IMA Journal of Complex Networks (2021) Page 1 of 16
doi:10.1093/comnet/xxx000

Role of mitochondrial genetic interactions in determining adaptation to high altitude in human population around the globe

RAHUL K VERMA
kumarvrahul16@gmail.com

KALYAKULINA ALENA
kalyakulina.alena@gmail.com

ANKIT MISHRA
ankitphy0592@gmail.com

MIKHAIL IVANCHENKO
ivanchenko.mv@gmail.com

AND

SARIKA JALAN
sarika@iiti.ac.in

*Corresponding author: *kumarvrahul16@gmail.com*, *sarika@iiti.ac.in*

[Received on 22 June 2021]

Physiological and haplogroup studies performed to understand high-altitude adaptation in humans are limited to individual genes and polymorphic sites. Due to stochastic evolutionary forces, the frequency of a polymorphism is affected by changes in the frequency of a near-by polymorphism on the same DNA sample making them connected in terms of evolution. Here, first we provide a method to model these mitochondrial polymorphisms as “co-mutation networks” for three high-altitude populations, Tibetan, Ethiopian and Andean. Then, by transforming these co-mutation networks into weighted and undirected gene-gene interaction (GGI) networks, we were able to identify functionally enriched genetic interactions of *CYB* and *CO3* genes in Tibetan and Andean populations, while NADH dehydrogenase genes in Ethiopian population playing a significant role in high altitude adaptation. These co-mutation based genetic networks provide insights into the role of different set of genes in high-altitude adaptation human sub-populations.

Keywords:

Co-mutation networks, High-altitude population, mtDNA

1. Introduction

Paramount success of network science banks heavily on influence of pair-wise interactions among the constituents of the corresponding complex systems on the evolution and performance of the systems. The current work provides a model to capture mitochondrial genomic variations by considering them as co-mutation networks, and analyse the role of polymorphic cohorts in convergent evolution for high altitude human population. The term ‘convergent evolution’ is defined as development of the same or similar phenotypic adaptations under a similar external environmental condition as a consequence of natural selection. We consider three high altitudes geographical regions, namely, (i) Andean Altiplano,

(ii) Qinghai–Tibetan Plateau and (iii) the Ethiopian Highlands to understand mechanisms behind the adaptation of human life for high altitudes. It is now believed that about 25% of the mtDNA protein sequence variations [1–3], 10 to 20% of the tRNA variations, and at least few of the rRNA variations have altered the mitochondrial coupling efficiency [4], thereby leading to an increase in the heat production at the expense of ATP production eventually permitting humans to adapt to the colder climates. However, association and mutual role of polymorphisms in other regions of mtDNA are subject to investigation in establishing a conclusive role. Complex systems have been extensively studied through networks[5] to infer role of proteins through protein-protein interaction (PPI) networks[6], to understand biological functions of neurons in various developmental stages of *C. elegans* by analyzing the underlying protein-protein interactions [7], to identify crucial structural patterns in diabetes mellitus II [9], and to understand genes responsible for evolution of specific characteristics of human sub-population [10]. Networks allow modeling of real-world complex systems by a very simple yet effective framework which consists of nodes and edges. Analysing the structural and other features of the underlying network properties based on the information available of the edges (interactions) could reveal many system level properties of corresponding complex systems [11].

Indigenous human populations residing on the Tibetan, Ethiopian and Andean Plateaus are believed to be descendants of colonizers who arrived at most ~25,000, 16000 and 11,000 years ago, respectively [12]. The three high-altitude populations currently can be viewed as a outcome of separate replications of a natural experiment in which an ancestral founding population moved from low to high altitude, and its descendants have been exposed for millennia to the opportunity for natural selection to improve their functions under high-altitude hypoxia. However, the three populations are believed to be evolved differently at the physiological levels [13]. Less oxygen in the inhaled air due to high altitude results in less oxygen to diffuse into the bloodstream flowing to the cells for oxygen-requiring energy-producing metabolism in the mitochondria. Humans do not store oxygen due to its rapid and destructive reactions with other molecules. Therefore, oxygen must be supplied without interruption to the mitochondria and to the $\geq 1,000$ oxygen-requiring enzymatic reactions in various cells and tissues [14]. Based on factors contributing to arterial oxygen content like hemoglobin content, oxygen saturation, hemoglobin affinity, etc., there are abundant evidences of the Andean-Tibetan difference in high-altitude adaptation while investigation of the Ethiopian population lies at its nascent stage [13]. There exist certain polymorphisms belonging to genes ND3 and CYTB which are believed to be associated with high-altitude adaptation in Tajiks population native to China [15]. It has been observed that genetic variations rarely impart their effect at phenotypic levels individually but as a cohort of multiple interactions together [16]. Single nucleotide polymorphisms (SNPs) have been displayed to have a small effect on the heritability of few complex diseases [17]. Additionally, various studies have indicated that interactions of SNPs [18–20] or mutations in proteins [21] are one of the key factors in manifestation of such complex diseases. To select a particular cohort of the variations and their interactions responsible for manifestation of complex phenotypes, various computational methods have been developed and implemented such as principal component analysis to evaluate SNP correlations [22], integrative scoring system based on their deleterious effects [23], and Pareto-optimal approach [24]. There exist other approaches based on pair-wise interaction such that two variations significantly interacting through logic regression [25], predictive rule inference [26], and shrunken methodology [27]. Networks of variable sites were used to identify angiogenesis genes associated with breast cancer [28], time-dependent weight dynamics in chickens [29], feed efficiency in duroc and landrace pigs [30], altitude-dependent interactions of mitochondrial genes in Asian population [31], etc.

