



OPEN Complete mitochondrial genome sequence analysis revealed double matrilineal components in Indian Ghoongroo pigs

Pranab Jyoti Das¹✉, Satish Kumar¹✉, Manasee Choudhury^{1,4}, Seema Rani Pegu², K. Meera¹, Rajib Deb², Sunil Kumar³, Santanu Banik¹ & Vivek Kumar Gupta²

This research aimed to characterize the mitochondrial genome of the Ghoongroo (GH) pig, a notable breed in India, along with its crossbred varieties, to elucidate their matrilineal components, evolutionary history, and implications for conservation. Seven pigs (5 GH, 2 crossbred, namely Rani and Asha) were sequenced for complete mitochondrial genome, while 24 pigs (11 GH, 6 Rani, and 7 Asha) were sequenced for the complete *D-loop* of the mitochondrial genome. The genome size of these pigs was determined to be 16,690 bp. Analysis of the mitochondrial sequences and phylogenetics uncovered two distinct matrilineal components within the GH population, a phenomenon also observed in its crossbred counterparts, Rani and Asha. Phylogenetic analysis demonstrated a clear clustering of GH sequences into two clades, indicating the presence of two independent maternal lineages. The phylogenetic study using complete mitogenome also indicated that GH pigs were originated locally from Indian wild boar independently from Asian and European pig population. Haplotype analysis from complete *D-loop* sequences revealed 10 different haplotypes, with some sequences shared among GH, Rani, and Asha, while others differed due to varying matrilineal origins. The haplotype analysis using complete mitogenome sequences revealed 16 different haplotypes with some shared sequences among GH. Furthermore, examination of tRNA genes and nucleotide composition of different genes namely *rRNAs*, *COX1*, *COX2*, *ATP6*, *ND4*, *ND5*, *ND6*, *Cytb* offered insights into genetic diversity within these pigs. The findings suggest that geographical isolation and historical events likely contributed to the emergence of distinct maternal lineages within the GH breed. This study underscores the significance of mitochondrial DNA analysis in uncovering hidden genetic diversity within seemingly uniform populations. The molecular insights gained into the genetic makeup of GH pigs could aid in designing effective breeding programs for conservation efforts and highlight its significance in understanding the broader context of pig domestication in India.

Keywords Indian Ghoongroo pig, Mitochondrial genome, *D-loop*, Double matrilineal, Phylogeny

Abbreviations

ICAR	Indian council of agricultural research
NRCP	National research centre on pig
CPCSEA	Committee for the purpose of control and supervision of experiments on animals
IAEC	Intuition animal ethic committee
NFW	Nuclease free water
PCR	Polymerase chain reaction
NCBI	National center for biotechnology information
H strands	Heavy strand
L strands	Light strand
<i>D-loop</i>	Displacement loop

¹Animal Genetics and Breeding, ICAR-National Research Centre on Pig, Rani, Guwahati 781131, Assam, India. ²Animal Health, ICAR-National Research Centre on Pig, Rani, Guwahati 781131, Assam, India. ³Animal Reproduction, ICAR-National Research Centre on Pig, Rani, Guwahati 781131, Assam, India. ⁴Present address: Assam Don Bosco University, Tapesia, Sonapur 782402, Assam, India. ✉email: drpranabjyotidas@gmail.com; hilsa.satis2007@gmail.com

Temp	Temperature
GH	Ghoongroo
AWH	African warthog
IWB	Indian wild boar

Pigs (*Sus scrofa*) are one of the most ancient domesticated, socioeconomically valued and widely distributed livestock species across the world¹. The process of pig domestication occurred independently in various regions from its wild ancestors, with evidence suggesting occurrences in western Asia around 8500 BC^{2,3}, in China around 6500 BC⁴, and in Southeast Asia and Europe approximately 9000 years ago^{5,6}. India is recognised as one of the centres for the domestication of pigs and domestic pigs potentially originating from Indian wild boars separate from European and other Asiatic lineages⁷. The genomic analyses revealed distinct mitochondrial haplotypes in Indian pig populations that were present in wild boar populations of India, but not in pigs from Europe and the East, indicating a localized event for domestication⁶.

Pig farming holds significant importance in the livelihood of rural tribal communities¹. India possesses a rich diversity of pig genetic resources, with significant variations among populations. Fourteen indigenous pig breeds, documented in the country's breed database, contribute significantly to the socioeconomic upliftment of rural poor pig farmers. Among these breeds, the Ghoongroo (GH), earlier known as Ghungroo, stands out as one of the most prolific pig breeds, primarily found in West Bengal and Assam which has the potential to be used in various breeding programmes⁸. This is the first registered pig breed of India exhibits distinctive features such as a black coat, a characteristic bulldog-like face, a cylindrical body shape, and large, drooping ears (Fig. 1a,b)^{9,10}. Out of total 9.06 million pig population of the India, GH pig are 208,751 in number with 2.9% of total indigenous pigs of India. The highest population of GH is found in West Bengal (117.4 thousand) followed by Assam (82.9 thousand), and Uttar Pradesh¹¹. The northeastern region of India, particularly Assam, emerges as a key hub for pig production, with GH and Doom being the predominant indigenous breeds traditionally raised in low-input backyard farming systems. These indigenous breeds possess inherent traits such as early sexual maturity, adaptability to harsh climate and management conditions and requirement of low input, disease resistance, strong maternal instincts, and desirable meat quality makes them the best enterprise for the weaker sections of society and the progressive farmers as well¹². Being most prolific pig breed of India, GH pigs are very popular among the farmers traditionally rear pigs in low input backyard farming system¹³. However, to enhance growth and reproductive performance, breeds like Hampshire and Duroc have been introduced and crossbred with GH, resulting in crossbred pig varieties like Rani and Asha. Notably, these crossbred varieties retain the maternal genetic heritage of GH¹⁴.

Mammalian mitochondrial DNA genome (mtDNA) is a double-stranded molecule, composed of a H (heavy) strand and a L (light) strand and is approximately 16.5 kb in size that varies with the species viz. cattle 16.34 Kb, goat 16.64 Kb; sheep 16.61 Kb; buffalo 16.36 Kb and in pig 16.69 Kb¹⁵⁻¹⁹. MtDNA encodes crucial proteins of



Fig. 1. Images of adult (a) Ghoongroo Boar (b) Ghoongroo Sow (c) Asha Crossbred (d) Rani crossbred pigs.

the electron transport chain. The location of genes varies in different species but most genes are located on the H-strand and only one or two genes are located on the L-strand. The mtDNA also has 22 tRNAs and 2 rRNAs that are involved in mtDNA transcript production and processing. Its maternal inheritance pattern and faster base substitution evolutionary rate allow for the investigation of evolutionary relationships within and between species²⁰. Of the mtDNA genome, the control region *i.e.* *D-loop* was used for investigating the genetic population structure of closely related animals in restricted areas^{21,22}.

The complete mtDNA of GH and its crossbreds, along with the assessment of matrilineal components and genetic diversity, represents a vital step towards conservation and genetic improvement. Several studies have been conducted to characterize the mitochondrial diversity in Indian domestic and wild pigs^{7,16,23}. The maternal diversity in five Indian pig breeds (Mali, Niang Megha, Tenyi Vo, Ghoongroo, and Doom) alongside three exotic breeds (Hampshire, Duroc, and Large White Yorkshire) were identified using variations in 487 bp fragment of mitochondrial *D-loop*²³. The study identified unique haplotypes within indigenous pig populations. Sharma et al.²⁴ explored the maternal genetic diversity in Indian pig breeds by analysing the 464 bp fragments of mitochondrial *D-loop*. The study revealed 32 maternal haplotypes in which GH pigs indicated only two haplotypes with 10 variable sites. The earlier reports mainly focused on small fragments of mtDNA, however, the present study aimed to characterize the complete mtDNA of Indian GH pigs and its crossbred varieties, tracing domestication patterns based on maternal lineages. Phylogenetic analyses will shed light on relationships among GH and its crossbreds and to what extent they were affected by the modern commercial breeds (Duroc, Yorkshire and Landrace) in maternal lineage, thus providing valuable insights into the multiple matrilineal components and evolutionary history of GH pigs.

Methods and materials

Ethics statement

All blood sampling procedures were conducted with minimal invasiveness, adhering to humane endpoint guidelines to ensure the welfare of the pigs. Blood was collected only once from each pig, minimizing the stress and physiological impact on the animals. All experiments conducted in this study adhered to the guidelines set forth by the animal ethics committee of the institute, with approval no. NRCP/CPCSEA/1658/IAEC-20/2018, ensuring compliance with ethical standards for animal care and use.

Animals and sampling

This study focused on the indigenous pig breed GH and its crossbreds found in the Bengal and Assam regions of India. A total of five GH pigs were used for the characterisation of the complete mtDNA genomic sequence and 11 GH pigs were used for the characterization of the complete *D-loop* sequence of the mitochondrial genome. Additionally, crossbred varieties, namely Asha and Rani, with maternal components of GH, were included in the study. Seven animals from the Asha variety and six from the Rani variety were used for complete *D-loop* sequence analysis, while one animal each from Rani and Asha crossbreds was employed for complete mitogenome sequencing. The samples were taken from different locations in the breeding track of GH from West Bengal and Assam and care were taken in including only one sample from one location such that the sample used in the study represent the true population. However, samples of crossbreds were collected from Assam region only where the crossbreds were developed and distributed. Rani (Fig. 1d) is a crossbred pig variety developed by crossing ♀ GH and ♂ Hampshire pigs, with 50% blood inheritance of each breed¹⁴. Asha (Fig. 1c), on the other hand, is a crossbred pig variety obtained by crossing ♀ Rani with the terminal sire ♂ Duroc, thereby maintaining mitochondrial inheritance from GH pigs exclusively. Blood samples of 5 ml each were collected from the anterior vena cava using a sterile needle and BD vacutainer, and then stored at -20°C until DNA extraction.

DNA extraction

Genomic DNA was extracted from the blood samples using the standard phenol–chloroform method^{25,26}. The quality of extracted genomic DNA was checked on agarose gel electrophoresis and the DNA samples without any smearing and having intact bands were used for further study. The purity and concentration of DNA were determined using a NanoDrop spectrophotometer, with samples having an A260/A280 ratio falling between 1.7 and 1.9 considered suitable for downstream applications. The DNA samples having good quality and purity were stored at -20°C until further use.

PCR amplification of mitochondrial genome

The primers sequences, amplification temperatures and product size along with the amplification conditions were similar to our earlier study⁷. The sequences of all 30 pairs of primers along with annealing temperature and product size were presented in Supplementary Table 1. Briefly, 30 pairs of overlapping primers were used for the amplification of the complete mtDNA of GH and its crossbreds. The PCR was carried out in a 25 μl reaction mixture having 2.5 μl of 10 \times PCR buffer (with Mg^{2+}), 0.5 μl of forward and reverse primer each (10 pm/ μl), 0.5 μl of dNTPs (10 mM), 0.2 μl *Taq* polymerase (1 unit), 1 μl DNA samples (50 ng/ μl), and 19.8 μl of NFW. The PCR condition was conducted in a thermocycler (Applied Biosystems) involving an initial denaturation at 95°C for 7 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at T_a for 30 s (T_a for different set of primers were either 58 or 58.5 or 59°C as mentioned in supplementary table 1), and extension at 72°C for 30 s followed by a 5 min final extension at 72°C . The PCR products were then analyzed by 2% agarose gel electrophoresis. The PCR products showing intact specific bands and without any smearing was used for downstream works and processed for sequencing.

