Contrasting population genetics of cattle- and buffalo- derived *Theileria annulata* causing tropical theileriosis

Umer Chaudhry[1‡]*, Qasim Ali[2‡], Lynn Zheng[1], Imran Rashid[3], Muhammad Zubair Shabbir[3], Muhammad Nauman[4], Kamran Ashraf[2], Mike Evans[1], Shahzad Rafiq[3], Muhammad Oneeb[3], Ivan Morrison[1], Liam Morrison[1], Neil D. Sargison[1]*

[1] University of Edinburgh, Royal (Dick) School of Veterinary Studies and Roslin Institute, UK
[2] Gomal University, Dera Ismail Khan, Khyber Pakhtunkhwa, Pakistan
[3] University of Veterinary and Animal Sciences Lahore, Pakistan
[4] Disease Diagnostic Laboratory, Livestock and Dairy Department, Okara, Pakistan

‡ Contributed equally

*Corresponding authors:

Umer Chaudhry
Email: uchaudhr@exseed.ed.ac.uk, Tel: 00441316519244
University of Edinburgh, The Roslin Institute, Easter Bush Veterinary Centre, UK, EH25 9RG

Neil Sargison
Neil.Sargison@ed.ac.uk, +44 (0)131 651 7300
University of Edinburgh, R(D)SVS, Easter Bush Veterinary Centre, UK, EH25 9RG

Abstract

The present study was designed to improve understanding of *Theileria annulata* in sympatric water buffalo and cattle in the Punjab province of Pakistan. The prevalence of tropical theileriosis is high, buparvaquone resistance is widespread, and vaccine protection is poor in the field. Better understanding is, therefore, needed of the factors that influence the genetics of *T. annulata* populations both within its hosts and in its overall populations. Here we utilise a panel of six satellites and a mitochondrial cytochrome b marker to explore the multiplicity of *T. annulata* infection and patterns of emergence and spread of different parasite genotypes. Parasite materials were collected from infected animals in defined regions, where water buffalo and cattle are kept together. Our results show that *T. annulata* is genetically more diverse in cattle- than in water buffalo-derived populations (the mean numbers of unique satellite alleles were 13.3 and 1.8 and numbers of unique cytochrome b locus alleles were 65 and 27 in cattle- and water buffalo- derived populations, respectively). The data show a high level of genetic diversity among the individual host-derived populations (the overall heterozygosity (H_e) indices were 0.912 and 0.931 in cattle, and 0.874 and 0.861 in buffalo, based on satellite and cytochrome b loci, respectively). When considered in the context of high parasite transmission rates and frequent animal movements between different regions, the predominance of multiple *T. annulata* genotypes, with multiple introductions of infection in the hosts from which the parasite populations were derived, may have practical implications for the spread of parasite genetic adaptations; such as those conferring vaccine cross-protection against different strains affecting cattle and buffalo, or resistance to antiprotozoal drugs.

Key Words: Tropical theileriosis, *T. annulata*, buparvaquone resistance, antiprotozoal drugs.

1. Introduction

Vector-borne haemoprotozoa impact globally on the health, welfare, and production of livestock (Jabbar et al., 2015). The genus *Theileria* comprises of two important species, *Theileria annulata* and *Theileria parva* (Lawrence, 1979). *T. parva* typically causes East Coast Fever, Corridor Disease, or January Disease in southern and eastern Africa (Uilenberg et al., 1982), whereas, *T. annulata* causes tropical theileriosis in North Africa and South Asia (Nourollahi-Fard et al., 2015). Tropical theileriosis amongst the most important neglected tropical parasitic diseases of livestock (Sivakumar et al., 2014). Cattle and buffalo become infected with *T. annulata* following the transmission of sporozoites through ixodid ticks of the genus *Hyalomma*. These stages invade lymphocytes and develop by schizogony. Micromerozoites released from the lymphocytes enter erythrocytes and multiply by binary fission to produce merozoites. Piroplasms ingested during tick feeding, migrate to the arthropod gut where gametogeny and production of zygotes occurs. These enter the haemolymph and are carried to the salivary gland, where further replication by binary fission gives rise to sporozoites. The life cycle is completed when the next instar of the tick feeds (Gharbi and Darghouth, 2015). Clinically infected cattle and buffalo show high fever and lymphadenopathy, sometimes accompanied by haemolytic anaemia, respiratory and ocular lesions (Al-Hosary et al., 2010; Mahmmod et al., 2011).

Avoidance, removal or suppression of *Hyalomma* ticks or prophylactic use of antiprotozoal drugs are generally impractical in the control of tropical theileriosis. The unsustainability of these disease control measures highlights the potential for effective *T. annulata* vaccines. Attenuated vaccines against *T. annulata* have been used in many countries where tropical theileriosis is endemic (Brown, 1990; Darghouth et al., 1999; Tait and Hall, 1990). However, there are practical constraints to the widespread use of live *T. annulata* vaccines in control programmes: (i) a requirement for distribution in liquid nitrogen, which accounts for approximately 30% of the cost (Bouslikhane et al., 1998); (ii) the need to use the vaccine immediately after thawing; (iii) the difficulty and cost of quality control to ensure consistent efficacy and the absence of other pathogens; (iv) problems with post-vaccination reactions, which have been recorded in 3% of animals immunised with Chinese, Moroccan, Iranian or Tunisian stocks (Darghouth, 2008); and (v) the possibility of reversion to virulence. To overcome these problems, research for the development of a killed subunit vaccine against *T. annulata* has focused on antigens present on the surface of sporozoite and merozoite stages of the parasite. High levels of genetic diversity are a major hurdle for the development of subunit vaccines for use in different host species; for example,

where cross-protection between cattle- and buffalo-derived *T. parva* strains may be incomplete (Young et al., 1973). Studies of *T. parva* in African buffalo have shown high genetic diversity when compared to cattle (Oura et al., 2005); but while *T. annulata* can infect both cattle and Asian buffalo (Nourollahi-Fard et al., 2015), genetic comparisons between strains infecting sympatric host populations have not been reported. Understanding of allelic variations of buffalo- and cattle-derived *T. annulata* and their circulation is needed to inform antigenicity and immunogenicity in vaccine development.

Control of *T. annulata* is heavily dependent on the prophylactic and therapeutic use of a single anti-protozoal drug, buparvaquone, but is now compromised by the emergence of resistance (Mhadhbi et al., 2015). Buparvaquone resistance has been reported with increasing frequency and now represents a serious challenge to efficient livestock production (Mhadhbi et al., 2010). Current understanding of the mechanisms and genetic basis of buparvaquone resistance in *T. annulata* is limited (Chatanga et al., 2019; Mhadhbi et al., 2015; Sharifiyazdi et al., 2012), with no functional link between mutations in the candidate cytochrome b locus and resistance phenotype having been demonstrated. Although previous studies have described multiple genotypes and high levels of genetic diversity (Al-Hamidhi et al., 2015; Weir et al., 2011), there is a lack of information on changes in response to drug selection pressure, fitness costs of mutations, and gene flow. It is, therefore, important to understand the genetic diversity and population sub-structure of *T. annulata*, with reference to the emergence and spread of buparvaquone resistance mutations. There is a need for a better understanding of the transmission of different *T. annulata* genotypes present in individual cattle and buffalo hosts as a proxy for multiplicity of infection when considering sustainable control strategies.

Population genetic studies of *T. annulata* have previously been performed using panels of satellite makers (Weir et al., 2007), but there are no reports based on mitochondrial cytochrome b sequence analysis using a post-genomic next-generation sequencing approach. Deep amplicon sequencing of metabarcoded mtDNA affords a practical and rhigh-throughput method when compared to conventional Sanger sequencing, to investigate genetic diversity between and within parasite populations. The Illumina MiSeq platform can provide 100,000 or more reads of up to 600 bp of loci of interest, depending on primer design. Recently, we have used these methods to study the population genetics of *Calicophoron daubneyi* (Sargison et al., 2019) and *Fasciola gigantica* infection in livestock keep in the United Kingdom and Pakistan.

In this paper, we describe the use of a mitochondrial cytochrome b locus and six polymorphic satellite markers investigate the population genetics of buffalo- and cattle-derived *T. annulata*. The results may help to: inform future rational vaccine design for the control of tropical theileriosis; predict the likely emergence and spread of genetic adaptations such as buparvaquone resistance; and identify patterns of infection with different genotypes.

2. Materials and Methods

2.1. Parasite resources and gDNA isolation

Blood samples were collected from the piroplasm-positive cattle and buffalo in veterinary clinics throughout the Punjab province of Pakistan between 2017 and 2019. The procedures involved jugular venipuncture and withdrawal of 5 ml of intravenous blood into EDTA tubes, followed by storage at -20^{o}C. Samples were collected by para-veterinary staff under the supervision of local veterinarians following consent from the animal owners. The study was approved by the Institutional Review Board of the University of Veterinary and Animal Sciences (UVAS-24817). Peripheral blood smears were prepared and stained with 4% Giemsa, then visualised under 1000x oil immersion to detect piroplasm infection. 'Haemoprotobiome' sequencing was performed on piroplasm-positive blood samples to confirm the presence of *T. annulata* (Chaudhry et al., 2019). Genomic DNA was isolated from *T. annulata* positive samples by lysis with GS buffer and proteinase K as described in the TIANamp Blood DNA Kit (TIANGEN Biotech Co. Ltd, Beijing) and stored at -20°C.

Four cell-line stocks of *T. annulata* positive controls originally derived from Turkey, India, Tunisia, and Morocco (Katzer et al., 1994) were available at the University of Edinburgh Roslin Institute, UK.

2.2. Adapter and barcoded PCR amplification of T. annulata cytochrome b

A 517 bp region of the cytochrome b locus was selected for deep amplicon sequencing. One µl gDNA of each of 31 buffalo, 54 cattle and the 4 positive control derived *T. annulata* samples was used as templates for the 1st round adapter PCR amplification. A *de novo* primer set (Supplementary Table S1) was used under the following conditions: 13.25 µl ddH_2O, 1 µl gDNA, 0.75 µl of 10 mM dNTPs mix, 5 µl of 5X HiFi Fidelity Buffer, 0.5 µl of 0.5U DNA polymerase enzyme (KAPA Biosystems, USA), and 0.75 µl of 10 µM forward adaptor primers and reverse

adaptor primers. The thermocycling conditions were: 95°C for 2 min, 35 cycles at 98°C for 20 sec, 55°C for 15 sec, and 72°C for 2 min, followed by a final extension of 72°C for 2 min. The PCR product was purified with AMPure XP Magnetic Beads (1X) (Beckman coulter Inc., USA). A barcoded primer set was used in the 2nd round of PCR amplification to add a fragment of unique sequence into each purified product (Supplementary Table S2) under the following conditions: 13.25 µl ddH_2O, 2 µl bead purified PCR products, 0.75 µl of 10 mM dNTPs mix, 5 µl of 5X HiFi Fidelity Buffer, 0.75 µl of 0.5U DNA polymerase enzyme (KAPA Biosystems, USA), and 1.25 µl of 10µM forward and reverse primers. The thermocycling conditions were: 98°C for 2 min, 7 cycles at 98°C for 20 sec, 63°C for 20 sec, and 72°C for 2 min. The PCR products were purified with AMPure XP Magnetic Beads (1X) (Beckman coulter Inc., USA).

