PLOS ONE

RESEARCH ARTICLE

# Indigenous cattle of Sri Lanka: Genetic and phylogeographic relationship with Zebu of Indus Valley and South Indian origin

**Lokugalappatti L. G. S.[1,2], Saumya Wickramasinghe[2], Alexander P. A. B. D.[2], Kamran Abbas[1,3], Tanveer Hussain[3], Saravanan Ramasamy[4], Vandana Manomohan[1,4], Arnaud Stephane R. Tapsoba[1,5], Rudolf Pichler[1], Masroor E. Babar[6], Kathiravan Periasamy[1]***

**1** Animal Production and Health Section, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Vienna, Austria, **2** Faculty of Veterinary Medicine and Animal Science, University of Peradeniya, Peradeniya, Sri Lanka, **3** Department of Molecular Biology, Virtual University of Pakistan, Lahore, Pakistan, **4** Department of Animal Genetics and Breeding, Veterinary College and Research Institute, Namakkal, Tamil Nadu Veterinary and Animal Sciences University, Chennai, India, **5** Laboratoire de Biologie et de Santé Animale (LaBioSA), Institut de l'Environnement et de Recherches Agricoles (INERA), Ouagadougou, Burkina Faso, **6** Department of Animal Sciences, The University of Agriculture, Dera Ismail Khan, Khyber Pakhtunkhwa, Pakistan

* kathirvet@yahoo.co.in, K.Periasamy@iaea.org





**Citation:** L. G. S. L, Wickramasinghe S, P. A. B. D. A, Abbas K, Hussain T, Ramasamy S, et al. (2023) Indigenous cattle of Sri Lanka: Genetic and phylogeographic relationship with Zebu of Indus Valley and South Indian origin. PLoS ONE 18(8): e0282761. https://doi.org/10.1371/journal.pone.0282761

**Editor:** Gyaneshwer Chaubey, Banaras Hindu University, INDIA

**Received:** February 20, 2023

**Accepted:** July 30, 2023

**Published:** August 16, 2023



**Data Availability Statement:** Sequence data generated in the study are available at NCBI-GenBank accession nos. MZ501954-MZ502165.

**Funding:** LOK was awarded a fellowship for the study by International Atomic Energy Agency (IAEA) under the Technical Cooperation Project SRL5046 (https://www.iaea.org/) PAL was awarded a Technical Cooperation project SRL5046 for the study (https://www.iaea.org/) PAL and LOK were awarded a research grant from University of

## Abstract

The present study reports the population structure, genetic admixture and phylogeography of cattle breeds of Sri Lanka viz. Batu Harak, Thawalam and White cattle. Moderately high level of genetic diversity was observed in all the three Sri Lankan zebu cattle breeds. Estimates of inbreeding for Thawalam and White cattle breeds were relatively high with 6.1% and 7.2% respectively. Genetic differentiation of Sri Lankan Zebu (Batu Harak and White cattle) was lowest with Red Sindhi among Indus Valley Zebu while it was lowest with Hallikar among the South Indian cattle. Global F statistics showed 6.5% differences among all the investigated Zebu cattle breeds and 1.9% differences among Sri Lankan Zebu breeds. The Sri Lankan Zebu cattle breeds showed strong genetic relationships with Hallikar cattle, an ancient breed considered to be ancestor for most of the Mysore type draught cattle breeds of South India. Genetic admixture analysis revealed high levels of breed purity in Lanka White cattle with >97% Zebu ancestry. However, significant taurine admixture was observed in Batu Harak and Thawalam cattle. Two major Zebu haplogroups, I1 and I2 were observed in Sri Lankan Zebu with the former predominating the later in all the three breeds. A total of 112 haplotypes were observed in the studied breeds, of which 50 haplotypes were found in Sri Lankan Zebu cattle. Mismatch analysis revealed unimodal distribution in all the three breeds indicating population expansion. The sum of squared deviations (SSD) and raggedness index were non-significant in both the lineages of all the three breeds except for I1 lineage of Thawalam cattle ($P<0.01$) and I2 lineage of Batu Harak cattle ($P<0.05$). The results of neutrality tests revealed negative Tajima's D values for both the lineages of Batu Harak ($P>0.05$) and White cattle ($P>0.05$) indicating an excess of low frequency polymorphisms and demographic expansion. Genetic dilution of native Zebu cattle germplasm observed in the study is a cause for concern. Hence, it is imperative that national breeding organizations

Peradeniya for the study RG/AF/2013/45/V (https://www.pdn.ac.lk/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

consider establishing conservation units for the three native cattle breeds to maintain breed purity and initiate genetic improvement programs.

## 1. Introduction

Cattle represent a major proportion of livestock population in Sri Lanka. Cattle genetic resources in Sri Lanka comprise indigenous Zebu (*Bos indicus*), exotic Zebu, commercial taurine (*Bos taurus*) and various crosses of these three types. Historically, cattle rearing in Sri Lanka dates back to 543 BC with the arrival of "Indo Aryans" and since then, periodic introduction of various cattle breeds has taken place through time [1–4]. The ancient "archaic cattle" believed to be introduced by the early settlers are considered as probable ancestors of the locally adapted indigenous cattle [5]. Since 1930s, a number of exotic breeds including Indo-Pakistan Zebu (Red Sindhi, Tharparkar, Sahiwal, Khillari, etc.), European/American taurine cattle (Friesian, Ayrshire, Shorthorn, Jersey etc.) and commercial composites (Australian Friesian Sahiwal (AFS), Australian Milking Zebu (AMZ), Girolando, New Zealand Kiwi etc.) have been introduced to improve the productivity of local cattle [1, 2, 4]. Apart from introgression of exotic germplasm, genetic and demographic processes over time have also shaped the genetic composition, population structure and diversity of extant cattle breeds in the country [1, 2, 6, 7].

The indigenous Zebu represents more than 50% of the total cattle population, counting 1.08 million heads in the country [8]. Although milk productivity per animal is low (2–4 liters per day), the indigenous Zebu is known for their characteristics such as draught power, ability to subsist on poor quality feed, adaptability to tropical heat and resistance to diseases [2, 3]. At least four distinct breeds/types of Sri Lankan indigenous Zebu have been described [2, 9, 10]. "Batu Harak" has varying shades of black to brown coat color and are distributed in the country's dry zone. They do not possess a prominent hump or dewlap and are small and compact in body size with an average adult weight of 160 kg [3]. "White cattle" or "Thamankaduwa" is predominantly found in the Eastern province and has white to grey coat color [10]. "Thawalam cattle" is an isolated population of indigenous Zebu which are used as pack animals in the Central and Uva provinces [2]. "Cape" or "Hatton" was a synthetic breed evolved by crossing taurine males (brought from Cape of Good Hope in South Africa) with local female Zebu cattle. This breed was a small-sized locally adapted dairy type breed but got deteriorated subsequently due to lack of scientific breeding programs and is now believed to be extinct [2, 11, 12]. Cross breeding Zebu with exotic breeds have resulted in genetic dilution of indigenous Sri Lankan cattle [2]. As a result, the Zebu breeds are now threatened due to significant genetic erosion and loss of potentially significant gene combinations and variability.

Although details of morphological and phenotypic characteristics, production system environment and management practices of indigenous Sri Lankan cattle is available, information on their diversity, genetic distance, admixture and phylogeography is limited [9]. Evaluation of diversity within and between breeds provides insights into population structure and genetic relationships necessary for establishing priorities and strategies for conservation, genetic improvement and sustainable utilization. Molecular tools such as autosomal short tandem repeat DNA markers have been helpful for estimations of diversity, genetic differentiation and admixture [13–15]. Extra-nuclear mitochondrial DNA D-Loop polymorphisms (maternally inherited) are used as the markers of choice for domestication studies, including assessment of maternal lineages and their geographic origins [16–18]. Hence, the present study was undertaken with the objectives of (i) estimating genetic diversity and relationship of indigenous Sri

Lankan cattle with Zebu of Indus Valley and South Indian origin (ii) evaluating population structure and genetic admixture of Sri Lankan cattle (iii) identifying mitochondrial DNA maternal lineage and assess phylogeography of Sri Lankan native cattle.

## 2. Material and methods

### 2.1. Animal ethics statement

Blood samples were collected from jugular vein in EDTA vacutainer tubes. Sampling was performed by local veterinarians in the respective native breed tracts following the standard good animal practice. The blood samples were collected as part of routine veterinary surveillance and consent was obtained from farmers for the usage of samples in the study. Therefore, no further approval from the "Institutional Committee for Care and Use of Experimental Animals" of the University of Peradeniya, Sri Lanka, Virtual University of Pakistan and the Joint FAO/IAEA Division, International Atomic Energy Agency, Vienna, Austria was required.

### 2.2. Sampling and genomic DNA extraction

Blood samples were collected from 349 cattle representing six Sri Lankan and three Indus Valley Zebu breeds. The number of samples collected from each of the breeds were: 51 from Batu Harak, 51 from White cattle, 25 from Thawalam, 53 from Holstein-Friesian, 27 from Ayrshire, 33 from Jersey, 11 from Frisian-Jersey crosses, 30 from Red Sindhi, 38 from Sahiwal and 30 from Tharparkar. Samples from Red Sindhi, Tharparkar and Sahiwal were collected from Indus Valley regions located in Pakistan. Additionally, short tandem repeat genotype and sequence data on two South Indian breeds, Hallikar (36) and Kangayam (51) generated in a previous study [19] was included for comparative analysis. A map indicating the sampling locations of Sri Lankan Zebu cattle is provided in Fig 1. Briefly, a stratified random sampling procedure was followed to collect samples from unrelated cattle, based on the information provided by local farmers. DNA was extracted from whole blood using MasterPure DNA Purification Kit (Biozym, Illumina Inc, USA). DNA quality and quantity estimation was done by agarose gel electrophoresis and spectrophotometry. DNA samples were then stored at -20˚C until further processing.

### 2.3. Microsatellite genotyping and mitochondrial DNA sequencing

Twenty seven FAO recommended microsatellite markers [20] with forward primers conjugated to one of the three fluorescent dyes (FAM, HEX and ATTO550) were used for genotyping. The details of microsatellite markers, dye conjugated, annealing temperature, allele size range and PCR conditions are described elsewhere [19]. The PCR products were capillary electrophoresed after multiplexing in an automated DNA analyzer ABI3100 (Applied Biosystems, USA) with ROX500 (Applied Biosystems, USA) as an internal lane control. The allele size data for each sample was extracted using GeneMapper v.4.1 software (Applied Biosystems, USA).