Sequencing of amplicon and structure analysis mitogenome

The purified PCR products were subjected to Sanger sequencing using 3500 Series Genetic Analyzers (Applied Biosystems). Sanger Sequencing not only sequences individual DNA fragments sequentially, but also guarantees full coverage of the reference genome. This is achieved through the utilization of overlapping fragments, effectively eliminating any gaps, and ensuring a comprehensive 100% coverage of the targeted sequence²⁷. The obtained sequences from Sanger sequencing were trimmed and edited using DNASTAR and Megalign software version 15.0.0 (<https://www.dnastar.com/software/lasergene/megalign-pro/>). The annotation of complete mtDNA sequences was finalized using MITOS2 tools in the Galaxy (https://usegalaxy.org/?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fiuc%2Fmitos2%2Fmitos2%2F2.1.9%2Bgalaxy0&version=latest) webserver Platform^{28–30}. The circular structure of the complete mitogenome was constructed using Proksee (<https://proksee.ca/>) Server³¹. The structure of the tRNA sequence identified in the complete mitogenome was predicted using the tRNAscan-SE 2.0 (<https://trna.ucsc.edu/tRNAscan-SE/>) web server³². The Nucleotide frequencies, G + C content, and A + T content of mitogenome were determined using EditSeq of Laser gene (DNA STAR Inc.). The skewness of protein-coding genes in the mitogenomes was calculated using the formula: GC skew = (G – C)/(G + C) and AT skew = (A – T)/(A + T)³³.

Phylogenetic analysis and genetic distance analysis

Phylogenetic analysis, using the complete mtDNA sequences of pigs generated in this study, was conducted, along with sequences from Indian wild boar, Asian pig breeds (5), European Pig breeds (4) and African warthog (AWH) downloaded from NCBI GenBank for comparison. The nucleotide sequences were aligned using the MUSCLE algorithm³⁴ of Molecular Evolutionary Genetics Analysis (MEGA) 11 (https://megasoftware.net/download_win_beta)³⁵. Detailed information regarding each substitution model, including their respective BIC, AICc values, Maximum Likelihood value (lnL), and the number of parameters, is provided in Supplementary Table 2. The substitution model utilized for the alignment was thoroughly assessed, and the Tamura-Nei model with gamma distribution rates among sites (TN93 + G), which yielded the lowest Bayesian information criterion (BIC) score, was selected as the most suitable for both alignment and phylogenetic analysis. This analysis involved 18 nucleotide sequences. There were a total of 18,320 positions in the final dataset. Subsequently, the aligned sequences were employed to construct a phylogenetic tree using the maximum likelihood (ML) method with 1000 bootstrap replications in MEGA 11, aimed at elucidating the matrilineal components in GH and its crossbred pig varieties.

Apart from the complete mitogenome sequence, complete *D-loop* sequences were also used for the phylogenetic and genetic distance analysis. For this purpose, the complete *D-loop* sequences of European pig breeds, which contributed to the development of crossbred varieties of GH pigs, were retrieved from the NCBI GenBank (<https://www.ncbi.nlm.nih.gov/>). The list of sequences used for the phylogenetic analysis is provided with their accession numbers in Supplementary Tables 3a,b. The nucleotide sequences were aligned using the MUSCLE package of MEGA 11 employing the HKY + G model as the best fit, which yielded the lowest Bayesian information criterion (BIC) score, and accounts for varying nucleotide frequencies and differing rates of transitions and transversions. Details of each substitution model including their respective BIC, AICc values, Maximum Likelihood value (lnL), R, f and the number of parameters, are provided in Supplementary Table 4. The phylogenetic analysis encompassed 28 nucleotide sequences and a total of 1330 positions in the final dataset. These aligned sequences were used to construct a phylogenetic tree via the ML method with 1000 bootstrap replications.

The phylogenetic tree constructed in MEGA was visualized using the FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) software³⁶. The genetic distances among the sequences were calculated using the maximum composite likelihood (MCL) model in MEGA 11^{35,37}. Since all pairwise distances in a distance matrix have correlations due to the phylogenetic relationships among the sequences, the sum of their log-likelihoods is a composite likelihood. The pairwise distances and the related substitution parameters could be accurately estimated by maximizing the composite likelihood³⁸. The AWH was selected as an outgroup for phylogenetic tree analysis because it is well known to be different from Eurasian wild boars and this pig has been commonly employed in past phylogenetic investigations of pigs^{39–41}. The haplotypes were identified from sequences of complete mitogenome and complete *D-loop* sequences using DNA Sequence Polymorphism software package (DnaSP v.6) (<http://www.ub.edu/dnasp/>)⁴². The network of haplotype was generated by the minimum spanning network method (epsilon = 0) using PopART v.1.7 (<https://popart.maths.otago.ac.nz/download/>)⁴³. The nucleotide diversity, no. of segregating sites, Tajima's D value and haplotype frequency were also analysed using DnaSP v.6⁴².

Results and discussion

This study was done to characterise the mitochondrial genome of GH and its crossbreds. The complete mitogenome and *D-loop* sequences were used for phylogenetic analysis and genetic distance estimation and to access the matrilineal components in these pigs. The *D-loop* region of the mitochondrial genome has high mutation rate and variable than any other region of the nuclear or mitochondrial genome⁴⁴ and thus important region for the phylogenetic analysis and evolution of animal breeds⁴⁵.

Sequencing and submission of complete mtDNA genome

The entire mtDNA of 5 GH and one each of the Rani and Asha crossbred varieties were amplified using 30 pairs of overlapping primers and sequenced by Sanger sequencing. All the fragments, including the *D-loop*, 2 rRNA, 22 tRNA, 13 coding genes, and repeat regions, were aligned to obtain the complete mtDNA genome of each pig and the sequences were deposited into the NCBI GenBank database and assigned accession numbers *MT501674*, *MZ703184*, *OM617468*, *OM634652*, *MZ647672* for GH breed; *ON706057* for Rani; and *ON715893* for Asha.

Apart from the complete mtDNA genome, complete *D-loop* sequences were amplified and sequenced using Sanger sequencing from GH, Asha and Rani pigs. The sequences were trimmed and edited using DNASTar and Megalign software and these sequences were submitted to NCBI GenBank database and received the accession numbers for GH (OP185718, OP185719, OP185720, OP185721, OP185722, and OP185723); Asha (ON934748, ON934749, ON934750, ON934751, ON934752, and ON934753) and Rani (OP352470, OP352471, OP352472, OP352473, and OP352474).

The base composition of the mtDNA genome

The complete mtDNA of all the GH and its crossbred viz. Rani and Asha were found to be of size 16,690 bp and for all the mitogenome, viz. GH, Asha and Rani, the approximate base composition was 34.7% for Adenine (A), 25.8% for Thymine (T), 13.3% for Guanine (G) and 26.17% for Cytosine(C), while the G + C content was, 39.5% for all the breeds studied. The composition of different nucleotide base compositions of the mitochondrial genome is depicted in Table 1, which shows that majority of the nucleotides in the mitogenomes were AT-rich with 60.52% of total bases. The nucleotide composition and organization of the mitogenome of GH and its crossbreds was like other pigs viz. Min pig¹⁸. The size of the mitogenome in these pigs was comparable with the earlier reports of pigs in Asian as well as European pig breeds^{7,16,41,46,47}. The nucleotide composition in mtDNA of I Pig was 34.66% for A, 26.24% for C, 13.35% for G and 25.75% for T, the A + T content was 60.41%³³ while in Indian wild boar was 26.25, 13.40, 25.79 and 34.56% for C, G, T and A, respectively and A + T content was 60.35⁷.

Annotation of complete Mitogenome

The mitogenome has one non-coding control region viz. displacement loop (*D-loop*), 22 transfer RNAs (tRNAs), 2 ribosomal RNAs (rRNAs), and 13 protein-coding genes which were similar to other pig mitogenomes⁴⁸. The typical circular structure of the mitogenome of GH, Asha and Rani is depicted in Fig. 2a–c which depicts the H and L strands, position of all the genes, tRNAs, rRNAs and *D-loop* along with GC content and Skewness of GC or AT. The annotation of the complete mitogenome of the GH breed is shown in Table 2. The *D-loop* was 1254 bp long and located between *tRNA-Pro* and *tRNA-Phe* having repeat regions. The two rRNAs viz. 12S and 16S rRNAs were 960 and 1570 bp, respectively. The size of tRNAs was varied from 59 (*tRNA-Ser-1*) to 75 bp (*tRNA-Leu*, *tRNA-Asn*) in size. The mitogenome had a total of 6 overlaps among all the genes ranging from 1 to 43 bp long. Additionally, there were 11 non-coding spaces ranging from 1 to 32 bp in length. The protein-coding genes had a total sequence length of 11,409 bp ranging from 204 (*ATP8*) to 1821 (*ND5*), which is 68.36% of the total mitogenome. The H-strand of the mitogenome consists of all the protein-coding genes, tRNAs, rRNAs and *D-loop*, except the *ND6* gene and 8 tRNAs (*tRNA-Ala*, *tRNA-Asn*, *tRNA-Cys*, *tRNA-Tyr*, *tRNA-Ser*, *tRNA-GLU*, *tRNA-Pro*) which were encoded on L-strand (Fig. 2).

