Mitochondrial coding and control region variants are associated with Type-2 Diabetes in Pakistani population

Saad Farooq
  University of Karachi

Syed Faizan-ul-Hasan Naqvi
  University of Karachi

Shahid Muhammad
  Dow University of Health Sciences

Agnieszka Piotrowska-Nowaka
  University of Warsaw

Syed Ali Raza Shah Bukhari (✉ sraza.mtech@gmail.com)
  Aga Khan University

Ahsan Javed
  Rawal Institute of Health Sciences

Muhammad Irfan
  Jamil-ur-Rahman Center for Genome Research, International Center for Chemical and Biological sciences, University of Karachi

Ishtiaq Ahmad Khan
  Jamil-ur-Rahman Center for Genome Research, International Center for Chemical and Biological sciences, University of Karachi

---

Research Article

Additional Declarations: No competing interests reported.

Abstract

Background

Sequence changes of human mitochondrial DNA (mtDNA) are involved in many human diseases. Mitochondrial DNA variants have been associated with development of type 2 diabetes, which is becoming more prevalent in the Pakistani population. We conducted a case-control study to investigate the role of mtDNA variants associated with diabetes in the Pakistani population.

Results

Analysis of the HVS2 region showed two variants m.309_310insCT and m.315dup were associated with diabetes. By analyzing complete mtDNA, no variant was found to have significantly different distribution between groups. However, comparison of our diabetic samples' variants with 1000 Genome Project variants showed eight highly significant variations in mitochondrial genome, four in non-coding region i.e. (m.513G > A, m.195T > C, m.16189T > C, m.16265A > C) and four in coding regions i.e. m.9336A > G (CO-III gene), m.11935T > C (ND4 gene), m.14766C > T (CYB gene) and m.7193T > C (CO-I gene) the last one being a rare mitochondrial variant also. We also found one novel variant m.570C > CACCC in the diabetic group.

Conclusion

We found specific variations in the mitochondrial genome are associated with type 2 diabetes in the Pakistani patients. These findings suggest that mtDNA variations may play a role in the development of type 2 diabetes in the Pakistani population.

Background

Human mitochondrial DNA (mtDNA) is a matrilineal, circular double-stranded molecule consisting of 16,569 bp with up to several thousand copies per cell, where number of mtDNA molecules varies depending on their number per mitochondrion or the number of mitochondria per cell (1, 2). mtDNA comprises two sections: one that encodes 13 protein subunits of four oxidative phosphorylation complexes (OXPHOS), 22 tRNAs and 2 rRNAs required for biosynthesis of its proteins (3). The non-coding region of 1.1 kb, so called control region, is responsible for the regulation of replication and transcription of mitochondrial genome (3). The majority of proteins involved in mitochondrial functions are encoded by the nuclear genome (4). The mitochondrial genome has played a central role in the understanding of evolution, migration patterns of individuals and is an informative marker to identify people around different regions of the world (5). Genetic variations in the mitochondrial genome exist in population as well as within individuals. In case of more than one form of mitochondrial genome existing within the same mitochondrion or between mitochondria of an individual, a condition of heteroplasmy occurs (that is, > 0% and < 100% allele frequency per cell) (2).

While the sequence variations in mitochondrial genome such as SNVs, insertions and deletions (indels) may be inherited and colonized in cells, unique variations also arise during life in independent cellular lineages (6). The polyploid nature of mtDNA allows inherited as well as acquired mutations to exist together in heteroplasmic state (7). There are more than 11,000 polymorphic sites present within the mitochondrial genome reported by public databases (8). Many polymorphisms of mtDNA are population-specific and they tend to reside in the form of clusters in lineages. These clusters of specific SNVs of mtDNA or mitochondrial haplotypes, constitute the haplogroups that are groups of similar haplotypes with several common polymorphisms. Typically present with a frequency of over 5% in a population, haplogroups emerge in geographically localized indigenous populations (9). In addition to common variants, many mtDNA haplogroup

determinant variants were found to be linked with certain human pathologies (10, 11). Cellular reactive oxygen species (ROS) levels are associated with mutations in mtDNA and their frequency increases with age (12). Increased mutational loads of mtDNA have been linked with aging, neurodegenerative disorders (13), heart diseases (14) and diabetes (15). The association between mtDNA haplogroups and a wide range of diseases of metabolism, infectious diseases, degenerative and autoimmune disorders, and predisposition to various cancers is widely studied (16).

Approximately 537 million adults are affected by diabetes worldwide, of which type 2 diabetes accounts for 90% of all the reported diabetes cases and 541 million are at risk of developing type 2 diabetes (17). The role of mtDNA variants as potential genetic causes in the incidence of type 2 diabetes is evidently observed in a number of patients by studying the dysfunction of energy production machinery of mitochondria (15, 16). Such variants cause mitochondrial anomalies and oxidative stress which affects pancreatic β-cells, thus leading to disease pathology (18). Pedigree analysis in families affected with diabetes showed that inheritance in the maternal line was more frequently observed than in the paternal line (19), which suggests the role of maternally inherited mtDNA in pathogenesis of diabetes. In Pakistan, the prevalence of type 2 diabetes is much higher than that estimated previously as reported to be 16.98% (95% CI 16.44–17.51) in a community-based national study (20).

Sequencing of coding and control regions of the mitochondrial genome can provide a detailed view of variants in diabetic subjects as well as individuals susceptible to diabetes. The aim of present study was to establish a clear link between incidence of type 2 diabetes and variations in mitochondrial genome. Genetic susceptibility to type 2 diabetes in relation to mtDNA variants present in Pakistani population was not studied previously. As the whole mitochondrial genome was sequenced through Next Generation Sequencing (NGS) technology in our study, it is possible to infer the genetic susceptibility of individuals to any mitochondrial disease.

Methods

Sample Collection

This is a case control study. After ethical approval (IRB-1135/DUHS/Approval/2018/), blood samples from 57 diabetic patients were recruited from National Institute of Diabetes and Endocrinology, Dow University of Health Sciences and 37 volunteer healthy controls from Karachi after taking their informed consent. Samples were collected in EDTA anticoagulated tubes and used for DNA isolation and sequencing.

Both males and females aged 17 years and above were included in the study. For the control group, subjects without a history of any disease including genetic disorders were included in the study. In the case of diabetic subjects, persons with known history of diabetes in accordance with T2DM WHO 1999 Diagnostic Criteria were included in the study. The HbA1c cut point for disease diagnosis was set as > 6.5% at the time of diagnosis. Pregnant females, patients with mental illness or having history of any other disease particularly genetic disorders were excluded from the study. Height and weight were measured, and body mass index (BMI) was calculated. The details of samples used in this study are given in Supplementary Table S1.

Isolation and amplification of DNA

From the collected blood samples, DNA was isolated using QIAamp DNA Mini Kit (cat# 51306, Qiagen). The concentration and purity of samples was determined through NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™). Isolated DNA was amplified using GoTaq PCR master mix (cat# M7122, Promega) following manufacturer's instructions. PCR primers validated for the amplification of mtDNA HVS2 were used in this study (21).

Hypervariable Segment 2 (HVS2) Sequencing and Data Analysis

HVS2 analysis was performed for 55 of 57 diabetic patients and 32 of 37 control subjects. PCR products were purified using AMPure XP magnetic beads according to the manufacturer's instructions (cat#A63800, Beckman Coulter, CA). 10–20 ng of purified DNA was used for Sanger sequencing. Briefly, purified PCR products were mixed with BigDye™ Terminator Sequencing reaction buffer and BigDye™ Terminator v3.1 Ready Reaction Mix (cat# 4337455). Forward primers were used for Sanger sequencing and the samples were subjected to capillary electrophoresis on the ABI 3500 genetic analyzer system (Thermo Scientific™). Sequence data was evaluated for quality and used for alignment with fully annotated mitochondrial genome sequence available at NCBI (NC_012920.1). Data was further processed for mutations and variation analysis using MEGA-X software and online service GEAR Genomics (https://www.gear-genomics.com/).