The mitochondrial D-Loop region was amplified using the primers BTMTD1-F (5′ AGGACAACCAGTCGAACACC 3′) and BTMTD1-R 5' (GTGCCTTGCTTTGGGTTAAG 3′) as reported elsewhere [19]. The amplicon included complete D-loop (control region) flanked by partial cytochrome B, complete t-RNA-Thr and complete t-RNA-Pro coding sequences at 5' end while partial t-RNA-Phe sequence flanked the 3'end. Polymerase chain reaction was performed in a total reaction volume of 40μl with the following cycling conditions: initial denaturation at 95˚C for 15 min followed by 30 cycles of 95˚C for 1 min; 58˚C for 1 min; 72˚C for 1 min with final extension at 72˚C for 10 min. Purified PCR products were sequenced using Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems, U.S.A) on an automated Genetic Analyzer ABI 3100 (Applied Biosystems, U.S.A).

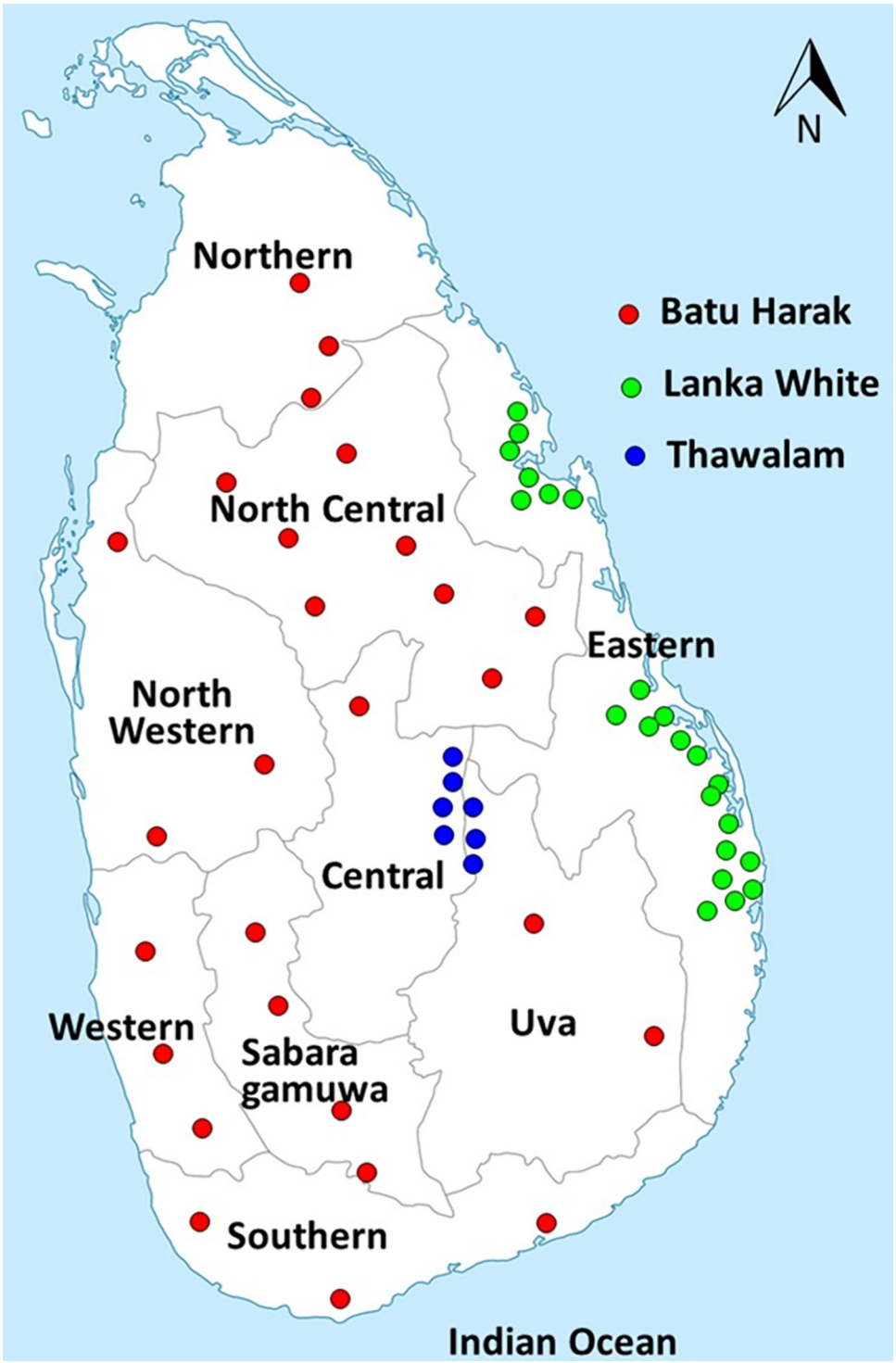


**Fig 1. Geographical distribution of Sri Lankan zebu cattle breeds included in the study (reprinted from https://commons.wikimedia.org/wiki/File:Map_of_Sri_Lanka_Provinces.png under the creative commons attribution-share alike 3.0 unported license (https://creativecommons.org/licenses/by-sa/3.0/deed.en) with the generic online permission issued by the original author Alex R.L (2011) under the terms of the GNU free documentation license, version 1.2 published by the free software foundation, 2008).** The licensor does not endorse the changes made in the reprinted version).

https://doi.org/10.1371/journal.pone.0282761.g001

### 2.4. Statistical analysis of data

To perform data quality control, the microsatellite genotypes were pruned for (i) missingness (ii) null alleles and (iii) deviations from selective neutrality. Individuals with >20% missing data (missing five or more genotypes out of 27 investigated loci) were excluded from the analysis. The presence of null alleles in the dataset was checked using MicroChecker version 2.2.3 [21]. The neutrality of the microsatellites used in this study was evaluated by comparing the markers against neutral expectations in a distribution of $F_{ST}$ vs. heterozygosities under an island model of migration using LOSITAN version 1 [22]. Basic diversity indices including observed number of alleles, observed and expected heterozygosity and pairwise Nei's genetic distances were calculated using Microsatellite Analyzer (MSA) version 3.15 [23]. Pairwise Nei's genetic distance was utilized to construct the NeighborNet tree using SplitsTree version 4.14.6 [24]. F statistics for each locus [25] was calculated and tested using the program FSTAT version 2.9.3.2. The exact tests for Hardy-Weinberg Equilibrium to evaluate both heterozygosity deficit and excess at each marker locus in each population (HWE) were performed using GENEPOP v4.0.9 software [26]. Analysis of molecular variance (AMOVA) was performed using ARLEQUIN version 3.5.3 [27]. Non-metric multidimensional scaling display of pairwise $F_{ST}$ was conducted using SPSS version 13.0 (SPSS Inc, Chicago, IL, USA). The extent of population sub-structuring was further explored using STRUCTURE [28] with the assumption of different clusters K = 2 to K = 8. The number of burn in periods and MCMC repeats used for these runs was 200000 [28]. The optimal 'K' was identified based on ΔK, the second order rate of change in LnP(D) following the procedure mentioned elsewhere [29].

Mitochondrial DNA sequences were checked and edited by using Codon Code Aligner version 3.7.1. (CodonCode Corporation, Centerville, MA, USA). A total of 212 sequences generated in the present study was deposited to NCBI-GenBank (accession nos. MZ501954-MZ502165). Additionally, 87 mtDNA sequences from two South Indian cattle breeds, Hallikar (NCBI-GenBank accessions MW319954-MW319988; MW320341) and Kangayam (NCBI-GenBank accessions MW319989-MW320038; MW320342) were included for comparative analysis [19]. Mitochondrial DNA diversity parameters including number of polymorphic, singleton and parsimony informative sites, nucleotide diversity, haplotype diversity and average number of nucleotide differences were calculated using DnaSP, version 4.10 [30]. Haplotype frequency, haplotype sharing and pairwise $F_{ST}$ among studied cattle populations were calculated using ARLEQUIN 3.5.3 [27]. Pairwise $F_{ST}$ derived from mtDNA haplotype frequency were utilized to perform non-metric multidimensional scaling (MDS) analysis using SPSS version 13.0. MEGA version 6.0 [31] was utilized to identify optimal DNA substitution model and construct maximum-likelihood phylogeny. Reference sequences from various taurine (T1-T5) and indicine (I1 and I2) maternal lineages [19] were used to construct phylogeny. HKY+I+G was identified as the optimal model for the current sequence dataset based on Akaike information criterion (AIC). To examine the possibility of a past demographic expansion, a mismatch analysis was performed by calculating frequency distributions of pairwise differences between sequences using ARLEQUIN 3.5.3. The tests for selective neutrality viz. Tajima's D and Fu's $F_S$ were conducted using ARLEQUIN 3.5.3. Median Joining (MJ) network of Sri Lankan cattle haplotypes was reconstructed using NETWORK 4.5.1.2 [32].

## 3. Results

### 3.1. Genetic variability in Sri Lankan native Zebu cattle

The basic diversity measures estimated in Sri Lankan native cattle breeds are presented in Table 1. The mean number of observed alleles was high in Sri Lankan native Zebu breeds and ranged from 6.96 (Thawalam) to 8.26 (Batu Harak). The mean observed heterozygosity was

**Table 1. Summary diversity statistics for Sri Lankan, Indus Valley, South Indian Zebu and commercial taurine cattle.**

| Origin | Species | Breed name | Breed code | N | $n_o$ | $H_o$ | $H_e$ | $F_{IS}$ | Test for HWE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | $H_e$ Deficit | $H_e$ Excess |
| Indus Valley | *B.indicus* | Red Sindhi | PRS | 30 | 6.96 | 0.608 | 0.692 | 0.124 | 6 | 0 |
| Indus Valley | *B.indicus* | Sahiwal | PSH | 38 | 6.81 | 0.636 | 0.671 | 0.053 | 6 | 0 |
| Indus Valley | *B.indicus* | Tharparkar | PTP | 30 | 5.67 | 0.586 | 0.623 | 0.060 | 7 | 1 |
| South India | *B.indicus* | Hallikar | IHL | 36 | 7.11 | 0.675 | 0.705 | 0.027 | 5 | 1 |
| South India | *B.indicus* | Kangayam | IKA | 50 | 6.19 | 0.598 | 0.637 | 0.063 | 7 | 2 |
| Sri Lanka | *B.indicus* | Batu Harak | LBH | 36 | 8.26 | 0.713 | 0.749 | 0.048 | 5 | 1 |
| Sri Lanka | *B.indicus* | Lanka White | LWC | 38 | 8.07 | 0.668 | 0.720 | 0.072 | 6 | 1 |
| Sri Lanka | *B.indicus* | Thawalam | LTM | 24 | 6.96 | 0.665 | 0.708 | 0.061 | 6 | 0 |
| Sri Lanka | *B.taurus* | Ayrshire | LAY | 27 | 5.19 | 0.680 | 0.680 | 0.000 | 2 | 1 |
| South India | *B.taurus* | Jersey | IJR | 33 | 4.56 | 0.619 | 0.627 | 0.013 | 2 | 1 |
| Sri Lanka | *B.taurus* | Holstein-Friesian | LHF | 53 | 6.85 | 0.678 | 0.716 | 0.054 | 6 | 1 |
| Sri Lanka | *B.taurus* | Friesian X Jersey Crossbred | LFJ | 11 | 5.22 | 0.693 | 0.718 | 0.037 | 4 | 1 |

https://doi.org/10.1371/journal.pone.0282761.t001

0.713, 0.668 and 0.665 in Batu Harak, Sri Lankan White and Thawalam cattle respectively while the expected heterozygosity was 0.749, 0.720 and 0.708 respectively. The overall mean inbreeding estimate ($F_{IS}$) among all the investigated cattle breeds was 0.054±0.016. Among Zebu cattle alone, the overall $F_{IS}$ was 0.063±0.020 and ranged from -0.053±0.023 (TGLA126) to 0.415±0.053 (HEL5) across different loci (S1 File). The mean overall $F_{IS}$ was higher than the values reported for East Indian [33] and Pakistani [34] cattle breeds while lower than reported for northwestern Indian [35] cattle breeds. In general, $F_{IS}$ estimates were positive at 16 of the investigated loci indicating heterozygosity deficit while remaining six loci showed evidence for heterozygosity excess (S1 File). The breed wise mean inbreeding estimate was 0.048, 0.072 and 0.061 in Batu Harak, White cattle and Thawalam respectively.

