Supplementary Figures and Tables

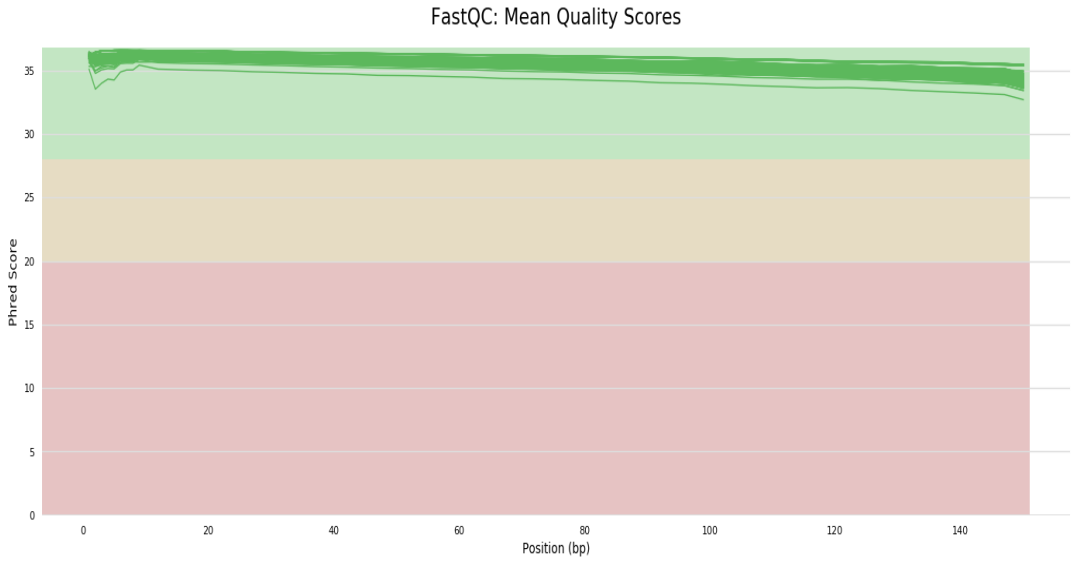
Sample Name	Total Data (bp)	Mapped data to Mito_Genome (bp)	X coverage
AGENOME-ZPMS-HV2a-1 (Nanopore)	24620822729 (24.6 Gb)	23156357 (23mb)	1447.27
AGENOME-ZPMS-HV2a-1 (Nanopore)	15157201611 (15.15 Gb)	7718168 (7.7 mb)	482.38

Sample Name	Total (Reads)	Total data in GB	Mitochondrial Reads in data	Mitochondrial coverage (mb)	X coverage
AGENOME-ZPMS-HV2a-1 (Illumina)	320987263	96.24	229095	6.8	4295

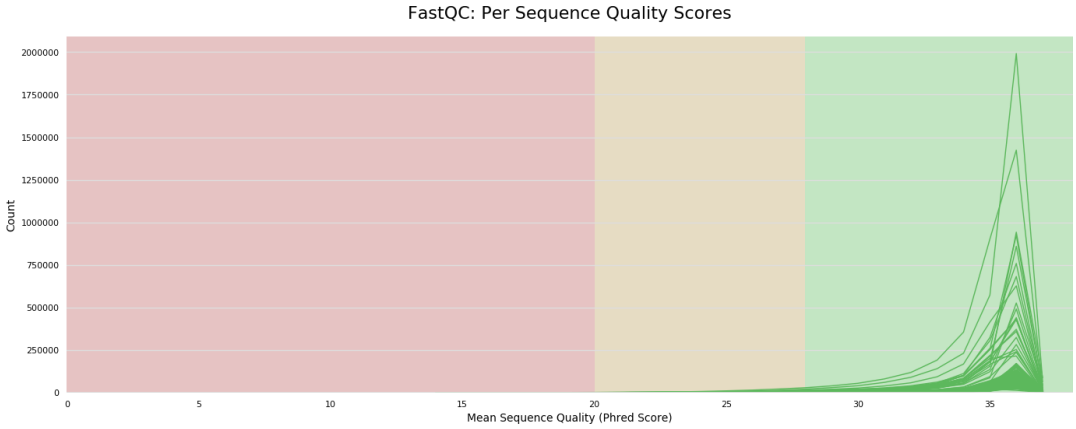
Supplementary Figure 1: QC data of the *de novo* Zoroastrian Parsi Mitochondrial Reference Genome (AGENOME-ZPMRG-HV2a-1)