2.3. Illumina Mi-Seq run and data handling

A pooled library was prepared with 10 µl of barcoded bead purification PCR product from each *T. annulata* sample and sent to Edinburgh Genomics, UK for deep amplicon sequencing. The size of the amplicon was measured by qPCR library quantification (KAPA Biosystems, USA), before running on an Illumina Mi-Seq sequencer using a 600-cycle pair-end reagent kit (Mi-Seq Reagent Kits v2, MS-103-2003) at a concentration of 15 nM with addition of 15% Phix control v3 (Illumina, FC-11-2003). The Illumina Mi-Seq post-run processing uses the barcoded indices to split all sequences by sample and generate FASTQ files. These were analysed using Mothur v1.39.5 software (Schloss et al., 2009) with modifications in the standard operating procedures of Illumina Mi-Seq (Kozich et al., 2013) in the Command Prompt pipeline. Briefly, the raw paired read-ends were run into the 'make.contigs' command to combine the two sets of reads for each sample. The command extracted sequence and quality score data from the FASTQ files, creating the complement of the reverse and forward reads, and then joining the reads into contigs. After removing the too long, or ambiguous sequence reads, the data were aligned with the *T. annulata* cytochrome b reference sequence library prepared from the positive controls (for more details Supplementary Data S1 and section 2.1) using the 'align.seqs' command. Any sequences that did not match with the *T. annulata* cytochrome b reference library were removed and the 'summary.seqs' command was used to summarise the 517 bp sequence reads of the *T. annulata* cytochrome b locus. The sequence reads were further run on the 'screen.seqs' command to generate the *T. annulata* cytochrome b FASTQ file. Once the sequence reads were classified as *T. annulata*, a count list of the consensus sequences of each population was created using the

'unique.seqs' command. The count list was further used to create FASTQ files of the consensus sequences of each population using the 'split.groups' command (for more details Supplementary Data S2).

2.4. Bioinformatics data analysis

The consensus sequences of *T. annulata* cytochrome b locus were aligned using the MUSCLE alignment tool in Geneious v10.2.5 software (Biomatters Ltd, New Zealand) and then imported into the FaBox 1.5 online tool (birc.au.dk) to collapse the sequences that showed 100% base pair similarity after corrections into a single genotype. The genotype frequency of each sample was calculated by dividing the number of sequence reads by the total number of reads. A split tree was created in the SplitTrees4 software (bio-soft.net) by using the neighbour-joining method in the JukesCantor model of substitution. The appropriate model of nucleotide substitutions for neighbour-joining analysis was selected by using the jModeltest 12.2.0 program (Posada, 2008). The tree was rooted with the corresponding cytochrome b sequence of *T. parva*. The branch supports were obtained by 1000 bootstraps of the data. The genetic diversity of cytochrome b was calculated within and between populations by using the DnaSP 5.10 software package (Librado and Rozas, 2009), and the following values were obtained: Heterozygosity (H_e), the number of segregating sites (S), nucleotide diversity (π), the mean number of pairwise differences (k), the mutation parameter based on an infinite site equilibrium model, and the mutations parameter based on segregating sites (S_θ).

2.5. Satellite genotyping and bioinformatics data analysis of T. annulata

Six satellite markers (TS6, TS8, TS20, TS31, TS12, TS16) were selected for use on each population as previously described by Weir et al. (2007). A summary of primer sequences and allele ranges is given in Supplementary Table S3. Eighteen buffalo and 35 cattle derived *T. annulata* populations were successfully analysed from six markers. PCR amplification was performed using a master mix containing 17.2 µl ddH_2O, 0.3 µl of 100 µM dNTPs, 2.5 µl of 1X thermopol reaction buffer, 0.3 µl of 1.25U Taq DNA polymerase (New England Biolabs) and 0.3 µl of 0.1µM forward and reverse primers. The thermo-cycling parameters were 94^{o}C for 2 min followed by 35 cycles of 94^{o}C for 1 min, staying at the annealing temperatures of 60^{o}C (TS6, TS20), 55^{o}C (TS8, TS12, TS16), or 50^{o}C (TS31) for 1 min and 65^{o}C for 1 min, with a single final extension cycle of 65^{o}C for 5 min. The forward primer of each microsatellite primer pair was 5'

end labelled with fluorescent dye (IDT, UK), and the ROX 400 internal size standard was used on the ABI Prism 3100 genetic analyser (Applied Biosystems, USA).

Individual chromatograms were analysed using Peak Scanner software version 2.0 (Thermo Fisher Scientific, USA) to determine the size of the alleles. These were combined across six markers to generate a multilocus genotype (MLG). From the MLG data, allelic variation (A) and heterozygosity (H_e) for individual populations were calculated using Arlequin 3.11 software (Guo and Thompson, 1992). Significance levels were calculated using the sequential method of Bonferroni correction for multiple comparisons in the same dataset (Rice, 1989). Analysis of molecular variance (AMOVA) was estimated through the partition of satellite diversity between and within populations (Excoffier et al., 1992) and fixation index (pairwise F_{ST}) values were calculated using Arlequin 3.11 to provide a measurement of population genetic sub-structure. Principal coordinate analysis (PCoA) was performed using GenALEx software to illustrate the extent of genetically distinct features of individual populations with plot coordinates (Peakall and Smouse, 2012).

The likelihood ratio test statistics (G-test) were calculated using Arlequin 3.11 software (Excoffier et al., 2005) to estimate genetic linkage equilibrium and deviations from Hardy-Weinberg equilibrium. There was no evidence to support linkage disequilibrium for any combination of satellite loci in the individual *T. annulata* populations of buffalo and cattle, indicating that alleles at these loci were randomly associating and not genetically linked (data on file). There was some significant departure from Hardy-Weinberg equilibrium, even after Bonferroni correction, in addition to relative P values for 108 loci combinations for buffalo and 210 loci combinations for cattle *T. annulata* populations. The bottleneck software (version 1.2.02) was, therefore, used to search for the evidence of heterozygosity excess and mode-shift. The Wilcoson signed-rank test was used to evaluate the statistical significance of any possible genetic drift equilibrium (Lefterova et al., 2015). The data showed that there was no heterozygosity excess according to the Sign Test and Wilcoxon Test (data on file). The mode shift analysis established most populations had a normal L-shaped distribution.

3. Results

3.1. Allelic diversity between cattle- and buffalo- derived T. annulata populations

The allelic diversity data show that *T. annulata* is more diverse in cattle as compared to water buffalo. The mean numbers (±SD) of satellite alleles in the cattle- and buffalo- derived populations were 34.7 (±9.1) and 23.7 (±5.2), respectively. The number of alleles (A_n) for each satellite marker in the cattle- and buffalo-derived populations ranged from 22 to 46 and 17 to 32, respectively. The mean numbers (±SD) of cattle- and buffalo-specific *T. annulata* satellite alleles (A_u) per marker were 13.3 (±6.3) and 1.8 (±2.04), respectively (Table 1 and Fig. 1). Seventy-nine and 41 cytochrome b locus alleles (A_n) were identified in the cattle- and buffalo-derived *T. annulata* populations. The numbers of cattle- and buffalo-specific *T. annulata* alleles (A_u) were 65 and 27, respectively (Table 1 and Fig. 1).

3.2. Genetic diversity within T. annulata populations derived from cattle and buffalo

Overall, the mean heterozygosities (H_e) (±SD) of the cattle- and buffalo-derived *T. annulata* populations were 0.912 (±0.056) and 0.874 (±0.028) respectively (Table 2). The heterozygosity (H_e) data reveal high levels of genetic diversity within each population of *T. annulata*; with mean values (±SD) for the six satellite markers ranging from 0.8 (±0.082) to 0.959 (±0.017) in individual cattle- and 0.738 (±0.196) to 0.956 (±0.034) in buffalo-derived populations (Table 2).

High levels of heterozygosity (H_e) were seen at the *T. annulata* cytochrome b locus, ranging from 0.556 to 0.923 (overall mean value 0.931) in cattle-derived populations and from 0.545 to 0.900 (overall mean value 0.861) in buffalo-derived populations (Table 3).

3.3. Genetic sub-structure between T. annulata populations derived from cattle and buffalo

Genetic differentiation was shown by comparing the fixation indices (F_{ST}) of two populations with each other in a pairwise manner. The F_{ST} values indicated a low level of genetic differentiation between *T. annulata* populations ranging from 0.001 to 0.114 in cattle and 0.003 to 0.132 buffalo, respectively (Supplementary Table S4).

An AMOVA was conducted using the panel of six satellite markers to estimate the genetic variation within and between the populations. This showed that genetic variation was distributed 97.91% and 97% and within; and 2.09% and 3% between cattle- and buffalo-derived *T. annulata* populations, respectively (data on file). PCoA was performed to illustrate as a two-dimensional plot the extent to which populations are genetically distinct. The two axes accounted for 72.5% (25.67 + 46.83) and 73.93% (27.6 + 46.33) of the variation in the cattle- and buffalo-derived

303 populations, respectively; and showed that populations from different regions formed overlapping
304 clusters, hence were not geographically sub-structured (Fig. 2).
305
306 *3.4. Genotype distribution of T. annulata populations derived from cattle and buffalo*
307 *T. annulata* satellite data were found to be highly polymorphic in each population, with overall
308 numbers of genotypes per population ranging from 2 to 26 in cattle and 2 to 17 in buffalo. The
309 mean numbers (±SD) of *T. annulata* genotypes in cattle- and buffalo-derived populations were
310 8.60 (±2.52) and 6.25, respectively (Table 4).
311 Seventy-nine *T. annulata* cytochrome b genotypes from individual cattle-derived populations
312 were analysed separately. Fifteen populations had a single genotype at high frequencies ranging
313 from 90 to 100%. These comprised of ten populations containing only one genotype; two
314 populations containing two genotypes; one population with three genotypes; one population with
315 four genotypes; and one population with seven genotypes (Fig. 3A). Thirty-nine populations had
316 multiple genotypes at high frequencies. These comprised of five populations containing two
317 genotypes; eight populations containing three genotypes; ten population containing four genotypes;
318 seven populations with five genotypes; four populations with six genotypes; one population with
319 seven genotypes; two populations with eight genotypes; one populations with 12 genotypes; and
320 one with 14 genotypes (Fig. 3A). Phylogenetic analysis of the 79 cytochrome b genotypes of the
321 cattle-derived *T. annulata* populations, showed that two genotypes for 22.9% and 10.8% of the
322 total number of the sequence reads, being present in 36 and 8, populations, respectively. Seven
323 genotypes accounted for between 5.0% and 7.0% of the sequence reads and 45 genotypes
324 accounted for less than 1.0% of the sequence reads (Fig. 4A). The split tree shows that eight
325 genotypes are shared between populations derived from cattle holdings in multiple locations of
326 Lahore, Chakwal, Gujranwala, Okara, and Sahiwal. Seven genotypes are shared between
327 populations derived from cattle holdings in any two locations of Lahore, Gujranwala, Okara, and
328 Sahiwal. Sixty-one genotypes are present in a single location (Gujranwala=24; Qadirabad=12;
329 Okara =14; Lahore=11 genotypes) (Fig. 4A).
330 Forty-one *T. annulata* cytochrome b genotypes from individual buffalo-derived populations
331 were analysed separately. A single genotype predominated in 15 populations at a frequency of
332 between 90 and 100%. These comprised of eight populations containing only one genotype; three
333 populations containing two genotypes; one population with three genotypes; one population with
334 five genotypes; one population with six genotypes; and one population with seven genotypes (Fig.