Here, in this work we first constructed the co-mutation network by selecting significantly interacting variations of the mitochondrial genome, thereafter, a gene-gene interaction (GGI) network was

constructed from the corresponding co-mutation networks and functional enrichment analysis was performed based on significantly interacting gene-sets. These networks were found to follow the small-world behaviour with high modularity. The weak ties which are nodes with low degree and high betweenness centrality were found only in the Tibetan network and acting as haplogroup markers. Through GGI networks, it was found that interactions of CYB and CO3 genes played important role for high-altitude adaptation in Tibetan and Andean population while ND genes in Ethiopian population.

2. Methods

2.1 Sequence Acquisition

Complete human mitochondrial genome sequences were downloaded from the Human mitochondrial Database (HmtDB) [32] for the Ethiopia and Andes regions which are situated ~3000m, and ~3500m above sea level, respectively. For Andes region, we have downloaded sequences from Peru region since it inhabits indigenous Andean (Aymara and Quechua) population. Tibetan sequences (~4000m) were downloaded from the GenBank. All the sequences were aligned globally and mapped with a master sequence rCRS (revised Cambridge Reference Sequence)[33].

2.2 Construction of Co-mutation Network and Gene-Gene Interaction (GGI) network:

We constructed two types of networks, first, the co-mutation network where nodes are variable sites, and second, the weighted GGI network where nodes were genes (Fig 1) for each of the high-altitude population.

STEP 1 (CO-MUTATION NETWORK): Any position having more than one allele within the samples is considered as a variable site. The variable sites were extracted from the aligned sequences for each region separately. For genomic equality, ambiguous nucleotides such as X, M, Y, etc were replaced with 'N' for all the sequences and triallelic sites were not considered.

STEP 2: To construct network for each high-altitude region, a node was represented by the position of a variable site and the edge was represented by co-mutation frequency between a pair of the nodes. We defined co-mutation frequency for a pair of nodes, C_{ij} , as

$$C_{ij} = \frac{(m_{ij})^2}{m_i m_j} \quad (2.1)$$

where (m_{ij}) represents number of times the minor alleles occur together (defined as co-mutation pair) at i^{th} and j^{th} positions, m_i and m_j indicate total number of times the minor allele occurs at i^{th} and j^{th} position, respectively.

STEP 3 (p-VALUE CALCULATION): For checking the significance of any co-mutation pair, we calculated the p -value as,

$$p - value = \frac{\#[(C_{ij}^r) \geq (C_{ij})]}{Total\ random\ simulations} \quad (2.2)$$

where (C_{ij}^r) is the co-mutation frequency calculated after permuting the alleles at the i^{th} and j^{th} positions randomly. 10,000 random simulations were generated and p -value was set to ≤ 0.05 (standard value) to consider a co-mutation pair significant.

STEP 4 (GENE-GENE INTERACTION (GGI) NETWORK): For each co-mutation network, nodes were mapped to their corresponding genes to achieve one weighted gene-gene interaction network for each high-altitude region. Since, two or more co-mutation pairs may belong to the same gene or a gene-pair, each link is counted as many times it is found and this number was considered as weight for that gene-pair. For example, the co-mutation pairs (3352-7623) and (4125-8054) yield two *ND1-CO2* gene-pairs, so this pair was counted twice and so on, similarly the co-mutation pairs (3352-3489) would yield a self-loop for *ND1* gene. The weight of a gene-pair was normalized by the sum of total length of both the genes for that gene-pair. Additionally, we have constructed GGI networks using the co-mutations of the (i) common nodes and (ii) exclusive nodes for all the three regions. For (i), we took the nodes commonly present in all the networks and scanned for their co-mutating partners, and mapped these co-mutating nodes to respective genes. Similarly, exclusive nodes for each region were used to construct the corresponding GGI network.

2.3 Null Model and Significant interactions

Random networks were generated corresponding to each co-mutation network to compare with the real networks by taking the same number of the nodes as in the real network, we connect them with a probability α such that total number of connections the random networks has is the same as of the corresponding real world;

$$\alpha = \frac{N_c}{N(N-1)/2} \quad (2.3)$$

where, N_c is total number of connections and N is total number of nodes in the real network. Further, these random networks were used to get the random GGI networks as described in *Step 4*. We compared the real GGI networks with the corresponding random GGI networks and considered only those pairs significant whose weights fell two standard deviations away from the corresponding random one.