Supplementary Figure 2: QC analysis of 100 Zoroastrian-Parsi mitochondrial genome sequences

A

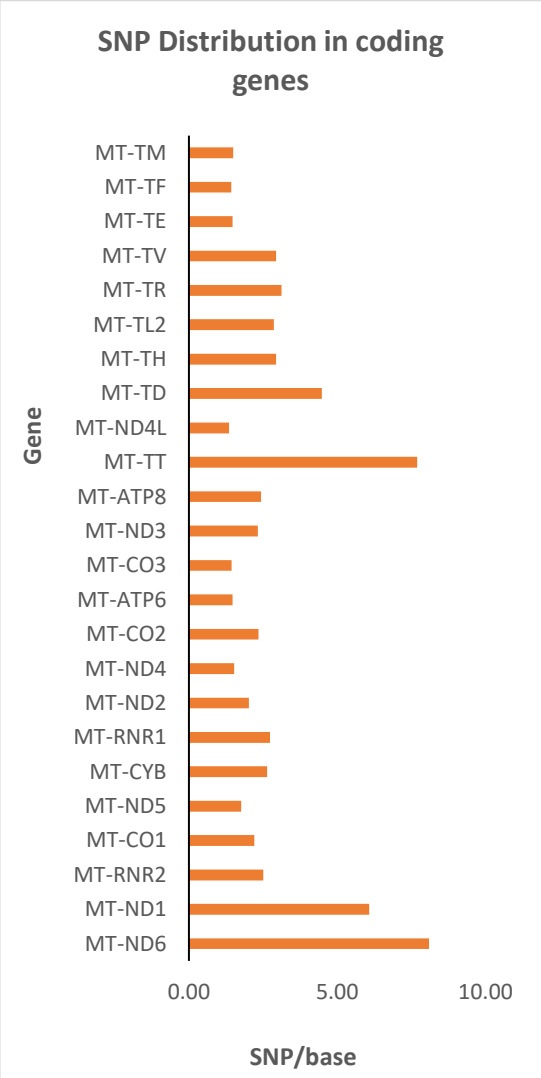


B



Supplementary Figure 2: QC analysis of 100 Parsi mitochondrial genomes (A) Frequency of mean PHRED score per read (150 read length) for 100 mitochondrial sample (B) Frequency of mean PHRED score per sequence for 100 mitochondrial samples