335 3B). Sixteen populations had high frequencies of multiple genotypes. These comprised of seven
336 populations contained 2 genotypes; two populations contained 3 genotypes; one population
337 contained 4 genotypes; three populations had 5 genotypes; two populations had 7 genotypes and
338 one population contained a maximum of 8 genotypes (Fig. 3B). Phylogenetic analysis of the 41
339 cytochrome b genotypes of the buffalo-derived *T. annulata* populations, showed that three
340 genotypes accounted for 39.1%, 10.8%, and 5.5% of sequence reads being present in 26, 15 and 3
341 populations, respectively. Eighteen genotypes accounted for between 1.0% and 3.0% of sequence
342 reads and 20 genotypes accounted for less than 1.0% of sequence reads (Fig. 4B). The split tree
343 shows that two genotypes are shared between *T. annulata* populations derived from buffalo
344 holdings in multiple locations of Gujranwala, Hafizabad, Okara, and Sahiwal. Five genotypes are
345 present in the populations derived from buffalo holdings in Gujranwala and Okara. One genotype
346 is present in the populations derived from buffalo holdings in Hafizabad and Okara and one
347 genotype from Okara and Sahiwal. Thirty-two genotypes are present in a single location (16
348 genotypes in Gujranwala; 13 genotypes in Okara; and 3 in Hafizabad) (Fig. 4B).
349
350 **4. Discussion**
351 *Theleria annulata* is considered to be the most economically important protozoan parasite of
352 water buffalo and cattle in Asia and North Africa, causing high mortality and morbidity. The
353 annual economic impact of the pathogen is estimated to be billions of USD, with a significant
354 impact on both meat and milk production (Jabbar et al., 2015). Our results reveal a high level of
355 genetic diversity within *T. annulata* infecting individual hosts, and between different host-derived
356 populations; being greater in cattle- than in water buffalo-derived populations. *Theileria annulata*
357 genotype distribution in the Punjab province of Pakistan is consistent with high parasite
358 transmission rates and frequent animal movements between different regions. These results may
359 have practical implications for the spread of parasite genetic adaptations.
360
361 *4.1. Potential impact of T. annulata allelic diversity on immunisation cross-protection*
362 The satellite and cytochrome b data reported in the present study show that *T. annulata* is
363 genetically more diverse with more circulation of alleles in cattle as compared to water buffalo.
364 The satellite data were informative due to the numbers of alleles present, showing more unique
365 alleles in cattle-derived than in Asian water buffalo-derived *T. annulata* populations. The
366 cytochrome b locus provided further evidence of allelic diversity and differences between cattle-

and buffalo- populations. Similarly high levels of parasite heterogeneity have been reported in *T. parva*, albeit with higher allelic diversity in African buffalo- than in cattle-derived parasite populations (Oura et al., 2011). Such allelic variation may reflect the capacity of parasite to undergo transmission. For example, the differences between *T. annulata* and *T. parva* could have arisen if the buffalo-derived *T. annulata* transmission is more restricted than transmission between cattle-derived populations in the Punjab province of Pakistan; while *T. parva* transmission may be more efficient between infected buffalo than cattle in endemic east coast fever regions of Africa (Oura et al., 2011).

The high level of allelic variation described in *T. annulata* might have practical implications for the development of a sporozoite and merozoite antigen subunit vaccine. Immunisation of cattle with different subunit versions of the sporozoite stage surface protein (SPAG1, TaSP) has provided a degree of protection against cattle-derived *T. annulata*, probably through a reduction in the level of parasitaemia (Schnittger et al., 2002). However, little is known about the antigenic diversity of *T. annulata* candidate loci, which may be important in vaccine development (MacHugh et al., 2011), or about the impact of differences between cattle- and buffalo-derived populations. Our results highlight the need for better understanding of antigenic diversity in the development of a subunit vaccine for tropical theileriosis, capable of providing cross-protection against cattle- and buffalo-derived *T. annulata* genotypes across different regions.

4.2. Potential impact of T. annulata genetic diversity on the emergence of buparvaquone resistance

Our satellite and cytochrome b locus data show high levels of genetic diversity in both cattle- and buffalo-derived *T. annulata* populations, reflecting high effective population sizes (Al-Hamidhi et al., 2015) and implying high mutation rates across the *T. annulata* genome. The 8.35 Mb nuclear genome sequence of *T. annulata* spans four chromosomes that range from 1.9 to 2.6 Mb, with 3,792 putative protein-coding genes. In addition, a total of 49 transfer RNA and five ribosomal RNA genes were identified (Pain et al. (2005). The high level of genetic diversity in *T. annulata* may practical implications for the emergence of antiprotozoal drug resistance mutations. Only one drug, buparvaquone, is available for the treatment of tropical theileriosis: Pakistani livestock are treated frequently with buparvaquone and anecdotal evidence suggests that resistance is widespread.

399 *4.3. Potential impact of the multiplicity of T. annulata infection on control strategies for tropical*
400 *theileriosis*

401 The presence of a single cytochrome b genotype in 15 cattle and 9 buffalo- derived populations
402 suggests a single emergence and subsequent spread of *T. annulata* infection. The satellite data
403 provide further evidence of low levels of genetic differentiation among cattle-and buffalo-derived
404 populations and show overlapping clusters that are not geographically sub-structured, consistent
405 with the high levels of gene flow due to livestock movements or translocations of ticks to a new
406 region. This could influence the spread of drug-resistant alleles. Studies of the global genetic sub-
407 structure of *T. annulata* have shown a high level of genetic differentiation between Turkey and
408 Tunisia (Weir et al., 2007). Studies have also shown a low level of genetic differentiation in *T.*
409 *annulata* within-countries including Turkey, Tunisia, Oman, China and Portugal (Al-Hamidhi et
410 al., 2015; Gomes et al., 2016; Weir et al., 2011; Yin et al., 2018), reflecting high levels of animal
411 movement, or translocation of ticks.

412 The predominance and the high proportions of multiple cytochrome b genotypes in 39 cattle-
413 and 16 buffalo-derived populations implies multiple emergence of infection. The numbers of
414 satellite genotypes per population provide further evidence for this pattern of emergence of
415 different genetically adapted strains. Random cross mating of gametes and genetic recombination
416 of *T. annulata* in ticks gives rise to the formation of new genotypes (Al-Hamidhi et al., 2015). The
417 multiplicity of infection of *T. annulata* may be influenced by variations in the intensity of
418 transmission due to levels of different tick species infestation, or prevalence of tick infection (Yin
419 et al., 2018). *Hyalomma scupense* is the most common and economically important species and
420 the major vector of *T. annulata*, but *Hyalomma marginatum* and *Hyalomma anatolicum* are also
421 found in Tunisia (Bouattour et al., 1996). There are four *Hyalomma* species responsible for the
422 transmission of *T. annulata* in Turkey, of which *H. anatolicum* is the major vector for *T. annulata*
423 (Aktas et al., 2004). Two tick species, *Hyalomma lusitanicum*, and *H. marginatum* are the vectors
424 of *T. annulata* in Portugal (Estrada-Pena and Santos-Silva, 2005). There are a few reports that *H.*
425 *anatolicum* may be a vector of *T. annulata* in the Punjab province of Pakistan (Karim et al., 2017).
426 The multiplicity of *T. annulata* infection may also be influenced by population bottlenecking
427 effects, arising from seasonal effects of climatic conditions on tick transmission and completion of
428 the parasite's lifecycle. The *T. annulata* transmission season in Turkey is between May and
429 September, with the peak of clinical cases occurring in mid-summer (Sayin et al., 2003). In a
430 region of endemic stability of tropical theileriosis in Tunisia, there is a high level of multiplicity

of *T. annulata* infection, but clinical disease is rare; while in a region of endemic instability, a proportion of the population becomes infected all year round, and the clinical disease occurs particularly in adult cattle (Gharbi et al., 2011).

Control strategies for tropical theileriosis need to consider factors such as: the reproductive isolation of parasite populations; management of host movements; control of tick vectors; mitigation of the impacts of climate change; and consequences of parasite exposure to antiprotozoal drugs. Understanding of the multiplicity of infection and of high levels of gene flow is, therefore, important in the educational dissemination and implementation of advice on sustainable parasite control.

Acknowledgement

The study was financially supported by the Carnegie Trust Scotland and Biotechnology and Biological Sciences Research Council (BBSRC). Work at the University of Veterinary and Animal Science Pakistan uses facilities funded by the Higher Education Commission of Pakistan. The authors of this study would like to thank Professor Dr. Masood Rabbani, Vice-Chancellor of the University of Veterinary and Animal Science Lahore Pakistan, for his great support in the arrangements for sample collections.

Conflict of interest

None

References

Aktas, M., Dumanli, N., Angin, M., 2004. Cattle infestation by Hyalomma ticks and prevalence of Theileria in Hyalomma species in the east of Turkey. Vet Parasitol 119, 1-8.

Al-Hamidhi, S., M, H.T., Weir, W., Al-Fahdi, A., Johnson, E.H., Bobade, P., Alqamashoui, B., Beja-Pereira, A., Thompson, J., Kinnaird, J., Shiels, B., Tait, A., Babiker, H., 2015. Genetic Diversity and Population Structure of Theileria annulata in Oman. PLoS One 10, e0139581.

Al-Hosary, A., Abdel-Rady, A., Ahmed, L.S., Mohamed, A., 2010. Comparison between Using of BUPAQUONE ® and Other Compounds in Treatment of Bovine Theileriosis. International Journal for Agro Veterinary and Medical Sciences 4, 3-7.

Bouattour, A., Darghouth, M.A., Ben Miled, L., 1996. Cattle infestation by Hyalomma ticks and prevalence of Theileria in H. detritum species in Tunisia. Vet Parasitol 65, 233-245.

