group
204 (except in AS; S4 Table and S3 Fig). A similar criterion of filtering network
205 edges has been earlier used to construct gene co-expression network (Jackson,
206 et al., 2018). It should be noted that α values for each genome groups were
207 different (Table 2).

208 *2.4. Detection of module structures in co-mutation networks*

209 We used the Louvain algorithm, a modularity maximization algorithm, for
210 community detection for our networks (Blondel, et al., 2008). The Louvain
211 method was a simple, efficient, and easy-to-implement method for identifying
212 communities in large networks. The python package of Louvain algorithm
213 was used to enumerate module structures (Blondel, et al., 2008), and Gephi
214 software was used for visualization (Bastian, et al. , 2009).

215 **3. Results**

216 Analysis of polymorphic sites is provided in supplementary materials and
217 summarised in Table 1 and Fig 3. Codon position (CP) 2 showed fewer
218 polymorphisms as compared to CP 1 and CP 3. Genes *ATP6* and *ATP8*
219 demonstrated a higher level of polymorphisms at all three CPs. Similarly, all
220 three HVS regions have displayed a higher level of polymorphisms, whereas
221 genes of rRNA and tRNA have shown lower levels of polymorphisms.

222 *3.1. Evolution of mitochondrial co-mutations*

223 *3.1.1. Co-mutations displaying intra- and inter- genomic loci biases*

224 Analysis of pairs of co-mutations provides insight into the relationship
225 between two distinct genome locations. Co-mutations can be formulated

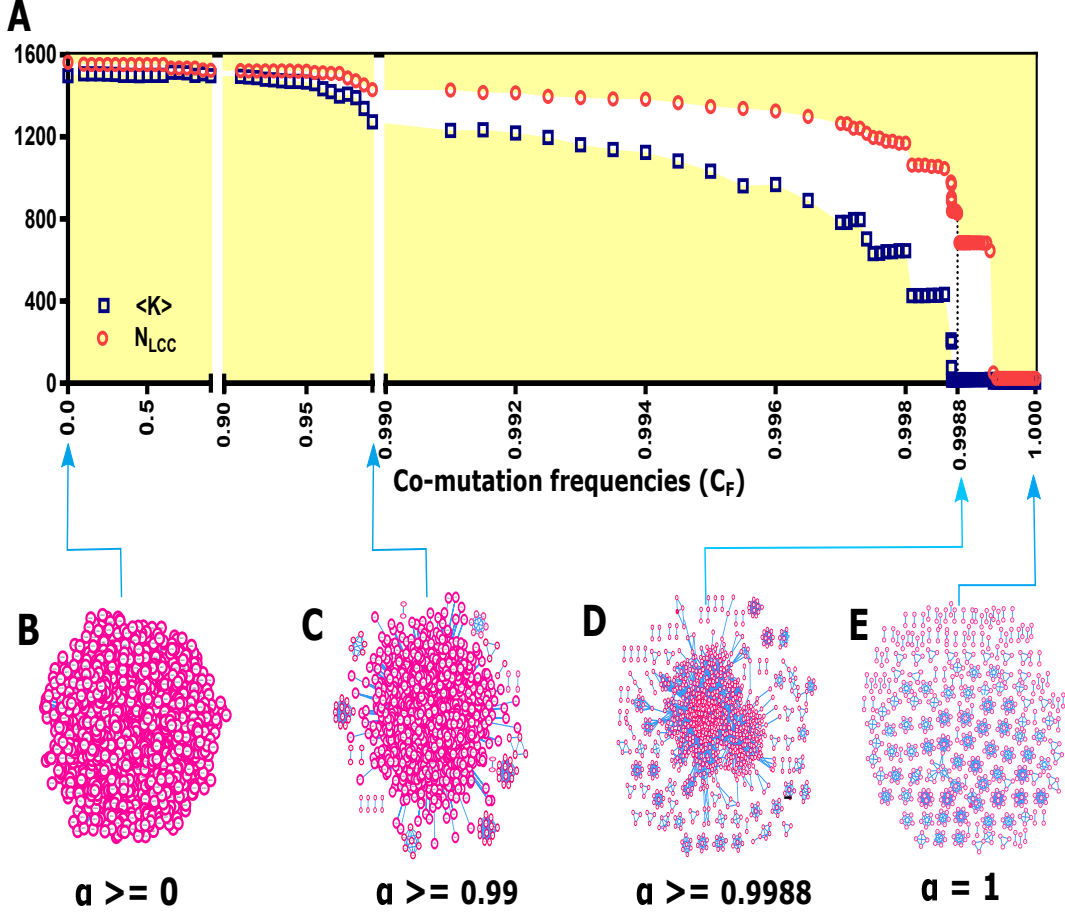


Figure 2: **Evolution of mitochondrial co-mutation network.** (A) The relative size of the largest component and the average degree of the largest component are plotted against co-mutation frequencies (C_F). The figure illustrates that at a particular α value (for OC, $\alpha = 0.9988$) co-mutation network has both a smaller value of the average degree and the number of nodes in the largest component are sufficiently in large number. We picked this C_F value for network construction. (B-E) A sample network at different C_F values show how it evolves from a globally connected network to the network with many disconnected components.

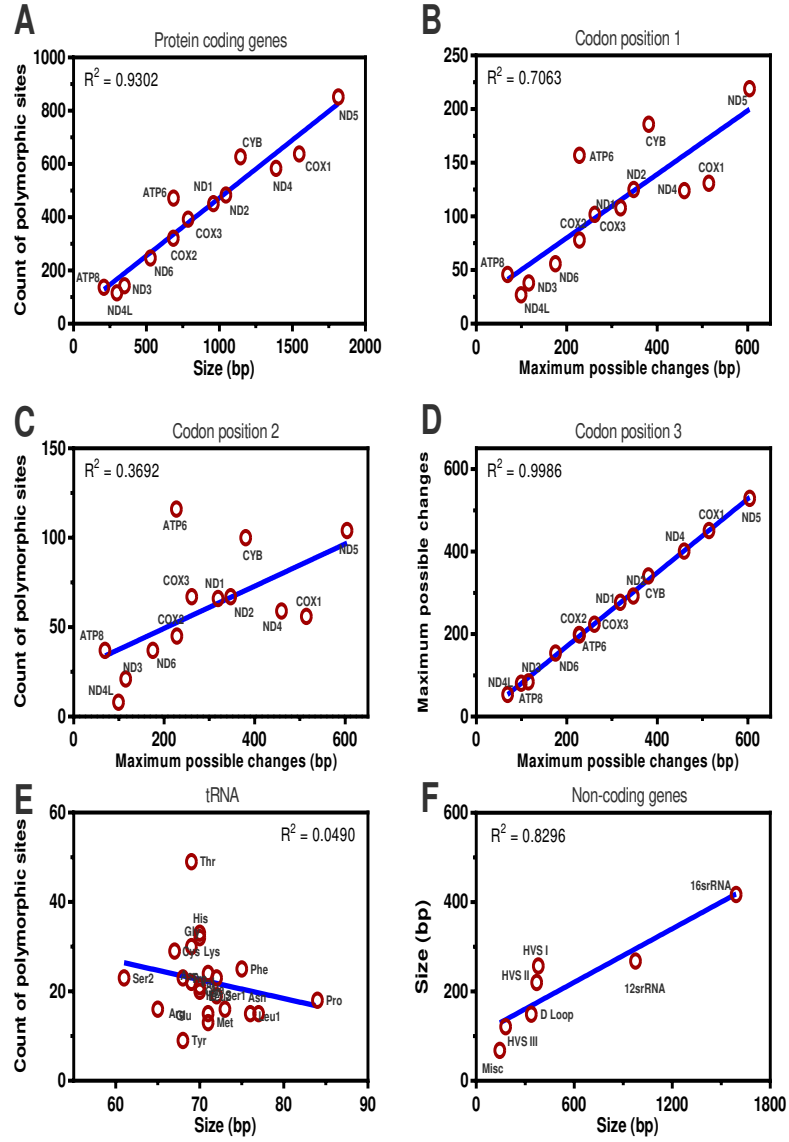


Figure 3: **Diversity among individual genome regions.** Correlation between the observed polymorphic positions and the gene size (bp) or maximum possible changes in (A) the 13 protein-coding genes, (B-D) the codon positions 1, 2 and 3 among the 13 protein-coding genes, (E) tRNA genes and (F) non-coding genes.