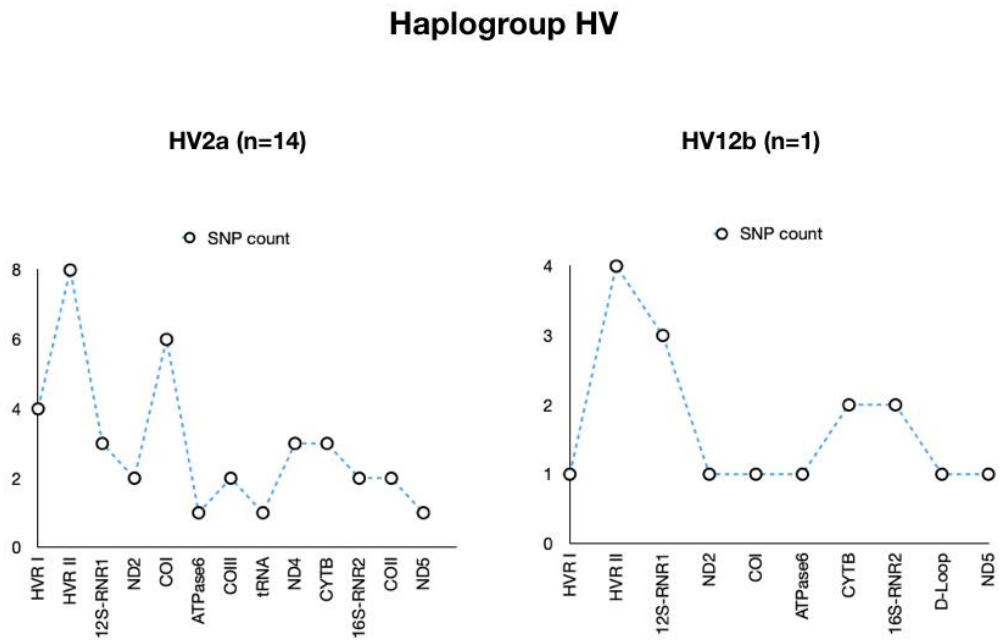
Supplementary Figure 3: Distribution of 420 variants across coding genes normalized for gene length



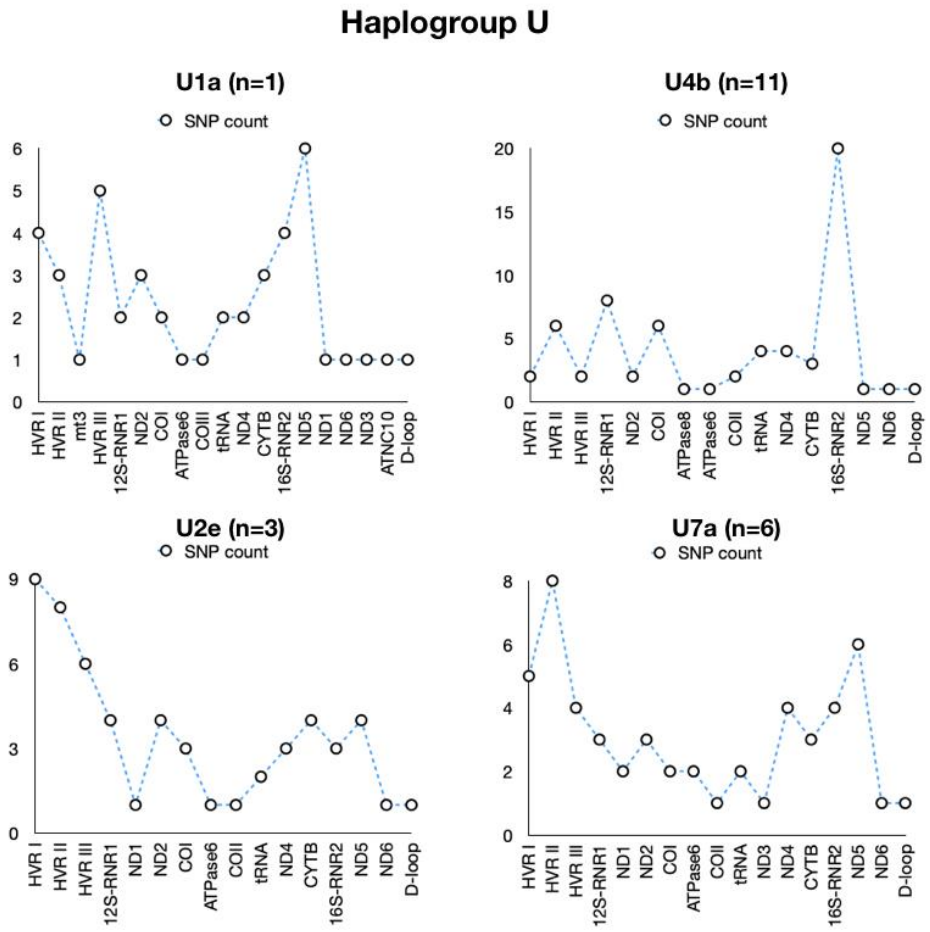
Gene	SNP/gene length
MT-ND6	8.10
MT-ND1	6.07
MT-RNR2	2.50
MT-CO1	2.21
MT-ND5	1.77
MT-CYB	2.63
MT-RNR1	2.73
MT-ND2	2.02
MT-ND4	1.53
MT-CO2	2.34
MT-ATP6	1.47
MT-CO3	1.44
MT-ND3	2.32
MT-ATP8	2.43
MT-TT	7.69
MT-ND4L	1.35
MT-TD	4.48
MT-TH	2.94
MT-TL2	2.86
MT-TR	3.13
MT-TV	2.94
MT-TE	1.47
MT-TF	1.43
MT-TM	1.49