To further assess the level of heterozygosity deficit, a total of 324 breed x locus combinations were tested for Hardy Weinberg equilibrium, of which 62 (19.14%) deviated significantly ($P<0.05$). Among the Sri Lankan native cattle 17 breed x loci combinations (20.99%) deviated significantly ($P<0.05$) due to heterozygosity deficit. The percent HWE deviation observed in the study is comparable to Brazilian (19.5% breed x locus combinations [36]) and Indian Zebu cattle (17.3% breed x locus combinations [37]). This could be due to factors such as presence of null alleles, selective forces operating at certain loci, intense artificial selection and/or use of few breeding bulls in the region, assortative mating, population sub-division and inbreeding within cattle herds. The MicroChecker analysis of Sri Lankan cattle genotypes at all loci revealed no significant presence of null alleles. Further, to evaluate the influence of selection on the microsatellite loci under study, a $F_{ST}$ outlier approach was followed. Loci with an unusually high $F_{ST}$ are putatively under directional selection, while loci with low $F_{ST}$ value are considered to be under stabilizing selection. Among the 27 loci investigated, ETH152 was deviating from selective neutrality due to positive selection while TGLA53 was deviating significantly possibly under the influence of balancing selection (S2 File). Both these loci deviated from selective neutrality when assumed under both infinite allele and step wise model of mutations and have been reported to be candidates influenced by selection [38, 39]. Hence, ETH152 and TGLA53 were removed from further analysis in the study.

## 3.2. Genetic differentiation among cattle breeds

Global F-statistics was performed to estimate the genetic differentiation among Sri Lankan Zebu, Indus Valley Zebu, South Indian Zebu and commercial taurine breeds, the results of

**Table 2. Pairwise $F_{ST}$ (upper triangle) and pairwise Nei's genetic distance (lower triangle) among Sri Lankan, Indus Valley, South Indian Zebu and commercial taurine cattle.**

| | PRS | PSH | PTP | IHL | IKA | LBH | LWC | LTM | LAY | IJR | LHF | LFJ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PRS | - | 0.050 | 0.061 | 0.048 | 0.084 | 0.025 | 0.026 | 0.054 | 0.209 | 0.245 | 0.196 | 0.203 |
| PSH | 0.130 | - | 0.113 | 0.086 | 0.124 | 0.048 | 0.053 | 0.075 | 0.226 | 0.252 | 0.211 | 0.220 |
| PTP | 0.127 | 0.198 | - | 0.082 | 0.108 | 0.064 | 0.063 | 0.080 | 0.255 | 0.298 | 0.244 | 0.262 |
| IHL | 0.133 | 0.178 | 0.168 | - | 0.091 | 0.042 | 0.033 | 0.059 | 0.214 | 0.241 | 0.194 | 0.199 |
| IKA | 0.141 | 0.198 | 0.169 | 0.161 | - | 0.079 | 0.057 | 0.085 | 0.246 | 0.284 | 0.239 | 0.251 |
| LBH | 0.105 | 0.125 | 0.158 | 0.128 | 0.155 | - | 0.015 | 0.028 | 0.176 | 0.202 | 0.160 | 0.159 |
| LWC | 0.087 | 0.129 | 0.127 | 0.105 | 0.105 | 0.093 | - | 0.015 | 0.203 | 0.227 | 0.186 | 0.185 |
| LTM | 0.138 | 0.179 | 0.149 | 0.157 | 0.154 | 0.113 | 0.086 | - | 0.197 | 0.229 | 0.183 | 0.183 |
| LAY | 0.495 | 0.485 | 0.547 | 0.509 | 0.552 | 0.415 | 0.501 | 0.449 | - | 0.137 | 0.066 | 0.074 |
| IJR | 0.522 | 0.513 | 0.569 | 0.501 | 0.555 | 0.422 | 0.510 | 0.465 | 0.245 | - | 0.097 | 0.042 |
| LHF | 0.462 | 0.458 | 0.528 | 0.469 | 0.531 | 0.369 | 0.466 | 0.426 | 0.151 | 0.186 | - | 0.017 |
| LFJ | 0.534 | 0.515 | 0.587 | 0.505 | 0.573 | 0.412 | 0.513 | 0.462 | 0.194 | 0.128 | 0.088 | - |

https://doi.org/10.1371/journal.pone.0282761.t002

which are presented in S1 File. The global $F_{ST}$ was estimated at 0.145±0.010 indicating 14.5% of the total variation being explained by between breed differences and 85.5% of the variation due to within breed differences. The between breed differences was estimated at 6.5% when all the Zebu breeds were considered while it was only 1.9% when Sri Lankan Zebu (Batu Harak, White cattle and Thawalam) alone were considered. The pairwise $F_{ST}$ estimates among Zebu cattle breeds ranged from 0.015 (LWC-LBH; LWC-LTM) to 0.085 (LTM-IKA) (Table 2). Genetic differentiation of Sri Lankan Zebu (Batu Harak and White cattle) was lowest with Red Sindhi among Indus Valley breeds while it was lowest with Hallikar among the South Indian breeds. Thawalam cattle was showing relatively higher pairwise $F_{ST}$ estimates with both Indus Valley and South Indian breeds. The pairwise Nei's genetic distance estimates showed a similar trend among Sri Lankan, Indus Valley and South Indian Zebu cattle (Table 2). The Neighbor-Net tree derived from pairwise Nei's genetic distance revealed distinct clustering of commercial taurine cattle and Zebu breeds (Sri Lankan, Indus Valley and South Indian Zebu) as expected. Within the taurine cluster two distinct sub clusters could be recognized one with Ayrshire and Holstein-Friesian and the other with Friesian X Jersey crossbreds and purebred Jersey. This clearly positioned the Friesian x Jersey crossbreds between purebred Jersey and Holstein-Friesians. Among the Zebu, Sri Lankan White and Thawalam cattle clustered with Hallikar cattle (South Indian Zebu) while Batu Harak cattle clustered with Indus Valley (Sahiwal) breeds (Fig 2). Genetic relationship among investigated cattle breeds was elucidated by implementing multidimensional scaling display of pairwise $F_{ST}$. The normalized raw stress was 0.0019 indicating excellent goodness of fit and representation of data in the model [40]. The Tucker's coefficient of congruence was estimated to be 0.999 implying >99% of variance in the model being accounted by the two predicted dimensions [41]. The MDS plot showed distinct clusters of Zebu and commercial taurine cattle. All the three native Sri Lankan cattle (White cattle, Batu Harak and Thawalam) were placed in between Zebu breeds of Indus Valley and South Indian origin (Fig 3).

## 3.3. Genetic relationships and population structure of Sri Lankan cattle

|Analysis of molecular variance (AMOVA) was performed to further assess the underlying genetic relationship among the investigated Zebu cattle populations and to test and validate the results of phylogeny and MDS analysis. When no grouping of Zebu cattle was assumed, AMOVA revealed 93.93% of total genetic variation was due to differences among individuals

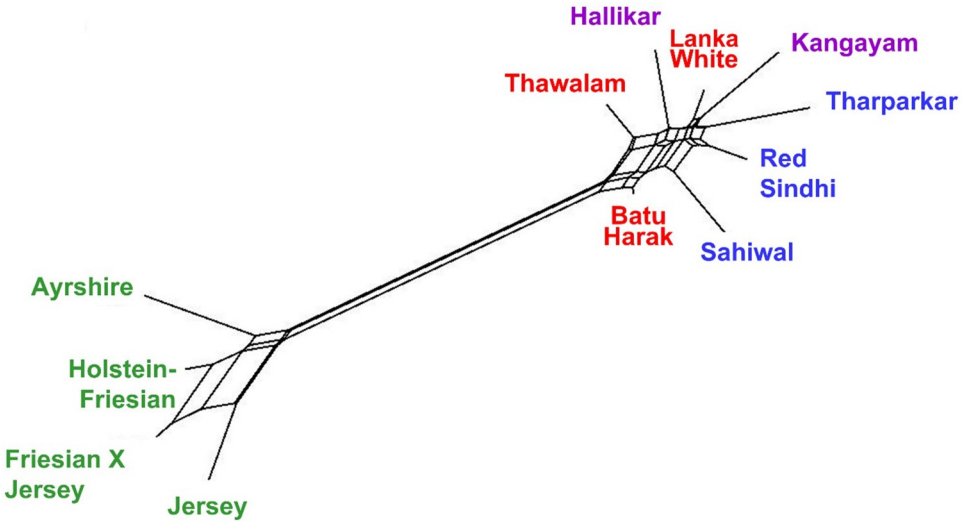


**Fig 2. NeighborNet (below) tree derived from pairwise Nei's genetic distance among Sri Lankan, Indus Valley, South Indian Zebu and commercial taurine cattle.**

https://doi.org/10.1371/journal.pone.0282761.g002

within breeds and only 6.07% was accounted for by differences between breeds (Table 3). When Zebu breeds were grouped according to geographical origin (Grouping I–Indus Valley, South Indian and Sri Lankan Zebu), among group variance was 1.0% ($P>0.05$) while between breed differences within groups accounted for 5.29% ($P<0.05$). When Sri Lankan Zebu were grouped with Indus Valley Zebu (Grouping II), among group variance was not significant ($P>0.05$). Similarly, when Sri Lankan Zebu was grouped alongside South Indian cattle (Grouping III), among group variance slightly decreased and not significant ($P>0.05$). When Red Sindhi cattle (Indus Valley) alone was grouped with Sri Lankan Zebu (as visualized in MDS plot of pairwise $F_{ST}$), among group variance was only 0.86% ($P>0.05$). However, when Hallikar cattle alone (South India) was grouped with Sri Lankan Zebu, among group variance was 2.54% and significant ($P<0.01$). The results clearly indicated strong genetic relationship of Sri Lankan Zebu cattle with Hallikar, an ancestral Mysore type South Indian breed known for its excellent draught characteristics. Bayesian clustering of individuals without prior population information was performed to assess the cryptic genetic structure and degree of admixture among the investigated cattle breeds. Among K = 2 to 14, our dataset was best described with K = 2 genetic clusters, at which the second order rate of change in LnP(D) was maximum (S3 File). When K = 2 was assumed, the investigated breeds got clustered along taurine-indicine divide (Fig 4). When K = 4 was assumed, Sahiwal cattle clustered separately from other Zebu cattle. At K = 5, Kangayam cattle clustered separately while at K = 6, Hallikar and Tharparkar formed separate clusters (Table 4).