Gene Names (OMIM Ids)	Gene (%)	CP 1(%)	CP 2(%)	CP 3(%)
<i>ATP6</i> (516060)	69	69	51	88
<i>ATP8</i> (516070)	66	67	54	78
<i>COX1</i> (516030)	41	25	11	88
<i>COX2</i> (516040)	47	34	20	87
<i>COX3</i> (516050)	50	39	26	86
<i>CYB</i> (516020)	55	49	26	90
<i>ND1</i> (516000)	47	34	21	87
<i>ND2</i> (516001)	46	36	19	84
<i>ND3</i> (516002)	41	33	18	73
<i>ND4</i> (516003)	42	27	13	87
<i>ND4L</i> (516004)	39	27	8	82
<i>ND5</i> (516005)	47	36	17	88
<i>ND6</i> (516006)	47	32	21	88

Table 1: **Gene- and codon-wise polymorphisms among 13 protein-coding genes.** The observed polymorphisms in each of 13 protein-coding genes show mutational biases at codon positions. ATP genes contained the most polymorphisms. CP 2 showed fewer polymorphisms as compare to CP 1 and CP 3. *COX1* and *ND4* had the lowest proportion of observed polymorphic sites, and *ATP6* had the largest proportion.

226 within a particular mitochondrial functional region (intra-loci) or between two
 227 functional regions (inter-loci). We enumerated co-mutations present among
 228 nine mt-genome functional regions. The number of polymorphic sites was
 229 normalized by the total number of co-mutating polymorphic sites in a genome
 230 group and used to construct Circos plots (Fig 4). Nine mt-genome functional
 231 regions, comprising of four oxidative phosphorylation (OXPHOS) complexes,
 232 two RNA and three non-coding regions, displayed different preferences to
 233 co-mutate with other functional regions. In particular, OXPHOS complexes
 234 I, IV and HVS functional regions have a large contribution to the overall co-
 235 mutation configuration in each network. To know more on how each functional
 236 region has contributed in forming co-mutations, we plotted the number of
 237 co-mutations in each functional region against the corresponding functional

238 region size for intra- and inter- loci (Fig 4). It was observed that co-mutations
 239 among functional regions were evenly distributed among both intra- and
 240 inter- loci in AM and AS. However, intra-loci were more evenly distributed as
 241 compared to inter-loci. Interestingly, we reported few functional regions found
 242 to be outside the 95% confidence intervals in both intra- and inter-loci (Fig
 243 4). For intra-loci, rRNA was an outlier in all populations, HVS in AF and
 244 OC whereas COX in AM and EU. For inter-loci, HVS was an outlier in AF,
 245 EU, and OC whereas COX in AM and AS. ATP and miscellaneous regions
 246 were outliers in AM, tRNA in AS and rRNA in OC. These statistical outlier
 247 regions should have an assertive evolutionary role in a population. To explore
 248 this further, we studied how these groups were separated from each other. We
 249 calculated Frobenius distances between each pair of five co-mutation matrices
 250 and then performed hierarchical clustering. A dendrogram clearly showed the
 251 separation of five genome groups into two main branches *i.e.* {AM, AS} and
 252 {AF, EU, and OC} (Fig 5).

253 To investigate global level co-mutation preferences between functional
 254 regions, we analyzed unique co-mutations from all the genome groups. Fewer
 255 co-mutation pairs were formulated among intra-loci than inter-loci. This
 256 relationship between co-mutations and the spatial proximity is shown to be
 257 conserved in the mt-genome since all 13 protein-coding genes formed many
 258 interactions with OXPHOS complexes (Wong, et al., 1975; Thompson, et
 259 al., 1994). However, co-mutation pairs formed among OXPHOS complex I
 260 or ND genes which make 38% of total mt-genome participated in 31% of
 261 inter-loci co-mutations but only 13% of intra-loci co-mutations. Both D-Loop
 262 and all three hypervariable regions displayed a tendency to co-mutate with
 263 almost all other mt-genome loci (Fig 4). The rRNA genes make-up 15%
 264 of total mt-genome but they participated in only 9% of co-mutating sites.
 265 All 22 tRNA genes, which make 9% of total mt-genome, participated in
 266 10% of co-mutating sites. Overall, co-mutations dispersed among mt-genome
 267 functional regions showed that formation of co-mutations was driven mainly

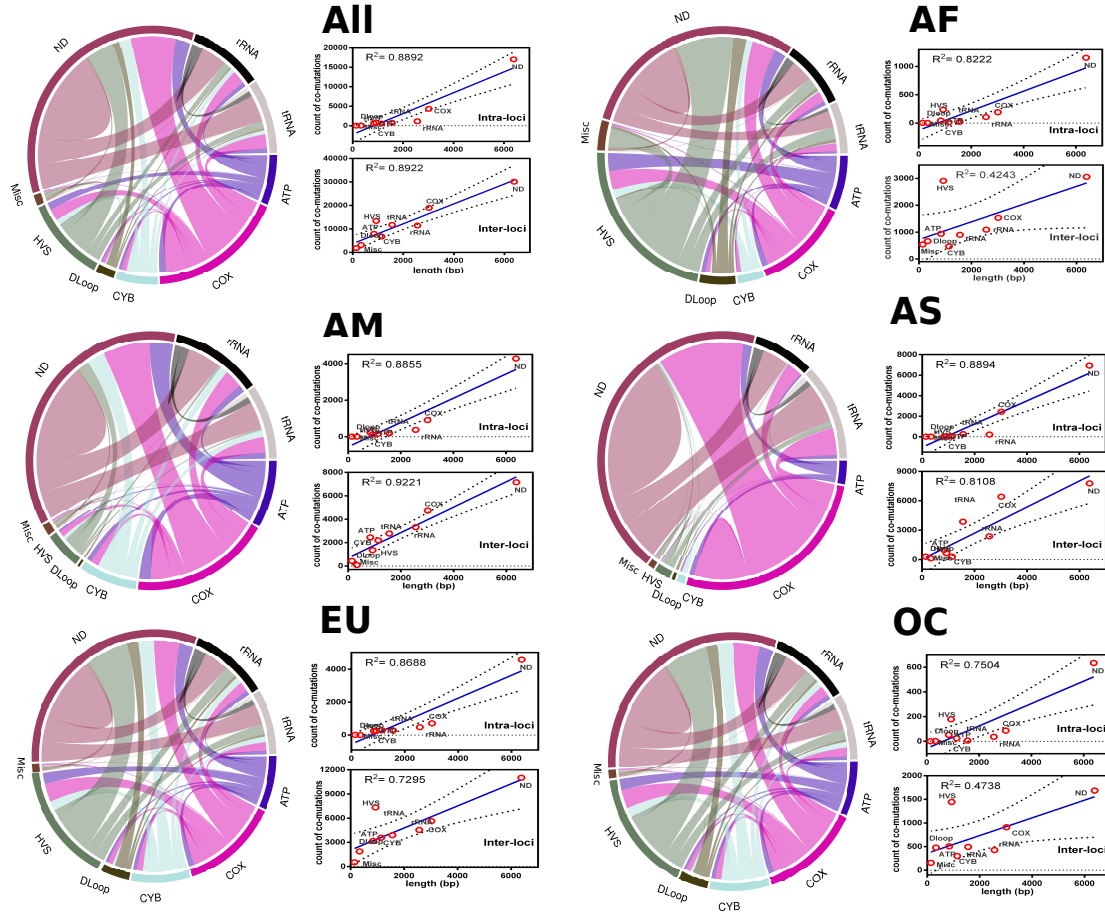


Figure 4: **Comparison of polymorphism among genomic loci.** A co-mutation configuration in the human mt-genome co-mutation network consisting of nine functional regions. These nine regions were four mitochondrial complexes (ND, COX, ATP, and CYB), three non-coding regions (DLoop, HVS, Miscellaneous) and two RNA regions (rRNA and tRNA). Links or ribbons represent the frequency of CO pairs between two genomic loci. The four functional regions make the mitochondrial oxidative phosphorylation machinery. In large part, mtDNA-specified proteins are components of respiratory complexes: Complex I (NADH dehydrogenase), Complex III (cytochrome c), Complex IV (cytochrome c oxidase) and Complex V (ATP synthase). The regression line is shown in blue (rigid) colour whereas 95% confidence interval is shown with black (dotted) colour. Circular maps were constructed using the *rcircize* package in R.