Supplementary Figure 3: Distribution of 420 variants across coding genes normalized to gene length (variants/gene length (in kb))

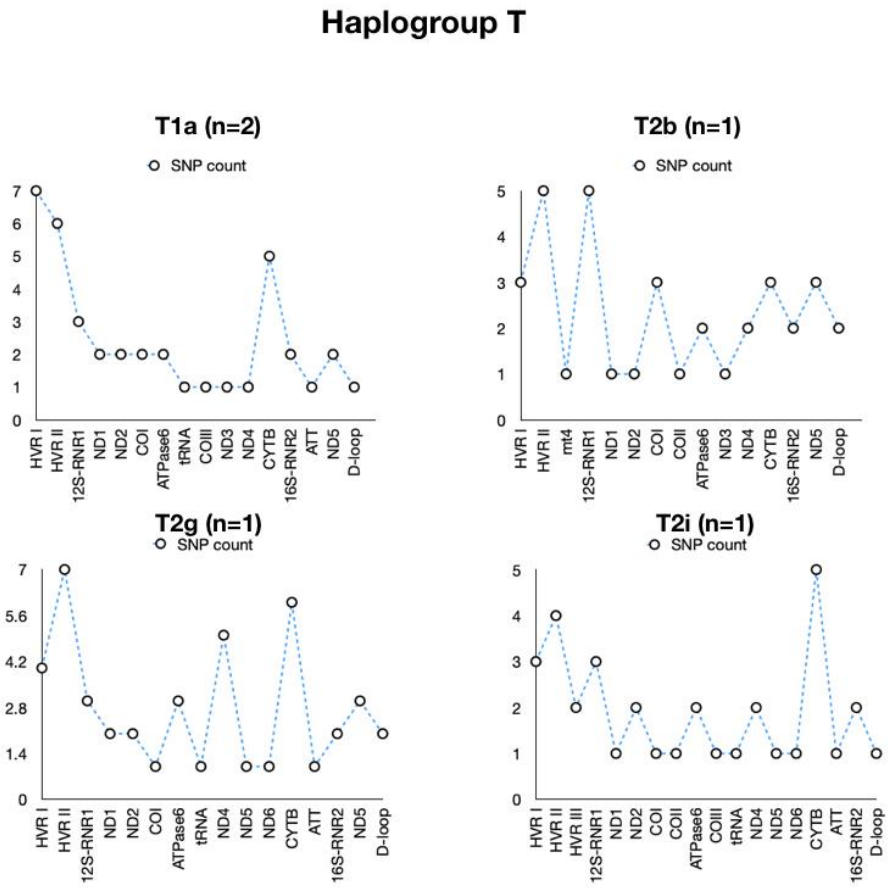
Supplementary Figure 4: Sub-haplogroup specific breakdown of 420 variants