## 3.4 Mitochondrial DNA diversity in Sri Lankan cattle

The haplogroup classification and diversity of Sri Lankan Zebu cattle are presented in Table 5. Two major Zebu haplogroups, I1 and I2 were observed in Sri Lankan Zebu with the former predominating the later in all the three breeds. A total of 112 haplotypes were observed in the studied breeds, of which 50 haplotypes (34 belonged to I1 lineage, 12 belonged to I2 lineage and two belonged to T3 lineage) were found in Sri Lankan Zebu cattle. The haplotype diversity was highest in Batu Harak cattle (0.958) while it was lowest in White cattle (0.802) among the Sri Lankan Zebu breeds. The haplotype diversity of I1 lineage was higher than that of I2 lineage in all the three Sri Lankan Zebu cattle. Among the haplotypes observed in all the Zebu cattle (Sri Lankan,

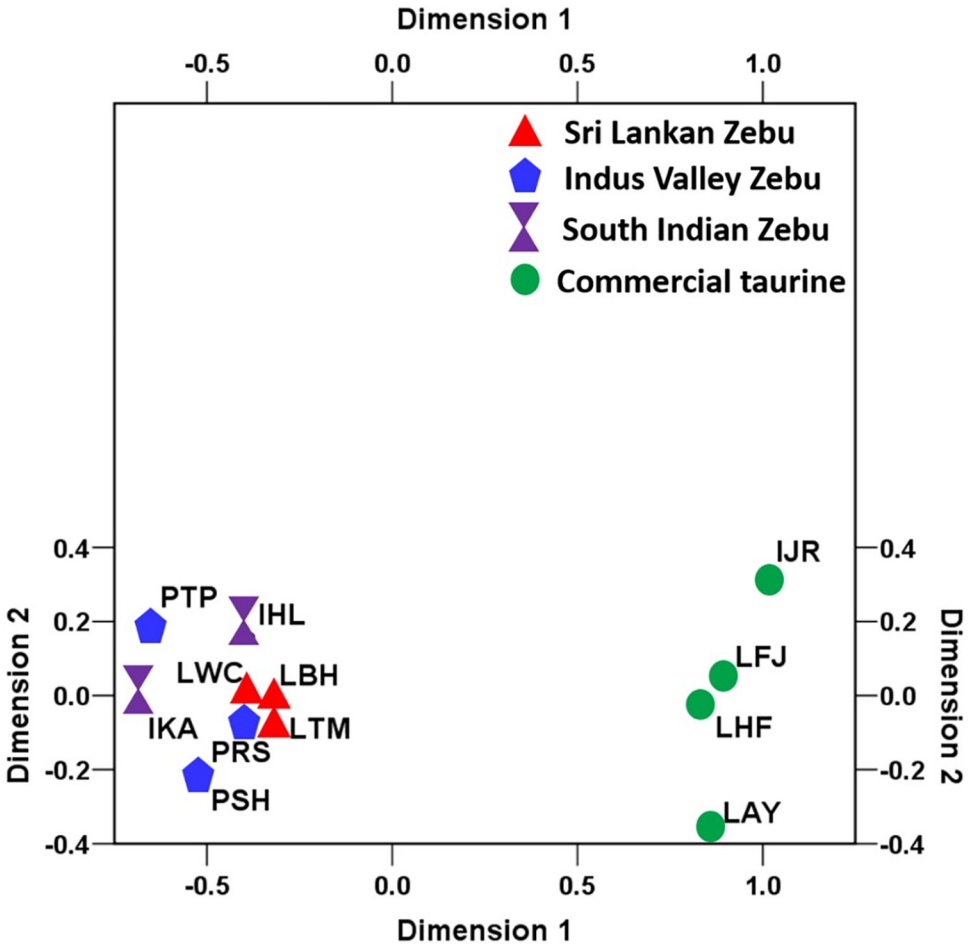


**Fig 3. Multi-dimensional scaling display of pairwise $F_{ST}$ among Sri Lankan, South Indian and Indus valley zebu cattle.**

https://doi.org/10.1371/journal.pone.0282761.g003

Indus Valley and South Indian), 16 were found to be shared across 2 or more breeds, of which 10 belonged to I1 lineage and 6 belonged to I2 lineage (S4 File). A total of 69 singleton haplotypes were observed, of which 44 belonged to I1 lineage and 25 belonged to I2 lineage. Among the I1 lineage, haplotype BWRSPHK_H19_I1 was shared among all the investigated breeds except Thawalam cattle. In case of I2 lineage, haplotype WTRPHK_H106_I2 was shared among all the studied breeds with the exception of Batu Harak and Sahiwal cattle. The diversity of mtDNA control region in the studied cattle populations is presented in Table 6. Overall, 103 polymorphic sites were observed, of which 31 were singleton sites and 70 were parsimony informative sites. The overall nucleotide diversity was 0.009, 0.004 and 0.003 in Batu Harak, Thawalam and White cattle respectively. The nucleotide diversity of I1 lineage in these three breeds was 0.004, 0.002 and 0.001 respectively while the nucleotide diversity of I2 lineage was 0.003, 0.001 and 0.001 respectively. The average number of nucleotide differences was high in Batu Harak cattle (8.22) while it was low in Thawalam (3.26) and White cattle (3.12).

## 3.5. mtDNA phylogeny and evolutionary relationship

Phylogenetic analysis was performed on all the 112 unique haplotypes along with reference sequences of each lineage, the results of which is presented in S5 File. The phylogenetic tree conformed to haplogroup classification with the formation of two major indicine clusters I1 and I2.

**Table 3. Analysis of variance (AMOVA) among Sri Lankan, South Indian and Indus Valley Zebu cattle.**

| Grouping | Source of variation | d.f. | Sum of squares | Variance components | Percentage of variation | P-value |
|---|---|---|---|---|---|---|
| No Grouping | Among populations | 7 | 298.98 | 0.499 | 6.07 | 0.000 |
| | Within populations | 556 | 4296.56 | 7.728 | 93.93 | 0.000 |
| Grouping I (Geography): 1-Indus Valley Zebu; 2-South Indian Zebu; 3-Sri Lankan Zebu | Among groups | 2 | 111.32 | 0.083 | 1.00 | 0.014 |
| | Among pops within groups | 5 | 187.66 | 0.436 | 5.29 | 0.000 |
| | Within populations | 556 | 4295.54 | 7.728 | 93.71 | 0.000 |
| Grouping II: 1-Indus Valley and Sri Lankan Zebu; 2-South Indian Zebu | Among groups | 1 | 69.57 | 0.105 | 1.27 | 0.105 |
| | Among pops within groups | 6 | 229.41 | 0.448 | 5.41 | 0.000 |
| | Within populations | 556 | 4296.56 | 7.728 | 93.32 | 0.000 |
| Grouping III: 1-Indus Valley Zebu; 2-South Indian and Sri Lankan Zebu | Among groups | 1 | 58.61 | 0.073 | 0.88 | 0.056 |
| | Among pops within groups | 6 | 240.37 | 0.461 | 5.58 | 0.000 |
| | Within populations | 556 | 4296.56 | 7.728 | 93.54 | 0.000 |
| Grouping IV (MDS plotting): 1-Indus Valley Zebu; 2-South Indian Zebu; 3-Sri Lankan Zebu and Red Sindhi | Among groups | 1 | 108.12 | 0.071 | 0.86 | 0.134 |
| | Among pops within groups | 6 | 190.86 | 0.447 | 5.42 | 0.000 |
| | Within populations | 556 | 4296.56 | 7.728 | 93.72 | 0.000 |
| Grouping IV (Phylogeny): 1-Indus Valley Zebu; 2-South Indian Zebu; 3-Sri Lankan Zebu and Hallikar | Among groups | 2 | 145.54 | 0.210 | 2.54 | 0.003 |
| | Among pops within groups | 5 | 153.43 | 0.349 | 4.21 | 0.000 |
| | Within populations | 556 | 4296.56 | 7.728 | 93.25 | 0.000 |

https://doi.org/10.1371/journal.pone.0282761.t003

No major sub-clusters were found within each of these two clusters, but three minor clades (one in I1 and two in I2) each with three or more haplotypes were observed. The first minor clade within I2 cluster included unique haplotypes of Batu Harak, Hallikar, Kangayam and Sahiwal cattle while the second minor clade included unique haplotypes of Lanka White and Red Sindhi cattle. The minor clade within I1 cluster included three unique haplotypes of Hallikar and Red Sindhi cattle. The pairwise $F_{ST}$ among Zebu cattle from Sri Lanka, Indus Valley and South India was estimated utilizing the frequency of mtDNA haplotypes (Table 7). Analysis of I1 haplotypes revealed pairwise $F_{ST}$ varying between zero (Hallikar-Red Sindhi; Sahiwal-Tharparkar) and 0.202 (Thawalam-Lanka White) while I2 haplotypes showed pairwise $F_{ST}$ varying between zero (Hallikar-Sahiwal) and 0.178 (Thawalam-Tharparkar). Among the Sri Lankan Zebu breeds, Batu Harak showed low $F_{STs}$ with Tharparkar ($F_{ST}$ = 0.030) and Red Sindhi ($F_{ST}$ = 0.039) among I1 haplogroup while it showed lowest $F_{ST}$ with Sahiwal ($F_{ST}$ = 0.021) among I2 haplogroup. Lanka White cattle showed low $F_{STs}$ with Sahiwal ($F_{ST}$ = 0.030) and Hallikar ($F_{ST}$ = 0.050) among I1 haplogroup while it showed low $F_{STs}$ with Red Sindhi ($F_{ST}$ = 0.020) and Hallikar ($F_{ST}$ = 0.053) among I2 haplogroup. Interestingly, I1 haplogroup of Thawalam cattle showed high $F_{STs}$ not only with Indus Valley and South Indian cattle ($F_{ST}$ = 0.108–0.199) but also with other Sri Lankan Zebu ($F_{ST}$ = 0.168–0.202) breeds. However, among I2 haplogroup, Thawalam cattle showed low $F_{STs}$ with Sahiwal ($F_{ST}$ = 0.020) and Red Sindhi ($F_{ST}$ = 0.039).