Network	α	N_P	N	N_C	$\langle K \rangle$	$\langle C \rangle$	r	Q_{Real}	Q_{Rand}
AF	0.99970	3716	2310	13721	12	0.33	-0.36	0.51	0.20
AM	0.99959	3581	2412	30283	25	0.31	-0.61	0.22	0.10
AS	0.999854	5405	2293	32705	29	0.22	-0.18	0.21	0.09
EU	0.999760	4557	2456	47952	39	0.18	-0.62	0.30	0.07
OC	0.998800	1565	1208	7304	12	0.54	-0.25	0.54	0.21

Table 2: **Data statistics and the properties of final co-mutation networks.** Here, α , N_P , N , N_C , $\langle K \rangle$, $\langle C \rangle$, r , Q represent the co-mutation frequency, number of polymorphic sites, number of nodes, number of edges, the average degree, the clustering coefficient, the assortativity coefficient and modularity coefficient for both real-world co-mutation networks and random networks, respectively. All five networks were sparse, disassortative and modular in nature. Network statistics of the largest connected component and the disconnected components are given in S1 Table and S2 Table. The number of nodes and edges forming final co-mutation networks were found to be different for each genome group. 1000 degree sequence preserved random networks are constructed for comparison of modularity in each co-mutation network and standard deviation was found to be less than 0.002 in average Q values of corresponding random networks.

by local preferences within each group.

3.1.2. Co-mutation networks exhibited similar network properties

Pair-wise co-mutations were not sufficient to fully reveal the underlying structure of functionally related nucleotide positions. As described in Fig 1, a co-mutation network was constructed for each genome group where polymorphic sites forming co-mutations constituted nodes, and edges represented co-mutating nucleotide positions. All five networks exhibited high average clustering coefficient, $\langle C \rangle$ values (Table 2), suggesting that the nodes of these networks are densely connected. Most real-world networks, particularly social networks, characterized by high $\langle C \rangle$ value, suggesting movie actors tend to create tightly knit groups by high compact ties (Sarkar, et al., 2016). In addition, all five networks displayed a highly negative degree-degree coefficient (r) (Table 2), suggesting that co-mutation networks were dis-assortative where high degree nodes, on average, prefer to link to low degree nodes (Newman,

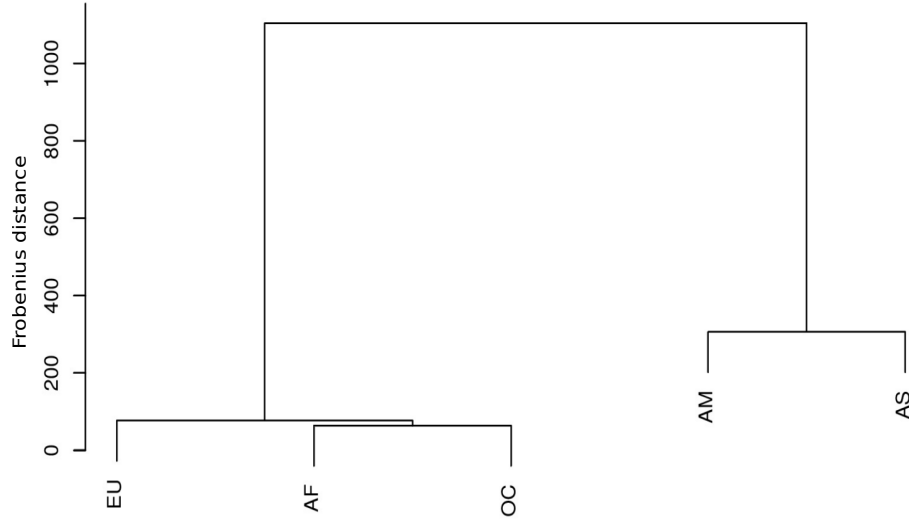


Figure 5: **Relationships among genome groups.** These relationships are predicted based on polymorphisms shown by their functional genomic loci. Five genome groups were classified as two main branches of the dendrogram , *i.e.* {AM, AS} and {AF, EU, and OC}. Here, Frobenius distance between co-mutation configuration matrices of different genome groups were used to define height of dendrogram. The branch separations shown in plot supports the routes of human migrations earlier discovered using global mt-genome mutational phylogeny. In particular, Asian haplogroup M and European haplogroup N arose from the African haplogroup L3 (Wallace, et al., 1999). Haplogroup M gave rise to the haplogroups A, B, C, D, G, and F (Wallace, et al., 1999) in which Haplogroups A, B, C, and D populated East Asia and the Americas. In Europe, haplogroup N led to the European haplogroups H, J, T, U, and V (Torroni, et al., 1996) whereas Haplogroups S, P, and Q are found in Oceania (Ruiz-Pesini, et al., 2006).

et al., 2003). Many biological and social networks have negative r values, suggesting that lack of a high degree node in a disassortative network has a large effect on the connectedness of the network (Newman, et al., 2003; Shinde, et al., 2015; Sarkar, et al., 2016). Overall, co-mutation networks have shown both the properties of high clustering and disassortative nature. This suggests the presence of dense subgraphs within the network and the presence of hierarchical structures. To explore more about the local interaction patterns in co-mutation networks, we investigated module structures within these networks.

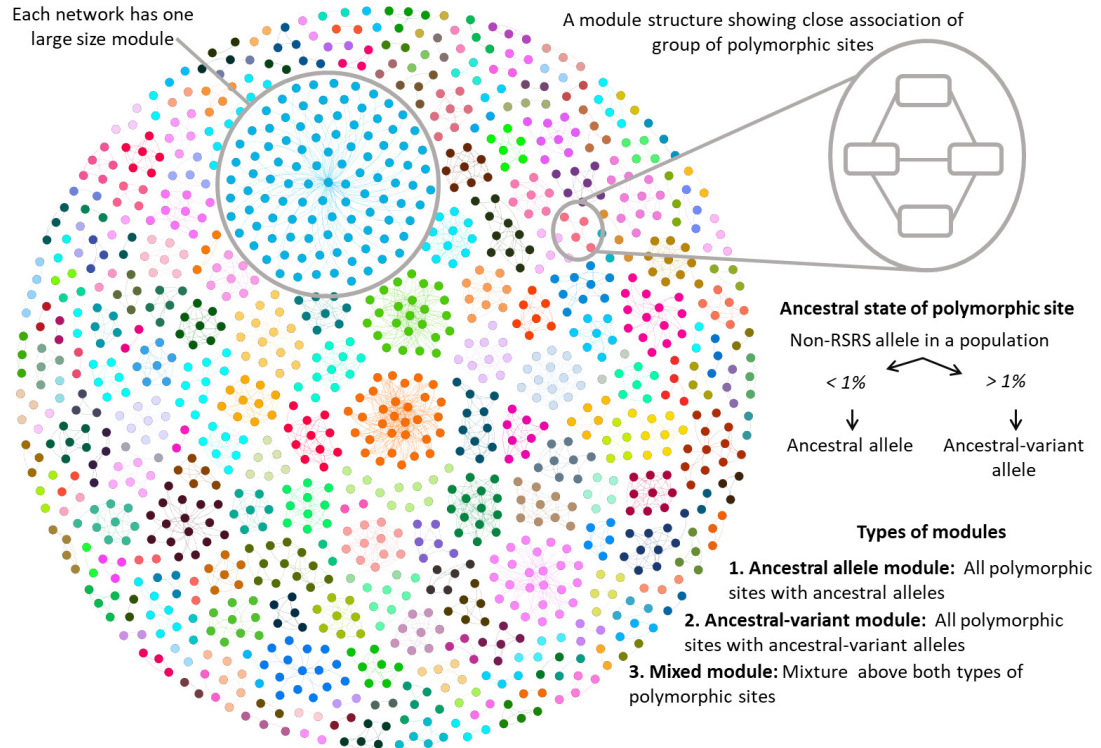


Figure 6: **Identification and characterization of network modules.** Network modules are identified using Louvain modularity algorithm. Here, the network view of modules is shown for OC network. Each network comprised of one large size module. Each polymorphic site in a module was classified as an ancestral or ancestral-variant polymorphic site. Further, modules are classified into three categories: ancestral allele module (ancestral polymorphic sites), ancestral-variant module (ancestral-variant polymorphic sites) and mixed module (both ancestral and ancestral-variant polymorphic sites).