2.4 Detection of Community and Role of nodes

By calculating the modularity we detected the communities using algorithm given in [34] which is implemented in Python using *community module*. It is a modularity maximization algorithm. The role of each node in the communities is determined by its within-module degree, Z score and the participation coefficient P . The within-module degree quantifies the node's intra-modular connectivity and was calculated as the Z-score-transformed degree of centrality within the module. For a given node i , Z_i is defined as,

$$Z_i = \frac{k_i - \bar{k}}{\sigma} \quad (2.4)$$

where k_i is degree of the i^{th} node in its own community, $\bar{k}$ is the average of k_i for all the nodes of that community, and σ is the standard deviation. Z_i takes a high value if within the cluster, degree of i^{th} node is high and vice versa.

Different roles of a node can also be deduced based on the number of connections the node makes with the nodes in the modules other than its own. For example, two nodes with the same Z -score will play different roles if one of them is connected to several nodes in other modules while the other is not. We define the participation coefficient P of node i as,

$$P_i = 1 - \sum_{s=1}^{N_m} \left(\frac{k_i}{K_i} \right)^2 \quad (2.5)$$

K_i is the total degree of the node i in the whole network. S is the community and N_m is the total number of communities. The participation coefficient of a node is therefore close to 1 if its links are uniformly distributed among all the modules and 0 if all its links are within its own module.

FIG. 1. Construction of Co-mutation network and Gene-Gene Interaction (GGI) network for each high-altitude region as explained in *Methods* section.

3. Results

3.1 Identification of Significant Interactions

We constructed the co-mutation networks for each of the high altitude human sub-population with mtDNA variants to analyse and compare the co-evolution of the variable sites for three underlined regions. The Andes region has the highest number of the samples among all the three regions (Table 1). More number of the samples rendered more number of interactions to be statistically significant which is observed in comparison of number of connections (L) at $C_{ij} > 0$ between the Andes and Ethiopia. Both the regions are having almost equal number of connections before applying the threshold. When significant pairs (with $p\text{-value} \leq 0.05$) were considered, the number of connections were decreased by $\sim 23\%$ in Andes, $\sim 49\%$ in Ethiopia and $\sim 53\%$ in Tibet. It was further noted that above the threshold value, the number of co-mutations with low C_{ij} values (≤ 0.2) were less while co-mutations with high C_{ij} (> 0.2) remained unaffected. This observation signifies the role of $p\text{-value}$ in determining the signif-

icant co-mutations pairs. Further, the distribution of C_{ij} in Andes population showed a heterogeneous distribution of the variations within the samples, i.e, there exist a fewer co-mutations for $C_{ij} > 0.8$, indicating minor alleles in question are not always present in the same samples. The number of nodes and edges for all the three regions are also mentioned in Table 1. To get an overall idea for these three regions, we scanned for the common nodes and the common connections among these three co-mutation networks (Fig 2a,b).

FIG. 2. Distribution of nodes (a) and co-mutation pairs (b) across all the three regions. (c) Nodes which are participating in network construction were mapped to their respective genes and genes were counted and plotted on y-axis with their lengths on x-axis. Note that t-RNA genes are not shown.

Table 1. Co-mutation networks

V represents the number of nodes, L represents the number of connections in largest connected component, C_{ij} is the co-mutation frequency, we considered those pairs where $C_{ij} > 0$, p -value is the measure of statistical significance. We have performed 10,000 random permutations to generate the p -value and considered only those pairs with p -value ≤ 0.05 .

Regions	Sample size	V	L
Tibet	86	398	3459
Ethiopia	119	838	13770
Andes	496	1197	20224

3.2 Structural Properties of Co-mutation Networks

The nodes in these co-mutation networks consist of nucleotide positions which vary among the mitochondrial genome samples for each region. In the co-mutation networks, the degree provides information on the frequency of co-mutation of any given variable site with the other variable sites. A node with high degree corresponds to a variable site undergoing high number of co-mutation with many variable

Table 2. Global properties of co-mutation networks

Network property ↓	Tibet	Ethiopia	Andes
Average degree	17	33	34
Clustering coefficient	0.8	0.7	0.8
Modularity	0.7	0.5	0.4
Avg path length	4.3	2.8	2.3
Degree-degree correlation	0.4	0.2	-0.4

sites. The nodes with the highest degree in the Tibetan (12308, tRNA-Leu) and Andes (10398, ND3) co-mutation networks were commonly present in all the three networks while that of the Ethiopian co-mutation network, the highest degree node (4104, ND1) was present in Andean and was absent in Tibetan network. It is noteworthy that these highest degree nodes were found to be haplogroup markers such that 12308 in Tibetan for K and U haplogroups of N lineage, 4104 in Ethiopian for L0, L1, L2 and L5 haplogroups of L lineage, and 10398 in Andean for multiple haplogroups of L, M and N lineages. The revelation of these haplogroup markers as high degree nodes suggested the dominance of specific haplogroup background for the co-mutation of mtDNA variable sites for each region. Among the high degree nodes, A15301G (Cyb gene) node was found to be common among all the three regions. The corresponding variable site characterizes lineage L and N, and was suggested to be a candidate site for functional analyses and data association [36]. Further, all the three networks were found to have small-world property characterized by high clustering coefficient $\langle C \rangle_{real} / \langle C \rangle_{rand} \gg 1$ and small diameter $L_{real} / L_{rand} \sim 1$ as for many other real-world networks [38–40]. The small-world behaviour shown by the brain networks suggests swift flow of information in minimal steps from one region to another. Similarly, in co-mutation networks the information of change in allele frequency of a certain nucleotide at one position is transferred to another nucleotide at other position in the same mtDNA sample. Although, for these co-mutation networks, it is a subject of further investigation that whether the two nodes which are connected with each other through more than one step are also sharing the information of change in allele frequency or not. In terms of co-mutation this provides evidence for the fixation and inheritance of variations as a single cohort in the human mitochondrial genome.