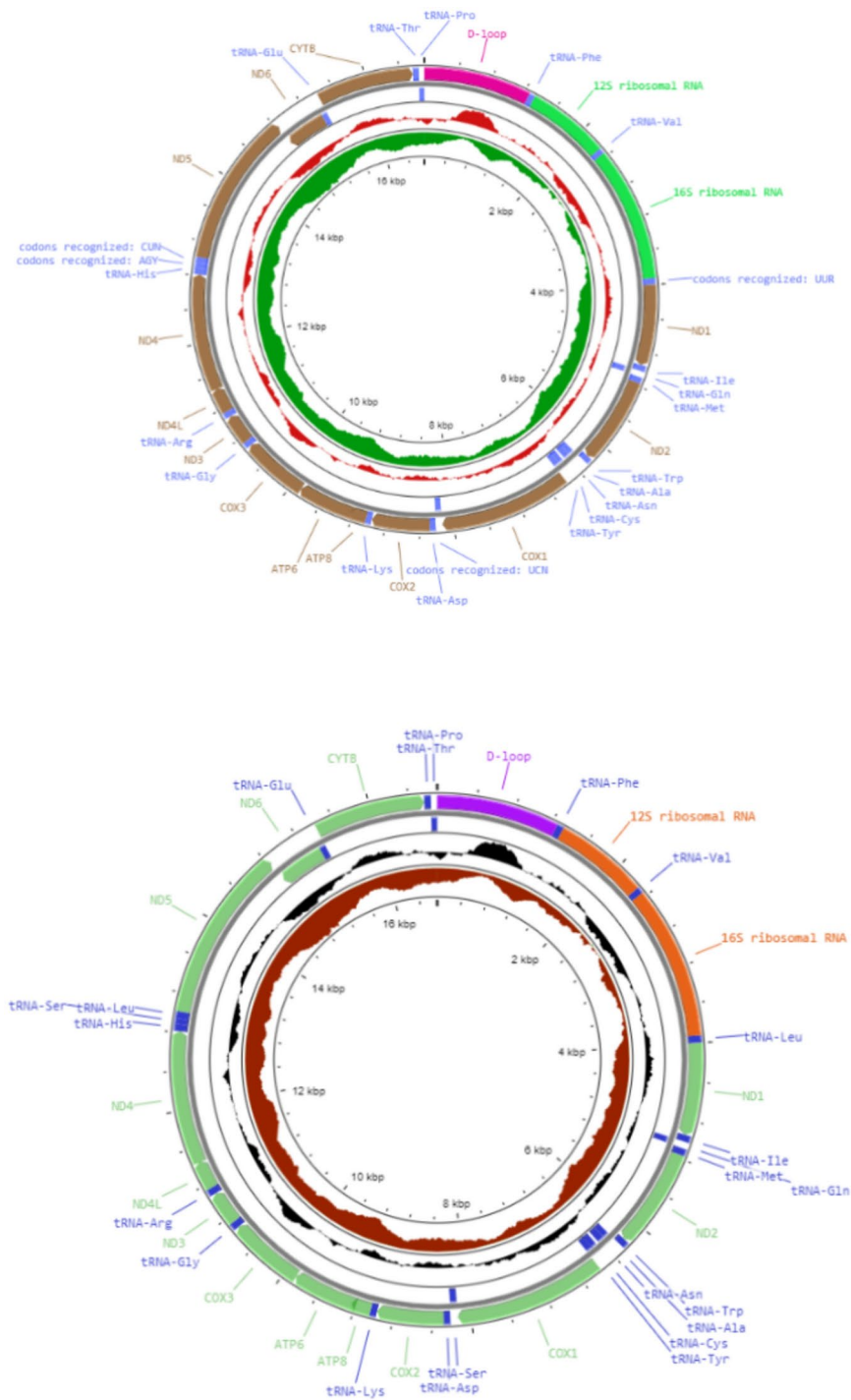
Consistent with our findings, the length of 13 protein-coding genes ranged from 204 to 1821 bp, and a total of 11,413 bp in I pig³³. Furthermore, studies in various pig populations^{33,48,49}, birds⁵⁰, and humans⁵¹ have identified one protein-coding gene and eight tRNA genes encoded on the L-strand. In contrast, prior research on the mitochondrial DNA (mtDNA) of wild and domestic pigs revealed only one protein-coding gene (*ND6*) and seven tRNA genes (*tRNA-Gln*, *tRNA-Ala*, *tRNA-Asn*, *tRNA-Cys*, *tRNA-Pro*, *tRNA-Tyr*, *tRNA-Glu*) encoded on the L-strand⁵². However, another study indicated two protein-coding genes (*COX3*, *ND6*) and two tRNA genes (*tRNA-Pro*, *tRNA-Glu*) on the L-strand⁵³. Notably, the complete mitochondrial genome of Daweizi and Ningxiang pigs revealed that all mitochondrial genes were encoded on the L-strand, except for three tRNA genes (*tRNA-Ile*, *tRNA-Asp*, *tRNA-Leu*), which were encoded on the H-strand^{54,55}. The arrangement and orientation of genes in the present study align with earlier reports on vertebrate and pig mitogenomes^{16,33,48,52,56}.

Each protein-coding gene begins with the start codon ATG, except for *ND4L*, which starts with GTG, while *ND3*, *ND2*, and *ND5* start with ATA. The termination codons in six protein-coding genes (*ND3*, *COX2*, *COX3*, *ND1*, *ND2* and *ND4*) were incomplete and were subsequently completed by the addition of 3' A residues to the mRNA during post-transcriptional polyadenylation. The annotation and codon sequences in our study are consistent with those found in the mitogenomes of other pig breeds from India and South-East Asia like I, Nicobari^{16,33,41,48,52}. In contrast to our findings, Singh et al.⁵³ reported that all protein-coding genes had ATG as the start codon except for *ND3*, *ND4*, *ND5*, *ND1*, and *ND2*, which started with ATA. In the mitochondrial genome of Daweizi pig, *ND2*, *ND3*, and *ND5* began with ATA, *ND4L* with GTG, and *ND6* with ATT, while the remaining proteins initiated with ATG⁵⁴. In Swedish pig breeds, the start codon for *ND2* and *ND4L* was ATT and GTG, respectively⁵⁷.

The nucleotide composition bias in mitogenome was estimated by GC and AT skews, and it showed that all genes on H stand coding for protein and rRNA were negatively GC-skewed and positively AT-skewed, which

Breed/crossbred	Total base	A (n, %)	G (n, %)	T (n, %)	C (n, %)	% A+T	% G+C
MT501674, GH 1	16690	5792, 34.7	2221, 13.31	4309, 25.82	4368, 26.17	60.52	39.48
MZ703184, GH 2	16690	5792,34.7	2221, 13.31	4309, 25.82	4368, 26.17	60.52	39.48
OM617468, GH 3	16690	579034.69	2222, 13.31	4307, 25.81	4371, 26.19	60.5	39.5
OM634652, GH 4	16690	579234.7	2220, 13.3	4310, 25.82	4368, 26.17	60.53	39.47
MZ647672, GH 5	16690	579234.7	2221, 13.31	4309, 25.82	4368, 26.17	60.52	39.48
ON706057, Rani	16690	578934.69	2222, 13.31	4308, 25.81	4371, 26.19	60.5	39.5
ON715893.1, Asha	16690	579234.7	2219, 13.3	4310, 25.82	4369, 26.18	60.53	39.47

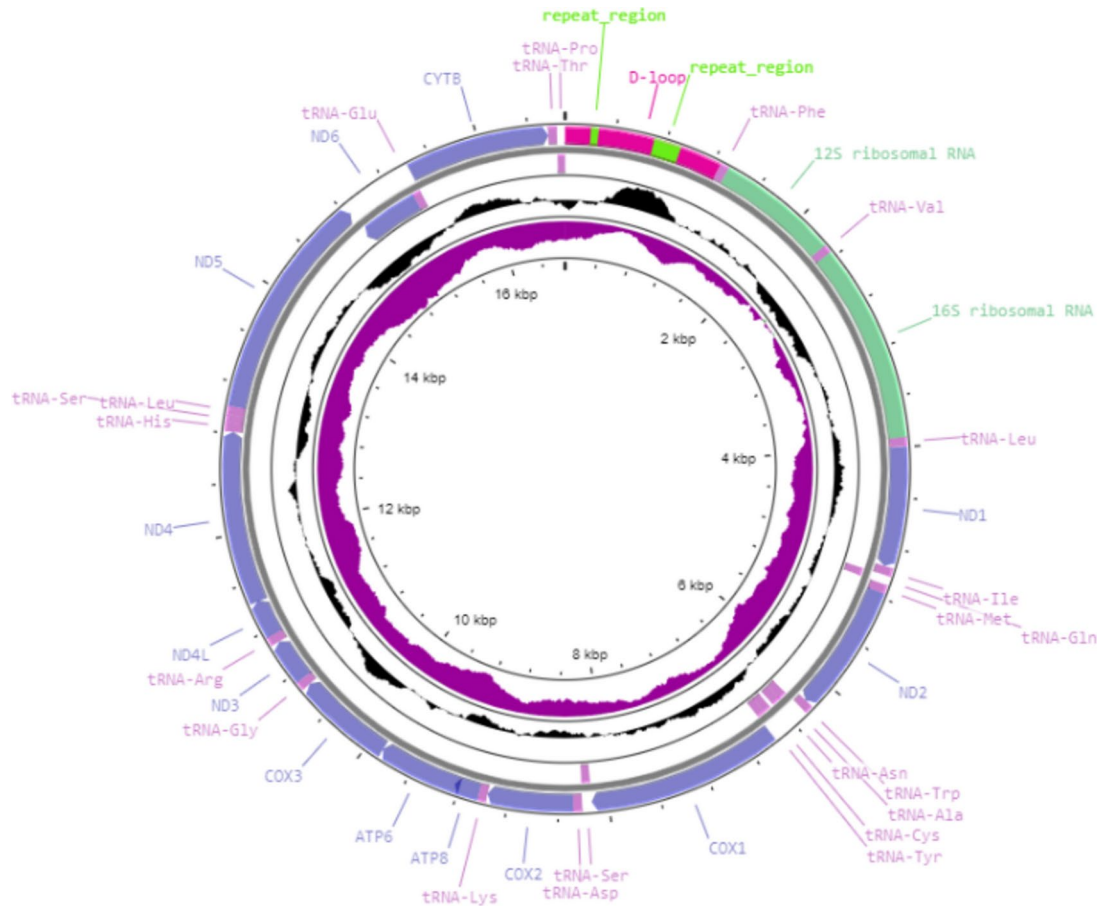
Table 1. Nucleotide base composition of the mitochondrial genome of Ghoongroo and its crossbred pigs.



Complete Mitochondrial Genome of Rani Crossbred

Fig. 2. Circular structure of complete mitochondrial genome of (a) Ghongroo (b) Rani and (c) Asha pigs.

denoted cytosine and adenine biasedness. The most cytosine bias was observed in ATPase 8 (-0.52), while the least cytosine bias was observed in *16S rRNA* (-0.11). In contrast, the *12S rRNA* and *16S rRNA* genes had the most adenine bias (0.25), while the *COX1* gene had no bias for AT content, which means that its adenine content is equal to the thymine content. The *ND6* gene was the only gene coded in the L strand that was positively skewed for GC content (0.55), whereas it was negatively skewed for AT content (-0.36) (Table 2). The AT content of the *D-loop* was 57.81% , whereas the GC skew and AT skew were found to be -0.26 and 0.14 , respectively. The AT content of the *D-loop* (57.81%) was lesser than that of the I pig (60.09%), Ningxiang pig (60.52%) and Wild



Complete mitochondrial genome of Asha Crossbred

Figure 2. (continued)

pig (60.65%)^{33,52}. The AT content, AT skew, and GC skew were used to determine the nucleotide composition behaviour of mitochondrial genomes as well as related to phylogenetics^{58,59}.

The nucleotide composition bias within the mitogenome was assessed using GC and AT skews, revealing that all genes on the H strand encoding for proteins and rRNA exhibited negative GC-skew and positive AT-skew, indicating a bias towards cytosine and adenine. Among these, *ATPase8* displayed the highest cytosine bias (-0.52), while *16S rRNA* showed the least (-0.11). Conversely, *12S rRNA* and *16S rRNA* exhibited the highest adenine bias (0.25), while *COX1* showed no bias in AT content, implying equal adenine and thymine content. Notably, the *ND6* gene, the sole gene encoded on the L strand, demonstrated positive GC skew (0.55) and negative AT skew (-0.36) (Table 2). The AT content of the *D-loop* was determined to be 57.81%, with corresponding GC and AT skews of -0.26 and 0.14, respectively. This AT content was lower than that observed in I pig (60.09%), Ningxiang pig (60.52%), and Wild pig (60.65%)^{33,52}. Utilizing AT content, AT skew, and GC skew aids in understanding the nucleotide composition patterns within mitochondrial genomes and their relevance to phylogenetics^{58,59}.

Structure of tRNA

The mitogenome of GH and its crossbred had 22 tRNA genes encoded for 22 tRNAs which consists of 2 tRNAs for Leucine and Serine amino acids while one tRNA for each of the other amino acids. All the tRNAs except eight were encoded on the H strand. The tRNA structure predicted by the tRNAscan-SE server revealed that all the tRNAs have typical cloverleaf structures except Ser-I which was devoid of D-arm. The structure and number of tRNAs in our study were similar to earlier reports in pigs³³. The structure of all the tRNAs is depicted in Fig. 3.