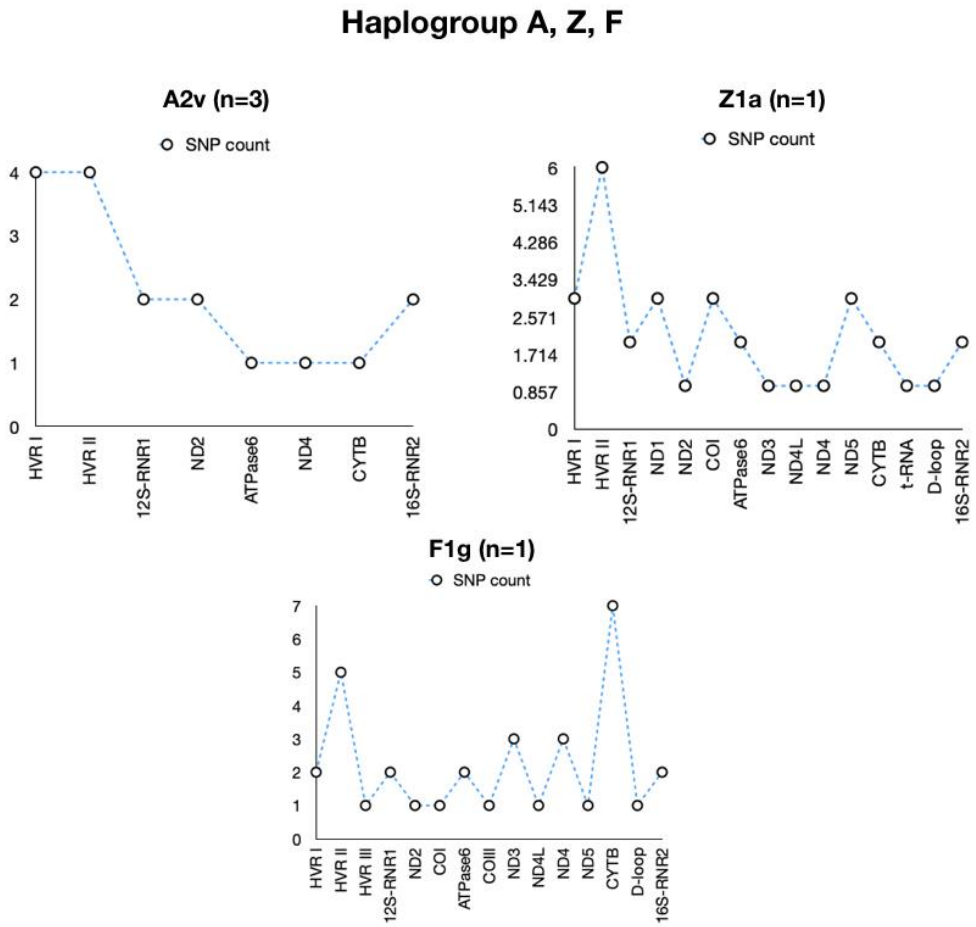
Supplementary Figure 4A: Distribution of Variants across gene loci in the HV haplogroup consisting of HV2a (n=14 subjects and HV12b (n=1 subject)



Supplementary Figure 4B: Distribution of Variants across gene loci in the U haplogroup consisting of U1a, U4b, U2e and U7a