Bouslikhane, M., M., K., H., O. 1998. La theiléríose bovine au Maroc, investigations épidémiologiques et étude de l'impact sur la productivité des élevages. In: Résumé de communication orale, 15 ème congrès vétérinaire maghrébin.

469 Brown, C.G., 1990. Control of tropical theileriosis (Theileria annulata infection) of cattle.
470 Parassitologia 32, 23-31.
471 Chatanga, E., Mosssad, E., Abdo Abubaker, H., Amin Alnour, S., Katakura, K., Nakao, R., Salim,
472 B., 2019. Evidence of multiple point mutations in Theileria annulata cytochrome b gene
473 incriminated in buparvaquone treatment failure. Acta Trop 191, 128-132.
474 Chaudhry, U., Ali, Q., Rashid, I., Shabbir, M.Z., Abbas, M., Numan, M., Evans, M., Ashraf, K.,
475 Morrison, I., Morrison, L., Sargison, N.D., 2019. Development of a deep amplicon
476 sequencing method to determine the proportional species composition of piroplasm
477 haemoprotozoa as an aid in their control. bioRxiv.
478 Darghouth, M.A., 2008. Review on the experience with live attenuated vaccines against tropical
479 theileriosis in Tunisia: considerations for the present and implications for the future.
480 Vaccine 26 Suppl 6, G4-G10.
481 Darghouth, M.A., Bouattour, A., Kilan, M., 1999. Tropical theileriosis in Tunisia: epidemiology
482 and control. Parassitologia 41 Suppl 1, 33-36.
483 Estrada-Pena, A., Santos-Silva, M.M., 2005. The distribution of ticks (Acari: Ixodidae) of
484 domestic livestock in Portugal. Experimental & applied acarology 36, 233-246.
485 Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software
486 package for population genetics data analysis. Evolutionary bioinformatics online 1, 47-50.
487 Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from
488 metric distances among DNA haplotypes: application to human mitochondrial DNA
489 restriction data. Genetics 131, 479-491.
490 Gharbi, M., Darghouth, M.A., 2015. Control of tropical theileriosis (Theileria annulata infection in
491 cattle) in North Africa. Asian Pacific Journal of Tropical Disease 5, 505-510.
492 Gharbi, M., Touay, A., Khayeche, M., Laarif, J., Jedidi, M., Sassi, L., Darghouth, M.A., 2011.
493 Ranking control options for tropical theileriosis in at-risk dairy cattle in Tunisia, using
494 benefit-cost analysis. Revue scientifique et technique (International Office of Epizootics)
495 30, 763-778.
496 Gomes, J., Salgueiro, P., Inacio, J., Amaro, A., Pinto, J., Tait, A., Shiels, B., Pereira da Fonseca, I.,
497 Santos-Gomes, G., Weir, W., 2016. Population diversity of Theileria annulata in Portugal.
498 Infect Genet Evol 42, 14-19.
499 Guo, S.W., Thompson, E.A., 1992. Performing the exact test of Hardy-Weinberg proportion for
500 multiple alleles. Biometrics 48, 361-372.
501 Jabbar, A., Abbas, T., Sandhu, Z.U., Saddiqi, H.A., Qamar, M.F., Gasser, R.B., 2015. Tick-borne
502 diseases of bovines in Pakistan: major scope for future research and improved control.
503 Parasites & vectors 8, 283.
504 Karim, S., Budachetri, K., Mukherjee, N., Williams, J., Kausar, A., Hassan, M.J., Adamson, S.,
505 Dowd, S.E., Apanskevich, D., Arijo, A., Sindhu, Z.U., Kakar, M.A., Khan, R.M.D., Ullah,
506 S., Sajid, M.S., Ali, A., Iqbal, Z., 2017. A study of ticks and tick-borne livestock pathogens
507 in Pakistan. PLoS neglected tropical diseases 11, e0005681.
508 Katzer, F., Carrington, M., Knight, P., Williamson, S., Tait, A., Morrison, I.W., Hall, R., 1994.
509 Polymorphism of SPAG-1, a candidate antigen for inclusion in a sub-unit vaccine against
510 Theileria annulata. Mol Biochem Parasitol 67, 1-10.
511 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of
512 a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence
513 data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112-5120.
514 Lawrence, J.A., 1979. The differential diagnosis of the bovine theilerias of Southern Africa.
515 Journal of the South African Veterinary Association 50, 311-313.

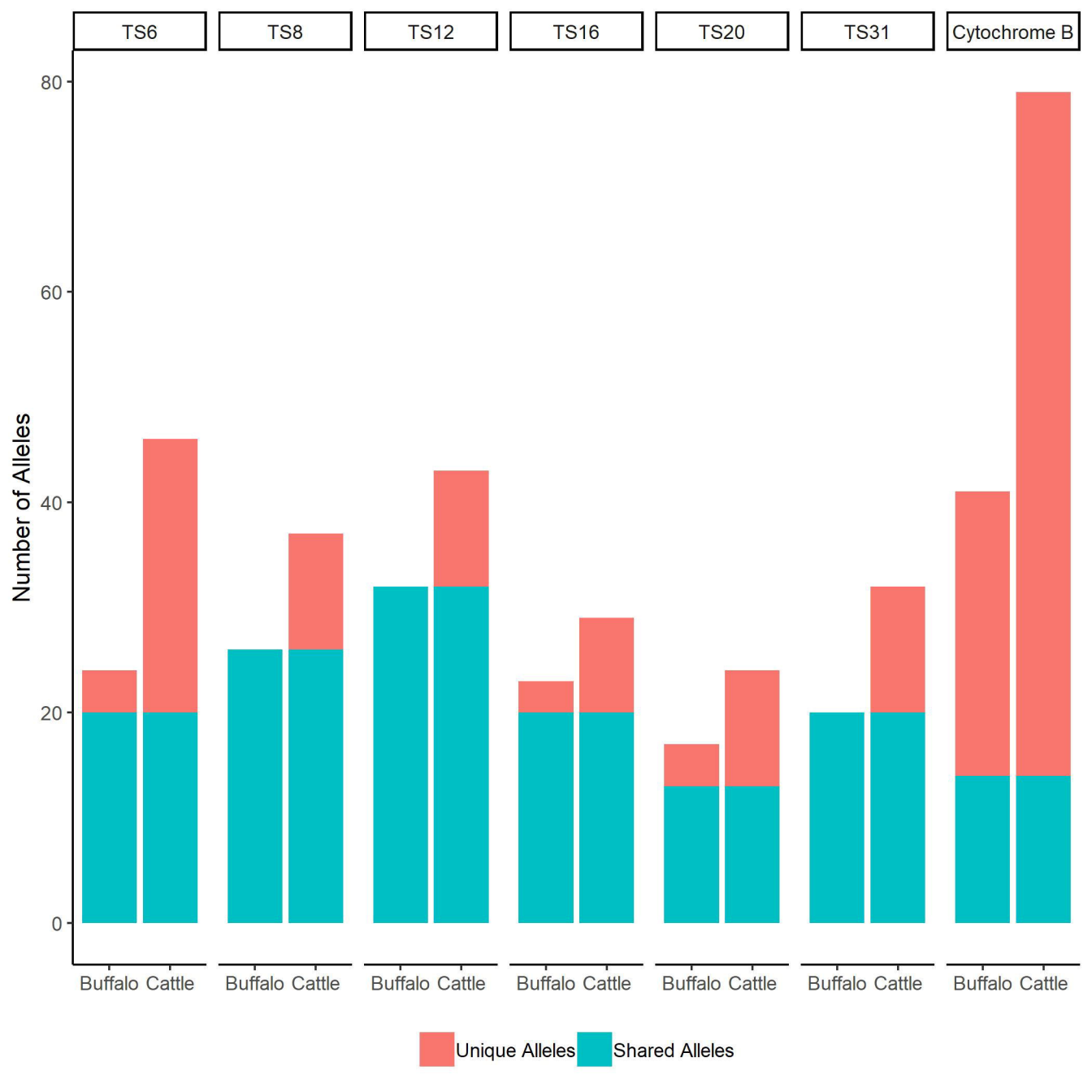
516 Lefterova, M.I., Budvytiene, I., Sandlund, J., Farnert, A., Banaei, N., 2015. Simple Real-Time
517 PCR and Amplicon Sequencing Method for Identification of Plasmodium Species in
518 Human Whole Blood. J Clin Microbiol 53, 2251-2257.
519 Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA
520 polymorphism data. Bioinformatics 25, 1451-1452.
521 MacHugh, N.D., Weir, W., Burrells, A., Lizundia, R., Graham, S.P., Taracha, E.L., Shiels, B.R.,
522 Langsley, G., Morrison, W.I., 2011. Extensive polymorphism and evidence of immune
523 selection in a highly dominant antigen recognized by bovine CD8 T cells specific for
524 Theileria annulata. Infection and immunity 79, 2059-2069.
525 Mahmmod, Y.S., Elbalkemy, F.A., Klaas, I.C., Elmekkawy, M.F., Monazie, A.M., 2011. Clinical
526 and haematological study on water buffaloes (Bubalus bubalis) and crossbred cattle
527 naturally infected with Theileria annulata in Sharkia province, Egypt. Ticks Tick Borne
528 Dis 2, 168-171.
529 Mhadhbi, M., Chaouch, M., Ajroud, K., Darghouth, M.A., BenAbderrazak, S., 2015. Sequence
530 Polymorphism of Cytochrome b Gene in Theileria annulata Tunisian Isolates and Its
531 Association with Buparvaquone Treatment Failure. PloS one 10, e0129678.
532 Mhadhbi, M., Naouach, A., Boumiza, A., Chaabani, M.F., BenAbderazzak, S., Darghouth, M.A.,
533 2010. In vivo evidence for the resistance of Theileria annulata to buparvaquone. Vet
534 Parasitol 169, 241-247.
535 Nourollahi-Fard, S.R., Khalili, M., Ghalekhani, N., 2015. Detection of Theileria annulata in blood
536 samples of native cattle by PCR and smear method in Southeast of Iran. Journal of
537 parasitic diseases : official organ of the Indian Society for Parasitology 39, 249-252.
538 Oura, C.A., Asiimwe, B.B., Weir, W., Lubega, G.W., Tait, A., 2005. Population genetic analysis
539 and sub-structuring of Theileria parva in Uganda. Mol Biochem Parasitol 140, 229-239.
540 Oura, C.A., Tait, A., Asiimwe, B., Lubega, G.W., Weir, W., 2011. Haemoparasite prevalence and
541 Theileria parva strain diversity in Cape buffalo (Syncerus caffer) in Uganda. Vet Parasitol
542 175, 212-219.
543 Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C.A., Weir, W., Kerhornou, A., Aslett,
544 M., Bishop, R., Bouchier, C., Cochet, M., Coulson, R.M., Cronin, A., de Villiers, E.P.,
545 Fraser, A., Fosker, N., Gardner, M., Goble, A., Griffiths-Jones, S., Harris, D.E., Katzer, F.,
546 Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neil, S.,
547 Price, C., Quail, M.A., Rabbinowitsch, E., Rawlings, N.D., Rutter, S., Saunders, D., Seeger,
548 K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, A.R., Woodward, J., Dobbelaere,
549 D.A., Langsley, G., Rajandream, M.A., McKeever, D., Shiels, B., Tait, A., Barrell, B.,
550 Hall, N., 2005. Genome of the host-cell transforming parasite Theileria annulata compared
551 with T. parva. Science 309, 131-133.
552 Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic
553 software for teaching and research--an update. Bioinformatics 28, 2537-2539.
554 Posada, D., 2008. jModelTest: phylogenetic model averaging. Molecular biology and evolution 25,
555 1253-1256.
556 Rice, W., 1989. Analyzing tables of statistical tests. Evolution 43.
557 Sargison, N.D., Shahzad, K., Mazeri, S., Chaudhry, U., 2019. A high throughput deep amplicon
558 sequencing method to show the emergence and spread of Calicophoron daubneyi rumen
559 fluke infection in United Kingdom cattle herds. Veterinary parasitology.
560 Sayin, F., Dincer, S., Karaer, Z., Cakmak, A., Inci, A., Yukari, B.A., Eren, H., Vatansever, Z.,
561 Nalbantoglu, S., 2003. Studies on the epidemiology of tropical theileriosis (Theileria
562 annulata infection) in cattle in Central Anatolia, Turkey. Tropical animal health and
563 production 35, 521-539.