Phylogenetic analysis and double matrilineal within Ghoongroo pigs

The aligned complete mitochondrial genome sequence of GH and its crossbred pigs identified 27 polymorphic sites (Table 3), with one insertion and one deletion in the rRNA region of the Rani mitogenome. Out of these polymorphic sites, other than the insertion and deletion, 9 were found in the control region of the *D-loop* while 2 in the rRNA region, 3 each in *ND4L* and *ND5* gene, 2 each in *COX1* and *COX2* gene, and one each in *ND4*,

Gene name	Strand	Start	End	Bases	Space (+)/overlap (-)	Start codon	Stop codon	Anti codon	GC skew	AT skew
<i>D-loop</i>	H	1	1254	1254	0				- 0.26	0.13
<i>tRNA-Phe</i>	H	1255	1324	70	0			GAA		
<i>12S rRNA</i>	H	1325	2284	960	0				- 0.15	0.25
<i>tRNA-Val</i>	H	2285	2352	68	0			TAC		
<i>16S rRNA</i>	H	2353	3922	1570	0				- 0.11	0.25
<i>tRNA-Leu</i>	H	3923	3997	75	0			TAA		
<i>ND1</i>	H	4000	4954	955	2	ATG	T-		- 0.42	0.14
<i>tRNA-Ile</i>	H	4955	5023	69	0			GAT		
<i>tRNA-Gln</i>	L	5021	5093	73	- 3			TTG		
<i>tRNA-Met</i>	H	5095	5164	70	1			CAT		
<i>ND2</i>	H	5165	6206	1042	0	ATA	T-		- 0.49	0.24
<i>tRNA-Trp</i>	H	6207	6274	68	0			TCA		
<i>tRNA-Ala</i>	L	6281	6348	68	6			TGC		
<i>tRNA-Asn</i>	L	6350	6424	75	1			GTT		
<i>tRNA-Cys</i>	L	6457	6522	66	32			GCA		
<i>tRNA-Tyr</i>	L	6522	6587	66	- 1			GTA		
<i>COX1</i>	H	6589	8133	1545	1	ATG	TAA		- 0.2	0
<i>tRNA-Ser</i>	L	8137	8207/8205	71	3			TGA		
<i>tRNA-Asp</i>	H	8213	8280	68	5			GTC		
<i>COX2</i>	H	8281	8968	688	0	ATG	T-		- 0.31	0.12
<i>tRNA-Lys</i>	H	8969	9035	67	0			TTT		
<i>ATP8</i>	H	9037	9240	204	1	ATG	TAA		- 0.52	0.19
<i>ATP6</i>	H	9198	9878	681	- 43	ATG	TAA		- 0.46	0.11
<i>COX3</i>	H	9878	10661	784	- 1	ATG	T-		- 0.31	0.04
<i>tRNA-Gly</i>	H	10662	10730	69	0			TCC		
<i>ND3</i>	H	10731	11076	346	0	ATA	T-		- 0.46	0.15
<i>tRNA-Arg</i>	H	11078	11146	69	1			TCG		
<i>ND4L</i>	H	11147	11443	297	0	GTG	TAA		- 0.33	0.01
<i>ND4</i>	H	11437	12814	1378	- 7	ATG	T-		- 0.48	0.12
<i>tRNA-His</i>	H	12815	12883	69	0			GTG		
<i>tRNA-Ser-1</i>	H	12884	12942	59	0			GCT		
<i>tRNA-Leu</i>	H	12943	13012	70	0			TAG		
<i>ND5</i>	H	13013	14833	1821	0	ATA	TAA		- 0.46	0.15
<i>ND6</i>	L	14817	15344	528	- 17	ATG	TAA		0.55	- 0.36
<i>tRNA-GLU</i>	L	15345	15413	69	0			TTC		
<i>Cytb</i>	H	15418	16557	1140	4	ATG	AGA		- 0.39	0.09
<i>tRNA-Thr</i>	H	16558	16625	68	0			TGT		
<i>tRNA-Pro</i>	L	16626/25	16689	64	0			TGG		

Table 2. Annotation of the complete mtDNA genome of Indian Ghoongroo and its crossbreds.

ND6, *ATP6* and *Cytb* (Table 3). Out of 27 polymorphic sites, 16 sites were present in the coding region of the mitogenome with mostly synonymous mutation and do not cause any change in the amino acid. Apart from complete mitogenomes, a complete *D-loop* region was also sequenced and aligned and a total of 18 polymorphic sites were observed in 1254 bp long *D-loop* (Table 4 and Supplementary Table 6). The phylogenetic analysis from the complete mitogenome sequence of GH and its crossbred along with the European pig breeds (Duroc, Large white Yorkshire, Hampshire, Landrace) and Asian pig breeds of different regions apart from India (Meishan, Banna Mini, I Pig, Ningxiang, Tibetan) was done using the AWH sequence as an outgroup. The phylogenetic analysis showed that different sequences of GH breeds were clustered in two different clades indicating two independent maternal lineages existing within GH. In one clade GH sequences were clustered with Rani crossbred while in other clade GH sequences were clustered with the Asha crossbred variety (Fig. 4). The presence of GH in two different clades indicated that GH breeds may have different matrilineal components and this may be the reason for their grouping in different clades. This hypothesis stands good if we see the position of Asha and Rani crossbreds, both crossbreds have mitochondrial inheritance of GH breed, Rani was clustered in one clade while Asha is in another clade which indicated that Rani crossbred used in this study may have developed from GH which have similar matrilineal components while Asha was developed from GH which have different matrilineal components. The phylogenetic analysis indicated that GH and its crossbred were present in separate clades from European and Asian pig breeds which shows independent domestication events in Indian Ghoongroo pig breed from other Asian and European pig breeds. The Indian Ghoongroo pigs might have originated from Indian

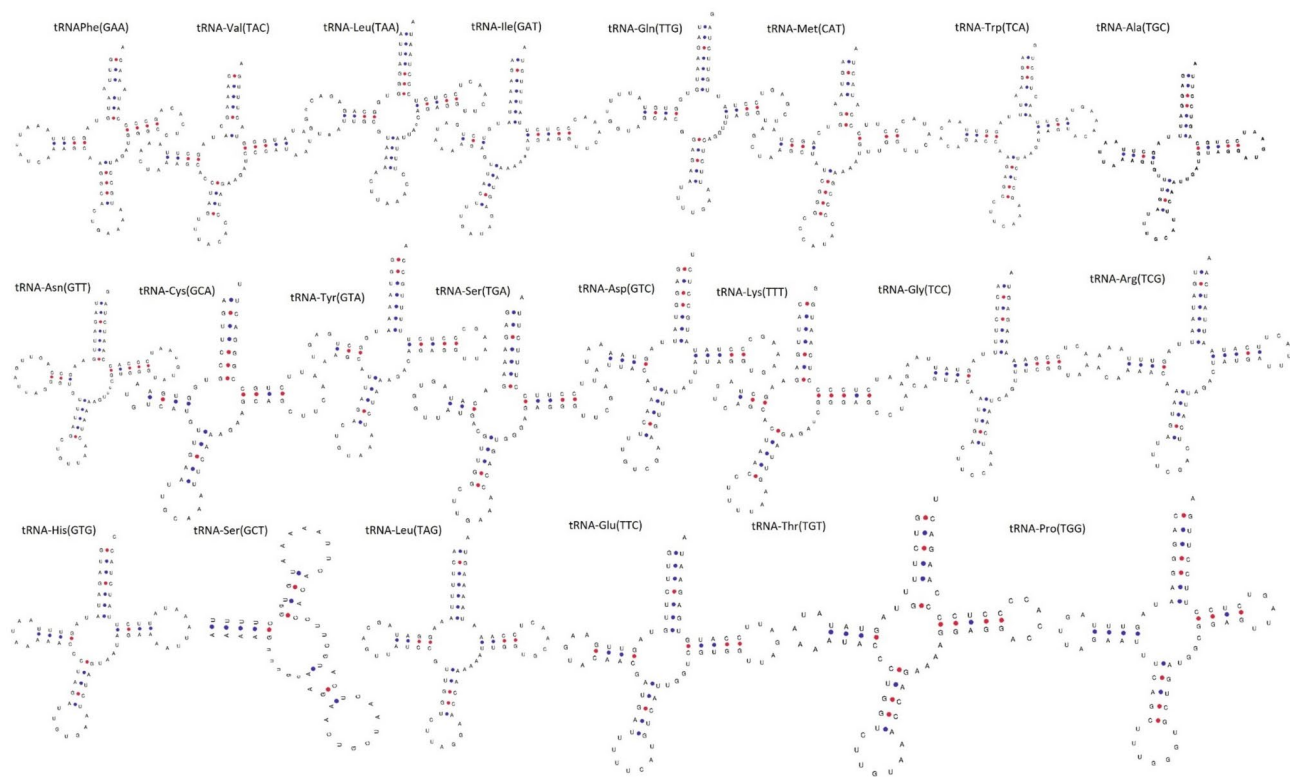


Fig. 3. tRNA structure of all 22 tRNAs found in the complete mitogenome of GH and its crossbreds.

Breed name/sample	2	2	2	2	4	6	7	7	1	1	1	2	2	6	7	8	8	9	1	1	1	1	1	1	1	1	1	1		
MT501674.1 GH 1	C	T	C	A	T	G	A	G	A	-	A	G	A	G	A	T	T	T	T	G	A	T	G	G	C	A	T			
MZ703184.1 GH 2	-
MZ647672.1 GH 5	-
OM634652.1 GH 4	T	C	T	G	C	A	G	A	G	-	T	.	A	A	T	T	G	C			
OM617468.1 GH 3	T	C	T	G	C	A	G	A	G	-	.	.	G	A	G	C	C	C	C	T	G	A	A	T	T	G	.			
ON706057.1 Rani	T	C	T	G	C	A	G	.	G	T	-	A	G	A	G	C	C	C	C	T	G	A	A	T	T	G	.			
ON715893.1 Asha	T	C	T	G	C	A	G	A	G	-	T	.	.	.	A	T	.	.	C		
	D-loop									rRNA	rRNA	COX1	COX2	ATP6	ND4L	ND4	ND5	ND6	Cytb											

Table 3. Polymorphic sites of complete mitochondrial genome sequence of Ghongroo and its crossbreds. Nucleotide positions are numbered according to the reference sequence GenBank ON706057. Sequences identical to the first sequence (GH1) are denoted by dots (.). GH Ghongroo.

wild boar locally independent from other Asian and European pig breeds. The same finding was also observed in studies by^{6,7} where independent domestication centres was reported in Indian, Asian and European regions. The haplotype analysis using a complete mitogenome also shows the same result (Fig. 5b, Supplementary Table 5b). There was a total of 1838 polymorphic or segregating sites, including 1539 singleton sites and 299 parsimony informative sites. The number of haplotypes generated from complete mitogenome sequences was 16 with haplotype diversity 1.00 ± 0.022 . This represents a very high level of haplotype diversity and suggests a highly diverse population with a wide variety of haplotypes and few common ones. It was found that GH were present in three different haplotypes while Rani and Asha were present in separate haplotype. The haplotype of Rani has only 1 mutation different from one of the GH haplotypes while Asha haplotype has 3 mutations different from another GH haplotype (Fig. 5b and Supplementary Table 5b). It was also evident that GH and crossbreds had separate haplotype which may be due to the different matrilineal components within them as evident from the phylogenetic analysis. The nucleotide diversity in the complete mitogenome sequences were measured as 0.01855 ± 0.00903 while Tajima's D was -1.96091 with statistical significance of $P < 0.05$, which suggest that population is experiencing non-neutral evolutionary process. The nucleotide diversity indicating