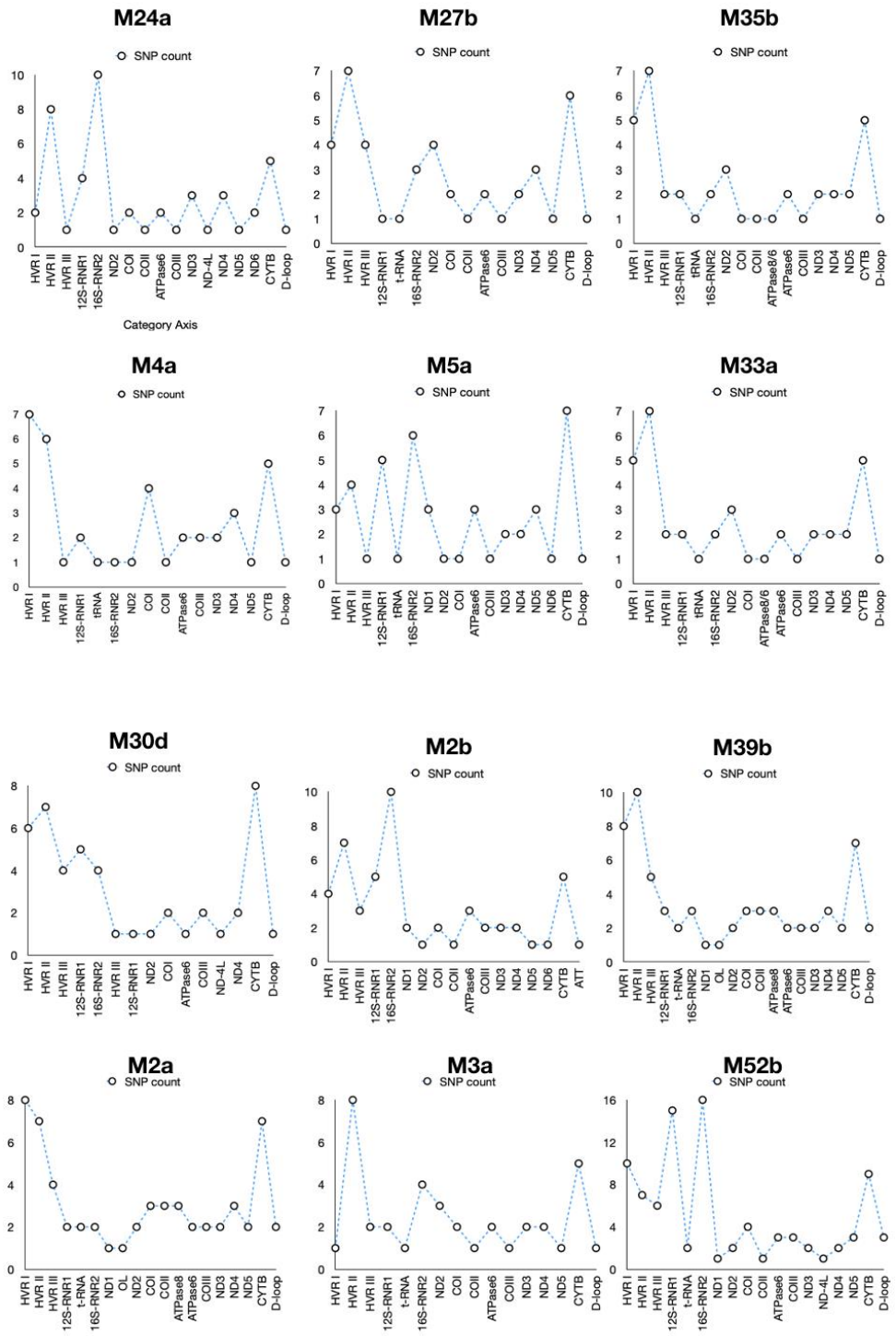


Supplementary Figure 4C: Distribution of Variants across gene loci in the T haplogroup consisting of T1a, T2b, T2g and T2i