564 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski,
565 R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., 2009. Introducing mothur: open-source,
566 platform-independent, community-supported software for describing and comparing
567 microbial communities. Appl. Environ. Microbiol. 75, 7537-7541.
568 Schnittger, L., Katzer, F., Biermann, R., Shayan, P., Boguslawski, K., McKellar, S., Beyer, D.,
569 Shiels, B.R., Ahmed, J.S., 2002. Characterization of a polymorphic Theileria annulata
570 surface protein (TaSP) closely related to PIM of Theileria parva: implications for use in
571 diagnostic tests and subunit vaccines. Molecular and biochemical parasitology 120, 247-
572 256.
573 Sharifiyazdi, H., Namazi, F., Oryan, A., Shahriari, R., Razavi, M., 2012. Point mutations in the
574 Theileria annulata cytochrome b gene is associated with buparvaquone treatment failure.
575 Vet Parasitol 187, 431-435.
576 Sivakumar, T., Hayashida, K., Sugimoto, C., Yokoyama, N., 2014. Evolution and genetic
577 diversity of Theileria. Infect Genet Evol 27, 250-263.
578 Tait, A., Hall, F.R., 1990. Theileria annulata: control measures, diagnosis and the potential use of
579 subunit vaccines. Rev Sci Tech 9, 387-403.
580 Uilenberg, G., Perie, N.M., Lawrence, J.A., de Vos, A.J., Paling, R.W., Spanjer, A.A., 1982.
581 Causal agents of bovine theileriosis in southern Africa. Tropical animal health and
582 production 14, 127-140.
583 Weir, W., Ben-Miled, L., Karagenc, T., Katzer, F., Darghouth, M., Shiels, B., Tait, A., 2007.
584 Genetic exchange and sub-structuring in Theileria annulata populations. Mol Biochem
585 Parasitol 154, 170-180.
586 Weir, W., Karagenc, T., Gharbi, M., Simuunza, M., Aypak, S., Aysul, N., Darghouth, M.A.,
587 Shiels, B., Tait, A., 2011. Population diversity and multiplicity of infection in Theileria
588 annulata. Int J Parasitol 41, 193-203.
589 Yin, F., Liu, Z., Liu, J., Liu, A., Salih, D.A., Li, Y., Liu, G., Luo, J., Guan, G., Yin, H., 2018.
590 Population Genetic Analysis of Theileria annulata from Six Geographical Regions in
591 China, Determined on the Basis of Micro- and Mini-satellite Markers. Frontiers in genetics
592 9, 50.
593 Young, A.S., Brown, C.G., Burridge, M.J., Cunningham, M.P., Kirimi, I.M., Irvin, A.D., 1973.
594 Observations on the cross-immunity between Theileria lawrencei (Serengeti) and Theileria
595 parva (Muguga) in cattle. Int J Parasitol 3, 723-728.
596

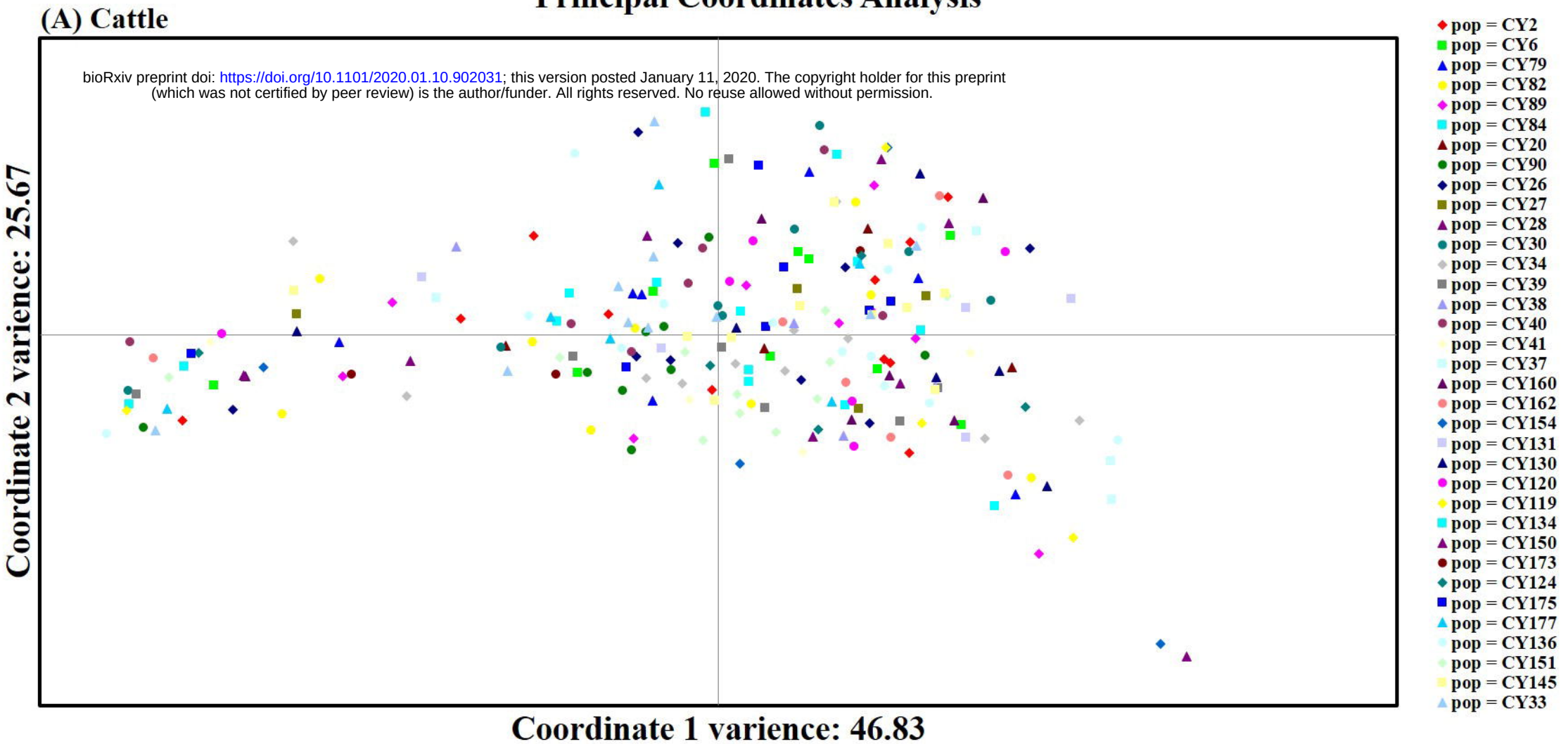
597 **Figure Legends**

598 **Fig. 1.** The mean number of alleles in *T. annulata* derived from buffalo and cattle based on a panel
599 of six satellite and and a cytochrome b marker. The bar of each marker shows the proportion of
600 alleles in buffalo and cattle. The Y-axis shows the number of alleles of each marker. The unique
601 and shared alleles of buffalo and cattle are represented by a different colour.
602
603 **Fig. 2.** Principal coordinate analysis using a panel of six satellite markers to represent 35 cattle-
604 and 18 buffalo-derived *T. annulata* populations.
605
606 **Fig. 3.** Relative genotype frequencies of the cytochrome b locus of 54 cattle-derived (2A) and 31
607 buffalo-derived (2B) *T. annulata* populations, collected from the Punjab province of Pakistan.
608 Each genotype (CHA and BHA) is represented by a different colour in the individual population
609 (CY). The population distribution and the frequency of sequence reads generated per population