291 *3.2. High cohesiveness and hierarchical organization of co-mutation commu-*
292 *nities*

293 Is real-world network organization driven by the non-random character, at
294 least to some extent, by modules present in the network? If this is the case,
295 it is expected that modules would be overrepresented in original co-mutation
296 networks compared to their counterparts such as random networks of the
297 same size (Prill, et al., 2005). To test this, we generated random networks,
298 referred to as a configuration model, with the same degree sequence as the
299 original co-mutation network (Csardi, and Nepusz., 2006). Random networks
300 lack organizing principles, therefore, the presence of modules in a random
301 network is determined by the density of edges (Itzkovitz, et al., 2003).

302 The major challenge for identifying modules in a hierarchical organization
303 is to decide the depth to decompose the network, as the Louvain algorithm
304 fragments networks and subsequently modules until it finds the greatest
305 partition (Meunier, et al., 2009). In order to avoid large numbers of smaller
306 modules (size 2), the size of the second largest connected component was used
307 to decipher submodules among each hierarchy of parent modules. The size
308 of the second largest connected component was 11, 8, 9, 6, and 12 for AF,
309 AM, AS, EU, OC genome groups respectively. We calculated the modularity
310 coefficient (Q) for five final co-mutation networks and also for corresponding
311 random networks (Table 2). Q value ranges between -1 and 1, where it takes
312 positive values if there are more edges between same-group vertices than
313 expected, and negative values if there are less (Blondel, et al., 2008). We
314 tested the hypothesis that the average Q of random networks equals that of
315 the co-mutation network. Q value was clearly reduced in the randomized
316 networks (t-test, $p < 0.001$, for all co-mutation networks), relative to the
317 original data, indicating that our results on real-world co-mutation networks
318 were not trivially reproduced in random networks. A high Q value will
319 manifest if networks are modular in nature. There were 557, 571, 552, 622,
320 and 227 modules obtained for AF, AM, AS, EU, and OC genome groups

321 respectively. The full list of modules is provided in S2 File.

322 In these networks, small sized modules (size less than 20) were predominant
323 alongside one or two large sized modules *i.e.* AF (size of 119), AM (270), AS
324 (217 and 216), AS (294) and OC (104) (S5 Fig). This signature of large-sized
325 modules found in co-mutation networks was not displayed by corresponding
326 random networks. Interestingly, large sized modules were only comprised
327 of polymorphic sites from non-coding regions (except in OC). Similarly to
328 co-mutations, we also noted that polymorphic sites among each module could
329 be from any of mt-genome loci. For example, in the OC population, module
330 59 had polymorphic sites only from *COX1* gene, whereas module 3 had all
331 polymorphic sites from different genes (S2 File). We noted that protein-coding
332 functional regions have a predominant role in the formation of modules (S6
333 Table and S7 Table). Particularly, ND and COX participated in >65% and
334 >40% of modules in each of the five networks, respectively. Additionally, we
335 also observed a total of 391 modules out of a total of 2529 modules where
336 all polymorphic sites in the module were from a single functional group.
337 Such mono-functional region modules were also prevailed by ND and COX
338 functional regions, 70% and 14% of total mono-functional region modules,
339 respectively (S6 Table).

340 3.3. *Modules of co-mutating polymorphic sites indicate ancestral relationships*

341 To investigate if the modules identified from the analysis of the network
342 structure were evolutionarily related, we examined polymorphic sites in
343 the individual modules for ancestral alleles from the Reconstructed Sapiens
344 Reference Sequence (RSRS). If a non-RSRS allele was present in more than
345 1% of samples in a genome group, we termed it an ancestral-variant allele.
346 Thus, we assigned ancestral-variant information to all of the network modules
347 and noted three distinct types of modules (Fig 6), which are explained in
348 following.

349 In the first and most common (more than 90% of total modules), all
350 polymorphic sites were closely related to ancestral alleles (Table 3) and we

	AF	AM	AS	EU	OC
	<i>Modules</i>				
Ancestral allele modules	501 (79%)	529 (76%)	488 (69%)	583 (75%)	205 (86%)
Ancestral-variant modules	26 (4%)	12 (2%)	11 (1%)	12 (1%)	12 (5%)
Mixed modules	30 (17%)	30 (22%)	53 (30%)	27 (24%)	10 (9%)
Total modules	557	571	552	622	227
	Ancestral lineage polymorphism				
ALPS ¹	163	192	136	177	122
Modules with atleast one ALPS	45	60	40	49	70
Modules with >1 ALPS	16	25	16	13	24
Modules with all ALPS	8	10	5	5	10

Table 3: **Statistics of modules and ancestral lineage polymorphism.** Count of modules among each genome group and percentage of nodes participating in those modules (brackets) is given. Mixed modules observed to the confined of the largest size modules. ¹ refers ancestral lineage polymorphic sites (ALPS)

351 termed them ancestral allele modules. All the polymorphic sites in these
 352 modules had ancestral alleles (or non-RSRS alleles present in < 1% of samples).
 353 Ancestral alleles were reported to be common throughout human mt-genome
 354 tree (Ruiz-Pesini, et al., 2006) and were also observed in large numbers
 355 in our genome group data (Table 3). In the second type of module, all
 356 the polymorphic sites were ancestral-variant alleles. We termed them as
 357 ancestral-variant modules and were of our particular interest because all
 358 polymorphic sites among these modules consist of the evolved character from
 359 RSRS. Ancestral-variant modules were observed the least out of three types of
 360 modules, both in terms of module count and the number of polymorphic sites
 361 present in these modules (Table 3). In the third type of module, polymorphic
 362 sites in a module were a mixture of ancestral and ancestral-variant alleles, and
 363 we termed them mixed modules. The polymorphic sites among these modules
 364 were hypothesized to have recently diverged. Mixed modules comprised of the

large-sized modules, therefore even though the module count was found to be lower, these mixed modules still possessed a higher number of nodes (Table 3). We also confirmed that count of each module type in all five co-mutation networks is significantly different than those of corresponding random networks (t-test, $p < 0.0001$; Fig S7, S8, S9). In addition, ancestral-variant modules were difficult to produce in random networks.

Modules were mapped to all known haplogroups, which showed that each polymorphic site contributed to one or many haplogroups, and thus entire module structure can be related to a single mt-genome haplogroup (S2 File). Further, we investigated the relationship between modules corresponding to ancestral haplogroup lineage markers (or top-level haplogroups). Information of ancestral lineage markers was taken from the Mitomap database, and polymorphic sites among each module were mapped to ancestral lineage markers. These ancestral lineage markers were observed to participate in the formation of entire module structures, and there were a total of 38 such modules structures obtained (Table 3; File S1). Out of the observed 38 modules, where all nodes were ancestral lineage polymorphic sites, 23 were ancestral-variant modules, 13 were ancestral modules, and two were mixed modules. Since all polymorphic sites among these 38 modules were the ancestral lineage markers, it would be reasonable to say that not only sub-level haplogroups but also top-level haplogroup markers have shown a tendency to be associated to each other.