## 3.6. Mismatch distribution, haplotype network and tests for selective neutrality

The demographic history of Sri Lankan Zebu cattle was explored by assessing mismatch distribution of nucleotide differences among haplotypes within I1 and I2 maternal lineages (Fig 5).

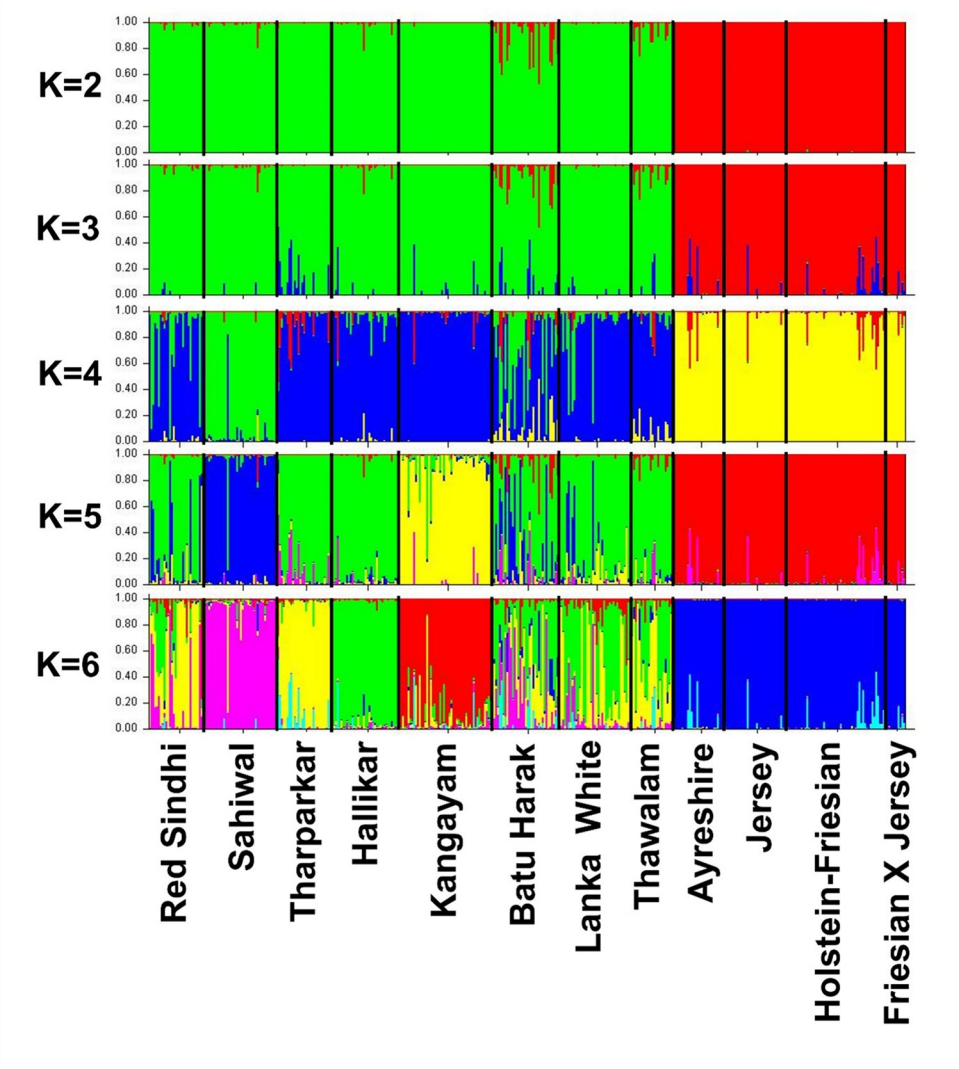


**Fig 4. Bayesian clustering of 406 cattle under assumption of 2 to 8 clusters without a priori population information.** The population names are given below the box plot with the individuals of different populations separated by vertical black lines.

https://doi.org/10.1371/journal.pone.0282761.g004

All the three breeds showed unimodal distribution indicating population expansion in both the lineages. The modal peak was observed at three and four nucleotide differences in Batu Harak cattle for I2 and I1 lineages respectively, while the modal peak was observed at two nucleotide differences in Thawalam cattle for both the lineages. In case of Lanka White cattle, haplotype pairs with no differences were abundant followed by pairs with one and two nucleotide differences in both I1 and I2 lineages. The sum of squared differences (SSD) was significant ($P<0.01$) for I1 lineage of Thawalam and I2 lineage of Batu Harak cattle (Table 8). Although high SSD was observed for I2 lineage of Thawalam cattle, it was not significant ($P>0.05$). Similarly, raggedness index was significantly high for I1 lineage of Thawalam cattle ($P<0.01$) and I2 lineage of Batu Harak cattle ($P<0.05$). The results of neutrality tests revealed negative Tajima's D values for both the lineages of Batu Harak ($P>0.05$) and White cattle ($P>0.05$) indicating an excess of low frequency polymorphisms and demographic expansion. Thawalam cattle showed positive Tajima's D for both I1 and I2 lineages although not

**Table 4. Proportion of membership of each pre-defined population of Sri Lankan Zebu in each of the assigned clusters when assumed under K = 2 and K = 6.**

| Breed | N | K = 2 | | K = 6 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| Red Sindhi | 30 | 0.010 | 0.990 | 0.070 | 0.197 | 0.009 | 0.538 | 0.181 | 0.006 |
| Sahiwal | 38 | 0.013 | 0.987 | 0.010 | 0.015 | 0.010 | 0.042 | 0.917 | 0.005 |
| Tharparkar | 30 | 0.005 | 0.995 | 0.024 | 0.025 | 0.002 | 0.838 | 0.011 | 0.100 |
| Hallikar | 36 | 0.014 | 0.986 | 0.013 | 0.931 | 0.011 | 0.018 | 0.013 | 0.015 |
| Kangayam | 50 | 0.003 | 0.997 | 0.853 | 0.041 | 0.004 | 0.068 | 0.015 | 0.018 |
| Batu Harak | 36 | 0.108 | 0.892 | 0.038 | 0.390 | 0.076 | 0.246 | 0.199 | 0.052 |
| Lanka White | 38 | 0.004 | 0.996 | 0.077 | 0.502 | 0.004 | 0.322 | 0.085 | 0.009 |
| Thawalam | 24 | 0.053 | 0.947 | 0.036 | 0.374 | 0.039 | 0.490 | 0.027 | 0.035 |
| Ayreshire | 27 | 0.997 | 0.003 | 0.002 | 0.003 | 0.933 | 0.003 | 0.003 | 0.056 |
| Jersey | 33 | 0.998 | 0.002 | 0.002 | 0.002 | 0.974 | 0.002 | 0.002 | 0.017 |
| Holstein-Friesian | 53 | 0.997 | 0.003 | 0.002 | 0.003 | 0.942 | 0.003 | 0.003 | 0.047 |
| Friesian X Jersey | 11 | 0.997 | 0.003 | 0.002 | 0.004 | 0.958 | 0.002 | 0.005 | 0.029 |

https://doi.org/10.1371/journal.pone.0282761.t004

significant (P>0.05). Fu's FS statistics was negative for both lineages in all the three Sri Lankan native cattle populations and significant (P<0.02). The results indicated the presence of excess number of alleles as would be expected from a recent population expansion or from genetic hitchhiking. Median Joining (MJ) network of control region haplotypes of Sri Lankan, Indus Valley and South Indian Zebu was reconstructed. The haplotype network results (Fig 6) revealed two distinct lineages, I1 and I2 as expected with expansion events radiating from ancestral nodes. Each of these two ancestral nodes consisted of haplotypes from all the three Sri Lankan native cattle (Batu Harak, Thawalam and White cattle). Interestingly, one of the low frequency nodes formed by haplotypes of Batu Harak and White cattle predated the ancestral and other internal nodes.

## 4. Discussion

### 4.1. Biodiversity and inbreeding estimates

Our study presents the first comprehensive analysis of diversity, genetic and phylogeographic relationship of island cattle of Sri Lanka with Zebu cattle of Indus Valley and South Indian origin. The existing cattle production system of Sri Lanka comprises of at least five categories of breed types that include native Zebu cattle, exotic Zebu, native Zebu x exotic Zebu crosses, exotic taurine, and native Zebu x exotic taurine crosses. The distribution of cattle breed types

**Table 5. Haplogroup classification and diversity of mtDNA control region haplotypes in Sri Lankan, Indus Valley and South Indian Zebu.**

| Breed | Haplogroup Count | | | | | Haplotype Count | | | | | Haplotype Diversity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ALL | I1 | I2 | T2 | T3 | ALL | I1 | I2 | T2 | T3 | ALL | I1 | I2 |
| IHL | 36 | 22 | 13 | 1 | - | 27 | 15 | 11 | 1 | - | 0.975 | 0.939 | 0.974 |
| IKA | 51 | 36 | 14 | 1 | - | 17 | 11 | 5 | 1 | - | 0.869 | 0.784 | 0.659 |
| LBH | 51 | 36 | 13 | - | 2 | 28 | 20 | 6 | - | 2 | 0.958 | 0.941 | 0.795 |
| LTM | 25 | 19 | 6 | - | - | 6 | 4 | 2 | - | - | 0.803 | 0.696 | 0.533 |
| LWC | 51 | 38 | 13 | - | - | 16 | 10 | 6 | - | - | 0.802 | 0.680 | 0.641 |
| PRS | 29 | 17 | 12 | - | - | 22 | 12 | 10 | - | - | 0.975 | 0.941 | 0.970 |
| PSH | 29 | 20 | 6 | - | 3 | 19 | 11 | 6 | - | 2 | 0.921 | 0.837 | 1.000 |
| PTP | 27 | 21 | 6 | - | - | 12 | 9 | 3 | - | - | 0.892 | 0.838 | 0.733 |
| Total/Overall | 299 | 209 | 83 | 2 | 5 | 112 | 69 | 37 | 2 | 4 | 0.941 | 0.896 | 0.894 |

https://doi.org/10.1371/journal.pone.0282761.t005

**Table 6. Polymorphism and diversity of mtDNA control region haplotypes in Sri Lankan, Indus Valley and South Indian Zebu.**