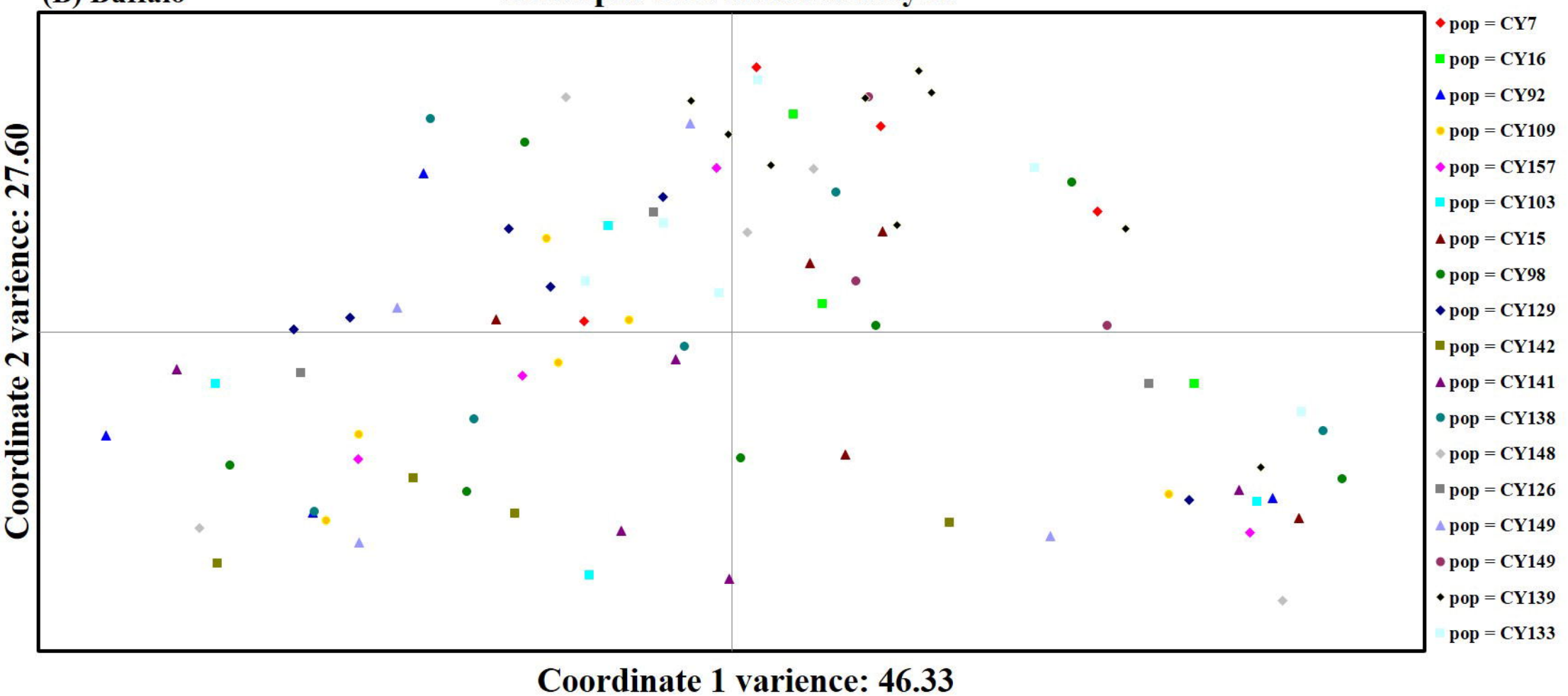
610 for each of the 79 cattle-derived and 41 buffalo-derived *T. annulata* genotypes is shown in the
611 insert table.
612
613 **Fig. 4.** Split tree analysis of 79 cattle-derived (3A) and 41 buffalo-derived (3B) *T. annulata*
614 genotypes collected from the Lahore (A), Gujranwala (B), Chakwal (C), Qadirabad (D), Okara (E),
615 Sahiwal (F), and Hafizabad (G) regions of the Punjab province of Pakistan. Each population (CY)
616 is represented by a different colour in the individual genotype (CHA and BHA). The pie chart
617 circles represent the distribution and percentage of sequence reads generated per genotype
618 identified in 54 cattle- and 31 buffalo-derived populations, as indicated in the insert table.
619
620

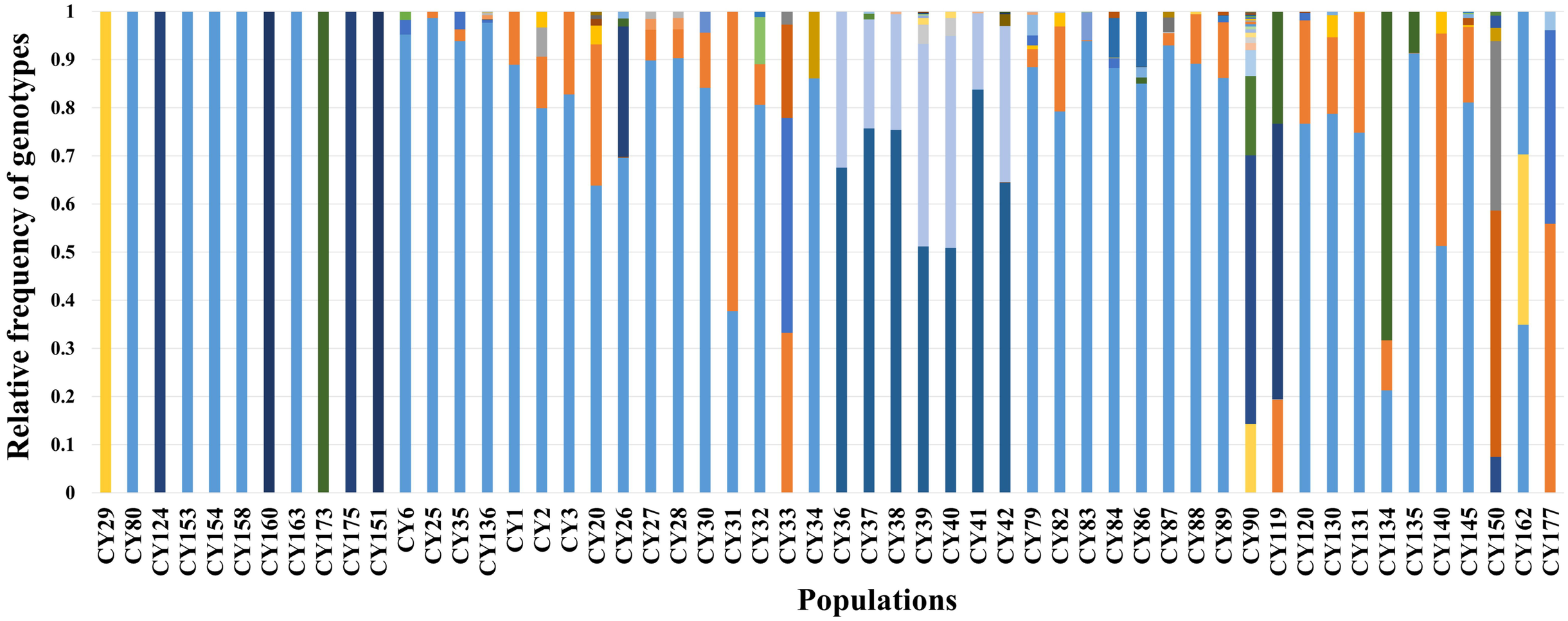

TS6
TS8
TS12
TS16
TS20
TS31
Cytochrome B
Number of Alleles
80
60
40
20
0
Buffalo Cattle
Unique Alleles
Shared Alleles