Similar to various biological networks [41], a high value of average clustering coefficient ($\langle C \rangle$) was observed here as well for all the co-mutation networks. This depicts high interactions (co-mutations) among the neighbors. Further, a decrease in the tendency of a variable site to form clusters with an increase in the degree of that node (Fig. 3) implies the presence of hierarchy in the network [42, 43]. mtDNA has acquired several variations since humans have first migrated outside Africa throughout the world [44]. Different variations have been enriched in various sub-populations at separate times giving rise to various haplogroups. This temporal and spatial co-evolution of variable sites in mtDNA becomes evident with the presence of high clustering and hierarchy in these co-mutation networks.

For a network, resilience is an important property, which is measured by the betweenness centrality. Usually, the nodes with high degree tend to have high betweenness centrality. However, it is observed here that a few nodes in spite of having low degree have high betweenness centrality. Which means that these nodes co-mutate with a few variable sites but from different modules. Thus, these sites are significantly important since their removal can result in breakdown of the network. We found four such nodes in the Tibetan co-mutation network which were 709, 15927, 16172 and 16362. Interestingly, all these nodes were again found to be haplogroup markers. This suggests that haplogroup markers not only provide necessary evolutionary background but also play a key role in assisting co-evolution of different

FIG. 3. Betweenness centrality and clustering coefficient are plotted as function of degree for all the three regions.

clusters. Moreover, 15927 node belongs to tRNA-Threonine which is one of the highly mutated tRNA among all the tRNAs in humans[45], and 709 belong to the 12S-rRNA gene. tRNAs and rRNAs play a central role in protein synthesis. This signifies that tRNAs and rRNAs might play decisive roles in the co-evolution of different mutational cohorts in Tibetan population. On the contrary, we did not find any such nodes in Ethiopian or Andean co-mutation networks. In these two networks the nodes with high betweenness centrality were those which have a large degree. This suggests that in Ethiopian and Andean populations the mtDNA has evolved through continuous co-evolution of different nodes (nodes with high β_c and k) of various cohorts while in Tibetan population the mtDNA has evolved through discrete cohorts of co-evolving variable sites (nodes with high β_c and low k). This supports the evidence of inheritance of multiple ancestral gene-pools from pre-LGM and post-LGM settlers in the current Tibetan population [46].

We applied Louvain community detection algorithm [34] on co-mutation networks which was implemented in python community module. The high modularity indicates formation of mutational cohorts pertaining to evolutionary constraints at whole genome level. We analyzed the genetic background of the communities formed in these co-mutation networks. On considering only the coding regions, it was observed that one or a few genes predominate each community in all the three co-mutation networks (Supplementary Table 1). Particularly, in Tibet region, *ATP6*, *CYB*, *ND5* genes predominated while in Ethiopia *ND5* gene, and in Andes regions *ND5*, and *CYB* genes predominated among all the com-