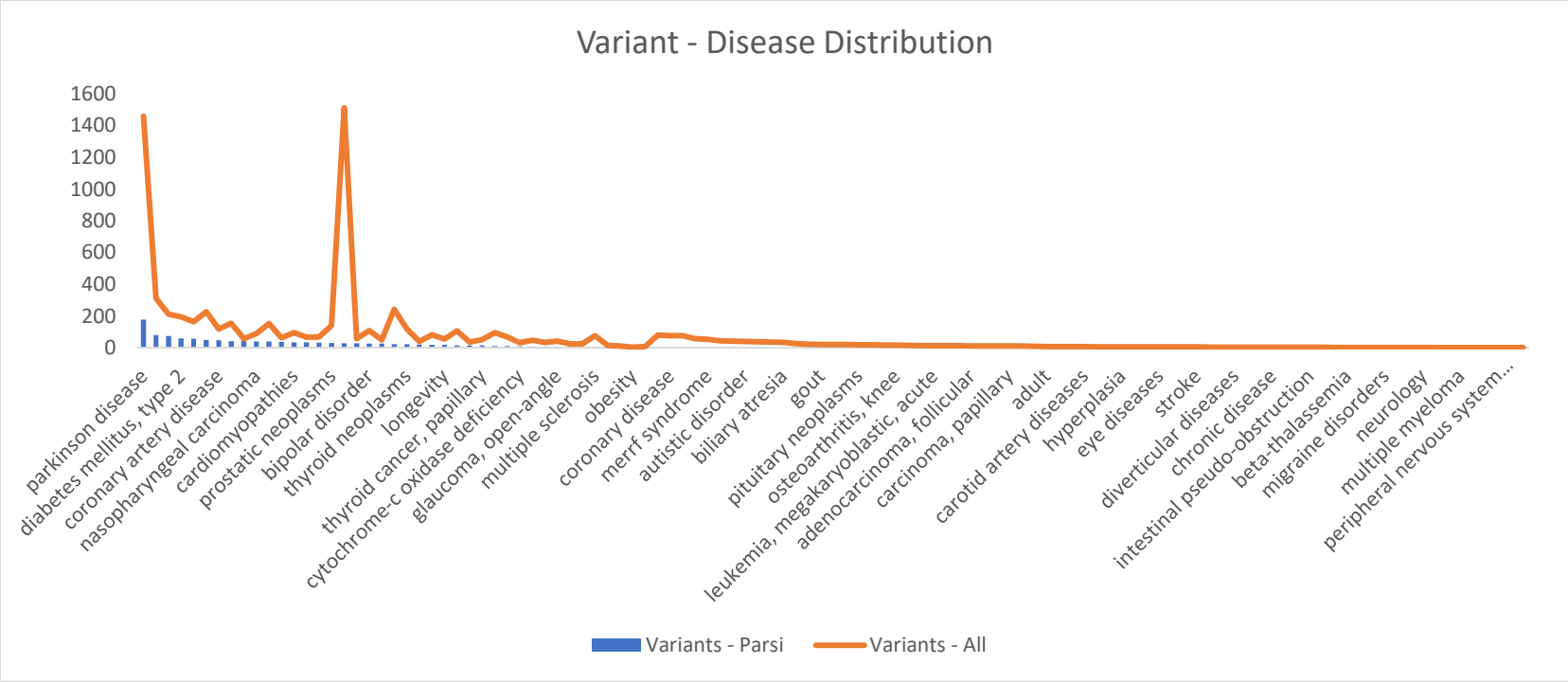
Supplementary Figure 4D: Distribution of Variants across gene loci in the A, Z and F haplogroup consisting of A2v, Z1a and F1g

Haplogroup M



Supplementary Figure 4E: Distribution of variants across gene loci across the M sub-haplogroups

Supplementary Figure 5: VarDiG® -R analysis of 420 variants indicates high association of Parsi specific variants with Parkinsons diseases



Supplementary Figure 5: Variant-disease distribution of 420 Parsi variants. Graph depicts the variant-disease distribution between Parsis (blue) and VarDiG®-R (Brown)

Supplementary Table 1: Description of primers used in validation of AGENOME-ZPMS-HV2a-1 by Sanger sequencing

Primer name	Primer type	Primer sequence	Amplicon size	Region of Interest
Hs_Mito_DL_15975	Forward	CTCCACCATTAGCACCCAAAGC	1198	D-loop HVR
Hs_Mito_DL_583	Reverse	GCTTTGAGGAGGTAAGCTAC		
Hs_Mito_3636	Forward	CCTAGCCGTTTACTCAATCC	3481	Other regions of mito genome
Hs_Mito_6997	Reverse	GGGTGTAGCCTGAGAATAG		

Supplementary Table 1: Table shows the list of primers sequences used for Sanger sequencing for validation of selected variants in the AGENOME-ZPMS-HV2a-1