4. Discussion

We used comparative genome analysis to investigate 24,167 mt-genomes and devised a network model comprising pairs of co-mutating nucleotides over the human mt-genome. The method presented here provides a perspective on epistatic interactions using only sequence information as well as serves as a comparative tool to understand intra-species variations. Our study showed the presence of heterogeneity in both epistatic mutations and functional

394 modules across investigated genome groups.

395 The comparison of observed polymorphisms with gene size clearly showed
396 two essential features in providing maximum functional level diversity with
397 the minimum level of genomic changes. First, genetic conservation at CP
398 2 but not at CP 3, was key to providing a protein diversity of mt-genome
399 complexes. Second, the restriction of mutations in tRNA and rRNA. These
400 two observations of biases against mutations at CP 2 and RNA genes were
401 earlier reported by Pereira et al. with 5140 human mt-genome sequences
402 (Pereira, et al., 2009). The similar biases of CP 2 and tRNA genes were
403 also reported among mt-genomes of other primates including *Macaca*, *Papio*,
404 *Hylobates*, *Pongo*, *Gorilla*, and *Pan* whereby the strength of selection was
405 determined in each lineage by the ancestral state of each codon position
406 (Kivisild, et al., 2006). Among non-coding genes, all three HVS regions have
407 displayed a higher level of polymorphisms, whereas genes of rRNA and tRNA
408 have shown lower levels of polymorphism. Our study, apart from providing the
409 detailed enlisting of diversity present among five genome groups, reiterated
410 that both codon level mutation bias and restriction of mutations among
411 RNA genes were more evident at the subpopulation level despite infrequently
412 reported to be at a global level. Furthermore, given the ubiquitous variation in
413 mt-genome, genetic flexibility may have evolved as a mechanism to maintain
414 OXPHOS under a range of environments.

415 As well as biases against polymorphisms at CP 2 in protein-coding genes,
416 our analysis indicates other biases with co-mutations. First, our results
417 with intra- and inter- loci preferences clearly suggested the dominance of
418 polygenic mutations in the human mt-genome. These polygenic mutations
419 are the outcome of a highly constrained organization of OXPHOS complexes
420 (Fonseca, et al., 2008) and also due to protein-protein interactions of the mt-
421 interactome (Schweppe, et al., 2017). Second, our analysis highlights regions of
422 the mt-genome which are rich in co-mutations and thus suggests the presence
423 of epistasis. In particular, a large portion of COX genes co-mutate in AS

424 and AM populations whereas in AF, EU and OC populations, there was
 425 greater co-mutation bias in functional regions of HVS. Although mitochondrial
 426 genome epistasis is largely described in the context of mitochondrial-nuclear
 427 interaction due to the closed assembly of OXPHOS complexes (Fonseca, et
 428 al., 2008; Picard, et al., 2018; Connallon, et al., 2018; Schweppe, et al., 2017),
 429 there are many reports describing the presence of mitochondrial-mitochondrial
 430 epistasis, for example, shared family features in Han Chinese family (Wang, et
 431 al., 2015) and homoplasmy guided by mt-tRNA genes (Moreno-Loshuertos, et
 432 al., 2011). Third, similar to polymorphic sites, co-mutations also showed biases
 433 at the subpopulation level. Genome group-wise comparison of co-mutations
 434 associated with mt-genome functional regions has helped in classifying these
 435 five human subpopulations into two prominent groups *i.e.* {AF, EU, OC}
 436 and {AS, AM}. This result was supported by a global mt-genome mutational
 437 phylogeny (Ruiz-Pesini, et al., 2006) showing the routes of human migrations
 438 (Fig 5). Overall, variations probed by epistatic interactions have provided
 439 local preferences among different mt-genome loci. These local preferences
 440 might have helped in not only forming the closed-assembly of OXPHOS
 441 complexes but also classifying subpopulations.

442 In our network model, the emergence of sparse networks was not a smooth,
 443 gradual process: the very dense largest connected component collapsed into
 444 a sparse largest connected component through a sudden change in the α
 445 curve (Fig 2). For all five genome groups, we encountered such a distinct
 446 phenomenon. A similar critical phenomenon was first observed by Erdős
 447 and Rényi through their random network model where the isolated nodes
 448 and tiny components observed for small $\langle K \rangle$ would collapse into one largest
 449 connected component (Erdos and Rényi., 1960). Interestingly, the nature
 450 of discontinuous transformations was earlier reported in biological networks
 451 (Fontana, and Schuster., 1998; Liu, et al., 2012) and have been hypothesized
 452 that biological processes follow discontinuous transformation during their
 453 evolution (Fontana, and Schuster., 1998).

454 We selected edges for inclusion in co-mutation networks based on their
 455 best fit to a network sparseness. Sparseness is one of the essential properties
 456 of biological networks since links are more difficult to create due to the
 457 evolutionary cost involved in forming more links. It is well known that
 458 co-mutational events are very selective and require a group of cooperative
 459 supporting mechanisms (Du, et al., 2008). Previously, modules of highly
 460 correlated genes were identified using similar edge-filtering based methods
 461 like weighted correlation network analysis (Jackson, et al., 2018). Network
 462 sparseness or similar data-driven approach avoids arbitrary selections of
 463 network edges and provides a uniform rationale that can be implemented to
 464 generate co-mutation network structures across different genome datasets.
 465 Therefore, it was reasonable to choose an α value where a network should
 466 have both the lowest value of $\langle K \rangle$ and the largest component with a higher
 467 count of nodes.

468 There was clear evidence for hierarchical modularity in our genome
 469 datasets, and the modular structure of the networks at all levels of the
 470 hierarchical patterns was reasonably similar across genome groups, suggesting
 471 that mt-genome functional modularity is likely to be a replicable phenomenon.
 472 A combination of co-mutations at different mitochondrial regions, that are
 473 closely linked, tend to be inherited together. This study provides a complete
 474 listing of the current knowledge of mt-genome variation in the human popu-
 475 lation, also with respect to their higher level associations with hierarchical
 476 modules. Every set of co-mutations found to originate from and remain part of
 477 a preceding single group of co-mutations. This nested hierarchy suggests the
 478 conservation of ancestral as well as inherited co-mutations throughout human
 479 lineages. Similar hierarchical modularity in brain network was related to func-
 480 tional regions in the brain and sub-set of brain functions have been reported
 481 to be associated among each hierarchy (Meunier, et al., 2009). Modularity is
 482 one of the main features of co-mutation networks, and evolutionary processes
 483 may favor the emergence of modularity by a combination of structural and

functional preferences in forming molecular interactions (Clune, et al., 2013). Therefore, it was reasonable to say that evolutionary processes may favor modularity by allowing both the specificity and autonomy of functionally distinct subsets of genomic positions. Overall, we demonstrated that molecular changes, such as mutations, were not randomly distributed across the genome, but instead concentrated within modules. In this sense, the concentration of genomic positions within modules provided a way to understand module integration, favoring distinct functional roles developed by genomic positions in distinct modules. In the human mt-genome, modules were associated with mitochondrial subcomplexes that act in distinct steps of the electron transport assembly and function. Thus, the closed assembly of mitochondrial complexes might favor the emergence of highly integrated genomic subunits, in which effects of pairwise interactions may also activate indirect effects on non-interacting genomic positions associated with the same function (Fonseca, et al., 2008). In addition, the OXPHOS system is intrinsically incapable of evolving to a fixed and general optimum state; therefore, both functional and genetic heterogeneity plays a vital role in providing robustness to the evolving OXPHOS system (Enriquez., 2016). Based on these results, we would expect that genome positions connecting modules were more conserved across evolution or, at least, less prone to failures that alter their function.