munities. *COI* and *ND2* genes were also found to dominate at least in one community in each of the three regions. We also analysed the highest degree nodes for the coding region in each community. These high degree nodes are considered as "community cores". We found that none of these community cores are common among the three regions. Although the three regions have certain number of common nodes (Fig 2), the community structures are derived by independent nodes. Upon individual inspection of the communities we found that in each of the three regions despite contributing few nodes *tRNA-Leu*, *tRNA-Lys*, and *tRNA-Gln* were found to be community cores in the Tibet, Ethiopia and Andes regions, respectively. Apart from that, the *CR* was evenly present in all the communities. We investigated localization properties of eigenvectors of these co-mutation networks. Localization of eigenvectors enjoy wide range of application in networks; in disease spreading[52], perturbation of propagation in mutualistics networks[53]. Other application of localization can also be found in [55, 56]. To quantify localization, we use correlation dimension (D_2) calculated by using box counting method for multifractal analysis of eigenvectors[56]. If $D_2 \rightarrow 0$, eigenvector is said to be localized while if $D_2 \rightarrow 1$, eigenvector is considered as delocalized. Thus, D_2 provides insight about the degree of localization of eigenvectors. We focus on the eigenvectors of eigenvalues nearer to zero and D_2 is average over all the eigenvectors inside width $d\lambda = 0.5$. Note that, slight increase or decrease in the width will not alter the results. Tibetan and Andean networks were found to be more localized with D_2 being 0.43 compared to Ethiopian network with D_2 0.65. In these networks the tendency of co-mutation is captured by D_2 in terms of localization. A co-mutation occurs when two minor alleles are present in considerable frequency in the population and the tendency of further co-mutation is limited by introduction of new alleles. In Andean and Tibetan populations this tendency have been limited compared to Ethiopia, which might be possible due to recent population admixture experienced by Ethiopian population[57]. For these networks, there exists no node with its degree distinctly very high than those of the other nodes, and hence importance of a node can not assigned based on its degree. Nevertheless, due to the presence of the high modularity, importance of a node can be determined, to a great extent, by its within-module degree and participation coefficient, which defines how a node is positioned in its own module and with respect to other modules [47–49]. Based on the within-module degree and the participation coefficient, nodes were categorized as module hubs and non-hubs. The highest degree nodes for the Tibetan co-mutation network were found in R3, non-hub connector category, while for the Ethiopian and Andean co-mutation networks the highest degree nodes were found in R6 connector hubs category. In the Tibetan co-mutation network, the nodes in R5 category were 3010 (*16S.rRNA*), 8414 (*ATP8*), 14668 (*ND6*) and 12361 (*ND5*). Variable site 3010 was shown to be a high-altitude marker in Tibetan population [50] and also form co-occurrence motifs with variable sites 8414 and 14668[31] while, 12361 was shown to be associated with nonalcoholic fatty liver disease [51]. It is noteworthy that in the Tibetan co-mutation network there were no nodes in R6 category, and in the Andean co-mutation network there were no nodes in R5 category while in the Ethiopian co-mutation network nodes were present in both the R5 and R6 categories. These co-mutation networks are then analysed at gene level through the construction of the gene-gene interaction networks discussed further.

3.3 Gene-Gene Interaction (GGI) Networks

First, we mapped the nodes to their corresponding genes and counted the occurrence of variable sites pertaining to each gene to access these occurrences with respect to their lengths for each region (Fig. 2c). It is reflected from the figure that except the *Control region (CR)*, the occurrence of variable sites for each gene increases with an increase in the length of genes. It is quite evident that certain genes are contributing more to the variable sites in the network construction than others in a particular region.

10 of 16

RKV, AK, AM, MI, SJ

FIG. 4. Roles of nodes in Z-P parameter space.

Especially, *ATP6*, *CO2* and *ND2* genes are contributing equally in the Tibet and Ethiopia networks while *12S-rRNA*, *16S-rRNA*, *CO3* and *ND4* are contributing equally in the Tibet and Andes. Among the coding genes, *ND5* gene showed highest difference of contributing nodes with minimum in Tibet and maximum in Andes (Fig. 2c). The contribution of each gene per 100 samples for the network construction was highest in Ethiopia among all the regions. The nodes pertaining to *CR* displayed largest participation in the network construction due to the fact that it is the highly variable part of mtDNA [35]. Contribution of the variable sites in each gene yields a partial information about the interaction of the genes. To overcome this limitation, we generated the gene-gene interaction networks through mapping the variable sites with the respective genes as discussed in the Methods section.

The gene-gene interaction networks provide a holistic and reductionist approach to investigate interactions in the three high altitude regions. We found significantly 17 gene-pairs in Tibetan, 23 gene-pairs in Ethiopian and 44 gene-pairs in Andes population after comparing with their corresponding random networks. Among these, the pair with the highest weight in the Tibetan network was *ATP6-CYB*, in the Ethiopian it was *ND4-ND5* and in the Andes it was *CYB-ND4*. Four significant gene-gene pairs were found to be commonly present in all the three populations, which were *CO1-CO2*, *ND2-ND4*, *ND3-ND4* and *ND4-ND5*. All these genes are involved in oxidative phosphorylation pathway (KEGG entry: 00190) and thermogenesis pathway (KEGG entry: 04714) [58], along with that these genes are also found to be involved in cellular respiration (GO:0045333) and response to abiotic stimulus (GO:0009628) [59]. Since the three populations are believed to share a similar physical environment and undergoing the process of convergent evolution, we analyzed the properties of all the common nodes (variable sites) and extracted their co-mutations information and mapped these co-mutations to corresponding genes for each population to obtain co-evolution network of genes. The common variable sites categorize these three populations under the same haplogroups while the co-mutations of the involved nodes apparently differ in these three populations. To capture this difference at the genetic level, significant genetic interaction (Fig 5) of the common nodes were extracted based on the functional enrichment analysis for GO terms and KEGG pathways using DAVID [60]. It was found that in the Tibetan population *CO3*, *CYB* and *ND5* genes, in the Andean population *ATP6*, *CO3*, *CYB*, *ND3* and *ND4* genes, and in the Ethiopian population *ATP6*, *CO1*, *CO2*, *ND1*, *ND2*, *ND4* and *ND5* genes were significantly interacting with other genes. The functional enrichment of these genes sets is shown in table in Fig 5. It is noteworthy that from cytochrome oxidase complex, the *CO3* sub-unit was interacting in the Tibetan and Andean populations while *CO1* and *CO2* sub-units were interacting in Ethiopian population. *CO3* sub-unit is the putative site for the entry of oxygen into the large cytochrome oxidase complex thereby regulating its activity under hypoxic conditions [61]. Even though, Ethiopian gene set has not shown