| Breed | No. Polymorphic Sites | | | Singleton Sites | | | Parsimony Informative Sites | | | Nucleotide Diversity | | | Avg. No. Nucleotide Differences | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ALL | I1 | I2 | ALL | I1 | I2 | ALL | I1 | I2 | ALL | I1 | I2 | ALL | I1 | I2 |
| IHL | 63 | 17 | 13 | 44 | 11 | 9 | 19 | 6 | 4 | 0.008 | 0.003 | 0.003 | 6.91 | 2.42 | 2.69 |
| IKA | 55 | 11 | 8 | 39 | 4 | 5 | 16 | 7 | 3 | 0.006 | 0.002 | 0.002 | 5.31 | 1.70 | 1.80 |
| LBH | 61 | 18 | 9 | 12 | 6 | 5 | 49 | 12 | 4 | 0.009 | 0.004 | 0.003 | 8.22 | 3.45 | 2.54 |
| LTM | 11 | 5 | 2 | 2 | 2 | 0 | 9 | 3 | 2 | 0.004 | 0.002 | 0.001 | 3.26 | 1.50 | 1.07 |
| LWC | 18 | 10 | 6 | 7 | 4 | 5 | 11 | 6 | 1 | 0.003 | 0.001 | 0.001 | 3.12 | 1.31 | 1.05 |
| PRS | 33 | 15 | 17 | 18 | 11 | 12 | 15 | 4 | 5 | 0.006 | 0.003 | 0.004 | 5.55 | 2.56 | 3.83 |
| PSH | 55 | 15 | 9 | 9 | 10 | 7 | 46 | 5 | 2 | 0.012 | 0.002 | 0.004 | 11.17 | 2.10 | 3.40 |
| PTP | 19 | 13 | 3 | 5 | 5 | 1 | 14 | 8 | 2 | 0.004 | 0.003 | 0.002 | 3.85 | 2.44 | 1.40 |
| Overall | 103 | 55 | 37 | 31 | 31 | 15 | 70 | 24 | 22 | 0.007 | 0.003 | 0.003 | 6.04 | 2.34 | 2.43 |

https://doi.org/10.1371/journal.pone.0282761.t006

varies with the agro-ecological zones with Zebu and Zebu crosses dominating the dry zone while purebred exotic taurine and native Zebu x exotic taurine crosses dominating the wet zone of the country. The intermediary zone consisted of both Zebu x taurine as well as native x exotic Zebu crosses [42]. Various breed improvement programs targeting productivity enhancement in indigenous Zebu cattle resulted in dilution of germplasm leading to concerns on breed purity and conservation. The present study showed allelic diversity and heterozygosity of Sri Lankan native Zebu cattle being slightly higher than the estimates observed for Indus Valley (Red Sindhi, Sahiwal and Tharparkar) and South Indian (Hallikar and Kangayam) breeds. Among the Sri Lankan Zebu, Batu Harak had the highest $n_o$, $H_o$ and $H_e$. The extant population of Batu Harak is believed to have resulted from extensive outcrossing of original Sri Lankan Zebu cattle with various Indo-Pakistan Zebu cattle brought to Sri Lanka [4, 43]. This might have resulted in higher levels of genetic variability in Batu Harak cattle as compared to other Sri Lankan Zebu breeds. The estimates of inbreeding in Sri Lankan native cattle breeds were relatively high and varied from 4.8% (Batu Harak) to 7.2% (White cattle). Such high estimates of $F_{IS}$ can be associated with heterogeneity of herds sampled within each breed, possibly resulting in Wahlund effect [44]. Assortative mating can also increase genetic relatedness within the livestock populations and this nonrandom mating pattern can result in increase in heterozygosity deficit. The Sri Lankan White cattle herds mainly consist of animals with similar phenotypes (body size, coat colour, etc.) and farmers preferred bulls with white coat color to ensure the uniformity of herd. The past and current farming practice in the Eastern province (native breed tract of White cattle) also indicates use of few breeding bulls within each herd. Similarly, Thawalam is an isolated population of local Zebu in Central and Uva

**Table 7. Pairwise $F_{ST}$ based on I1 (upper triangle) and I2 (lower triangle) haplogroups among Sri Lankan, Indus Valley and South Indian Zebu.**

| | IHL | IKA | LBH | LTM | LWC | PRS | PSH | PTP |
|---|---|---|---|---|---|---|---|---|
| IHL | - | 0,067 | 0,067 | 0,139 | 0,050 | 0,000 | 0,001 | 0,006 |
| IKA | 0,063 | - | 0,132 | 0,199 | 0,071 | 0,086 | 0,056 | 0,068 |
| LBH | 0,098 | 0,139 | - | 0,168 | 0,123 | 0,039 | 0,073 | 0,030 |
| LTM | 0,051 | 0,109 | 0,108 | - | 0,202 | 0,149 | 0,108 | 0,128 |
| LWC | 0,053 | 0,071 | 0,162 | 0,083 | - | 0,062 | 0,030 | 0,060 |
| PRS | 0,054 | 0,051 | 0,100 | 0,039 | 0,045 | - | 0,015 | 0,007 |
| PSH | 0,000 | 0,061 | 0,021 | 0,020 | 0,094 | 0,037 | - | 0,000 |
| PTP | 0,075 | 0,111 | 0,119 | 0,178 | 0,088 | 0,051 | 0,081 | - |

https://doi.org/10.1371/journal.pone.0282761.t007

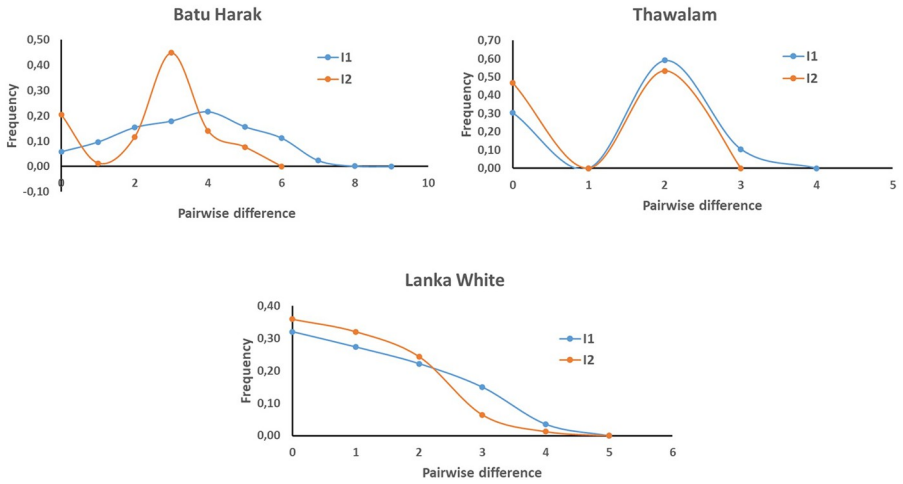


**Fig 5. Mismatch distribution of pairwise nucleotide differences among mtDNA control region haplogroups of Sri Lankan zebu cattle.**

https://doi.org/10.1371/journal.pone.0282761.g005

provinces with little gene flow from outside. This might have contributed to consanguineous mating and higher levels of inbreeding within these populations.

## 4.2. Breed history, genetic relationships and population structure

The present-day Batu Harak cattle are believed to represent the descendants of crosses between original archaic "Sri Lankan Zebu cattle" and various Indus Valley (Indo-Pakistan) Zebu breeds [3, 4]. The White cattle, also called as Thamankaduwa are believed to have originated from the mixture of cattle that existed in ancient Sri Lanka and Indian white cattle breeds [3]. Anecdotal evidence indicates White cattle probably originated from South India and were brought by tobacco planters in the north-eastern region of the country [3, 5, 45]. The Thawalam cattle are pack animals that are bred and reared in isolated pockets of Central and Uva provinces of Sri Lanka. Basically, they represent non-descript type, but several generations of breeding in isolation resulted them in emerging as a separate breed with distinct body conformation and capable of pulling loads in hilly regions [3]. The results of the present study

**Table 8. Tests for demographic expansion based on mismatch analysis and neutrality in Sri Lankan, Indus Valley and South Indian Zebu.**

| Test | Statistics | I1 haplogroup | | | I2 haplogroup | | |
|---|---|---|---|---|---|---|---|
| | | LBH | LTM | LWC | LBH | LTM | LWC |
| **Neutrality tests** | | | | | | | |
| Tajima's D test | Tajima's D | -0,68 | 0,14 | -1,36 | -0,49 | 1,03 | -1,68 |
| | Tajima's D p-value | 0,297 | 0,589 | 0,060 | 0,360 | 0,852 | 0,028 |
| Fu's FS test | FS | -26,08 | -27,91 | -28,28 | -14,22 | -6,35 | -22,06 |
| | FS p-value | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| **Mismatch analysis** | | | | | | | |
| Demographic expansion | Tau | 3,93 | 2,39 | 2,00 | 3,52 | 2,59 | 1,00 |
| | SSD | 0,003 | 0,203 | 0,002 | 0,100 | 0,204 | 0,006 |
| | Model (SSD) p-value | 0,241 | 0,003 | 0,873 | 0,007 | 0,141 | 0,980 |
| | Raggedness index | 0,021 | 0,688 | 0,024 | 0,263 | 0,787 | 0,042 |
| | Raggedness p-value | 0,399 | 0,001 | 0,973 | 0,018 | 0,243 | 0,992 |