Breed/sample	93	213	240	278	293	451	483	511	691	703	714	743	753	761	772	773	802	1175
Consensus	C	T	C	T	G	C	G	G	A	G	C	G	A	G	T	G	T	G
MT501674 GH1	.	C	T	C	A	T	.	.	G	A	.	.	G	A
MZ703184 GH2	.	C	T	C	A	T	.	.	G	A	.	.	G	A
OM617468 GH3
OM634652 GH4
MZ647672 GH5	.	C	T	C	A	T	.	.	G	A	.	.	G	A
OP185718 GH6	.	C	T	C	A	T	.	.	G	A	.	.	G	A
OP185719 GH7	A
OP185720 GH8
OP185721 GH9	.	C	T	C	A	T	.	.	G	A	.	.	G	.	A	.	C	A
OP185722 GH10	.	C	T	C	A	T	.	.	G	A	.	.	G	A
OP185723 GH11	A	.	.
ON715893 Asha1
ON934748 Asha2	.	C	T	C	A	T	A	.	G	A	.	.	G	A
ON934749 Asha3	.	C	T	C	A	T	A	.	G	A
ON934750 Asha4
ON934751 Asha5
ON934752 Asha6	A
ON934753 Asha7
ON706057 Rani1	G
OP352470 Rani2	G
OP352471 Rani3	A	C	T	C	A	T	.	.	G	A	.	.	G	.	.	A	.	A
OP352472 Rani4	.	C	T	C	A	T	.	C	G	A	.	.	G
OP352473 Rani5	G	.	.	A	.	.
OP352474 Rani6	A	.	G

Table 4. Polymorphic sites of complete *D-loop* sequence of Ghongroo and its crossbred. Nucleotide positions are numbered according to the reference sequence GenBank *OP185718*. Sequences identical to the first sequence (Consensus) are denoted by dots (.). *GH* Ghongroo.

substantial genetic variation within the mitogenome sequences suggesting that the population has a diverse set of mitochondrial DNA sequences, which could be due to a large effective population size, long history of divergence within population, or high mutation rates in the mitogenome⁶⁰. A high negative Tajima's D value with statistical significance generally indicates an excess of rare variants compared to what is expected under neutral evolution. This may happen due to recent population expansion or purifying (negative) selection where deleterious mutations are removed more quickly, leading to an excess of low-frequency alleles⁶¹. This suggests deviation from neutrality, potentially indicating population expansion, purifying selection, or a recent bottleneck⁶². The analysis of molecular variance of haplotypes indicated the fixation index Φ_{ST} value as 0.95746 ($P < 0.001$, a significant complete mtDNA variation was found within (0.4866% and among (99.51% the population sequences of the pigs. A high Φ_{ST} value of 0.99513 suggests that nearly all the genetic variation is present between population rather than within them. This means the populations are almost entirely distinct from each other genetically.

To further validate the results of double matrilineal components in GH with complete mitogenome samples, another phylogenetic analysis was conducted on the complete *D-loop* sequences of GH, Rani and Asha. We have also downloaded the complete *D-loop* sequences of Hampshire, Duroc and Landrace for phylogenetic analysis to see whether any influence of paternal mitochondrial components was present in Rani and Asha. The transition-to-transversion (Ts/Tv) ratio in the complete *D-loop* sequences from this study was notably higher (5.09:1) than the critical value of 2:1 to 4:1 reported by Perna and Kocher⁶³ and Knight and Mindell⁶⁴, indicating a high preference for transitions in GH and its crossbreds. This may reflect evolutionary pressures, mutational biases. Our study corroborates to earlier reports where the high ratio (3.85–10.5) was found in different pigs^{23,65}. In this study, Φ_{ST} were found to be 0.95746 ($P < 0.001$, a significant mtDNA *D-loop* variation was found within (4.25%) and among (95.75%) the breed sequences of the pigs. A high Φ_{ST} value of 0.95746 suggests that nearly all the genetic variation is present between population rather than within them. This means the populations are almost entirely distinct from each other genetically. Similar to the results of complete mitogenome analysis, GH and its crossbreds Rani and Asha were clustered in two distinct clades (Fig. 6). It was clearly indicated that some of the GH, Rani and Asha were clustered together in one clade while the rest of GH, Rani and Asha were clustered in another clade. It may be due to the fact that the animals within the same clade have the same matrilineal components while animals between the two clades have different matrilineal components. The European breeds which were used to produce crossbreds have clustered separately in the third clade which shows that GH crossbreds have no effect on the matrilineal components of European breeds, signifying European

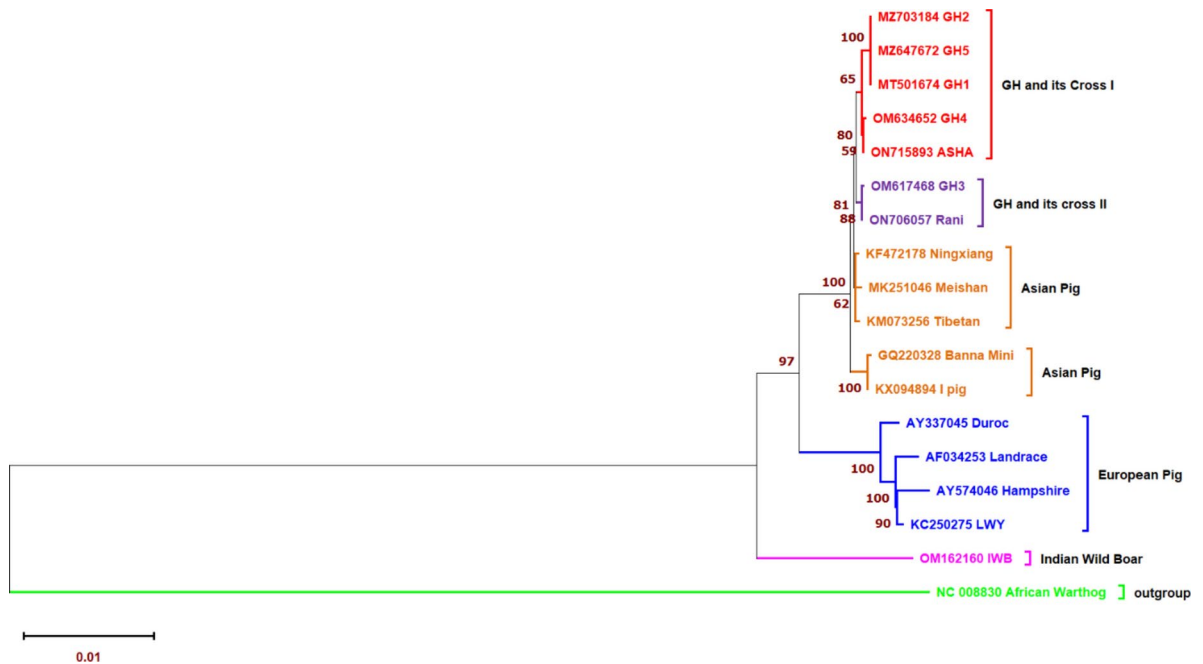


Fig. 4. Phylogenetic tree generated using complete mitogenome sequences of GH and its crossbreds using Maximum Likelihood methods using 1000 bootstrap replications in MEGA11 and Visualise using FigTree v1.4.4. Two distinct clusters can be seen of GH in red and violet colour where Rani is clustered in violet clade while Asha is clustered in red clade. Asian and European pig breeds are clustered separately from GH and its crossbred in brown and blue colour, respectively. Indian Wild boar is present separately from both the haplogroup. African Warthog was used as an outgroup.

breeds were used as sire components only and no mitochondrial inheritance was found in crossbreds from these European breeds.

To further explore the genetic differentiation of the population, the haplotype analysis was performed and a median-joining network profile was generated (Fig. 5a). The haplotype network has a correlation with the results of the phylogenetic tree. There was a total of 125 polymorphic or variable sites, including 104 singleton sites and 21 parsimony informative sites. The number of haplotypes generated from complete *D-loop* sequences was 10 with haplotype diversity 0.720 ± 0.079 . This indicates a moderately high level of haplotype diversity and suggests that the population has a fair amount of genetic variation, but there are some common haplotypes shared among individuals. Among the 10 haplotypes identified in this study, one haplotype was shared by GH and its crossbred, while other haplotypes were unique to GH, Rani, Asha and other European pig breeds. None of the haplotypes were shared between GH and its crossbreds with European pig breeds. It was found that GH, Asha and Rani shared a single largest haplotype (Fig. 5a and Supplementary Table 5a) which is obvious due to the same matrilineal components within them. It was also evident that GH and crossbreds had another separate haplotype which may be due to the different matrilineal components within them as evident from the phylogenetic analysis. Indian domestic breeds have unique and differentiated haplotypes and no haplotype sharing was found between Indian domestic pig breeds with exotic pig breeds²³ and between wild and domestic pig breeds²⁴. Larson et al.³⁹ had also reported the similar findings from other regions. Our findings corroborate with the earlier report where there were 2 haplotypes in GH pig population based on 464 bp *D-loop* fragments in 10 individuals²⁴. The nucleotide diversity measured 0.01333, while Tajima's *D* was -2.10929 . These values suggest non-neutral evolution, potentially indicating an abundance of rare alleles, suggestive of positive selection or a selective sweep⁶⁶. The nucleotide diversity found in our study is in line with the earlier reports where the nucleotide diversity was found to be 0.01²⁴.