FIG. 5. Significant gene-gene interactions of common node co-mutations and corresponding GO terms and KEGG pathway. (The node size depicts the degree of the node and edge size represents the edge weight.)

any feature related to the hypoxia adaptation in functional enrichment analysis, variations in *ND1* and *ND2* genes were shown to be associated with high-altitude hypoxia in Tibetan yak [62] and endemic Ethiopian rats [63]. Apart from the functional enrichment, the Tibetan and Andean gene-sets were also involved in non-alcoholic fatty liver disease (NAFLD). However, it has been shown that high altitude might improve the mitochondrion function and alleviate the NAFLD [64]. Further to explore the role of such genetic interactions at region specific level, we extracted the co-mutation pairs pertaining to the exclusive nodes of each region and constructed GGI networks. In GGI networks with these exclusive nodes, we found that there are certain interactions which were having significantly lower weight and certain others with significantly higher weight than the corresponding random ones (Fig. 6). It is readily observed from Fig. 6 that those genetic interactions which were significantly up in Andean and Tibetan population were significantly down in the Ethiopian population and vice versa.

This suggests that both Tibetan and Andean populations have evolved at high altitudes through the interactions of *CYB* and *CO3* genes while Ethiopian population deviated in sharing the mitochondrial genetic interactions with the other two populations. This dissimilarity could be explained based on two facts, one is that Ethiopia is situated at lowest altitude among all the three populations, and second that Ethiopia is believed to undergo frequent admixture in its gene pool from lower altitude populations[65].

FIG. 6. Significant gene-gene interactions of exclusive nodes co-mutations. The *Region_up* and *Region_down* represents those interactions which were showing higher and lower weight than the corresponding random networks for each region, respectively. (The node size depicts the degree of the node and edge size represents the edge weight.)

4. Conclusion

Although it is true that biologically a population can be defined in terms of haplogroups or specific mutations, with a firmly significant threshold co-mutations fortify the specific genetic interactions even in similar environmental backgrounds. As our result showed that in Andean and Tibetan populations, mtDNA has evolved similarly for these populations while Ethiopian population showed different signatures, suggesting that Ethiopian population has evolved at high-altitudes through some different mechanisms than Tibetan and Andean populations. In Tibetan and Andean population, *CYB* and *CO3* genes are commonly present in Tibetan and Andean population and are interacting with other genes namely *ND5* in Tibetan and *ATP6*, *ND3* and *ND4* in Andean. Whereas, in Ethiopian population four NADH dehydrogenase genes (*ND1*, *ND2*, *ND4*, *ND5*) showed interactions with *ATP6*, *CO1* and *CO2*. It is noticeable that in exclusive GGI networks, Ethiopian population showed the contrasting behaviour compared to the Tibetan and Andean population at the genetic level. Further, the D_2 analysis also showed that Tibetan and Andean populations are similar in their localization behavior compared to Ethiopian population. Comprehensively, our analysis showed that Ethiopian population has evolved heterogeneously than the

Tibetan and Andean populations. The method of co-mutation analysis provides information about co-evolving sets of genes based on their co-mutations to further investigate their role in environmental adaptations pertaining to convergent evolution at high-altitudes in human sub-populations.

Acknowledgments

SJ acknowledges support from SERB grant (EMR/2016/001921) and the council of scientific and industrial research grant (CSIR, 25(0293)/18/EMR-II) Govt of India. RKV gratefully acknowledges University Grant Commission for SRF fellowship (305089). MI acknowledges the Ministry of education and science of the Russian Federation megagrant (075-15-2019-871). We acknowledge Pramod Shinde for continuous discussions throughout and Cristina for useful discussions in the beginning of the project.