(A) Cattle



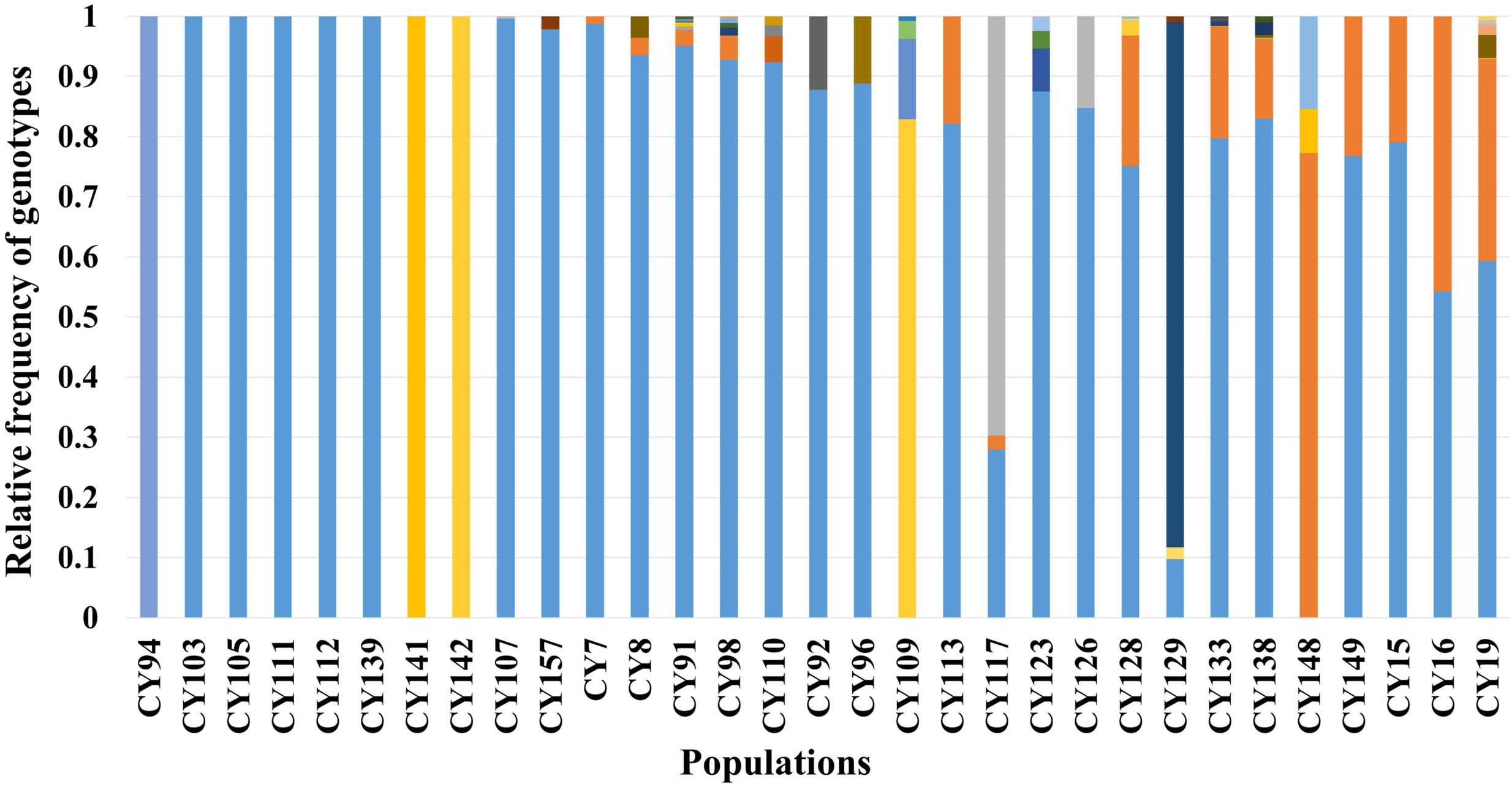
(B) Buffalo

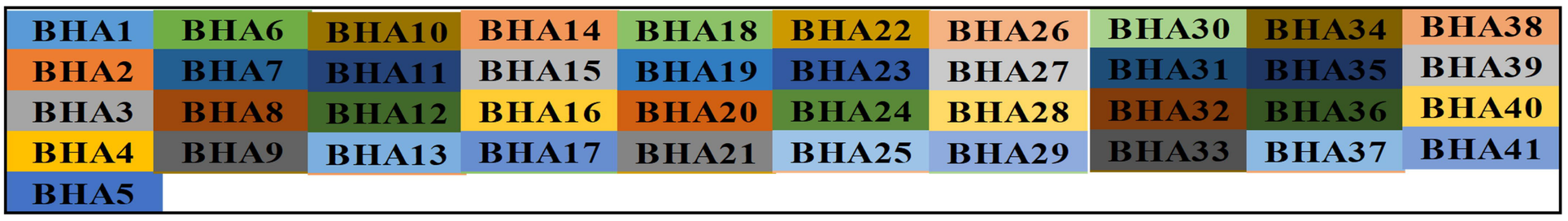




Populations	Sequence reads	Genotypes (%)
CY29	6442	CHA16(99.2),CHA4(0.2)
CY80	2997	CHA1(100)
CY124	3450	CHA65(100)
CY151	2032	CHA35(100)
CY153	4391	CHA1(100)
CY154	3483	CHA1(100)
CY158	1514	CHA1(100)
CY160	3621	CHA35(100)
CY163	2842	CHA1(100)
CY173	3393	CHA66(100)
CY175	3798	CHA65(1 00)
CY6	7361	CHA1(95.2), CHA5(3.0), CHA6(1.7), CHA7(0.1)
CY25	12422	CHA1(98.7), CHA2(1.2)
CY35	3780	CHA1(90.8), CHA2(3.4), CHA5(4.7)
CY136	109536	CHA1(97.6), CHA68(0.9), CHA69(0.4), CHA5(0.6), CHA33(0.1), CHA70(0.1), CHA71(0.1)
CY1	3055	CHA1(89), CHA2(11)
CY2	8032	CHA1(77.9), CHA2(10.6), CHA3(6.0), CHA4(5.3)
CY3	3573	CHA1(82.7), CHA2(17.2)
CY20	66615	CHA1 (60.8), CHA2(29.2), CHA4(6.9), CHA8(1.3), CHA9(0.8), CHA10(0.7)
CY26	8670	CHA1(69.6), CHA11(27.0), CHA12(1.6), CHA13(1.4), CHA8(0.1)
CY27	15741	CHA1(89.7),CHA2(6.4), CHA14(2.2), CHA15(1.5), CHA4(0.6)
CY28	9559	CHA1(90.3),CHA2(5.9), CHA14(2.3), CHA15(1.3)
CY30	8168	CHA1(81.1), CHA2(11.5), CHA17(6.3)
CY31	4219	CHA2(62.2), CHA1(37.7)
CY32	2304	CHA1(80.6), CHA2(8.4), CHA18(9.7), CHA19(1.1)
CY33	2162	CHA2(33.2), CHA5(43.6), CHA20(19.4), CHA21(4.5)
CY34	13252	CHA1(86.1), CHA22(13.8)
CY36	4812	CHA7(67.6), CHA23(32.3)
CY37	47941	CHA7(75.6), CHA23(22.6), CHA24(1.2), CHA25(0.4)
CY38	4812	CHA7(75.3), CHA23(24.0), CHA26(0.5), CHA8(0.2)
CY39	45065	CHA7(51.1), CHA23(42.0), CHA27(4.0), CHA28(1.4), CHA29(0.4), CHA30(0.2), CHA31(0.2), CHA32(0.2)
CY40	29245	CHA7(48.8), CHA23(44.0), CHA27(5.8), CHA28(1.2), CHA33(0.3)
CY41	77853	CHA7(83.7), CHA23(15.7), CHA13(0.1), CHA26(0.3)
CY42	57452	CHA7(62.4), CHA23(32.4), CHA34(4.3), CHA35(0.3), CHA36(0.2), CHA8(0.1)
CY79	48294	CHA1(85.4), CHA37(5.3), CHA2(4.6), CHA5(3.1), CHA4(0.7), CHA38(0.4), CHA39(0.1), CHA40(0.4)
CY82	23106	CHA1(79.2), CHA2(17.6), CHA4(2.8), CHA39(0.2)
CY83	64110	CHA1(90.8), CHA41(8.8), CHA2(0.1), CHA42(0.1)
CY84	77806	CHA1(86.3),CHA43(10.2), CHA5(2.0), CHA44(1.3), CHA40(0.3), CHA42(0.3)
CY86	60995	CHA1(85.0), CHA43(11.4), CHA13(2.1), CHA12(1.2), CHA33(0.1)
CY87	9526	CHA1(90.9), CHA45(4.1), CHA2(3.5), CHA39(0.1), CHA46(1.2)
CY88	5518	CHA1(89.1), CHA2(10.2), CHA40(0.5)
CY89	114872	CHA1(86.2), CHA2(11.5), CHA19(1.0), CHA44(0.7), CHA5(0.3), CHA33(0.1)
CY90	64051	CHA47(55.7), CHA48(16.4), CHA40(14.3), CHA49(5.4), CHA50(1.5), CHA51(1.1), CHA52(0.9), CHA53(0.7), CHA54(0.5), CHA55(0.5), CHA56(0.4), CHA57(0.3), CHA58(0.2), CHA59(0.2), CHA60(0.2), CHA61(0.2), CHA62(0.2), CHA63(0.2), CHA64(0.1), CHA45(0.1)
CY119	80952	CHA65(57.2), CHA66(23.2), CHA2(19.3), CHA39(0.1)
CY120	55003	CHA1(76.6),CHA2(21.4), CHA5(1.5), CHA33(0.7), CHA8(0.1)
CY130	26053	CHA1(78.8), CHA2(15.8), CHA67(0.7)
CY131	20896	CHA1(74.8), CHA2(24.9), CHA66(0.4, CHA4(0.1), CHA33(0.1)
CY134	3929	CHA66(68.3), CHA1(21.3),CHA2(10.3)
CY135	2083	CHA1(90.3), CHA66(9.5), CHA2(0.9)
CY140	4043	CHA1(50.3), CHA2(43.1), CHA4(6.5)
CY145	63270	CHA1(81.1), CHA2(15.7), CHA44(1.3), CHA67(0.8), CHA4(0.4), CHA72(0.3), CHA73(0.2)
CY150	25039	CHA74(48.1), CHA75(32.2), CHA47(7.4), CHA76(5.7), CHA77(5.4), CHA78(0.8)
CY162	3576	CHA1(34.8), CHA40(35.4),CHA55(29.6)
CY177	3917	CHA2(55.8), CHA5(40.1), CHA79(3.9)