The evolutionary divergence in the form of genetic distance between different breeds sequences was estimated using the Tamura-Nei model⁶⁷ based on complete mitogenome sequences and presented in Supplementary Table 7. Apart from the pairwise genetic distance between breeds, the genetic distance between the groups were also estimated (Supplementary Table 8). The analysis showed that the genetic distance between two different matrilineal components of GH was 0.11% which is lowest than other groups comparisons. The genetic distance between GH and Asian Pigs was less (0.17%) than the distance between GH and European Pigs (1.31%). The genetic distance between GH and Indian Wild Boar was lesser than the distance between European pigs and Indian wild boar. As African warthog was used as an outgroup, the highest genetic distance was observed between AWH and other groups of pigs. However, pairwise genetic distance between breeds revealed that GH and its crossbreds were more distant to European pig breeds than the other Asian pig breeds. Among European pig breeds, the genetic distance with GH was highest in Hampshire (1.45%) and subsequently decreases in Landrace (1.35%), LWY (1.24%) and Duroc (1.19%). Among Asian Pig breeds, the genetic distance was highest

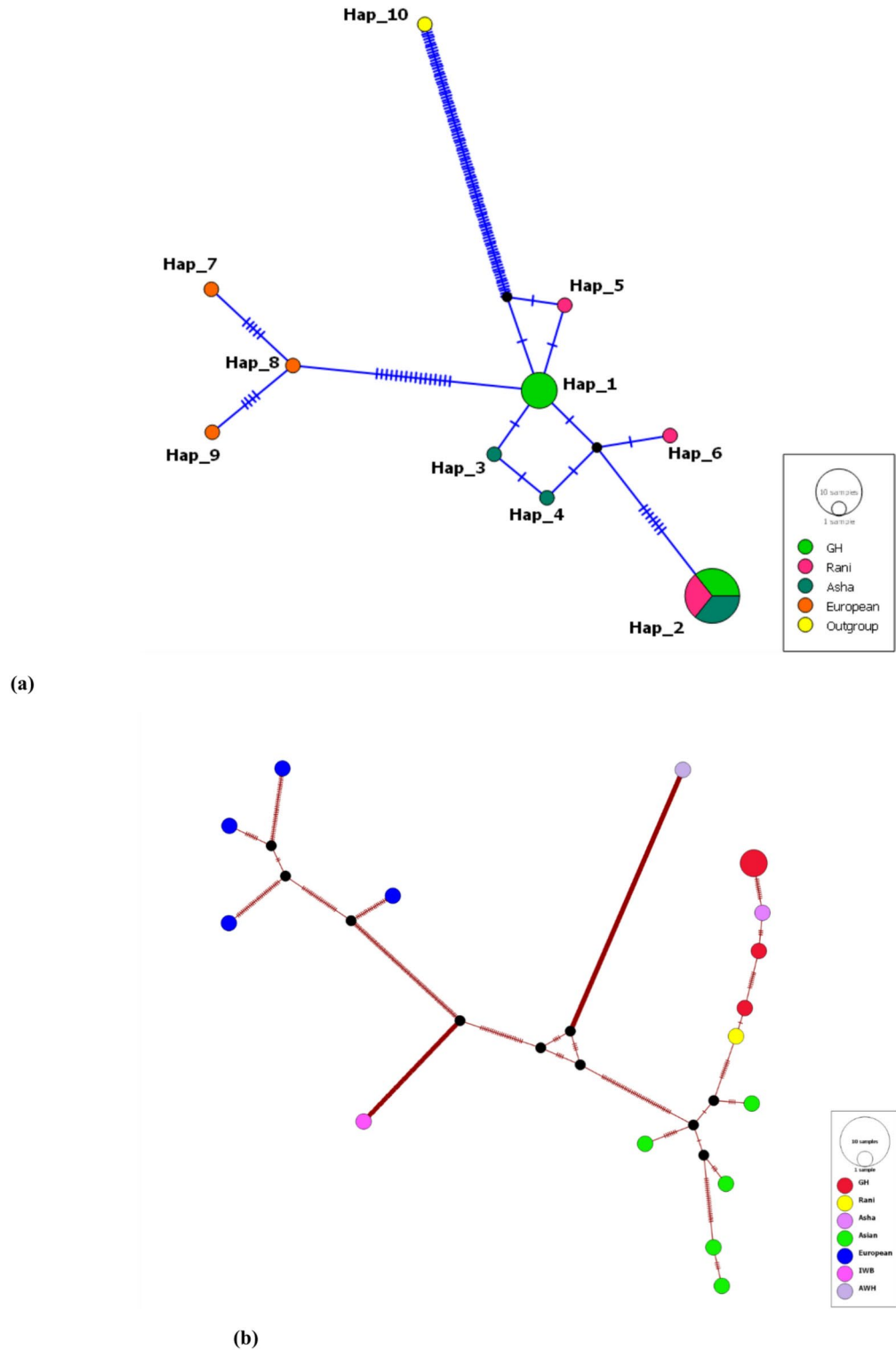


Fig. 5. Median Joining Haplotype Network created with PopArt v1.7 showing the genetic relationships between all individuals using (a) Complete *D-loop* sequences (b) Complete mitogenome sequences. Colours represent different pig populations, mutational steps between haplotypes are shown as hatchmark across connection lines. Each haplogroup is represented by a circle, with the area of the circle proportional to the haplogroup's frequency.

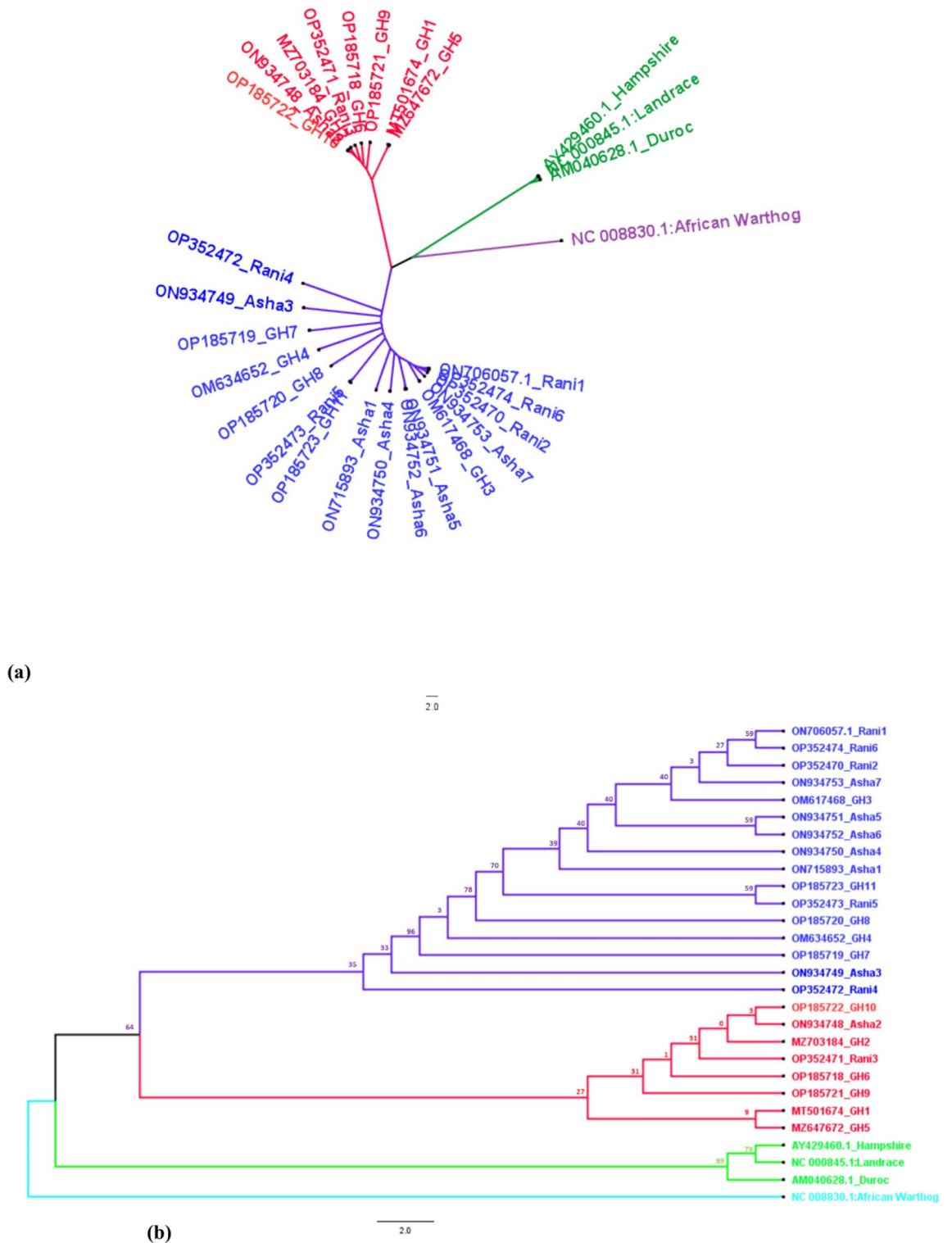


Fig. 6. Phylogenetic tree generated using complete *D-loop* sequences of GH and its crossbreds using Maximum Likelihood methods using 1000 bootstrap replications in MEGA11 and Visualise using FigTree v1.4.4. The frequency and branch length are presented above and below the branch, respectively. Three distinct clusters can be seen in blue, red and green colour where blue and red haplogroup represent the GH and its crossbred while green haplogroup represent European pigs. (a) Radial representation (b) rectangular representation.

	Standard error													
	1	2	3	4	5	6	7	8	9	10	11	12		
Genetic distance	1		0.21	0.21	0.06	0.10	0.20	0.08	0.09	0.48	0.33	0.43	1.73	MT501674 GH1
	2	0.47		0.00	0.22	0.19	0.06	0.24	0.22	0.51	0.39	0.48	1.74	OM617468 GH3
	3	0.47	0.00		0.22	0.19	0.06	0.24	0.22	0.51	0.39	0.48	1.74	ON715893 Asha1
	4	0.05	0.53	0.53		0.08	0.20	0.10	0.10	0.48	0.34	0.43	1.74	ON934748 Asha2
	5	0.16	0.42	0.42	0.10		0.20	0.13	0.11	0.49	0.36	0.44	1.75	ON934749 Asha3
	6	0.42	0.05	0.05	0.47	0.47		0.23	0.21	0.51	0.38	0.48	1.72	ON706057.1 Rani1
	7	0.10	0.58	0.58	0.16	0.26	0.53		0.13	0.50	0.36	0.45	1.74	OP352471 Rani3
	8	0.10	0.47	0.47	0.16	0.16	0.42	0.21		0.51	0.36	0.46	1.76	OP352472 Rani4
	9	1.27	1.39	1.39	1.33	1.39	1.39	1.33	1.39		0.18	0.27	2.15	AY429460 Hampshire
	10	0.85	1.11	1.11	0.90	1.01	1.06	0.95	0.95	0.30		0.15	1.78	AM040628 Duroc
	11	1.14	1.37	1.37	1.19	1.25	1.37	1.20	1.25	0.54	0.22		1.99	NC000845 Landrace
	12	6.02	6.08	6.08	6.08	6.20	6.02	6.08	6.15	7.48	6.22	6.98		NC008830 AWH

Table 5. Estimates of pairwise evolutionary divergence based on complete *D-loop* sequences of different pigs by the maximum composite likelihood method in MEGA11. The percent genetic distance and standard error are depicted in the lower triangle in the upper triangle. This analysis involved 12 nucleotide sequences. There were a total of 1330 positions in the final dataset. All ambiguous positions were removed for each sequence pair (pairwise deletion option). GH represent Ghoongroo while AWH represent African Warthog.

in Banna mini (0.25%) and subsequently decreases in I-Pig (0.24%), Meishan (0.14%), Tibetan (0.13%) and Ningxiang (0.12%). The results of genetic distance corroborated with the phylogenetic tree.

The pairwise percentage genetic distance between GH, GH crossbreds and European breeds and AWH was calculated using the maximum composite likelihood model³⁸ based on complete *D-loop* sequences and presented in Table 5. The animals used for distance calculation were one from each haplogroup as depicted in supplementary table 5a. The analysis showed that the genetic distance between GH having different matrilineal components was more (0.47%) than the distance between GH and Asha (0.05–0.16%) or Rani (0.10%) crossbreds. As crossbreds Rani and Asha were developed from the GH matrilineal components, the genetic distance based on the *D-loop* should be negligible but a significant genetic distance was found between two GH and some of the Rani and Asha while with others it was negligible. It may be because Rani and Asha may have matrilineal components different from the GH for which distance was calculated. The genetic distance between GH and its crossbreds against European breeds revealed that the highest distance was between GH and Hampshire (1.27–1.39%) followed by Landrace (1.14–1.37%) and Duroc (0.85–1.11%). The results of genetic distance corroborated with the phylogenetic tree. Furthermore, after placing all the GH and its crossbreds in one group, European pigs in 2nd group and AWH in 3rd group, the genetic distance between the GH and European group was found to be 1.21% while between GH and AWH group was 6.09% while the distance between European and AWH was 6.9%. Our study corroborates with the earlier reports where the genetic distance between Indian wild boar and domestic pigs was 3.29%⁷ and 3.5%²³ and European wild boars and domestic breeds were 1.16%⁴⁶.