REFERENCES

1. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., Sukernik, R. I., Olckers, A., Wallace, D. C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), 171–176.
2. Wallace D. C. (2011). Bioenergetic origins of complexity and disease. *Cold Spring Harbor symposia on quantitative biology*, 76, 1–16.
3. Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., Wallace, D. C. (2004). Effects of Purifying and Adaptive Selection on Regional Variation in Human mtDNA. *Science*, 303(5655), 223–226.
4. Ruiz-Pesini, E., Wallace, D. C. (2006). Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA. *Human mutation*, 27(11), 1072–1081.
5. Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N. et al. Using graph theory to analyze biological networks. *BioData Mining* 4, 10 (2011)
6. Vazquez, A., Flammini, A., Maritan, A. et al. (2003). Global protein function prediction from protein-protein interaction networks. *Nat Biotechnol* 21, 697–700.
7. Shinde, P., Jalan, S. (2015). A multilayer protein-protein interaction network analysis of different life stages in *Caenorhabditis elegans*, *EPL*, 112, 58001.
8. Rai, A., Pradhan, P., Nagraj, J., Lohitesh, K., Chowdhury, R., and Jalan, S. (2017). Understanding cancer complexome using networks, spectral graph theory and multilayer framework, *Sci. Rep.* 7, 41676.
9. Rai, A., Pawar A.K., Jalan, S. (2015). Prognostic interaction patterns in diabetes mellitus II: A random-matrix-theory relation, *Phys. Rev. E* 92, 022806.
10. Shinde, P., Sarkar, C., Jalan, S. (2018). Codon based co-occurrence network motifs in human mitochondria. *Sci. Rep.* 8, 3060.
11. Jalan, S., Sarkar, C. (2017). Complex Networks: an emerging branch of science. *Physics News* 47, 3–4.
12. Aldenderfer, M. (2003). Moving Up in the World: Archaeologists seek to understand how and when people came to occupy the Andean and Tibetan plateaus. *American Scientist*, 91(6), 542–549.
13. Beall C. M. (2006). Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integrative and comparative biology*, 46(1), 18–24.
14. Raymond, J., Segrè, D. (2006). The effect of oxygen on biochemical networks and the evolution of complex life. *Science (New York, N.Y.)*, 311(5768), 1764–1767.
15. Chen, Y., Gong, L., Liu, X. et al. (2020). Mitochondrial DNA genomes revealed different patterns of high-altitude adaptation in high-altitude Tajiks compared with Tibetans and Sherpas. *Sci Rep* 10, 10592.
16. Papp, B., Pál, C. (2011). Systems biology of epistasis: Sheddinglight on genetic interaction network “hubs”, *Cell Cycle*, 10:21, 3623–3624.
17. Jakobsdottir, J., Gorin, M. B., Conley, Y. P., Ferrell, R. E., Weeks, D. E. (2009). Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS genetics*, 5(2), e1000337.

18. Marchini, J., Donnelly, P., Cardon, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature genetics*, 37(4), 413–417.
19. Cordell H. J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human molecular genetics*, 11(20), 2463–2468.
20. Moore, J. H., Williams, S. M. (2009). Epistasis and its implications for personal genetics. *American journal of human genetics*, 85(3), 309–320.
21. Paris, L., Como, G., Vecchia, I., Pisani, F., Ferrara, G. (2020). The protein interaction network of the inherited central nervous system diseases reveals new gene candidates for molecularly unclassified myelin disorders. *Journal of Complex Networks*, 8(4).
22. Horne, B. D., Camp, N. J. (2004). Principal component analysis for selection of optimal SNP-sets that capture intragenic genetic variation. *Genetic epidemiology*, 26(1), 11–21.
23. Lee, P. H., Shatkay, H. (2009). An integrative scoring system for ranking SNPs by their potential deleterious effects. *Bioinformatics (Oxford, England)*, 25(8), 1048–1055.
24. Lee, P. H., Jung, J. Y., Shatkay, H. (2009). Functionally informative tag SNP selection using a pareto-optimal approach: playing the game of life. *BMC Bioinformatics*, 10(Suppl 13), O5.
25. Schwender, H., Ickstadt, K. (2008). Identification of SNP interactions using logic regression. *Biostatistics (Oxford, England)*, 9(1), 187–198.
26. Wan, X., Yang, C., Yang, Q., Xue, H., Tang, N. L., Yu, W. (2010). Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics (Oxford, England)*, 26(1), 30–37.
27. Liu, Y., Ng, M. (2010). Shrunk methodology to genome-wide SNPs selection and construction of SNPs networks. *BMC systems biology*, 4 Suppl 2(Suppl 2), S5.
28. Lin, H.Y., Amankwah, E.K., Tseng, T.S., Qu, X., Chen, D.T., et al. (2013). SNP-SNP Interaction Network in Angiogenesis Genes Associated with Prostate Cancer Aggressiveness. *PLOS ONE* 8(4): e59688.
29. Li, F.G., Li, H. (2019). A time-dependent genome-wide SNP-SNP interaction analysis of chicken body weight. *BMC Genomics* 20, 771.
30. Banerjee, P., Carmelo, V., Kadarmideen, H. N. (2020). Genome-Wide Epistatic Interaction Networks Affecting Feed Efficiency in Duroc and Landrace Pigs. *Frontiers in genetics*, 11, 121.
31. Verma, R. K., Kalyakulina, A., Giuliani, C., Shinde, P., Kachhvah, A. D., Ivanchenko, M., et al. (2021). Analysis of human mitochondrial genome co-occurrence networks of Asian population at varying altitudes. *Scientific reports*, 11(1), 133.
32. Clima, R., Preste, R., Calabrese, C., Diroma, M. A., Santorsola, M., Scioscia, G., Simone, D., Shen, L., Gasparre, G., Attimonelli, M. (2017). HmtDB 2016: data update, a better performing query system and human mitochondrial DNA haplogroup predictor. *Nucleic acids research*, 45(D1), D698–D706.
33. Andrews, R., Kubacka, I., Chinnery, P. et al. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147.
34. Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10).
35. Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L., Erlich, H. A. (1991). Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *American journal of human genetics*, 48(2), 370–382.
36. Tranah, G. J., Manini, T. M., Lohman, K. K., Nalls, M. A., Kritchevsky, S., Newman, A. B., Harris, T. B., Miljkovic, I., Biffi, A., Cummings, S. R., Liu, Y. (2011). Mitochondrial DNA variation in human metabolic rate and energy expenditure. *Mitochondrion*, 11(6), 855–861.
37. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *American journal of human genetics*, 84(6), 740–759.
38. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, 407(6804), 651–654.
39. Wagner, A., Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings. Biological sciences*, 268(1478), 1803–1810.