Amplification of whole mitochondrial genome for NGS

The entire mitochondrial genome was amplified from total DNA samples by REPLI-g Mitochondrial DNA Kit (cat# 151023, Qiagen). It amplifies the complete human mitochondrial genome using Multiple Displacement Amplification (MDA) technology which is suitable for NGS applications. Total 34 samples of diabetic subjects and 11 samples of healthy controls were processed for NGS of the whole mitochondrial genome.

Library preparation and sequencing

After the amplification of mtDNA, the samples underwent processing to become compatible with the Illumina sequencing platform. This was achieved through the utilization of the Nextera DNA sample preparation kit (cat# FC-141-1007, Illumina), following the instructions provided by the manufacturer. Qubit dsDNA assay (cat# Q32851, Thermo Scientific™) was used to measure the mtDNA concentrations for the Nextera tagmentation procedure. In a single step, tagmentation, enzymatic fragmentation and adapter sequence tagging of the input DNA were carried out through the Nextera transposome. Subsequently, the tagmented DNA was amplified through a PCR), which added unique barcodes (index adapters) and sequences necessary for cluster formation to each sample library (Nextera XT Index Kit, cat# FC-121-1012, Illumina). Next, amplified library DNA was purified and size-selected (short library fragments were removed) using magnetic beads. After purification, samples with different index tags were pooled together (2 µl each) and concentrated with a final volume of 15 µl. The libraries ready for sequencing were quantitated using Agilent 2100 Bioanalyzer High Sensitivity kit (cat# 5067 – 4626). The Illumina MiSeq system was used to sequence the DNA libraries.

Bioinformatics Analysis

Sample demultiplexing based on index sequences was carried out on raw data produced by Illumina MiSeq equipment and on-platform sequencing software which converted this data to FASTQ file format. After quality control, the paired-end FASTQ files underwent alignment with the reference sequence of mtDNA, the revised Cambridge Reference Sequence (rCRS) (GenBank NC_012920.1) (22), utilizing the short read sequence aligner BWA-MEM v.0.7.17 (23). These reads mapped to the mtDNA rCRS were sorted based on the start position and subsequently transformed into BAM files by employing SAMtools v.1.16 (24). Then marking and removing of optical duplicate reads was performed by using Picard tools (v.2.26.1; Broad Institute). The default thresholds applied for mapping quality, base quality, and alignment quality scores in the Phred scale were 20, 10, and 20, respectively. The identification of heteroplasmic and homoplasmic variants was executed by utilizing Mutserve v.1.2.1, a stand-alone version of the web based service mtDNA-server (25). The minimum level set for heteroplasmy was 0.05 - loci with a heteroplasmy level below this threshold were assigned homoplasmic wild-type alleles and loci with a heteroplasmy level above 0.95 were assigned as homoplasmic variants. Variant calling was performed using BCFtools v.1.15 (24), mpileup and norm to normalize and left align the sequences to the rCRS reference sequence. The resulting VCF files were then used for interpretation of variants. The flow chart of the data analysis performed in our study is presented in Fig. 1.

Comparison of mtDNA variant frequency with public databases

The frequencies of whole mtDNA variants identified in diabetic individuals were compared with the allele frequencies in related population data available in public databases. The gnomAD v3.1.2 database (https://gnomad.broadinstitute.org/) contains information on mitochondrial genomic variants from 56,434 samples so it is a valuable resource for the comparison of allele frequency and interpretation of variants (26). South Asian (SAS) population allele frequencies from the gnomAD database were compared with those observed in our diabetic subjects. The frequency of mtDNA variants found in diabetic individuals was also compared with the allele count in the Punjabi from Lahore Population (PJL) from the 1000 Genomes Project (https://www.internationalgenome.org/).

Variant interpretation

The filtered variants were interpreted and variants of interest were identified. Mitomap database (27) and mtDNA-server (25) were used for annotation of identified mtDNA variants providing details about the variant locus in the mitochondrial genome, translation effect, percentage among the present set of GenBank and gnomAD v3.1 sequences, frequency in the predicted haplogroups and phenotype reports. Variants were then filtered by their effect on translation and rare variants were defined as those with ≤ 0.05% GenBank frequency. Rare non-synonymous amino acid changing variants, rRNA and tRNA variants were further evaluated to determine their functional relevance using in silico predictive tools, publicly accessible mtDNA databases and scientific literature. MutPred scores were given to non-synonymous substitutions of diabetic subjects using mtDNA Server. Non-synonymous SNVs were defined as possibly deleterious and high confidence harmful based on MutPred score of ≥ 0.5 and ≥ 0.7 respectively as mentioned previously (28). Haplogroups of HVS2 sequenced samples were determined by using online web service Haplogrep 3 (29). FASTA files of diabetic and control samples were separately uploaded to Haplogrep 3 and a graphical output altogether for all samples of a group was created.

Statistical analyses

Mitochondrial genome variation was compared between diabetic subjects and the ethnically matched unaffected individuals. The statistical analysis was performed using Python. Fisher's exact test was performed to assess the association of variants of interest with diabetes. The significance of the observed differences was determined and when the p-value was less than 0.05, the results were considered statistically significant.

Results

The clinical characteristics and demographic details of our study participants are presented in Table 1. Group of diabetic patients (N = 57) consisted of 61.40% males and 38.60% females and a mean age at the time of sample collection was 50.52 years. The mean age of the control group (N = 37) was 31.3 years and it consisted of 78.8% males and 21.6% females. The mean Body Mass Index (BMI) of diabetic and healthy individuals was 25.69 and 23.3 respectively. All the participants of this study were recruited from diverse ethnic backgrounds that were similarly distributed in both study groups of diabetic and control subjects. Complete demographic details of these individuals are presented in Supplementary Table S2.

Table 1
Clinical and demographic details of study participants.

	Diabetic patients	Controls
Age (years)		
Mean ± SD	50.52 ± 10.0	31.3 ± 7.7
Gender		
Male	61.4%	78.8%
Female	38.6%	21.6%
BMI		
Mean ± SD	25.69 ± 4.8	23.3 ± 3.1
Years since onset of diabetes		
Mean ± SD	4.82 ± 4.3	N/A
Both Diabetic (N = 57) and control (N = 37) groups are composed of individuals from similar ethnicities commonly present in Sindh province. BMI: Body Mass Index, SD: Standard Deviation.		

Hypervariable Segment 2 (HVS2) Variants

The mtDNA HVS2 spans nucleotide positions m.57–372 and was sequenced by Sanger sequencing which revealed that total 81 variants were present in both controls and diabetic subjects. These variants along with their distribution in both groups, i.e. the number of exclusive variations (present in one group but absent in another) and variants shared between both groups are shown in Venn diagram (Fig. 2). We found that two variants in HVS2 had significantly different distributions between both groups. They are shown in Table 2. Haplogroups of the diabetic and control samples were determined by using Haplogrep 3 and they are shown in Fig. 3. Haplogroups H and K were observed quite frequently in diabetic subjects as together they constitute approximately 70% of diabetic samples. In control samples, haplogroups U, H and M were found frequently (nearly 70%) whereas haplogroup K was found in only one sample (3%).

Table 2
HVS2 variants found to have association with diabetes.

Variant	Count in Diabetic Patients (n = 55)	Count in Controls (n = 32)	p-value	Odds Ratio	95% Confidence Interval
m.315dup	19	2	0.003	7.91	1.70 to 36.76
m.309_310insCT	12	1	0.026	8.65	1.06 to 70.05
Fisher exact test, odds ratio and 95% confidence interval are given for the variants. p value less than 0.05 was considered statistically significant.					