The study first time used the complete mtDNA genome of GH and its crossbred varieties namely Rani and Asha pigs. The mtDNA genome sequencing was performed for GH pigs from different locations in their breeding tract from West Bengal to Assam. The study discovered that within the GH breed, a wide genetic diversity is present and it may have two subspecies, with exact same morphological characters, which is not possible to identify phenotypically. Genetic measures like sequencing and phylogenetic analysis of mtDNA have given the tools to identify the matrilineal components of GH pig and identification of such distinct clusters within a breed made possible. The complete mtDNA genome as well as complete *D-loop* sequences revealed the double matrilineal components within GH breeds. Moreover, the crossbred commercial varieties namely Rani (♀GH x ♂Hampshire) and Asha (♀Rani x ♂Duroc) where GH has been used as the maternal lineages also showed similar differentiation. The Rani and Asha were also differentiated into two different clusters confirming the double matrilineal components within the GH population. The complete mitochondrial genome data of five GH pigs as well as complete *D-loop* sequences of 11 pigs of this breed clearly validate that GH pig has maintained two different mitochondrial inheritances. Analogous patterns of gene flow parallel to these two maternal inheritances can also be seen in Rani and Asha, crossbred pigs developed using GH as the maternal lineage. In this connection, the phylogenetic analysis of the complete mitochondrial genome as well as complete *D-loop* sequences of Rani and Asha, validated the vertical inheritance of two mitochondrial genomes of GH subspecies among these crossbreds (Figs. 4 and 6). The GH pigs having two different matrilineage footprints might have encountered unusual natural/artificial selection pressure in the past, which led to the evolution of two distinct subspecies within the breed without affecting their phenotypic characters. The geographical isolation between Assam and West Bengal due to the Sankosh River, a tributary of the Brahmaputra River may be one of the reasons for the introduction of mitochondrial inheritance within GH from a related different ancestor. Therefore, from the study, it can be inferred that GH pigs have experienced a silent mitochondrial inheritance within the breed, which has changed their genetic framework but did not alter the phenotypic attributes of the animal.

This is the first study which revealed the two different matrilineal components within the same breed. As GH is distributed in a wide range of West Bengal and Assam states of India and wide variation is present in this

breed but phenotypically all are the same. The double matrilineal components within GH may be due to their origination from not a single ancestor but from closely related maternal ancestors. This may also be possible due to the derivation/dispersal from the matrilineal pool of one region to the other regions because of the migration. A similar pattern has been observed in previous studies^{23,68}. The GH breed might have originated locally from Indian Wild boars independent to other Asian and European pig breeds. The eight pig breeds of Shandong province of China had lower divergence and shared the same haplotype because of the possibility that they stemmed from closely related maternal ancestors, if not from a common ancestor⁶⁹. The Dapulian Black and Laiwu Black breeds had independent maternal lineage but other indigenous pigs have extensive gene flow with other breeds⁶⁹. The neighbour-joining tree analysis identified two distinct clades among individual pigs, with a Chinese domestic breed showing a scattered distribution across multiple breeds⁷⁰. The commercial European breeds like Landrace, Hampshire, Large White Yorkshire and Duroc were clustered together in a separate clade independent of Indian pigs. Our findings align with earlier studies indicating a distinct phylogenetic separation between Indian pigs and European-American and Asian pig clades⁵². The separate matrilineage observed in Indian wild boars may stem from the differentiation of wild boar populations that originated from Island South East Asia (ISEA) and subsequently migrated to the Indian subcontinent before further dispersal to East Asia and eventually across Eurasia³⁹. Additionally,⁶ reported that modern Indian domestic pigs have ancestral ties to local wild boar populations distinct from those found in Asia and Europe. Multiple domestication centres have been identified, including four for Chinese Native Pigs⁷¹ and six for native pigs in East Asia³⁹. The results of this study is based on the 7 complete mitochondrial genome sequences as well as 24 complete *D-loop* sequences of GH and its crossbreds. We have included some of the sequences from the NCBI GenBank but due to limited complete *D-loop* sequences of Indian pigs in the databank, the study is limited by the small sample size. However, to estimate the genetic diversity and phylogenetic analysis using complete mitochondrial or Ccomplete *D-loop* sequences, preliminary studies were conducted with smaller sample size. The genetic diversity in 17 indigenous Chinese pig breeds and 3 European breeds was determined based on one sample from each breed only⁷². Similarly, the phylogenetic relationships between Asian and European pig breeds were determined using one or two samples from 19 breeds of pig⁴⁶.

The GH pig is the most prolific pig breed in India and shows a wide range of variability in terms of reproductive and productive performance⁷³ which may be linked to the two different genetic subpopulations of GH having different maternal lineages. The absence of a structured breeding program for indigenous pig breeds like GH has contributed to a decline in their population⁷⁴. Consequently, it becomes crucial to characterize the mitochondrial DNA (mtDNA) of GH and its crossbred counterparts to assess genetic diversity and maternal lineages within these pigs. This information is pivotal for designing appropriate breeding programs aimed at conserving this significant pig breed. In contrast to many species where domestication has been one of the major causes for drastic change in their morphological parameters, in this study it has been particularly uncovered that GH pig due to multiple expansion events though maintaining two different maternal lines in their genome has experienced no phenotypic changes in their morphology. The mitochondrial genomic data of GH pigs having two matrilineage, annotated in this study, could be fine-tuned as a powerful tool to trace the origin of pig domestication, conserve their indigenous germplasm and identify the purity as well as elucidate the lineage of other nondescript pigs.

Conclusion

This study aimed to identify the molecular breed signatures of GH pigs through an integrative analysis of mtDNA expression and *D-loop* sequence data of GH pigs from different locations and their crossbred species (Rani and Asha). As evidenced by the mitochondrial genomic data GH pigs of this region have undergone multiple domestication events and have been carrying two different maternal lineages in their genome from the past, which is also apparent in their crossbred species. The study also underscores that GH pigs have undergone silent mitochondrial inheritance in their breed, without affecting their morphological attributes. The study identified two unique maternal legacies, which might be useful for their genotypic identification and thus work as an input for designing and implementing genetic strategies to conserve this important indigenous pig breed of India.

Data availability

The complete mitogenome sequences with gene annotation and complete D-loop sequences has been submitted to the NCBI GenBank. The details of accession numbers of all the sequence data utilized in this study can be found in the Supplementary Table 3.

Received: 11 June 2024; Accepted: 25 November 2024

Published online: 17 January 2025

References

- Bharati, J., De, K., Paul, S., et al. Mobilizing Pig Resources for Capacity Development and Livelihood Security. In *Agriculture, Livestock Production and Aquaculture: Advances for Smallholder Farming Systems*: Volume 2 (2022).
- Conolly, J. et al. Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry. *J. Archaeol. Sci.* **38**, 1. <https://doi.org/10.1016/j.jas.2010.10.008> (2011).
- Ervynck, A., Dobney, K., Hongo, H. & Meadow, R. Born free? New evidence for the status of "Sus scrofa" at Neolithic Çayönü Tepesi (southeastern Anatolia, Turkey). *Paléorient* **1**, 47–73 (2001).
- Cucchi, T., Hulme-Beaman, A., Yuan, J. & Dobney, K. Early Neolithic pig domestication at Jiahu, Henan Province, China: Clues from molar shape analyses using geometric morphometric approaches. *J. Archaeol. Sci.* **38**, 1. <https://doi.org/10.1016/j.jas.2010.07.024> (2011).
- Giuffra, E. et al. The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics* **154**, 1. <https://doi.org/10.1093/genetics/154.4.1785> (2000).