https://doi.org/10.1371/journal.pone.0282761.t008

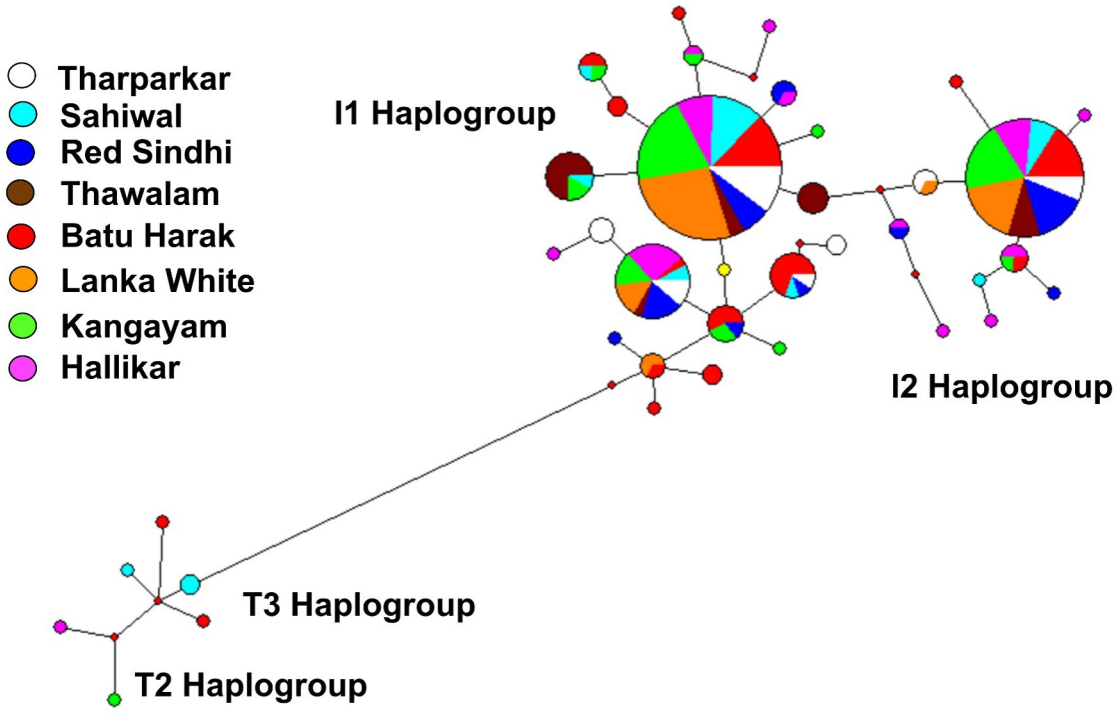


**Fig 6. Median joining network of mtDNA control region haplotypes of Sri Lankan, Indus valley and South Indian zebu (size of the circle indicates frequency of haplotypes).**

https://doi.org/10.1371/journal.pone.0282761.g006

confirmed the close genetic relationships of indigenous Sri Lankan cattle with South Indian and Indus Valley Zebu, particularly with Hallikar and Red Sindhi cattle respectively. The pairwise genetic differentiation of Sri Lankan Zebu cattle from these two breeds were relatively low as compared to other Indus Valley and South Indian breeds analyzed in this study. The analysis of molecular variance revealed significant among group variance when the Sri Lankan Zebu cattle were grouped with Hallikar cattle (Table 3). All the three Sri Lankan Zebu breeds shared significant levels of ancestry with Hallikar cattle breed, while admixture of Indus Valley breeds was also noticeable. The proportion of membership coefficient assigned to Hallikar cluster was 39.0%, 50.2% and 37.4% for Batu Harak, Lanka White and Thawalam cattle respectively when K = 6 was assumed under unsupervised Bayesian clustering analysis. Similarly, the proportion of membership coefficient assigned to Tharparkar/Red Sindhi cluster was 24.6%, 32.2% and 49.0% while the assignment to Sahiwal cluster was 19.9%, 8.5% and 2.7% in Batu Harak, Lanka White and Thawalam cattle respectively (Table 4). The White cattle showed a relatively high proportion of shared ancestry with South Indian Zebu while Thawalam cattle showed high levels of shared ancestry with Indus Valley Zebu. Batu Harak cattle showed more or less similar levels of shared ancestry with both South Indian and Indus Valley Zebu cattle. The results of genetic structure analysis revealed strong genetic relationships between Sri Lankan and South Indian Zebu cattle, while varied levels of admixture was observed with Indus Valley Zebu cattle. Introgression of Indus Valley Zebu in Sri Lankan cattle appears to be a recent phenomenon with the introduction of grading up scheme during 1950s in the mid and low country dry zones using Tharparkar, Red Sindhi and Sahiwal cattle. This was evident from the presence of Sahiwal introgression at specific locations in the native tract of Batu Harak cattle.

## 4.3. Breed purity and taurine introgression

Historically, with the arrival of Europeans, temperate *Bos taurus* breeds were brought into Sri Lanka and since then crossbreeding is used as one of the methods for genetic improvement of milk production in the country. However, indiscriminate breeding of local cattle using semen of taurine and crossbred bulls resulted in dilution of native germplasm. Bayesian clustering analysis was performed to assess the taurine admixture in Sri Lankan Zebu cattle using pure-bred Jersey, Ayrshire and Holstein-Friesian as reference genotypes. The proportion of membership coefficient obtained for individual animals in the inferred Zebu and taurine clusters were utilized to estimate taurine admixture and purity of the investigated Zebu cattle breeds [19]. The results revealed high levels of breed purity in Lanka White cattle with all the sampled individuals showing >97% Zebu ancestry in them. Interestingly, the farmers in the Eastern provincial region predominantly follow natural service for breeding their cattle and the artificial insemination coverage is less than 5% [42, 46]. It is also believed that the White cattle are descendants of the ancient royal herds and white colour is considered as a sign of cleanliness and prosperity [12]. Further, the farmers owning the White cattle herds are aware of the breed purity and use true to the type bulls for breeding, with white coat colour as a primary criterion for selection. With respect to Batu Harak cattle, >38% of animals showed at least 6.25% taurine admixture. Further, more than one-fifth of Batu Hark cattle sampled in the present study showed at least 25% taurine admixture although they had Zebu features morphologically. In case of Thawalam cattle, 29.2% of sampled individuals had at least 6.25% taurine admixture while 4.2% had at least 25% taurine admixture. The National Livestock Breeding Policy of Sri Lanka [12] recommends grading up of native cattle in the low and mid country dry zones using Ayrshire and Jersey breeds with up to 50% of taurine blood level in the crossbreds for improved milk production. The level of taurine admixture in Batu Harak and Thawalam cattle is relatively higher, possibly due to well established AI programs and readily available exotic taurine semen for insemination.

## 4.4. Maternal lineage and phylogeography

The mtDNA maternal lineages of indigenous Sri Lankan Zebu cattle belonged to I1 and I2 haplogroups typical to other South Asian Zebu cattle [12, 47]. The mtDNA diversity of Sri Lankan Zebu cattle in terms of nucleotide diversity, average number of nucleotide differences and haplotype diversity was relatively low for both I1 and I2 lineages with the exception of Batu Harak cattle. In case of Batu Harak cattle, the diversity parameters were comparable to that of Indus Valley, North and South Indian Zebu cattle [19, 37, 47]. The mismatch analysis to explore demographic history of Sri Lankan Zebu cattle indicated unimodal distribution and potential population expansion in all the three breeds. The SSD and raggedness index was non-significant in both the lineages of all the three breeds except for I1 lineage of Thawalam cattle ($P<0.01$) and I2 lineage of Batu Harak cattle ($P<0.05$). The SSD and raggedness index values were relatively high for I2 lineage of Thawalam cattle as well but were not significant ($P>0.05$). Non-significant values of SSD indicate the goodness of fit between observed and expected data (in a parametric bootstrap simulation analysis) under the population expansion model. Significant values of SSD indicate the deviation of observed data from expected data which is common in populations under demographic equilibrium. Similarly, the values of raggedness index will be larger for populations in demographic equilibrium while they will be small and non-significant for populations under demographic expansion. Interestingly, Tajima's D statistics was positive for both I1 and I2 lineages of Thawalam cattle, while it was negative in Batu Harak and White cattle. The negative Tajima's D was significant ($P<0.05$) only for I2 lineage of White cattle. The Tajima's D statistics was based on frequency spectrum of

mutations while Fu's FS was based on haplotype distribution. The negative Tajima's D signifies an excess of low frequency polymorphisms than expected, indicating either expansion of population size or purifying selection. Similarly, negative FS is evidence for an excess number of alleles as would be expected from a recent population expansion or genetic hitchhiking. The results of the present study thus indicated purifying selection among the Sri Lankan Zebu cattle breeds. Further, Fu's FS statistic, sensitive to the presence of singletons and devised more specifically to detect demographic expansion was significant ($P<0.01$) in all the investigated cattle breeds. However, it is worth to mention that analysis of additional samples from Thawalam cattle will give more insights into its demographic history as it shows signs of population equilibrium.

## 5. Conclusion

The present study revealed moderately high levels of genetic diversity in Sri Lankan Zebu cattle breeds. The estimates of inbreeding were relatively high in Thawalam and White cattle breeds. The Sri Lankan Zebu cattle breeds showed strong genetic relationships with Hallikar cattle, an ancient breed considered to be ancestor for most of the Mysore type draught cattle breeds of South India. Recent introgressions of Indus Valley Zebu like Red Sindhi, Tharparkar and Sahiwal was also evident in Sri Lankan Zebu, especially in Batu Harak and Thawalam cattle breeds. mtDNA analysis showed two maternal lineages, I1 and I2 in Sri Lankan Zebu cattle with relatively high level of haplotype diversity in Batu Harak cattle. Genetic admixture analysis revealed high levels of breed purity in Lanka White cattle while significant taurine admixture was observed in Batu Harak and Thawalam cattle. Genetic dilution of native Zebu germplasm is a cause for concern from biodiversity and conservation standpoint. It is recommended that national breeding organizations consider establishing conservation units for the three native cattle breeds to maintain breed purity and initiate genetic improvement programs. Considering the superior adaptability characteristics of these animals, such an initiative will be beneficial in the long run to develop locally adapted dual-purpose cattle for milk and meat production in Sri Lanka.

## Supporting information

**S1 File. Global F statistics among Sri Lankan, Indus Valley, South Indian Zebu and commercial taurine cattle.**
(DOCX)

**S2 File. The test for selective neutrality at 27 microsatellite marker loci.** Selection detection based on the $F_{ST}$ outlier approach using LOSITAN under the assumption of infinite allele mutation model.
(DOCX)

**S3 File. Determination of correct number of clusters in Bayesian STRUCTURE analysis (Evanno et al.** 2005) (a) Mean L (K) over 20 runs for each K value of 1 to 14 (b) Distribution of ΔK with the modal value (K = 2) indicating the true K or the uppermost level of structure.
(DOCX)

**S4 File. Frequency of mtDNA haplotypes and their sharing among different breeds of Sri Lankan, Indus Valley and South Indian Zebu.**
(DOCX)

**S5 File. Maximum likelihood tree of mitochondrial DNA haplotypes of Sri Lankan, Indus Valley and South Indian Zebu (B-Batu Harak; W-Lanka White; T-Thawalam; R-Red**

**Sindhi; S-Sahiwal; P-Tharparkar; H-Hallikar; K-Kangayam).**
(DOCX)

## Acknowledgments

We thank local farmers and veterinarians in Sri Lanka and Pakistan along with Livestock and Dairy Development Departments of different provinces for their help in sampling. The authors also acknowledge the support of S. Bandara, N. Shanjayan, R. Rupasinghe, D.C.A. Gunawardena and staff of Department of Basic Veterinary Sciences, University of Peradneiya, Sri Lanka for their support during the study.