Classification of whole mitochondrial genome variants

A total of 1,092 and 353 variants were identified in diabetic and control subjects, respectively, through next generation sequencing of complete mitochondrial genome. On average 33 variants per sample were observed in both groups (Fig. 4). The number of unique and non-iterating variants in diabetic subjects was found to be 341 and 145 in controls. The exclusive and shared variants between both groups are shown in Fig. 5.

Non-synonymous mtDNA variants identified in diabetic subjects

Non-synonymous variants can be a potential genetic predisposition to diseases, therefore such variants in diabetic individuals were the focus of this study. A total of 49 unique non-synonymous variants in the coding region of the mitochondrial genome were identified in diabetic subjects. Of these, 40 variants were present exclusively in diabetic subjects (Table 3). Four variants were previously found to be associated with diabetes: m.3316G > A, m.4216T > C, m.4833A > G and m.13204G > A. Variant m.4216T > C is linked with insulin resistance which is the underlying cause of type 2 diabetes in many cases (Hu Y, Liu J, He Y, Chen J, 2015) (30). Moreover, another variant m.16189T > C in the non-coding region of the mitochondrial genome, reported to be linked with diabetes was also found in diabetic samples only. The characteristics of all diabetes associated variants found only in diabetic subjects are shown separately in Table 4. These are the common variants and all are haplogroup markers except m.4833A > G. Variant m. 4216T > C showed MutPred score of 0.611 while the remaining variants were below threshold of a 0.5 score. All disease associated variants as reported in Mitomap identified in diabetic group as well as in control group can be found in supplementary Table S3.

Table 3
Non-synonymous variants identified in the mitochondrial genome of diabetic subjects.

Variant	Mutation type	MutPred score	Locus: Amino Acid Change	Patient Report
m.8584G > A	transition	0.553	ATPase6:A20T	-
m.8594T > C	transition	-	ATPase6:I23T	-
m.9067A > G	transition	-	ATPase6:M181V	-
m.8684C > T	transition	-	ATPase6:T53I	-
m.9142G > A	transition	0.695	ATPase6:V206I	-
m.8418T > C	transition	-	ATPase8:L18P	Mitochondrial Respiratory Chain Disorder
m.8502A > G	transition	-	ATPase8:N46S	-
m.8477T > C	transition	-	ATPase8:S38P	-
m.8396A > G	transition	-	ATPase8:T11A	-
m.6261G > A	transition	-	COI:A120T	Prostate Cancer, LHON
m.7389T > C	transition	-	COI:Y496H	-
m.7859G > A	transition	-	COII:D92N	Progressive Encephalomyopathy
m.9336A > G	transition	-	COIII:M44V	-
m.15317G > A	transition	-	Cytb:A191T	-
m.15287T > C	transition	-	Cytb:F181L	Possible DEAF helper mutation
m.15452C > A	transversion	-	Cytb:L236I	-
m.15458T > C	transition	-	Cytb:S238P	-
m.3316G > A	transition	-	ND1:A4T	*Diabetes*, LHON, PEO, vascular dementia
m.4232T > C	transition	0.609	ND1:I309T	-
m.4225A > G	transition	-	ND1:M307V	-

Amino acid substitution and patient report from Mitomap database is given along with MutPred scores. MutPred score of variants greater than 0.5 is shown.

Variant	Mutation type	MutPred score	Locus: Amino Acid Change	Patient Report
m.3397A > G	transition	0.723	ND1:M31V	ADPD, possibly LVNC-cardiomyopathy associated
m.3505A > G	transition	-	ND1:T67A	-
m.4216T > C	transition	-	ND1:Y304H	LHON, *Insulin Resistance*, miscarriage
m.5460G > A	transition	0.505	ND2:A331T	AD, PD, LHON
m.5442T > C	transition	-	ND2:F325L	-
m.4965A > G	transition	-	ND2:S166G	-
m.4833A > G	transition	-	ND2:T122A	*Diabetes* helper mutation AD, PD
m.5046G > A	transition	-	ND2:V193I	-
m.4491G > A	transition	-	ND2:V8I	High altitude pulmonary edema susceptibility
m.10084T > C	transition	-	ND3:I9T	-
m.10149A > T	transversion	-	ND3:M31L	-
m.11253T > C	transition	-	ND4:I165T	LHON, PD
m.11016G > A	transition	-	ND4:S86N	-
m.13135G > A	transition	-	ND5:A267T	Possible HCM susceptibility
m.13477G > A	transition	0.601	ND5:A381T	-
m.13781T > C	transition	-	ND5:I482T	-
m.14062A > G	transition	-	ND5:I576V	-
m.13651A > G	transition	0.626	ND5:T439A	-
m.13966A > G	transition	-	ND5:T544A	-
m.13204G > A	transition	-	ND5:V290I	Peripheral neuropathy of *T2 diabetes*

Amino acid substitution and patient report from Mitomap database is given along with MutPred scores. MutPred score of variants greater than 0.5 is shown.

Table 4
SNVs with reported association with Diabetes.

Variant	Count in Diabetic Patients (n = 34)	GB Frequency	Locus	Translation effect	Haplogroup Marker	Hetero-plasmy	MutPred score	Disease Reports
m.3316G > A	2	0.96%	ND1	Non-Syn; A = > T	Yes; R30a1c	**Yes**	0.463	**Diabetes**, LHON, PEO, vascular dementia
m.4216T > C	2	10.30%	ND1	Non-Syn; Y = > H	Yes; R2D	No	**0.611**	**Insulin Resistance**, LHON, possible adaptive high altitude variant, miscarriage
m.4833A > G	1	0.99%	ND2	Non-Syn; T = > A	No	No	0.488	**Diabetes** helper mutation AD, PD
m.13204G > A	1	0.07%	ND5	Non-Syn; V = > I	Yes; U7b	No	0.413	Peripheral neuropathy of **T2 diabetes**
m.16189T > C	5	25.020%	non-coding	**Non-coding**	Yes; U7b	**Yes**	NA	**Diabetes**, Cardiomyopathy, cancer risk, mtDNA copy nbr, Metabolic Syndrome, Melanoma patients

SNVs linked with **Diabetes** and **Insulin Resistance** as reported in Mitomap database present only in diabetic subjects. Findings of interest are shown in bold italics. m.4216T > C showed higher MutPred score (i.e. greater than 0.5).

By assigning MutPred scores to the non-synonymous mtDNA variants, their functional impact can be assessed. The MutPred algorithm assigns each non-synonymous variant in the protein-encoding regions of mitochondrial genome a pathogenicity score between 0 and 1 (28). Variants with Mutpred score greater than 0.75 observed in both groups are given in Table 5 and all those with score greater than 0.5 are listed in supplementary Table S4. Notably, variants with higher MutPred score (greater than 0.5) are transversions with high occurrence. mtDNA variants with score of 0.5 or greater were more frequently present in diabetic subjects than in controls subjects.

Table 5
Non-synonymous mtDNA variants with high Mutpred score (> 0.75) found in diabetic and control individuals.

Variant	Count in Diabetic Subjects (n = 34)	Count in Controls (n = 11)	Substitution type	Gene	Codon Position	Amino Acid position	Amino Acid change	MutPred Score
m.4651T > C	1	0	transition	MT-ND2	2	61	L->P	0.876
m.14624A > C	3	0	transversion	MT-ND6	2	17	V->G	0.801
m.6369T > C	1	1	transition	MT-CO1	1	156	S->P	0.799
m.14778T > A	1	0	transversion	MT-CYB	2	11	M->K	0.791
m.9041A > C	1	1	transversion	MT-ATP6	2	172	H->P	0.78
m.11343T > A	1	0	transversion	MT-ND4	2	195	M->K	0.78
m.6943T > C	1	1	transition	MT-CO1	2	347	L->P	0.776

Annotation of the variants using mtDNA Server was performed to determine the effect of non-synonymous amino acids substitution.