40. Jeong, H., Mason, S. P., Barabási, A. L., Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833), 41–42.
41. Brown, K. S., Hill, C. C., Calero, G. A., Myers, C. R., Lee, K. H., Sethna, J. P., Cerione, R. A. (2004). The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical biology*, 1(3-4), 184–195.
42. Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442.
43. Barabasi, A. L., Albert, R. (1999). Emergence of scaling in random networks. *Science (New York, N.Y.)*, 286(5439), 509–512.
44. Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., Costa, M. D., Musilová, E., Macaulay, V., Richards, M. B., Cerny, V., Pereira, L. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Molecular biology and evolution*, 29(3), 915–927.
45. Zifa, E., Giannouli, S., Theotokis, P., Stamatis, C., Mamuris, Z., Stathopoulos, C. (2007). Mitochondrial tRNA mutations: clinical and functional perturbations. *RNA biology*, 4(1), 38–66.
46. Lu, D., et al., (2016). Ancestral Origins and Genetic History of Tibetan Highlanders. *American journal of human genetics*, 99(3), 580–594.
47. Rives, A. W., Galitski, T. (2003). Modular organization of cellular networks. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 1128–1133.
48. Han, J. D., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J., Cusick, M. E., Roth, F. P., Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430(6995), 88–93.
49. Csermely, P., London, A., Wu, L. Y., Uzzi, B. (2013). Structure and dynamics of core/periphery networks, *Journal of Complex Networks*, 1(2), 93–123.
50. Luo, Y., Gao, W., Liu, F., Gao, Y. (2011). Mitochondrial nt3010G-nt3970C haplotype is implicated in high-altitude adaptation of Tibetans. *Mitochondrial DNA*, 22(5-6), 181–190.
51. Lu, M. Y., et al., (2012). Mitochondrial polymorphism 12361A₄G is associated with nonalcoholic fatty liver disease. *Translational research : the journal of laboratory and clinical medicine*, 159(1), 58–59.
52. A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, J. F. Mendes. (2012). Local-ization and spreading of diseases in complex networks, *Physical reviewletters* 109 128702.
53. S. Suweis, J. Grilli, J. R. Banavar, S. Allesina, A. Maritan. (2015). Effect of localization on the stability of mutualistic ecological networks, *Naturecommunications* 6 1–7.
54. R. Pastor-Satorras, C. Castellano. (2016). Distinct types of eigenvector localization in networks, *Scientific reports* 6 18847.
55. Priodyuti Pradhan, Alok Yadav, Sanjiv K. Dwivedi, and Sarika Jalan. (2017). Optimized evolution of networks for principal eigenvector localization, *Phys. Rev. E* 96, 022312.
56. Ankit Mishra, Jayendra N Bandyopadhyay and Sarika Jalan. (2021). Multifractal analysis of eigenvectors of smallworld networks. *Chaos, Solitons Fractals* 144, 110745.
57. Simonson T. S. (2015). Altitude Adaptation: A Glimpse Through Various Lenses. *High altitude medicine biology*, 16(2), 125–137. <https://doi.org/10.1089/ham.2015.0033>
58. Kanehisa, M., Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), 27–30.
59. The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research*, 47(D1), D330–D338.
60. Huang, d., Sherman, B. T., Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), 44–57.
61. Lau, G. Y., Mandic, M., Richards, J. G. (2017). Evolution of Cytochrome c Oxidase in Hypoxia Tolerant Sculpins (Cottidae, Actinopterygii). *Molecular biology and evolution*, 34(9), 2153–2162.
62. Shi, Y., Hu, Y., Wang, J., Elzo, M. A., Yang, X., Lai, S. (2018). Genetic diversities of MT-ND1 and MT-ND2 genes are associated with high-altitude adaptation in yak. *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, 29(3), 485–494.

- 63.** Bartáková, V., Bryjová, A., Nicolas, V., Lavrenchenko, L. A., Bryja, J. (2021). Mitogenomics of the endemic Ethiopian rats: looking for footprints of adaptive evolution in sky islands. *Mitochondrion*, 57, 182–191.
- 64.** Song, K., Zhang, Y., Ga, Q., Bai, Z., Ge, R. L. (2020). High-altitude chronic hypoxia ameliorates obesity-induced non-alcoholic fatty liver disease in mice by regulating mitochondrial and AMPK signaling. *Life sciences*, 252, 117633.
- 65.** Hodgson, J. A., Mulligan, C. J., Al-Meer, A., Raaum, R. L. (2014). Early back-to-Africa migration into the Horn of Africa. *PLoS genetics*, 10(6), e1004393.