It was expected that module level associations would reflect evolutionary relationships between underlying genomic positions as each module consisted of ancestrally similar genomic polymorphisms. Our results added that the distinction between ancestral and ancestral-variant mitochondrial polymorphisms was clear when the entire module was made up of ancestral-variant polymorphic sites. However, a large number of modules (more than 90% of total modules) were made-up of ancestral polymorphic sites. In addition, the large number of nodes in mixed modules were of ancestral origin (S2 File). Using the list modules of all five networks, it would be reasonable to assert that contemporary mt-genome nucleotide bases most closely resembled the

514 ancestral state and very few of them were ancestral-variants. This observation
 515 was in agreement with previous studies which found co-mutation among
 516 nucleotide positions to be higher between genetically similar taxa (Chaffron,
 517 et al., 2010). This fact was widely observed in our data as both sub-level, and
 518 top-level haplotype markers were associated with each other in a closed group
 519 of network modules. Furthermore, mitochondrial polymorphic positions adapt
 520 from ancestral state to ancestral-variant state (Keightley, and Jackson., 2018),
 521 is also demonstrated by a larger count of transient state mixed modules.
 522 Overall, these evolutionarily closed associations suggest that interactions
 523 between nucleotide positions might evolve within genetically related genomic
 524 polymorphic positions (more likely of having similar functionality) responding
 525 to intra-species biases (Du, et al., 2008).

526 Understanding the formation of the network would require an extension
 527 of the described approaches. Here, we used simplistic information possessed
 528 by each genome position in terms of their underlying ancestral markers.
 529 Previously, this ancestral marker information has been used in order to define
 530 taxa (precisely haplogroups) in mitochondrial phylotrees which have provided
 531 the exact mapping of mitochondrial signatures to infer the routes of human
 532 intra-species diversification events (Nakatsuk, et al., 2017; Derenko, et al.,
 533 2001). Similarly, our co-mutation modules have provided a detailed listing of
 534 mitochondrial co-mutations which were ancestrally associated together.

535 Consistent with the proposed importance of mt-genome variation in human
 536 adaptation (Wallace., 2015), regional haplotypes are generally founded by one
 537 or more functionally significant polypeptide, tRNA, rRNA, and control region
 538 variants. These variant traits, beyond being retained in the descendant popu-
 539 lation, manifest functional signatures about affecting phenotypic variations
 540 in the subpopulation. Particularly, a list of singular events involving both
 541 mt-genome variation and epistatic interactions have been evaluated in terms
 542 of affecting phenotypic variation in metabolism, fitness, and life-history traits
 543 (Wallace., 2015; Shlush, et al., 2008; Li, et al., 2015). Hence, it is intuitive

544 to observe epistatic patterns of both genotype and phenotype variants along
545 with life-history traits at the subpopulation level. However, at the between-
546 population level, the evidence in support of the relationship between human
547 mt-genome variation and the metabolic rate is compelling.

548 **5. Conclusion**

549 We constructed and investigated human mt-genome co-mutation networks
550 of continents using a combined framework of genomics and network theory.
551 Our principal result was that mitochondria undergo substantial levels of
552 co-mutational biases. Codon-level mutation bias, particularly at CP 2, and
553 restriction of mutations in RNA genes was also evident at the continental
554 level, which was earlier only reported in the global human population. The
555 analysis highlighted regions of mt-genome rich for co-mutations and thus
556 suggested the presence of epistasis. In particular, a large portion of COX and
557 ND genes found to be co-mutated in AS and AM populations whereas in AF,
558 EU, and OC populations, there was greater co-mutation bias between regions
559 of HVS and ND, thus our networks identified differences in co-mutation bias
560 between human populations. It was of great interest to investigate and verify
561 different co-mutation patterns of various geographical regions. Importantly, we
562 deduced hierarchical modular structures formed within co-mutation networks.
563 Downstream analysis of these modules suggested that contemporary human
564 population are dominated by ancestral states. In addition, ancestral-variant
565 module structures are found to be in a lesser number, and such modules
566 have found to be difficult to produce in corresponding random networks. The
567 analysis presented here can be extended to study the complexity of mt-genome
568 evolution by forming various geographical groups as well as to understand
569 alterations in personal traits leading to complexity in mt-genome evolution.

570 **Competing interests**

571 The authors declare that they have no competing interests.

572 Acknowledgments

573 We are grateful to our the editor and reviewers for helping us to improve
574 the manuscript. PS acknowledges Inspire fellowship [IF150200] from the
575 Department of Science and Technology (DST), Government of India. SJ
576 thanks the support by grant of DST [EMR/2016/001921], Government of
577 India. RKV acknowledges CSIR-NET fellowship [Roll No.: 305089] from
578 CSIR, Government of India.

579 Data Availability Statement

580 All data sources and related information is given in the manuscript and
581 associated supplementary material files.

582 Supplementary materials

583 **S1 File.** Information of genome samples considered in the study.

584 **S2 File.** Information of modules identified in the study.

585 **Table S1:** Network properties of Largest connected component (Lcc).

586 **Table S2:** Network properties of all disconnected components together except
587 Lcc.

588 **Table S3:** Level-wise community detection in five co-mutation networks.

589 **Table S4:** Comparison of polymorphism α_{pre} and α_{post} .

590 **Table S5:** List of hub nodes among co-mutation networks.

591 **Table S6:** Distribution of modules having at least one polymorphic site
592 among mt-genome functional groups.

593 **Table S7:** Modules comprising all nodes as ancestral lineage polymorphic
594 sites.

595 **Table S8:** Proportion of modules among co-mutation networks and corre-
596 sponding random networks.

597 **Table S9:** Proportion of nodes in modules among co-mutation networks and
598 corresponding random networks.

599 **Figure S1:** Distribution of variable sites across genome groups and the
600 distance between each genome from the reference sequence (RSRS).
601 **Figure S2:** Degree distribution of five co-mutation networks.
602 **Figure S3:** Gene-wise comparison of polymorphism before and after α .
603 **Figure S4:** Transition and transversion.
604 **Figure S5:** Schematic description of assortativity.
605 **Figure S6:** Distribution of module sizes.
606 **Figure S7:** Identification and characterization of network modules.
607 **Figure S8:** Distribution of modules count in random networks.
608 **Figure S9:** Distribution of modularity coefficient among random networks.
609 **Figure S10:** Count of different type of modules among random networks

610

611 References

- 612 Albert, R., Barabási, A. L. (2002). Statistical mechanics of complex networks.
613 *Rev Mod Phys.* 74:47. DOI: 10.1103/RevModPhys.74.47
- 614 Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan,
615 A. and et al. (2004). The COSMIC (Catalogue of Somatic Mutations in
616 Cancer) database and website. *British journal of cancer*, 91(2): 355. DOI:
617 10.1038/sj.bjc.6601894
- 618 Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: an open source soft-
619 ware for exploring and manipulating networks. *Third international AAAI*
620 *conference on weblogs and social media*.
- 621 Boles, R. G., Roe, T., Senadheera, D., Mahnovski, V., Wong, L. J. (1998).
622 Mitochondrial DNA deletion with Kearns Sayre syndrome in a child with Ad-
623 dison disease. *Eur J Pediatr.* 157(8):643-647. DOI: 10.1007/s004310050902
- 624 Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast

625 unfolding of communities in large networks. *J Stat Mech Theory Exp.* 10.
626 DOI: 10.1088/1742-5468/2008/10/P10008

627 Chaffron, S., Rehrauer, H., Pernthaler, J., von Mering, C. (2010). A global
628 network of coexisting microbes from environmental and whole-genome
629 sequence data. *Genome Res.* 104521. DOI: 10.1101/gr.104521.109

630 Chen, J. B., Chuang, L. Y., Lin, Y. D., Liou, C. W., Lin, T. K., Lee, W. C.
631 and et al. (2013). Preventive SNP-SNP interactions in the mitochondrial
632 displacement loop (D-loop) from chronic dialysis patients. *Mitochondrion.*
633 13(6):698-704. DOI: 10.1016/j.mito.2013.01.013

634 Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S., Wallace,
635 D. C. (1995). Analysis of mtDNA variation in African populations reveals
636 the most ancient of all human continent-specific haplogroups. *Am J Hum*
637 *Genet.* 57(1):133.

638 Cochran, W. G. (1997). Sampling techniques. John Wiley and Sons. ISBN:
639 0-471-16240-X

640 Connallon, T., Camus, M. F., Morrow, E. H., Dowling, D. K. (2018). Coadap-
641 tation of mitochondrial and nuclear genes, and the cost of mother's curse.
642 *Proc R Soc B.* 285(1871):20172257. DOI: 10.1098/rspb.2017.2257

643 Conte, M., Ostan, R., Fabbri, C., Santoro, A., Guidarelli, G., Vitale, G. and
644 et al. (2018). Human aging and longevity are characterized by high levels
645 of mitokines. *J Gerontol. Gly:*153.