Novel and rare variants identified in mtDNA

Mitochondrial DNA variants with less than 0.05 percent GenBank frequency, available from Mitomap, were considered as rare variants. In diabetic subjects, 40 non-iterating rare variants were detected while in control samples 24 such variants were identified. Novel mtDNA variants were defined as those that were not previously reported in public databases and not frequently observed in our samples. One novel variant was identified in our study that was present only in a diabetic subject (Table 6). This variant 570C > CACCC is the insertion found in the HVS3 region of mitochondrial genome. All the variants below 0.05 percent GenBank frequency are listed in supplementary Table S5.

Table 6
Novel variant found in mitochondrial genome of a diabetic individual.

mtDNA variant	Count in Diabetic Subjects (n = 34)	Count in Controls (n = 11)	Mutation type	Locus	GB Frequency FL or CR	gnomAD Frequency	PJL Frequency
m.570C > CACCC	1	0	insertion	CR: HVS3	NR	NR	NR

This variant was not found in the Mitomap and gnomAD databases, Punjabi Lahore (PJL) data from the 1000 Genomes project, nor was frequently observed in our samples.

Comparison of mtDNA variant frequency with public databases

The frequencies of whole mtDNA variants identified in diabetic individuals were compared with the allele frequencies of the variants in gnomAD v3.1 (26) and the allele count in the Punjabi from Lahore Population (PJL) from the 1000 Genomes Project. Fisher's exact test was performed to compare mtDNA variant frequency in diabetic subjects and PJL data. The variants possibly associated with diabetes are listed in Table 7. Among these, four variants m.7193T > C, m.9336A > G, m.11935T > C and m.14766C > T were present in the coding region of mitochondrial genome.

Table 7
Comparison of mtDNA variants between diabetic and Punjabi from Lahore (PJL) samples of 1000 Genomes Project.

Variant	Count in Diabetic Subjects (n = 34)	Count in PJL (n = 96)	p value	Odds Ratio	95% Confidence Interval
m.513G > A	6	1	0.001	20.35	2.35 to 176.27
m.195T > C	10	9	0.009	4.02	1.47 to 11.03
m.16189T > C	5	2	0.013	8.10	1.49 to 43.99
m.7193T > C	3	0	0.016	21.44	1.07 to 426.61
m.9336A > G	3	0	0.016	21.44	1.07 to 426.61
m.11935T > C	3	0	0.016	21.44	1.07 to 426.61
m.16265A > C	4	2	0.040	8.103	1.49 to 43.99
m.14766C > T	27	89	0.049	0.30	0.09 to 0.94

Variants found to have association with diabetes are listed. Fisher exact test, odds ratio and 95% confidence interval is given for the variants. The p value less than 0.05 was considered statistically significant.

Discussion

The prevalence of diabetes has risen spectacularly over the years in Pakistan. According to the International Diabetes Federation, in Pakistan, 26.7% of adults have diabetes and this number can be lower than the expected as a large number of patients remain undiagnosed (31). It was recently reported that the incidence of diabetes was much higher in urban areas of Pakistan than in the rural areas (32). Therefore, the participants in this study belonged to the urban area of Karachi, Pakistan.

In our case-control study, we first analyzed the HVS2 region of the mitochondrial genome through Sanger sequencing to identify potential variants associated with diabetes. We found that the distribution of two variants, m.315dup and m.309_310insCT, was significantly different between the two groups, with higher prevalence of these two variants in diabetic patients when compared to controls. The m.302−315 positions cover the interrupted C stretch of the HVS2 region. The m.309_310insCT and m.315dup are relatively common variants according to GenBank frequency (25.791% and 30.086%, respectively), however they are not reported in the gnomAD database. Both variants were also observed quite frequently in our NGS-tested samples, although differences between groups were not statistically significant, possibly because of the small sample size of control group. After finding variants in the HVS2 with a possible association with diabetes, these variants were further evaluated using NGS to gain more insight into their relationship with diabetes. The most common haplogroups found in diabetic samples were H (43.64%) and K (25.45%) whereas in control samples haplogroups U (25%), H (21.88%) and M (21.88%) constituted the highest portion (Fig. 3).

The non-coding region of mtDNA exhibited 329 variants when NGS was carried out in samples from diabetic subjects. This is the highest number compared to any other mtDNA (coding) region due to the highly polymorphic nature of the mtDNA control region (Fig. 6). These mutations can interfere with some regulatory elements in the mtDNA control region and affect the binding affinity of promoters modulating the transcription and replication processes (33). Of these, the m.16189T > C variant was quite frequently found (14.7%) exclusively in samples of diabetic subjects. In previous findings, m.16189T > C was reported to be associated with an increased risk of type 2 diabetes in Asians (34) and was recently reported to be associated with type 2 diabetes in the individuals from Pakistani population (35). This variant is considered as the genetic marker of type 2 diabetes along with other variants in Asians (36). Hence, our results are in agreement with these reports.

Protein-coding regions constitute the vast majority of the mitochondrial genome. We evaluated non-synonymous mtDNA variants found exclusively in diabetic samples. Most of these variants were found in genes encoding the mitochondrial respiratory complex I subunits (ND1-ND6). Some previous studies have reported the association of mitochondrial ND genes variants with type 2 diabetes (37) and other metabolic syndromes (38). Interestingly, all the mtDNA coding region variants linked to diabetes (according to Mitomap patient reports) identified in our patients were non-synonymous and found in the ND1, ND2 and ND5 subunit genes (Tables 3 and 4). Variant m.14766C > T was frequently occurring and present in both diabetic and control groups as being a haplogroup marker of various clusters.

The pathogenicity prediction of non-synonymous mtDNA variants with MutPred tool can allow for assessing their functional impact. We evaluated possibly deleterious variants with MutPred scores above 0.5 and literature survey was conducted to determine whether there is any association of these variants with diabetes. It is important to note that the substitution types of the variants with higher MutPred scores (greater than 0.75) were transversions with high occurrence. This is consistent with a well-known fact that transversions frequently disrupt protein coding genes and result in amino acid changes (39).

Variants identified in the whole mtDNA in diabetic patients were also compared with data from public databases. Mitomap and gnomAD databases are of particular importance as they include variant details from different populations, including South Asian and Pakistani populations. This helped in the identification of novel variant in our study. Mitochondrial DNA variants not observed frequently in this study and not reported in databases and scientific literature were considered novel variants. We identified one novel variant located in non-coding region of mtDNA (Table 6). When comparison of mtDNA variant frequency between our diabetic samples and PJL population from 1000 Genomes Project was performed, eight variants were found to be possibly associated with diabetes (Table 7). Four of these variants m.7193T > C, m.9336A > G, m.11935T > C and m.14766C > T were present in the coding mtDNA region. However, no such association was observed when we compared diabetic samples with control samples from this project. This can be due to the limited size of control population in our study as large sample size can reveal more interesting variants that can be compared.

There are some variations that are strongly associated with diabetic phenotype in various populations but are not even present in the Pakistani population. For instance, it is known that increasing levels of heteroplasmic transition mutation m.3243A > G, a known pathogenic variant, are responsible for the pathogenesis of the diseases such as diabetes, neuromuscular degenerative disorder, and perinatal lethality in some populations (40, 41). However, m.3243A > G was not found in any diabetic subject in our study, and it was also absent in a recent study of six diabetic individuals of a Pakistani family (42).

As shown in the pedigree analysis in families affected with diabetes, inheritance in the maternal line was more frequently observed than in the paternal line (19), which suggests the role of mtDNA in pathogenesis of diabetes. To establish this scenario in Pakistani population, type 2 diabetic patients with family history should be studied along with their close maternal relatives so that mitochondrial genetic variations within similar haplotypes could give a more detailed picture. There is also involvement of nuclear genomic variants in the onset of diabetes and cause aberrations in the mitochondrial metabolism. All these important phenomena, studied together, could show more insight into the molecular basis of diabetes. Furthermore, this study has some limitations as it reflects data from a limited Pakistani population. Therefore, more detailed studies on a large number of samples with closely matched case-control groups in terms of age and gender could be very useful.