6. Larson, G. et al. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc. Natl. Acad. Sci. U S A* **107**, 1. <https://doi.org/10.1073/pnas.0912264107> (2010).
7. Das, P. J. et al. Characterization of the complete mitochondrial genome and identification of signature sequence of Indian wild pig. *Gene* **897**, 1. <https://doi.org/10.1016/j.gene.2023.148070> (2024).
8. Banik, S. et al. Generation-wise performance evaluation of Ghongroo pig: An effort to improve the productivity. *Indian J. Anim. Sci.* **94**(11), 1. <https://doi.org/10.56093/ijans.v94i11.138253> (2024).
9. Banik, S. et al. Effect of different body measurements on body weight in Ghongroo pigs. *J. Anim. Sci.* **1**, 1. <https://doi.org/10.56093/ijans.v82i9.23679> (2012).
10. Bharati, J. et al. Ovarian follicle transcriptome dynamics reveals enrichment of immune system process during transition from small to large follicles in cyclic Indian Ghongroo pigs. *J. Reprod. Immunol.* **160**, 1. <https://doi.org/10.1016/j.jri.2023.104164> (2023).
11. DAHD. Breed-wise Report of Livestock and Poultry based on 20th Livestock Census¹ Department Of Animal Husbandry & Dairying. <https://dahd.nic.in/sites/default/files/BreedReportEnglish04.08.2022.pdf> (2022).
12. Rajak, S. K., Kumar, S., Bharati, J., et al. Pig production and livelihood security. In: Rana T, Soto-Blanco B, (eds) Good Practices and Principles in Pig Farming. Livestock Diseases and Management (Springer, Singapore, 2024)
13. Rajak, S.K., Bharati, J., Kumar, S., Bharti, R., & Preety, P. Pig farming and business opportunities for financial benefit. In: Rana, T., Soto-Blanco, B. (eds) Good Practices and principles in pig farming. Livestock Diseases and Management (Springer, Singapore, 2024).
14. Bharati, J. et al. Transcriptome profiling of different developmental stages of corpus luteum during the estrous cycle in pigs. *Genomics* **113**, 366–379. <https://doi.org/10.1016/j.ygeno.2020.12.008> (2021).
15. Arbizu, C. I. et al. The complete mitochondrial genome of a neglected breed, the peruvian creole cattle (*Bos taurus*), and its phylogenetic analysis. *Data* **7**, 1. <https://doi.org/10.3390/data7060076> (2022).
16. De, A. K. et al. Mitochondrial landscape of indigenous pig germplasm of Andaman and Nicobar Islands. *Mitochondrial DNA Part B Resour.* **4**, 1. <https://doi.org/10.1080/23802359.2019.1660240> (2019).
17. Di, H. X. & Gao, L. Z. The complete mitochondrial genome of domestic sheep *Ovis aries*. *Mitochondrial DNA* **27**, 1. <https://doi.org/10.3109/19401736.2014.953076> (2016).
18. Niu, T. et al. The complete mitochondrial genome of Min pig (Hebao) and a phylogenetic analysis. *Mitochondrial DNA Part B Resour.* **4**, 1. <https://doi.org/10.1080/23802359.2019.1678424> (2019).
19. Siddiki, A. et al. Complete mitochondrial genome sequence of Black Bengal goat (*Capra hircus*). *Mitochondrial DNA Part B Resour.* <https://doi.org/10.1080/23802359.2019.1623098> (2019).
20. Avise, J. C. *Phylogeography: the history and formation of species* (Harvard University Press, 2000).
21. Alves, E., Óvilo, C., Rodríguez, M. C. & Silió, L. Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations. *Anim. Genet.* **34**, 1. <https://doi.org/10.1046/j.1365-2052.2003.01010.x> (2003).
22. Ghivizzani, S. C. et al. Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA *D-loop* region. *J. Mol. Evol.* **37**, 1. <https://doi.org/10.1007/BF00170460> (1993).
23. Laxmivandana, R. et al. Genetic diversity in mitochondrial DNA *D-loop* region of indigenous pig breeds of India. *J. Genet.* **101**, 1. <https://doi.org/10.1007/s12041-021-01353-8> (2022).
24. Sharma, A. et al. Tracing the genetic footprints: India's role as a gateway for pig migration and domestication across continents. *Anim. Biotech.* **34**(9), 5173–5179. <https://doi.org/10.1080/10495398.2023.2268683> (2023).
25. Kumar, S. et al. Elucidation of novel SNPs affecting immune response to classical swine fever vaccination in pigs using immunogenomics approach. *Vet. Res. Commun.* **48**, 1. <https://doi.org/10.1007/s11259-023-10262-3> (2024).
26. Sambrook, J., & Russel, D. W. *Molecular cloning: A laboratory manual* vols 1,2, and 3 (2001).
27. Hagemann, I. S. Overview of technical aspects and chemistries of next-generation sequencing. In: *Clinical Genomics*, pp 3–19 (Elsevier, 2015).
28. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, 1. <https://doi.org/10.1093/nar/gkac247> (2022).
29. Al Arab, M. et al. Accurate annotation of protein-coding genes in mitochondrial genomes. *Mol Phylogenet Evol* **106**, 1. <https://doi.org/10.1016/j.ympev.2016.09.024> (2017).
30. Donath, A. et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res.* **47**, 1. <https://doi.org/10.1093/nar/gkz833> (2019).
31. Grant, J. R. et al. Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* **51**, 1. <https://doi.org/10.1093/nar/gkad326> (2023).
32. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. TRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 1. <https://doi.org/10.1093/nar/gkab688> (2021).
33. Nguyen, H. D. et al. The complete mitochondrial genome sequence of the indigenous i pig (*Sus scrofa*) in Vietnam. *Asian-Australasian J. Anim. Sci.* **30**, 1. <https://doi.org/10.5713/ajas.16.0608> (2017).
34. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **1**, 1. <https://doi.org/10.1093/nar/gkh340> (2004).
35. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 1. <https://doi.org/10.1093/molbev/msab120> (2021).
36. Rambaut, A. FigTree v1. 3.1. Institute of Evolutionary Biology (University of Edinburgh, Edinburgh, 2010).
37. Kumar, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1. <https://doi.org/10.1093/molbev/msy096> (2018).
38. Tamura, K., Nei, M. & Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U S A* **101**, 1. <https://doi.org/10.1073/pnas.0404206101> (2004).
39. Larson, G. et al. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1. <https://doi.org/10.1126/science.1106927> (2005).
40. Lucchini, V. et al. New phylogenetic perspectives among species of South-east Asian wild pig (*Sus* sp.) based on mtDNA sequences and morphometric data. *J. Zool.* **266**, 1. <https://doi.org/10.1017/S0952836905006588> (2005).
41. Yu, G., Xiang, H., Wang, J. & Zhao, X. The phylogenetic status of typical Chinese native pigs: Analyzed by Asian and European pig mitochondrial genome sequences. *J. Anim. Sci. Biotechnol.* **4**, 1. <https://doi.org/10.1186/2049-1891-4-9> (2013).
42. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 1. <https://doi.org/10.1093/molbev/msx248> (2017).
43. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1. <https://doi.org/10.1111/2041-210X.12410> (2015).
44. Nicholls, T. J. & Minczuk, M. In *D-loop*: 40 years of mitochondrial 7S DNA. *Exp. Gerontol.* **56**, 1. <https://doi.org/10.1016/j.exger.2014.03.027> (2014).
45. Chen, C. H. et al. Mitochondrial genome of Taiwan pig (*Sus Scrofa*). *Afr. J. Biotechnol.* **10**, 1 (2011).
46. Kim, K. I. et al. Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA *D-loop* sequence polymorphism. *Anim. Genet.* **33**, 1. <https://doi.org/10.1046/j.1365-2052.2002.00784.x> (2002).

47. Thom, B. T. et al. Morphological and mitochondrial genome characterization of indigenous Dong Khe pig in rural areas of Northeast Vietnam. *Livest. Res. Rural Dev.* **33**, 1 (2021).
48. BichVo, T. T. The Mitochondrial Genome and Phylogenetic Relationships of Muong Lay Black Pig (*Sus scrofa*) in Vietnam: Approaches Poultry. *Dairy Vet. Sci.* **5**, 1. <https://doi.org/10.31031/apdv.2018.05.000607> (2018).
49. Wang, Y. F. et al. The complete mitochondrial genome of Diqing wild boar (*Sus verrucosus* breed Diqing wild boar). *Mitochondrial DNA* **27**, 1. <https://doi.org/10.3109/19401736.2014.926540> (2016).
50. Liu, D. et al. Mitochondrial genome of the critically endangered Baer's Pochard, *Aythya baeri*, and its phylogenetic relationship with other Anatidae species. *Sci. Rep.* **11**, 1. <https://doi.org/10.1038/s41598-021-03868-7> (2021).
51. Taanman, J. W. The mitochondrial genome: Structure, transcription, translation and replication. *Biochim. Biophys. Acta Bioenerg.* **1410**, 1 (1999).
52. Kumar Jadav, K., Pratap Singh, A., Srivastav, A. B. & Sarkhel, B. C. Molecular characterization of the complete mitochondrial genome sequence of Indian wild pig (*Sus scrofa cristatus*). *Anim. Biotechnol.* **30**, 1. <https://doi.org/10.1080/10495398.2018.1469506> (2019).
53. Singh, A. P. et al. Complete mitochondrial genome sequencing of central Indian domestic pig. *Mitochondrial DNA Part B Resour.* **1**, 1. <https://doi.org/10.1080/23802359.2016.1197077> (2016).
54. Xu, D. et al. The complete mitochondrial genome of the Daweizi pig. *Mitochondrial DNA* **26**, 1. <https://doi.org/10.3109/19401736.2013.836514> (2015).
55. Xu, D. et al. The complete mitochondrial genome of the Ningxiang pig. *Mitochondrial DNA* <https://doi.org/10.3109/19401736.2013.834433> (2015).
56. Sarvani, R. K. et al. Characterization of the complete mitogenome of Indian Mouse Deer, *Moschiola indica* (Artiodactyla: Tragulidae) and its evolutionary significance. *Sci. Rep.* **8**, 1. <https://doi.org/10.1038/s41598-018-20946-5> (2018).
57. Ursing, B. M. & Arnason, U. The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *J. Mol. Evol.* **47**, 1. <https://doi.org/10.1007/PL00006388> (1998).
58. Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* **54**, 1. <https://doi.org/10.1080/10635150590947843> (2005).
59. Wei, S. J. et al. New views on strand asymmetry in insect mitochondrial genomes. *PLoS One* **5**, 1. <https://doi.org/10.1371/journal.pone.0012708> (2010).
60. Avise, J. C., Neigel, J. E. & Arnold, J. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.* **20**, 99–105 (1984).
61. Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*. **141**(1), 413–429 (1995).
62. Schmidt, D., & Pool, J. The effect of population history on the distribution of the Tajima's D statistic. Population English Edition. 1–8 (2002).
63. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
64. Knight, A. & Mindell, D. P. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of Fea's viper. *Syst. Biol.* **42**(1), 18–31 (1993).
65. Jiao, T. et al. Mitochondrial DNA D-loop diversity of Tibetan pig populations. *Philippine Agric. Sci.* **92**(4), 362–369 (2009).
66. Eckshtain-Levi, N., Weisberg, A. J. & Vinatzer, B. A. The population genetic test Tajima's D identifies genes encoding pathogen-associated molecular patterns and other virulence-related genes in *Ralstonia solanacearum*. *Mol. Plant. Pathol.* **19**, 1. <https://doi.org/10.1111/mpp.12688> (2018).
67. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
68. Wu, G. S. et al. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biol.* **8**, 1. <https://doi.org/10.1186/gb-2007-8-11-r245> (2007).
69. Wang, J. et al. Phylogenetic relationships of pig breeds from Shandong province of China and their influence by modern commercial breeds by analysis of mitochondrial DNA sequences. *Ital. J. Anim. Sci.* **9**, 1. <https://doi.org/10.4081/ijas.2010.e48> (2010).
70. Niu, L., Xie, J., Shi, K. & Zhong, T. Genetic diversity of mitochondrial DNA D-loop in wild and domestic pigs (*Sus scrofa*) in East Asia. *Folia Biol.* **71**, 1. https://doi.org/10.3409/fb_71-1.05 (2023).
71. Cai, Y. et al. Multiple domestication centers revealed by the geographical distribution of Chinese native pigs. *Animals* **9**, 1. <https://doi.org/10.3390/ani9100709> (2019).
72. Yang, J. et al. Genetic diversity present within the near-complete mtDNA genome of 17 breeds of indigenous Chinese pigs. *J. Heredity*. **94**(5), 381–385 (2003).
73. Boro, P. et al. Performances of Ghongroo pigs reared under farm condition. *J. Entomol. Zool. Stud.* **9**, 2265–2267 (2021).
74. Subalini, E., Silva, G. & Demetawewa, C. Phenotypic characterization and production performance of village pigs in Sri Lanka. *Trop Agric Res* **21**, 1. <https://doi.org/10.4038/tar.v21i2.2601> (2010).

Acknowledgements

The authors are thankful to the Director, Indian Council of Agricultural Research-National Research Centre on Pig for providing infrastructural facilities to carry out this project.

Author contributions

Conceptualization: Pranab Jyoti Das, Satish Kumar; Methodology: Pranab Jyoti Das, Satish Kumar; Formal Analysis and investigation: Satish Kumar, Pranab Jyoti Das, Manasee Choudhury; Writing original Manuscript: Satish Kumar, Pranab Jyoti Das; Review and Editing: Pranab Jyoti Das, Satish Kumar, Seema Rani Pegu, Meera K, Rajib Deb, Sunil Kumar; Supervision and Resources: Pranab Jyoti Das, Santanu Banik, Vivek Kumar Gupta; All authors revised and approved the final version of the manuscript.

Funding

This study was supported by Indian Council of Agricultural Research-National Research Centre on Pig Institutional Project Grant no. IXX13503 and ICAR-AICRP on Pig, NRCP-Unit to Pranab Jyoti Das.

Declarations

Competing interests

The authors declare no competing interests.

Compliance with ethical standards

All experiments conducted in this study adhered to the guidelines set forth by the animal ethics committee of the institute, with approval no. NRCP/CPCSEA/1658/IAEC-20/2018. Blood collection from the animals was performed with supreme care to minimize discomfort or harm. It is confirmed that authors complied with the ARRIVE guidelines.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-81205-4>.

Correspondence and requests for materials should be addressed to P.J.D. or S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025