646 Clune, J., Mouret, J. B., Lipson, H. The evolutionary origins of modularity.
647 (2013). *Proc R Soc B.* 280(1755):20122863. DOI: 10.1098/rspb.2012.2863

648 Csardi, G., Nepusz, T. The igraph software package for complex network
649 research. (2006). *InterJournal Complex Systems* 1695(5):1-9.

- Enriquez, J.A. Supramolecular organization of respiratory complexes. (2016)
Annu Rev Physiol, 78: 533-561. DOI: 10.1146/annurev-physiol-021115-
 105031
- da Fonseca, R. R, Johnson, W. E., O'Brien, S. J., Ramos, M. J., Antunes, A.
 (2008). The adaptive evolution of the mammalian mitochondrial genome.
BMC Genomics. 9(1):119. DOI: 10.1186/1471-2164-9-119
- Deng, L., Liu, M., Hua, S., Peng, Y., Wu, A., Qin, F. X. and et al. (2015).
 Network of co-mutations in Ebola virus genome predicts the disease lethality.
Cell Res. 25(6):753. DOI: 10.1038/cr.2015.54
- Derenko, M. V., Grzybowski, T., Malyarchuk, B. A., Czarny, J., Miścicka-
 Śliwka, D., Zakharov, I. A. (2001). The presence of mitochondrial hap-
 logroup X in Altaians from South Siberia. *Am J Hum Genet*. 69(1):237-241.
 DOI: 10.1086/321266
- Du, X., Wang, Z., Wu, A., Song, L., Cao, Y., Hang, H., Jiang, T. (2008). Net-
 works of genomic co-occurrence capture characteristics of human influenza
 A (H3N2) evolution. *Genome Res*. 18(1):178-187. DOI: 10.1101/gr.6969007
- Erdős, P, Rényi, A. (1960). On the evolution of random graphs. *Publications of
 the Mathematical Institute of the Hungarian Academy of Sciences*. 5(1):17-60
- Fan, S., Hansen, M. E., Lo, Y., Tishkoff, S. A. (2016). Going global by adapting
 local: A review of recent human adaptation. *Science*. 354(6308):54-59. DOI:
 10.1126/science.aaf5098
- Fontana, W., Schuster, P. (1998). Continuity in evolution: on the na-
 ture of transitions. *Science*. 280(5368):1451-1455. DOI: 10.1126/sci-
 ence.280.5368.1451
- Giuliani, C., Sazzini, M., Pirazzini, C., Bacalini, M. G., Marasco, E., Ruscone,
 G. A. and et al. (2018). Impact of demography and population dynamics

676 on the genetic architecture of human longevity. *Aging*. 10(8):1947. DOI:
677 10.18632/aging.101515

678 Goodman, J. E., Mechanic, L. E., Luke, B. T., Ambs, S., Chanock, S., Harris,
679 C. (2006). Exploring SNP-SNP interactions and colon cancer risk using
680 polymorphism interaction analysis. *Int J Cancer*. 118(7):1790-1797. DOI:
681 10.1002/ijc.21523

682 Haddad, R., Meter, B., Ross, J.A. (2018). The genetic architecture of
683 intra-species hybrid mito-nuclear epistasis. *Front Genet*, 9, 481. DOI:
684 10.3389/fgene.2018.00481

685 Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A.
686 (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase
687 of human genes and genetic disorders. *Nucleic acids res*, 33: D514-D517.
688 DOI: 10.1093/nar/gki033

689 Hartwig, F. P. (2013). SNP-SNP Interactions: focusing on variable coding
690 for complex models of epistasis. *J Genet Syndr Gene Ther*. 4(189):10-4172.
691 DOI: 10.4172/2157-7412.1000189

692 Ho, Y.Y., Cope, L.M. Parmigiani, G. (2014). Modular network construction
693 using eQTL data: an analysis of computational costs and benefits. *Front*
694 *Genet*, 5, 40. DOI: 10.3389/fgene.2014.00040

695 Hudson, G., Nalls, M., Evans, J. R., Breen, D. P., Winder-Rhodes, S.,
696 Morrison, K. E. and et al. (2013). Two-stage association study and meta-
697 analysis of mitochondrial DNA variants in Parkinson disease. *Neurolog*.
698 80(22):2042-2048. DOI: 10.1212/WNL.0b013e318294b434

699 Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., ContopoulosIoannidis DG.
700 (2001). Replication validity of genetic association studies. *Nat Genet*. 29:306-
701 309. DOI: 10.1038/ng749

702 Itzkovitz, S., Milo, R., Kashtan, N., Ziv, G., Alon, U., 2003. Subgraphs in
703 random networks. *Physical review E*. 68(2): 026127. DOI: 10.1103/Phys-
704 RevE.68.026127

705 Jackson, M. A., Bonder, M. J., Kuncheva, Z., Zierer, J., Fu, J., Kurilshikov,
706 A. and et al. (2018). Detection of stable community structures within gut
707 microbiota co-occurrence networks from different human populations. *PeerJ*.
708 6:e4303. DOI: 10.7717/peerj.4303

709 Keightley, P.D., Jackson, B.C. (2018). Inferring the probability of the derived
710 vs. the ancestral allelic state at a polymorphic site. *Genetics*. 209(3): 897-
711 906. DOI: 10.1101/257246

712 Kenney, M. C., Hertzog, D., Chak, G., Atilano, S. R., Khatibi, N., Soe, K.
713 and et al. (2013). Mitochondrial DNA haplogroups confer differences in
714 risk for age-related macular degeneration: a case control study. *BMC med*
715 *gen*. 14(1):4. DOI: 10.1186/1471-2350-14-4

716 Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K. and et al. (2006).
717 The role of selection in the evolution of human mitochondrial genomes.
718 *Genetics*. 172(1):373-387. DOI: 10.1534/genetics.105.043901

719 Lanave, C., Tommasi, S., Preparata, G., Saccone, C. (1986). Transition and
720 transversion rate in the evolution of animal mitochondrial DNA. *Biosystems*.
721 19:273-283. DOI: 10.1016/0303-2647(86)90004-3

722 Lane, H. Y., Tsai, G. E., Lin, E. (2012). Assessing gene-gene interactions
723 in pharmacogenomics. *Molecular diagnosis and therapy*. 16(1):15-27. DOI:
724 10.1007/BF03256426

725 Li, Y., Beckman, K. B., Caberto, C., Kazma, R., Lum-Jones, A., Haiman, C.
726 A. and et al. (2015). Association of genes, pathways, and haplogroups of the
727 mitochondrial genome with the risk of colorectal cancer: the multiethnic
728 cohort. *PloS One*. 10(9):e0136796. DOI: 10.1371/journal.pone.0136796