Conclusion

This study identified variants in HVS2 and complete mtDNA of control and diabetic subjects from Pakistani population. Two variants in HVS2 (m.309_310insCT and m.315dupC) were identified in Sanger-sequenced samples to have a possible association with diabetes in our study. Moreover, using the 1000 Genomes Project data for comparison we found that eight

mtDNA variations are significantly associated with diabetes. Although some mtDNA variants seem to be associated with type 2 diabetes in Pakistani population as found in our study, further research with increased sample sizes of studied groups can provide more insightful findings.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from Dow University of Health Sciences, Karachi to conduct study on human blood samples (IRB-1135/DUHS/Approval/2018/).

Consent for publication

Consent was taken from study participants for research and publications.

Availability of data and materials

The data have been deposited at NCBI GenBank with accession numbers PP134863-PP134909. The corresponding author can be contacted for further details, if any.

Competing interests

The authors declare no competing interests.

Authors' contributions

S.F., S.A.R.S.B.: The acquisition, analysis and interpretation of data for drafting and revising the content of manuscript.

M.S.: Recruitment of subjects in the study and managed compiling of their clinical parameters.

M.I.: Involved in experimental work and checked data quality.

S.F.H.N., A.P.N.: Reviewed the manuscript critically for important intellectual content.

I.A.K and A.J.: Critically read/edited the article provided and was partially involved in funds acquisition. All the authors have approved the final version of the manuscript to be published.

References

1. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290(5806):457–65.
2. Weerts MJA, Timmermans EC, Vossen RHaM, van Strijp D, Van den Hout–van Vroonhoven MCGN, van IJcken WFJ et al. Sensitive detection of mitochondrial DNA variants for analysis of mitochondrial DNA-enriched extracts from frozen tumor tissue. Sci Rep. 2018;8(1):2261.

3. Doimo M, Pfeiffer A, Wanrooij PH, Wanrooij S. mtDNA replication, maintenance, and nucleoid organization. The Human Mitochondrial Genome. Elsevier; 2020. pp. 3–33.

4. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. Nat Rev Genet. 2015;16(9):530–42.

5. Nesheva DV. Aspects of Ancient Mitochondrial DNA Analysis in Different Populations for Understanding Human Evolution. Balk J Med Genet. 2014;17(1):5–5.

6. Morris J, Na YJ, Zhu H, Lee JH, Giang H, Ulyanova AV, et al. Pervasive within-Mitochondrion Single-Nucleotide Variant Heteroplasmy as Revealed by Single-Mitochondrion Sequencing. Cell Rep. 2017;21(10):2706–13.

7. Kauppila TES, Kauppila JHK, Larsson NG. Mammalian Mitochondria and Aging: An Update. Cell Metab. 2017;25(1):57–71.

8. Clima R, Preste R, Calabrese C, Diroma MA, Santorsola M, Scioscia G, et al. HmtDB 2016: data update, a better performing query system and human mitochondrial DNA haplogroup predictor. Nucleic Acids Res. 2017;45(D1):D698–706.

9. Wallace DC. Mitochondrial DNA Variation in Human Radiation and Disease. Cell. 2015;163(1):33–8.

10. Poznyak AV, Ivanova EA, Sobenin IA, Yet SF, Orekhov AN. The Role of Mitochondria in Cardiovascular Diseases. Biology. 2020;9(6):137.

11. Sun D, Wei Y, Zheng HX, Jin L, Wang J. Contribution of Mitochondrial DNA Variation to Chronic Disease in East Asian Populations. Front Mol Biosci. 2019;6:128.

12. Hahn A, Zuryn S. The Cellular Mitochondrial Genome Landscape in Disease. Trends Cell Biol. 2019;29(3):227–40.

13. Grimm A, Eckert A. Brain aging and neurodegeneration: from a mitochondrial point of view. J Neurochem. 2017;143(4):418–31.

14. Ding Y, Gao BB, Huang JY. The role of mitochondrial DNA mutations in coronary heart disease. Eur Rev Med Pharmacol Sci. 2020;24(16):8502–9.

15. Li C, Xiang Y, Zhang Y, Tang D, Chen Y, Xue W, et al. A preliminary analysis of mitochondrial DNA atlas in the type 2 diabetes patients. Int J Diabetes Dev Ctries. 2022;42(4):713–20.

16. Pinti MV, Fink GK, Hathaway QA, Durr AJ, Kunovac A, Hollander JM. Mitochondrial dysfunction in type 2 diabetes mellitus: an organ-based analysis. Am J Physiol-Endocrinol Metab. 2019;316(2):E268–85.

17. Atlas ID. Diabetes around the world in 2021. Int Diabetes Fed. 2021.

18. Zhunina OA, Yabbarov NG, Grechko AV, Starodubova AV, Ivanova E, Nikiforov NG, et al. The Role of Mitochondrial Dysfunction in Vascular Disease, Tumorigenesis, and Diabetes. Front Mol Biosci. 2021;8:671908.

19. Lyssenko V, Groop L, Prasad RB. Genetics of Type 2 Diabetes: It Matters From Which Parent We Inherit the Risk. Rev Diabet Stud RDS. 2015;12(3–4):233–42.

20. Aamir AH, Ul-Haq Z, Mahar SA, Qureshi FM, Ahmad I, Jawa A, et al. Diabetes Prevalence Survey of Pakistan (DPS-PAK): prevalence of type 2 diabetes mellitus and prediabetes using HbA1c: a population-based survey from Pakistan. BMJ Open. 2019;9(2):e025300.

21. Daud S, Shahzad S, Shafique M, Bhinder MA, Niaz M, Naeem A, et al. Optimization and validation of PCR protocol for three hypervariable regions (HVI, HVII and HVIII) in human mitochondrial DNA. Adv Life Sci. 2014;1(3):165–70.

22. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23(2):147.

23. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25(16):2078–9.

25. Weissensteiner H, Forer L, Fuchsberger C, Schöpf B, Kloss-Brandstätter A, Specht G, et al. mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. Nucleic Acids Res. 2016;44(W1):W64–9.

26. Laricchia KM, Lake NJ, Watts NA, Shand M, Haessly A, Gauthier L, et al. Mitochondrial DNA variation across 56,434 individuals in gnomAD. Genome Res. 2022;32(3):569–82.

27. Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, et al. mtDNA variation and analysis using mitomap and mitomaster. Curr Protoc Bioinforma. 2013;44(1):1–23.

28. Pereira L, Soares P, Radivojac P, Li B, Samuels DC. Comparing Phylogeny and the Predicted Pathogenicity of Protein Variations Reveals Equal Purifying Selection across the Global Human mtDNA Diversity. Am J Hum Genet. 2011;88(4):433–9.

29. Schönherr S, Weissensteiner H, Kronenberg F, Forer L. Haplogrep 3 - an interactive haplogroup classification and analysis platform. Nucleic Acids Res. 2023;51(W1):W263–8.

30. Czech MP. Insulin action and resistance in obesity and type 2 diabetes. Nat Med. 2017;23(7):804–14.

31. Azeem S, Khan U, Liaquat A. The increasing rate of diabetes in Pakistan: A silent killer. Ann Med Surg. 2012. 2022;79:103901.

32. Adnan M, Aasim M. Prevalence of Type 2 Diabetes Mellitus in Adult Population of Pakistan: A Meta-Analysis of Prospective Cross-Sectional Surveys. Ann Glob Health. 2020;86(1):7.

33. Akouchekian M, Houshmand M, Hemati S, Ansaripour M, Shafa M. High rate of mutation in mitochondrial DNA displacement loop region in human colorectal cancer. Dis Colon Rectum. 2009;52(3):526–30.