## Author Contributions

**Conceptualization:** Lokugalappatti L. G. S., Alexander P. A. B. D., Kathiravan Periasamy.

**Data curation:** Arnaud Stephane R. Tapsoba.

**Formal analysis:** Arnaud Stephane R. Tapsoba, Kathiravan Periasamy.

**Funding acquisition:** Alexander P. A. B. D.

**Investigation:** Lokugalappatti L. G. S., Saumya Wickramasinghe, Kamran Abbas, Tanveer Hussain, Saravanan Ramasamy, Vandana Manomohan, Masroor E. Babar.

**Methodology:** Lokugalappatti L. G. S., Saumya Wickramasinghe, Kamran Abbas, Vandana Manomohan, Rudolf Pichler.

**Resources:** Tanveer Hussain, Masroor E. Babar.

**Supervision:** Kathiravan Periasamy.

**Writing – original draft:** Kathiravan Periasamy.

**Writing – review & editing:** Lokugalappatti L. G. S., Alexander P. A. B. D., Tanveer Hussain, Saravanan Ramasamy, Masroor E. Babar, Kathiravan Periasamy.

## References

1. Jalatge EFA. Crossbreeding of Cattle in Sri Lanka. 3rd World Congress on Genetics Applied to Livestock Production. 1986; 14. https://digitalcommons.unl.edu/wcgalp/14.
2. Chandrasiri ADN. The state of animal genetic resources in Sri Lanka. A joint publication of Department of Animal Production and Health and the office of the FAO representative in Sri Lanka and Maldives, Colombo, Sri Lanka. 2004.
3. Silva P, Herath HMSP, Chandrasiri AND. Indigenous cattle types and production systems in Sri Lanka. Economic Review. 2008; 34: 25–28.
4. Perera BMAO Jayasuriya MCN. The dairy industry in Sri Lanka: Current status and future directions for a greater role in national development. Journal of the National Science Foundation of Sri Lanka. 2008; 36: 115–126. https://doi.org/10.4038/jnsfsr.v36i0.8050
5. Felius M. Cattle Breeds: An Encyclopedia. Misset Uitgeverij bv, Postbus 4, 7000 BA Doetinchem, NL. 1995.
6. Tilakaratne N, Matsukawa T, Buvanendran V. Estimation of body weight in pure and crossbred Sinhala cattle using body measurements. Ceylon Veterinary Journal. 1974; XXII: 82–84.
7. Buvanendran V, Mahadevan P. Crossbreeding for milk production in Sri Lanka. World Animal Review. 1975; 15: 7–13.
8. FAOSTAT. https://www.fao.org/faostat/en/#data/QCL.2019.
9. Silva P, Jianlin H, Hanotte O, Chandrasiri AND, Herath HMSP. Indigenous Cattle in Sri Lanka: Production Systems and Genetic Diversity. In "Sustainable Improvement of Animal production and Health" edited by Odongo N. E., Garcia M. and Viljoen G. J., pp. 45–48, Food and Agriculture Organization of the United Nations, Rome. 2010.

10. Shanjayan N, Lokugalappatti LGS. A morphometric analysis of indigenous white cattle (Thamankaduwa breed) in the Eastern province of Sri Lanka with a description of a novel character. 5th International Symposium-IntSym 2015, SEUSL, 2015.
11. FAO. World Watch List for Domestic Animal Diversity. 3rd edition. Ed. Sherf BD. Food and Agriculture Organization of the United Nations. Viale delle Terme di Caracella 00100. Rome, Italy. 2000.
12. The National Livestock Breeding Policy guidelines and strategies for Sri Lanka. Department of Animal Production and Health, Ministry of Livestock and Rural Community Development No. 45, St. Micheals's Road, Colombo 3. 2010.
13. Canon J, Garcia D, Garcia-Atance MA, Obexer-Ruff G, Lenstra JA, Ajmone-Marsan P, et al. Geographical partitioning of goat diversity in Europe and the Middle East. Animal Genetics. 2006; 37: 327–334. https://doi.org/10.1111/j.1365-2052.2006.01461.x PMID: 16879341 ISBN No: 978-955-0355-01-3.
14. Vahidi SM, Tarang AR, Naqvi AN, Anbaran MF, Boettcher P, Joost S, et al. Investigation of the genetic diversity of domestic Capra hricus breeds reared within an early goat domestication area in Iran. Genetics Selection Evolution. 2014; 46: 27.
15. Ajmone-Marsan P, Colli L, Han JL, Achilli A, Lancioni H, Joost S, et al. The characterization of goat genetic diversity: towards a genomic approach. Small Ruminant Research. 2014; 121: 58–72.
16. Joshi MB, Rout PK, Mandal AK, Tyler-Smith C, Singh L, Thangaraj K. Phylogeography and origin of Indian domestic goats. Molecular Biology and Evolution. 2004; 21 (3): 454–462. https://doi.org/10.1093/molbev/msh038 PMID: 14660684.
17. Naderi S, Rezaei H-R, Taberlet P, Zundel S, Rafat S-A, Naghash H-R, et al. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. PLoS One. 2007; 2 (10): e1012, https://doi.org/10.1371/journal.pone.0001012 PMID: 17925860
18. Liu YP, Cao SX, Chen SY, Yao YG, Liu TZ. Genetic diversity of Chinese domestic goat based on the mitochondrial DNA sequence variation. Journal of Animal Breeding and Genetics. 2009; 126: 80–89. https://doi.org/10.1111/j.1439-0388.2008.00737.x PMID: 19207934
19. Manomohan V, Saravanan R, Pichler R, Murali N, Sivakumar K, Sudhakar K, et al., Legacy of draught cattle breeds of South India: Insights into population structure, genetic admixture and maternal origin. PLoS ONE. 2021; 16(5): e0246497. https://doi.org/10.1371/journal.pone.0246497 PMID: 34029341
20. FAO. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. No. 9. Rome; 2011.
21. Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 2004; 4: 535–538.
22. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. LOSITAN: a workbench to detect molecular adaptation based on a FST outlier method. BMC Bioinformatics. 2008; 9: 323. https://doi.org/10.1186/1471-2105-9-323 PMID: 18662398.
23. Dieringer D, Schlötterer C. MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for larger microsatellite data sets. Molecular Ecology Resources. 2003; 3 (1): 167–169. https://doi.org/10.1046/j.1471-8286.2003.00351.x
24. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution, 2006; 23(2): 254–267. https://doi.org/10.1093/molbev/msj030 PMID: 16221896.
25. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38: 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x PMID: 28563791.
26. Rousset F "genepop'007: a complete re-implementation of the genepop software for Windows and Linux." Molecular Ecology Resources, 2008; 8: 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x PMID: 21585727.
27. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 2010; 10(3):564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x PMID: 21565059.
28. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945–959. https://doi.org/10.1093/genetics/155.2.945 PMID: 10835412.
29. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology. 2005; 14: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739
30. Rozas J. DNA sequence polymorphism analysis using DnaSP. Methods in Molecular Biology. 2009; 537: 337–350. https://doi.org/10.1007/978-1-59745-251-9_17 PMID: 19378153.
31. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution. 2013; 30(12):2725–9. https://doi.org/10.1093/molbev/mst197 PMID: 24132122.

32. Bandelt HJ, Forster P, Rohl A. Median-Joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution. 1999; 16: 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250.

33. Sharma R, Maitra A, Singh PK, Tantia MS. Genetic diversity and relationship of cattle populations of East India: distinguishing lesser known cattle populations and established breeds based on STR markers. SpringerPlus. 2013; 2:359. https://doi.org/10.1186/2193-1801-2-359 PMID: 23961421.

34. Hussain T, Babar ME, Peters SO, Wajid A, Ali A, Azam A, et al. Microsatellite markers based genetic evaluation of Pakistani cattle breeds. Pakistan Journal of Zoology. 2016; 48(6): 1633–1641.

35. Sodhi M, Mukesh M, Mishra BP, Ahlawat SPS, Prakash B, Sobti RC. Microsatellite analysis of genetic population structure of Zebu cattle (Bos indicus) breeds from North-Western region of India. Animal Biotechnology. 2011; 22:16–29. https://doi.org/10.1080/10495398.2011.536091 PMID: 21328102

36. Egito AA, Paiva SR, Mariante AS, Almeida LD, Castro SR, Grattapaglia D. Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil. BMC Genetics. 2007; 8: 83. https://doi.org/10.1186/1471-2156-8-83 PMID: 18067665

37. Sharma R, Kishore A, Mukesh M, Ahlawat S, Maitra A, Pandey AK, et al., Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. BMC Genetics. 2015; 16:73. https://doi.org/10.1186/s12863-015-0221-0 PMID: 26123673.

38. Smith SB, Zembayashi M, Lunt DK, Sanders JO, Gilbert CD. Carcass traits and microsatellite distributions in offspring of sires from three geographical regions of Japan. Journal of Animal Science. 2001; 79:3041–3051. https://doi.org/10.2527/2001.79123041x PMID: 11811458

39. Li M-H, Iso-Touru T, Laurén H, Kantanen J. A microsatellite-based analysis for the detection of selection on BTA1 and BTA20 in northern Eurasian cattle (Bos taurus) populations. Genetics Selection Evolution. 2010; 42:32. http://www.gsejournal.org/content/42/1/32. https://doi.org/10.1186/1297-9686-42-32 PMID: 20691068

40. Kruskal JB. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964; 29: 115–129.

41. Moroke ND. Profiling Some of the Dire Household Debt Determinants: A Metric Multidimensional Scaling Approach. Journal of Economics and Behavioral Studies, 2014; 6 (11): 858–867.

42. Abeygunawardena H, Rathnayaka D, Jayatilake WMAP. Characteristics of cattle farming systems in Sri Lanka. Journal of the National Science Foundation of Sri Lanka. 1997; 25:25–38.

43. Perera ANF, Perera ANK, Perera ERK.Traditional cattle management systems and Indigenous Veterinary Medicine of Sri Lanka. Economic Review. 2010; (April/May): 63–67. http://dl.nsf.ac.lk/handle/1/14122

44. Hedrick PW. High inbreeding in sheep or erroneous estimation. Journal of Heredity. 2013; 104:298–299. https://doi.org/10.1093/jhered/ess139 PMID: 23293349

45. Nadheer MA. A preliminary investigation on indigenous cattle and farming system in South Eastern region of Sri Lanka. M.Sc. Thesis submitted to Postgraduate Institute of Agriculture, University of Peradeniya, Sri Lanka. 2005.

46. Sinniah J, Pollott GE. Breeding activities and adoption of artificial insemination amongst dairy herds in the dry zone of Sri Lanka. 2006; 18, Article #82.

47. Chen S, Lin B-Z, Baig M, Mtra B, Lopes RJ, Santos AM, et al., Zebu cattle are an exclusive legacy of the South Asia neolithic. Molecular Biology and Evolution. 2010; 27 (1): 1–6. https://doi.org/10.1093/molbev/msp213 PMID: 19770222.