- 729 Liu, R., Li, M., Liu, Z. P., Wu, J., Chen, L., Aihara, K. (2012). Identifying
730 critical transitions and their leading biomolecular networks in complex
731 diseases. *Sci Rep.* 2:813. DOI: 10.1038/srep00813
- 732 Lunetta, K. L., Hayward, L. B., Segal, J., Van Eerdewegh, P. (2004). Screening
733 large-scale association study data: exploiting interactions using random
734 forests. *BMC Genet.* 5:32. DOI: 10.1186/1471-2156-5-32
- 735 Meunier, D., Lambiotte, R., Fornito, A., Ersche, K., Bullmore, E. (2009). Hier-
736 archical modularity in human brain functional networks. *Front Neuroinform.*
737 3:37. DOI: 10.3389/neuro.11.037.2009
- 738 Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini,
739 S. and et al. (2003). Natural selection shaped regional mtDNA variation in
740 humans. *Proc Natl Acad Sci.* 100(1):171-176. DOI: 10.1073/pnas.0136972100
- 741 Moreno-Loshuertos, R., Ferrín, G., Acín-Pérez, R., Gallardo, M. E., Viscomi,
742 C., Pérez-Martos, A. and et al. (2011). Evolution meets disease: penetrance
743 and functional epistasis of mitochondrial tRNA mutations. *PLoS Genet.*
744 7(4):e1001379. DOI: 10.1371/journal.pgen.1001379
- 745 Morrow, E. H., Camus, M. F. (2017). Mitonuclear epistasis and mitochondrial
746 disease. *Mitochondrion.* 35:119-22. DOI: 10.1016/j.mito.2017.06.001
- 747 Nakatsuka, N., Moorjani, P., Rai, N., Sarkar, B., Tandon, A., Patter-
748 son, N. and et al. (2017). The promise of discovering population-specific
749 disease-associated genes in South Asia. *Nat Genet.* 49(9):1403. DOI:
750 10.1038/ng.3917
- 751 Newman, M. J. (2003). Mixing patterns in networks. *Phys Rev E.* 67:026126.
752 DOI: 10.1103/PhysRevE.67.026126
- 753 Newman, M. J., Girvan, M. (2004). Finding and evaluating commu-
754 nity structure in networks. *Phys Rev E.* 69:026113. DOI: 10.1103/Phys-
755 RevE.69.026113

DOI: 10.1103/PhysRevE.58.193

Pereira, L., Freitas, F., Fernandes, V., Pereira, J. B., Costa, M. D., Costa, S. and et al. (2009). The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet.* 84(5):628-640. DOI: 10.1016/j.ajhg.2009.04.013

Picard, M., McEwen, B. S. (2018). Psychological stress and mitochondria: a conceptual framework. *Psychosom Med.* 80(2):126-40. DOI: 10.1097/PSY.0000000000000544

Prill, R. J., Iglesias, P. A., Levchenko, A. (2005). Dynamic properties of network motifs contribute to biological network organization. *PLoS biology*, 3(11), e343. DOI: 10.1371/journal.pbio.0030343

Pritchard, J. K., Pickrell, J. K., Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 20(4):R208-15. DOI: 10.1016/j.cub.2009.11.055

Rai, A., Shinde, P., Jalan, S. (2018). Network spectra for drug-target identification in complex diseases: new guns against old foes. *Appl Netw Sci.* 3(1):51. DOI: 10.1007/s41109-018-0107-y

Raule, N., Sevini, F., Li, S., Barbieri, A., Tallaro, F., Lomartire, L. and et al. (2014). The co-occurrence of mt DNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. *Aging cell.* 13(3):401-407. DOI: 10.1111/accel.12186

Rubino, F., Piredda, R., Calabrese, F. M., Simone, D., Lang, M., Calabrese C. and et al. (2011). HmtDB, a genomic resource for mitochondrion-based human variability studies. *Nucleic Acids Res.* 40(D1):D1150-9. DOI: 10.1093/nar/gkr1086

Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D. and et al. (2006). An enhanced MITOMAP with a

783 global mtDNA mutational phylogeny. *Nucleic Acids Res.* 35:D823-8. DOI:
784 10.1093/nar/gkl927

785 Sarkar, C., Yadav, A., Jalan, S. (2016). Multilayer network decoding versatility
786 and trust. *EPL*. 113(1):18007. DOI: 10.1209/0295-5075/113/18007

787 Schrider, D. R., Kern, A. D. (2017). Soft sweeps are the dominant mode
788 of adaptation in the human genome. *Mol Biol Evol.* 34(8):1863-77. DOI:
789 10.1093/molbev/msx154

790 Schweppe, D. K., Chavez, J. D., Lee, C. F., Caudal, A., Kruse, S. E., Stuppard,
791 R. and et al. (2017). Mitochondrial protein interactome elucidated by
792 chemical cross-linking mass spectrometry. *Proc Natl Acad Sci.* 114(7):1732-
793 1737. DOI: 10.1073/pnas.1617220114

794 Shinde, P., Jalan, S. (2015). A multilayer protein-protein interaction network
795 analysis of different life stages in *Caenorhabditis elegans*. *EPL*. 112(5):58001.
796 DOI: 10.1209/0295-5075/112/58001

797 Shinde, P., Sarkar, C., Jalan, S. (2018). Codon based co-occurrence network
798 motifs in human mitochondria. *Sci Rep.* 8(1):3060. DOI: 10.1038/s41598-
799 018-21454-2

800 Shinde, P., Yadav, A., Rai, A., Jalan, S. (2015). Dissortativity and duplications
801 in oral cancer. *Eur Phys J B.* 88(8):197. DOI: 10.1140/epjb/e2015-60426-5

802 Shlush, L.I., Atzmon, G., Weissshof, R., Behar, D., Yudkovsky, G., Barzilai,
803 N., Skorecki, K. (2008). Ashkenazi Jewish centenarians do not demonstrate
804 enrichment in mitochondrial haplogroup J. *PLoS One*, 3(10): 3425. DOI:
805 10.1371/journal.pone.0003425

806 Shriner, D., Keita, S.O. (2016). Migration route out of Africa unresolved by
807 225 Egyptian and Ethiopian whole genome sequences. *Front Genet*, 7, 98.
808 DOI: 10.3389/fgene.2016.00098

809 Smith, D. J., Luskis, A. J. (2002). The allelic structure of common disease. *Hum*
810 *Mol Genet.* 11(20):2455-61. DOI: 10.1093/hmg/11.20.2455

811 Steinberg, B., Ostermeier, M. (2016). Environmental changes bridge evolu-
812 tionary valleys. *Sci Adv.* 2(1):1500921. DOI: 10.1126/sciadv.1500921

813 Sun, X., Lu, Q., Mukherjee, S., Crane, P.K., Elston, R. and Ritchie, M.D.
814 (2014). Analysis pipeline for the epistasis search-statistical versus biological
815 filtering. *Front Genet*, 5, 106. DOI: 10.3389/fgene.2014.00106

816 Thompson, J. N. (1994). The coevolutionary process. *University of Chicago*
817 *Press.* ISBN: 0226-79759-7. DOI: 10.7208/chicago/9780226797670.001.0001

818 Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari,
819 R. and et al. (1996). Classification of European mtDNAs from an analysis
820 of three European populations. *Genetics.* 144(4):1835-50.

821 Van Der Walt, J. M., Dementieva, Y. A., Martin, E. R., Scott, W. K.,
822 Nicodemus, K. K., Kroner, C. C. and et al. (2004). Analysis of European
823 mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett.*
824 365(1):28-32. DOI: 10.1016/j.neulet.2004.04.051

825 Wallace, D. C. (2015). Mitochondrial DNA variation in human radiation and
826 disease. *Cell.* 163(1):33-38. DOI: 10.1016/j.cell.2015.08.067

827 Wallace, D. C, Brown, M. D, Lott, M. T. (1999). Mitochondrial DNA variation
828 in human evolution and disease. *Gene.* 238(1):211-230. DOI: 10.1016/S0378-
829 1119(99)00295-4

830 Wang, F., Huang, G. D., Tian, H., Zhong, Y. B., Shi, H. J., Li Z. and et al.
831 (2015). Point mutations in KAL1 and the mitochondrial gene MT-tRNA
832 cys synergize to produce Kallmann syndrome phenotype. *Sci Rep.* 5:13050.
833 DOI: 10.1038/srep13050

- 834 Whitwell, H. J., Blyuss, O., Menon, U., Timms, J. F., Zaikin, A. (2018).
835 Parenclitic networks for predicting ovarian cancer. *Oncotarget*. 9(32): 22717.
836 DOI: 10.18632/oncotarget.25216
- 837 Wong, J. T. (1975). A co-evolution theory of the genetic code. *Proc Natl Acad*
838 *Sci*. 72(5):1909. DOI: 10.1073/pnas.72.5.1909
- 839 Zanellati, M.C., Monti, V., Barzaghi, C., Reale, C., Nardocci, N., Al-
840 banese, A., and et al. (2015). Mitochondrial dysfunction in Parkinson
841 disease: evidence in mutant PARK2 fibroblasts. *Front Genet*, 6, 78. DOI:
842 10.3389/fgene.2015.00078