34. Park KS. The Search for Genetic Risk Factors of Type 2 Diabetes Mellitus. Diabetes Metab J. 2011;35(1):12–2.

35. Association of Mitochondrial HVS-I Region Variants with. Type 2 Diabetes in Pakistani Diabetic Subjects. J Coll Physicians Surg Pak. 2023;33(07):980–5.

36. Gumilar GG, Purnamasari Y, Setiadi R. Mitochondrial DNA variant at HVI region as a candidate of genetic markers of type 2 diabetes. In: AIP Conference Proceedings. AIP Publishing; 2016 Feb 8 (Vol. 1708, No. 1).

37. Al-Ghamdi BA, Al-Shamrani JM, El-Shehawi AM, Al-Johani I, Al-Otaibi BG. Role of mitochondrial DNA in diabetes Mellitus Type I and Type II. Saudi J Biol Sci. 2022;29(12):103434.

38. Kraja AT, Liu C, Fetterman JL, Graff M, Have CT, Gu C, et al. Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits. Am J Hum Genet. 2019;104(1):112–38.

39. Guo C, McDowell IC, Nodzenski M, Scholtens DM, Allen AS, Lowe WL, et al. Transversions have larger regulatory effects than transitions. BMC Genomics. 2017;18:394.

40. Kopinski PK, Janssen KA, Schaefer PM, Trefely S, Perry CE, Potluri P, et al. Regulation of nuclear epigenome by mitochondrial DNA heteroplasmy. Proc Natl Acad Sci. 2019;116(32):16028–35.

41. Shen X, Du A. The non-syndromic clinical spectrums of mtDNA 3243A > G mutation. Neurosciences. 2021;26(2):128–33.

42. Abrar S, Muhammad K, Zaman H, Khan S, Nouroz F, Bibi N. Molecular genetic analysis of Type II diabetes associated m.3243A > G mitochondrial DNA mutation in a Pakistani family. Egypt J Med Hum Genet. 2017;18(3):305–8.
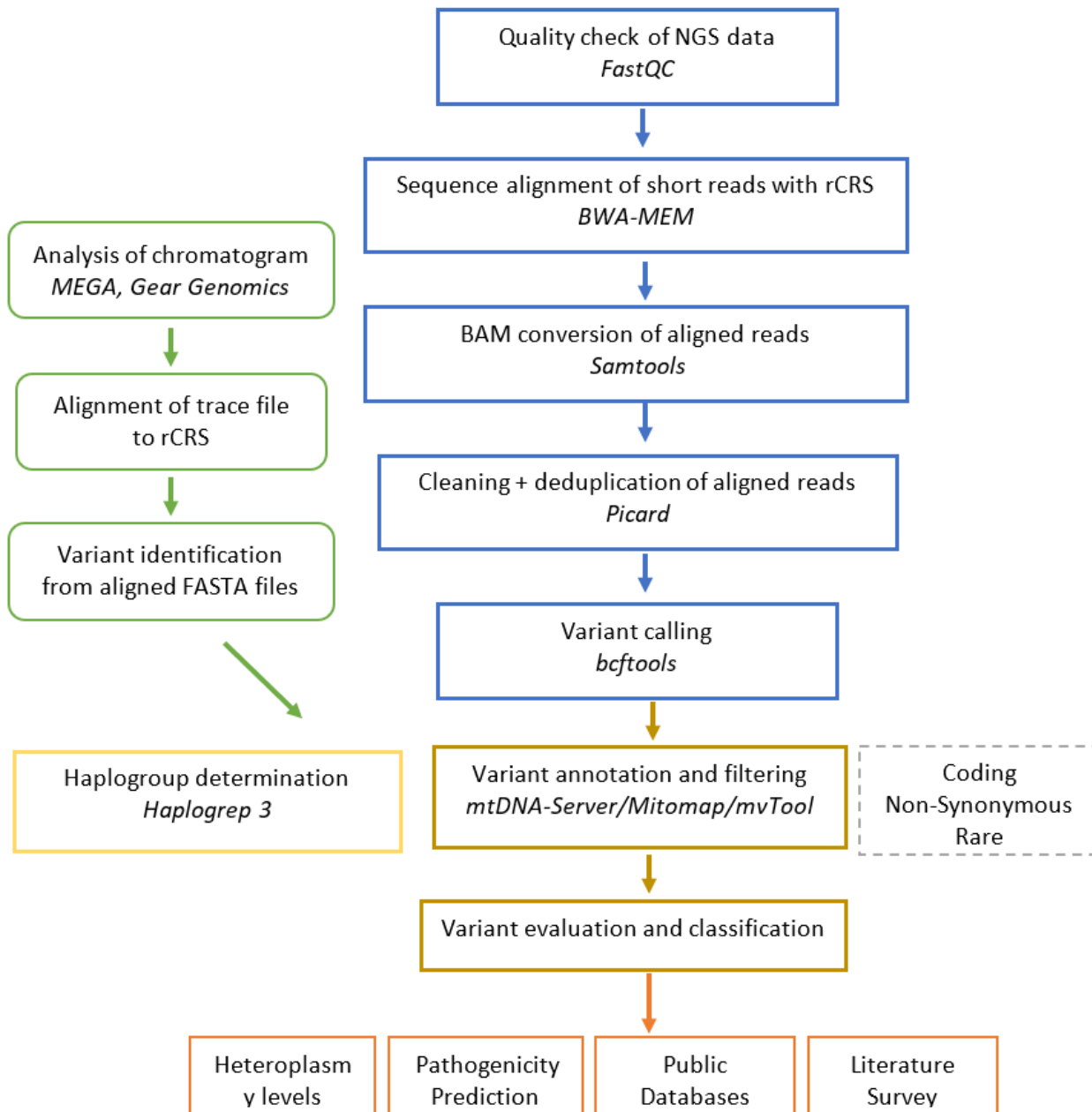
Figures

Figure 1

Workflow of methodology and bioinformatics analysis. DNA isolated from samples was PCR amplified, subjected to Sanger sequencing and analyzed (green boxes). NGS analysis pipeline used is depicted by blue boxes. Tools used for NGS data and variant analysis are written in italics. Variant annotation and classification was carried out for filtered variants (in dashed box) and in depth analysis was performed.
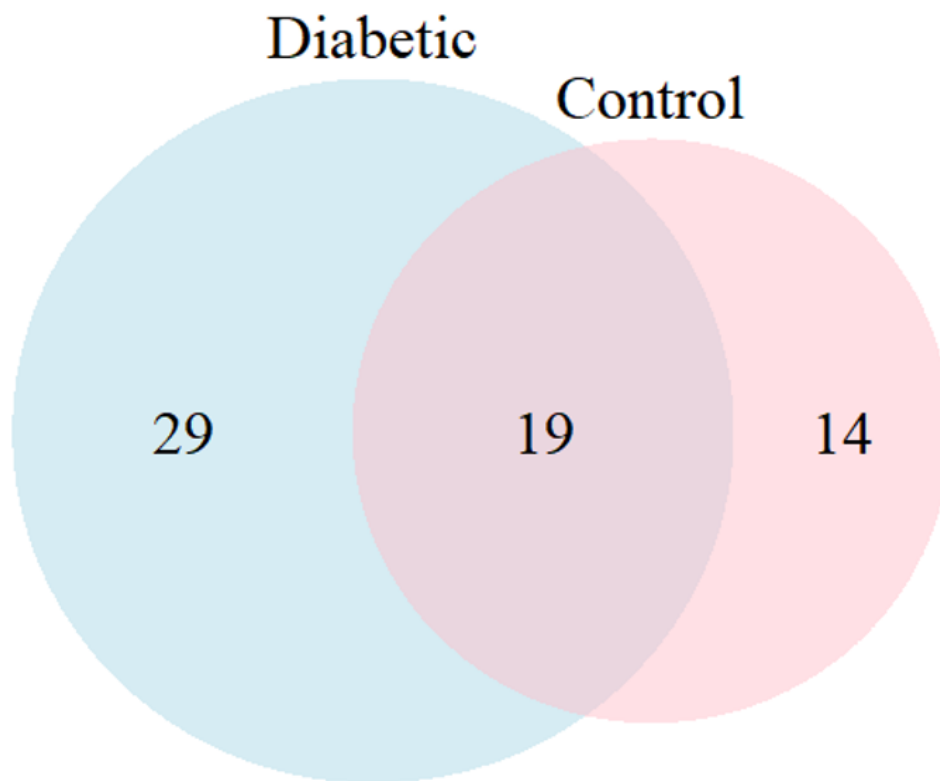
Figure 2

Distribution of HVS2 variants in diabetic and control groups analyzed by Sanger sequencing. 29 unique variable sites were exclusively present in diabetic subjects and 14 unique variable sites were present only in control samples. Moreover, 19 variants were found to be shared between both groups.
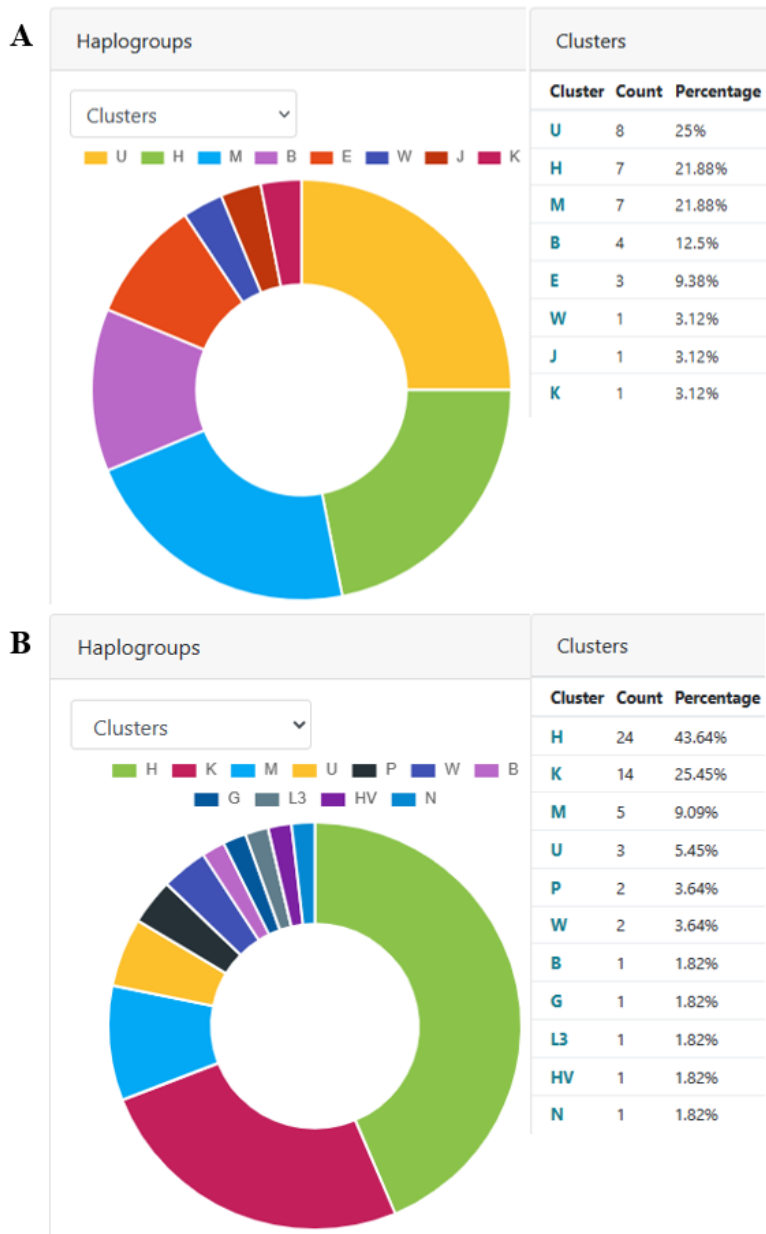
A

Haplogroups				Clusters		
Clusters ⌄				**Cluster**	**Count**	**Percentage**
▣ U ▣ H ▣ M ▣ B ▣ E ▣ W ▣ J ▣ K				U	8	25%
				H	7	21.88%
				M	7	21.88%
				B	4	12.5%
				E	3	9.38%
				W	1	3.12%
				J	1	3.12%
				K	1	3.12%

B

Haplogroups				Clusters		
Clusters ⌄				**Cluster**	**Count**	**Percentage**
▣ H ▣ K ▣ M ▣ U ▣ P ▣ W ▣ B				H	24	43.64%
▣ G ▣ L3 ▣ HV ▣ N				K	14	25.45%
				M	5	9.09%
				U	3	5.45%
				P	2	3.64%
				W	2	3.64%
				B	1	1.82%
				G	1	1.82%
				L3	1	1.82%
				HV	1	1.82%
				N	1	1.82%

Figure 3

Frequency of mitochondnrial haplogroups in the studied individuals. Percentages of the haplogroups found in (A) healthy controls (n=32) and (B) diabetic subjects (n=55) are shown. Haplogroups U, H and M were the top haplogroups present in controls while the hapogroups H and K accounted for the majority of the haplogroups in diabetic subjects.
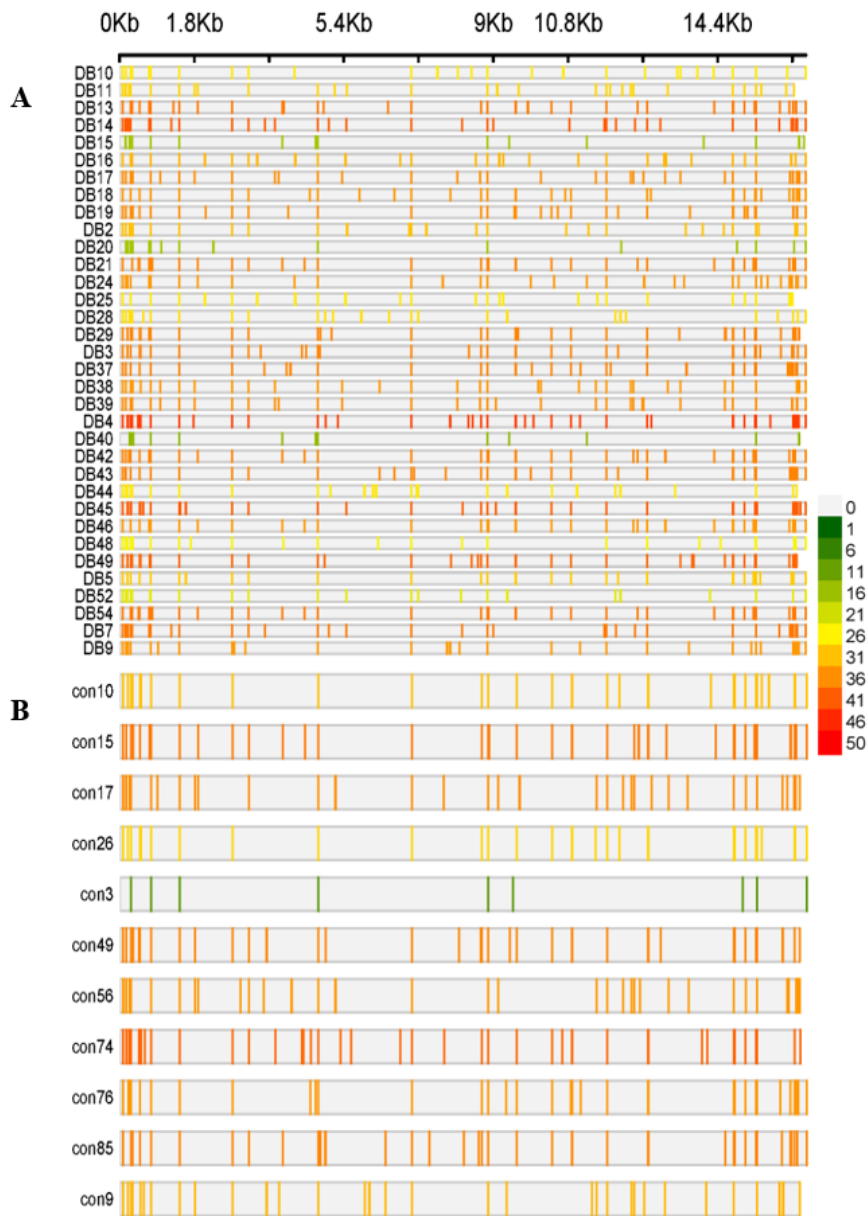
Figure 4

SNV density in the studied samples analyzed by NGS. Number and position of variants detected in (A) diabetic subjects (n=34) and (B) healthy controls (n=11) are shown. SNV density scale is given by grading of colors.
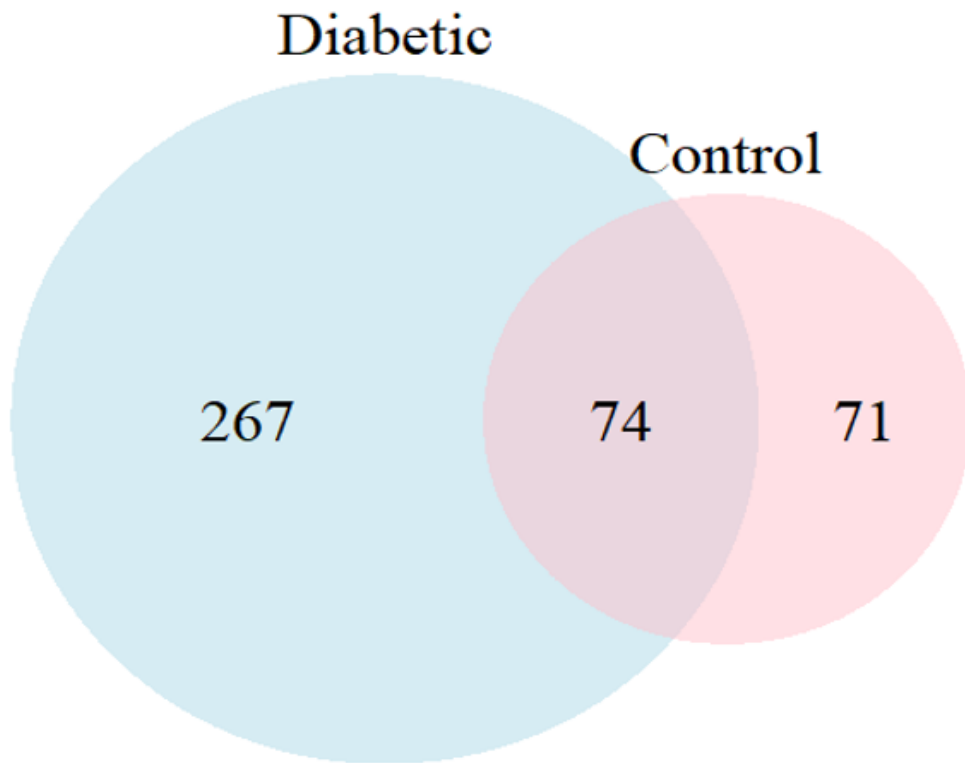
Figure 5

Venn diagram showing the number of mtDNA variants analyzed by NGS. The variants present exclusively in diabetic subjects or controls and shared between both groups are given.
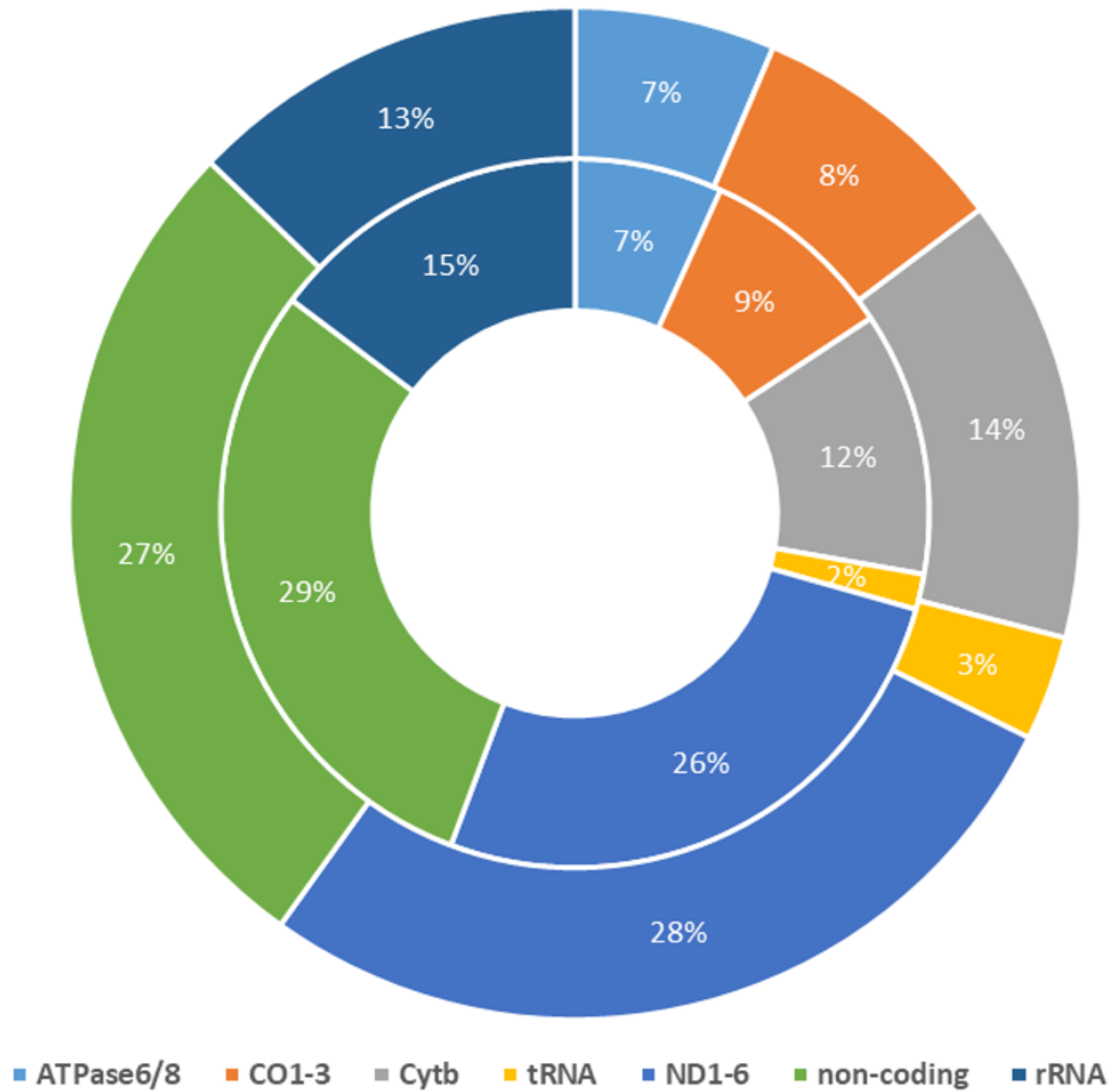
Figure 6

Distribution of variants found in different regions of mitochondrial genome of both groups. The outer side of this donut chart shows percentage of variants found in diabetic subjects while inner side - in control subjects.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS1.docx
- SupplementaryTableS1.xlsx
- SupplementaryTableS2.xlsx
- SupplementaryTableS3.xlsx
- SupplementaryTableS4.xlsx
- TableS5.docx