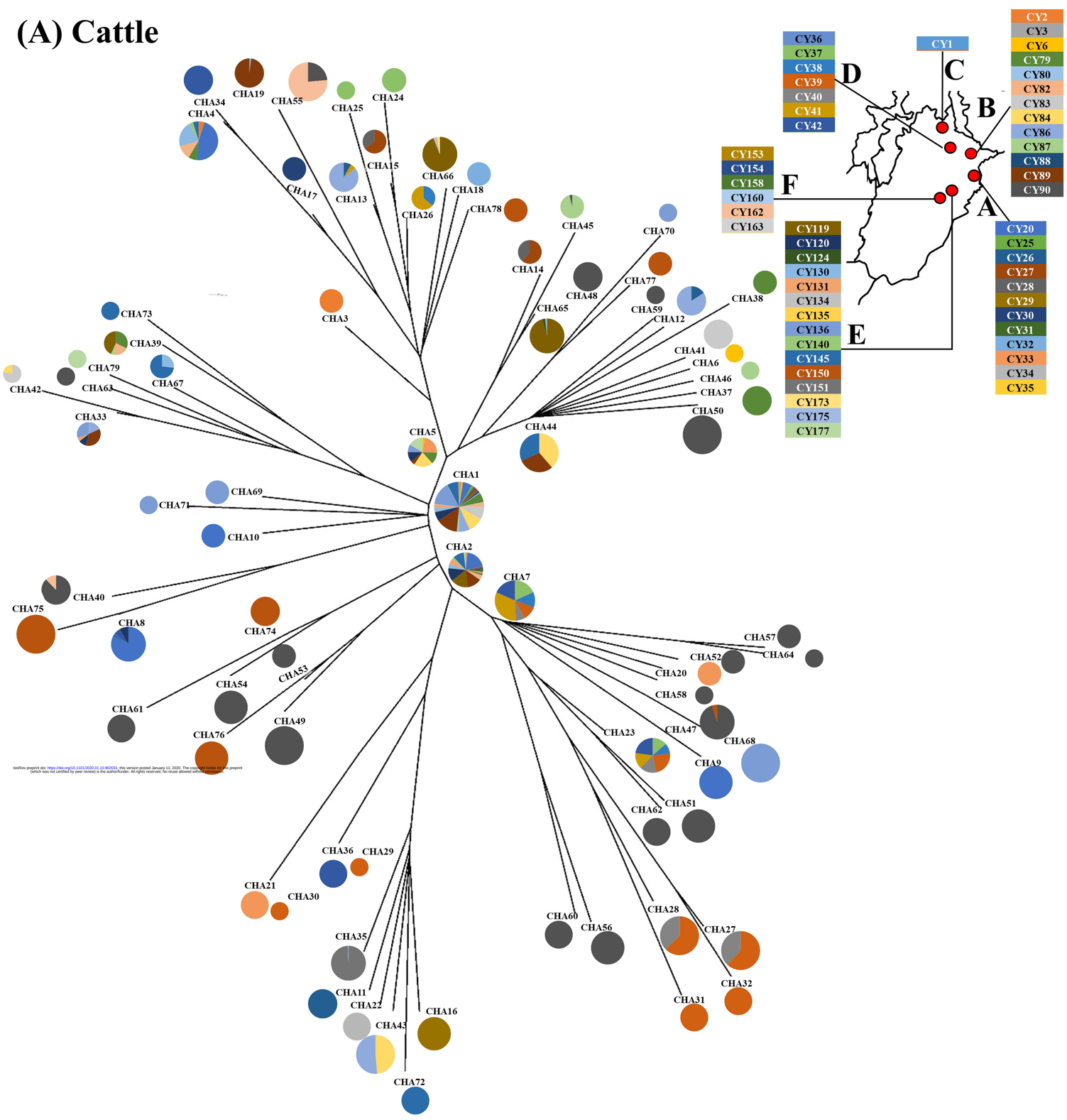
(B) Buffalo





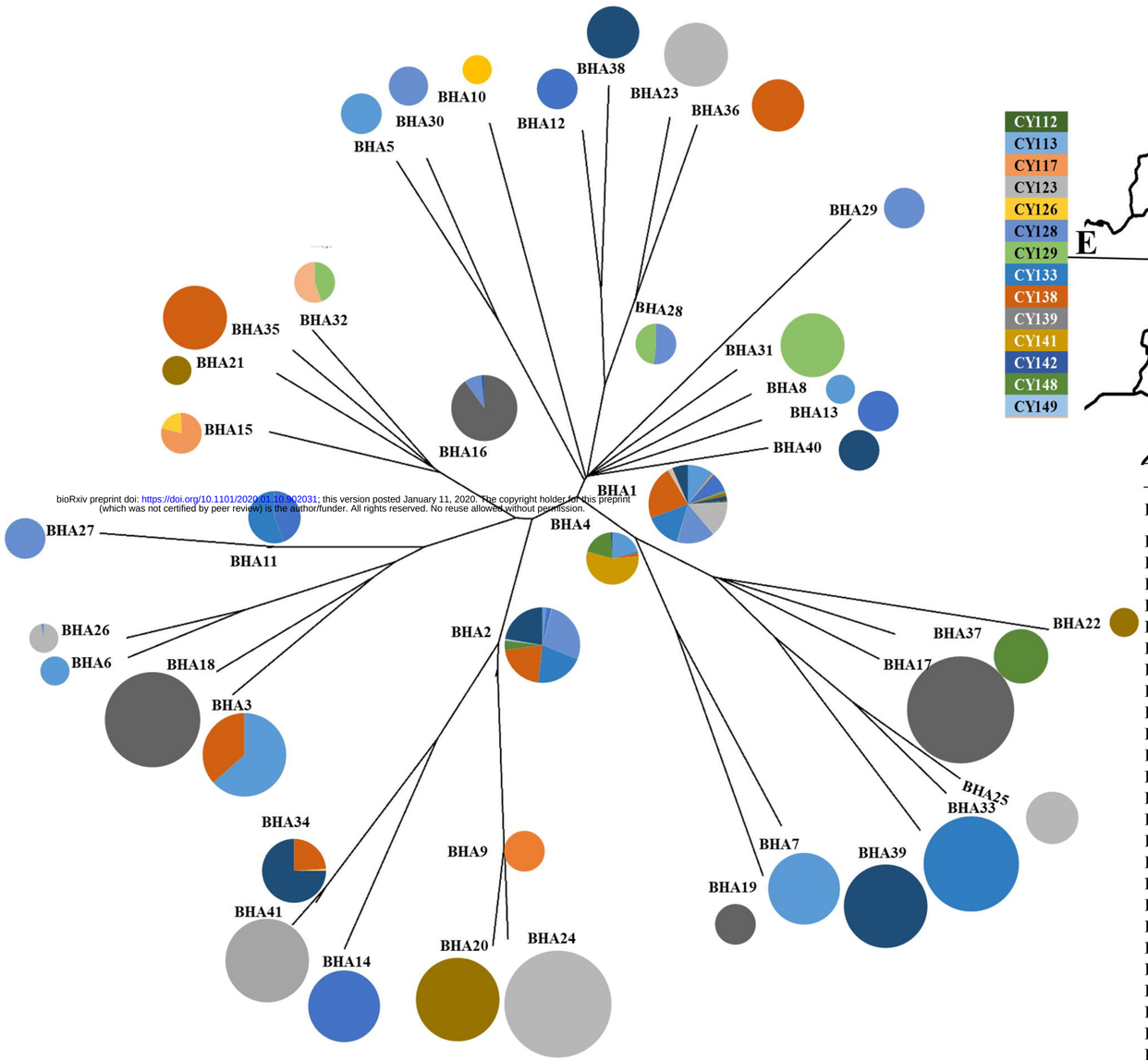
Populations	Sequence reads	Genotypes (%)
CY94	22831	BHA41(100)
CY103	14412	BHA1(100)
CY105	12823	BHA1(100)
CY111	51382	BHA1(100)
CY112	15822	BHA1(100)
CY139	18352	BHA1(100)
CY141	14892	BHA41(100)
CY142	12272	BHA16(100)
CY107	12712	BHA1(99), BHA15(0.3)
CY157	13062	BHA1(97.8), BHA32(2.1)
CY7	14892	BHA1(98.7), BHA2(1.2)
CY8	13682	BHA1(93.4), BHA2(2.9), BHA2(3.5)
CY91	29533	BHA1(95.1), BHA2(2.6), BHA3(0.5), BHA4(0.6), BHA6(0.2), BHA7(0.2), BHA8(0.1)
CY98	23295	BHA1(92.7), BHA2(4.0), BHA11(1.2), BHA12(0.8), BHA13(0.8), BHA14(0.2)
CY110	38681	BHA1(92.3), BHA20(4.3), BHA21(1.7), BHA22(1.4)
CY92	12361	BHA1(87.7), BHA9(12.2)
CY96	16971	BHA1(88.8), BHA10(11.9)
CY109	16723	BHA16(82.9), BHA17(13.3), BHA18(2.9), BHA19(0.7)
CY113	14081	BHA1(82.1), BHA2(17.8)
CY117	13502	BHA1(28), BHA15(69.8), BHA2(2.2)
CY123	43438	BHA1(87.5), BHA23(7.1),BHA24(2.9), BHA25(2.3), BHA26(0.1)
CY126	14262	BHA1(75.1), BHA15(15.2)
CY128	55010	BHA1(75.1), BHA2(21.7), BHA16(2.3), BHA26(0.3), BHA27(0.2), BHA28(0.2), BHA29(0.2), BHA30(0.1)
CY129	15470	BHA31(87.3), BHA1(9.7), BHA2(0.1), BHA28(1.9), BHA32(0.9)
CY133	49042	BHA1(79.7), BHA2(18.6), BHA4(0.2), BHA11(0.7), BHA33(0.8)
CY138	70250	BHA1(83.0), BHA2(13.2), BHA3(0.1), BHA4(0.2), BHA34(0.5), BHA35(2.0), BHA36(1.0)
CY148	12369	BHA2(77.2), BHA37(15.4), BHA4(7.3)
CY149	10771	BHA1(76.7), BAH2(23.2)
CY15	12621	BHA1(79.0), BHA2(20.9)
CY16	15922	BHA1(54.2), BHA2(45.7)
CY19	29321	BHA1(59.3), BHA2(33.6), BHA4(0.3), BHA34(3.9), BHA38(1.6), BHA39(0.7), BHA40(0.6)

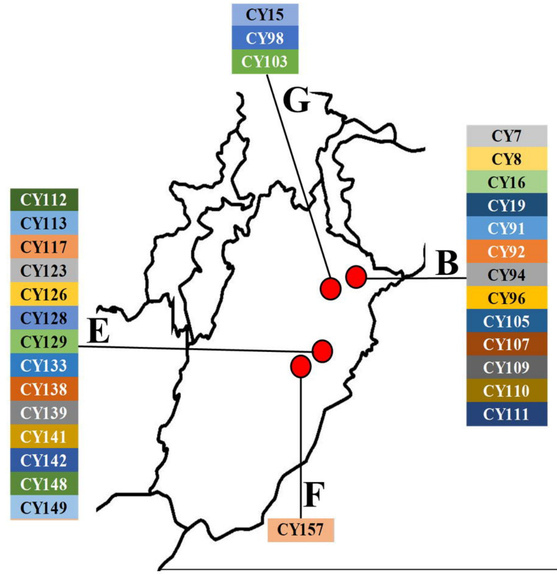
(A) Cattle



Genotypes	Sequence reads	Number of populations
CHA1	306152 (22.9)	36 (CY1, CY2, CY3, CY6, CY20, CY25, CY26, CY27, CY28, CY30, CY31, CY32, CY34, CY35, CY79, CY80, CY82, CY83, CY84, CY86, CY87, CY88, CY89, CY120, CY130, CY130, CY131, CY134, CY135, CY136, CY140, CY145, CY153, CY154, CY158, CY162, CY163)
CHA7	144834 (10.8)	8 (CY6, CY36, CY37, CY38, CY39, CY40, CY41, CY42)
CHA2	93391 (6.9)	27 (CY1, CY2, CY3, CY20, CY25, CY27, CY28, CY30, CY31, CY32, CY33, CY35, CY79, CY82, CY83, CY87,CY88, CY89, CY119, CY120,CY131, CY134, CY135, CY140, CY145, CY177)
CHA23	80938 (6.0)	7 (CY36, CY37, CY38, CY39, CY40, CY41, CY42)
CHA35	70229 (5.2)	3 (CY42, CY151,CY160)
CHA65	84581 (6.3)	3 (CY119, CY124, CY175)
CHA47	73571 (5.5)	2 (CY90, CY150)
CHA8	92376 (6.9)	5 (CY20, CY26, CY38, CY42, CY120)
CHA66	80068 (5.9)	5 (CY119, CY131, CY134, CY135, CY173)
CHA43	12608 (0.9)	2 (CY84, CY86)
CHA74	11807 (0.8)	1 (CY150)
CHA48	10559 (0.7)	1 (CY90)
CHA40	10464 (0.7)	5 (CY79, CY84, CY88, CY90, CY162)
CHA75	13825 (1.0)	1 (CY150)
CHA5	7836 (0.5)	9 (CY6, CY33, CY35, CY79, CY84, CY89, CY120, CY136, CY177)
CHA4	6560 (0.4)	10 (CY2, CY20, CY27, CY29, CY79, CY82, CY130, CY131, CY140, CY145)
CHA41	5724 (0.4)	1 (CY83)
CHA49	6472 (0.4)	1 (CY90)
CHA27	7923 (0.5)	2 (CY39, CY40)
CHA44	6823 (0.5)	3 (CY84, CY89, CY145)
CHA11	4346 (0.3)	1 (CY26)
CHA37	5130 (0.3)	1 (CY79)
CHA13	4519 (0.3)	3 (CY26, CY41, CY86)
CHA55	4388 (0.3)	2 (CY90, CY162)
CHA34	5381 (0.4)	1 (CY42)
CHA19	5243 (0.3)	2 (CY32, CY89)
CHA28	5018 (0.3)	2 (CY39, CY40)
CHA68	4006 (0.2)	1 (CY136)
CHA50	3964 (0.2)	1 (CY90)
CHA12	3922 (0.2)	2 (CY26, CY86)
CHA20	4758 (0.3)	1 (CY33)
CHA51	4746 (0.3)	1 (CY90)
CHA67	3729 (0.2)	2 (CY130, CY145)
CHA76	4689 (0.3)	1 (CY150)
CHA77	5621 (0.4)	1 (CY150)
CHA52	3594 (0.2)	1 (CY90)
CHA24	4588 (0.3)	1 (CY37)
CHA14	4674 (0.3)	2 (CY27, CY28)
CHA9	3536 (0.2)	1 (CY20)
CHA10	4516 (0.3)	1 CY20)
CHA3	5489 (0.4)	1 (CY2)
CHA53	3476 (0.2)	1 (CY90)
CHA26	4474 (0.3)	2 (CY38, CY41)
CHA69	5449 (0.4)	1 (CY136)
CHA16	4441 (0.3)	1 (CY29)
CHA33	3397 (0.2)	6 (CY40, CY86, CY89, CY120, CY131, CY136)
CHA15	3373 (0.2)	2 (CY27, CY28)
CHA17	4354 (0.3)	1 (CY30)
CHA54	5347 (0.4)	1 (CY90)
CHA56	4315 (0.3)	1 (CY90)
CHA45	5308 (0.3)	2 (CY87, CY90)
CHA39	4273 (0.3)	4 (CY79, CY82, CY87, CY119)
CHA18	5225 (0.3)	1 (CY32)
CHA78	4223 (0.3)	1 (CY150)
CHA38	3210 (0.2)	1 (CY79)
CHA57	3510 (0.2)	1 (CY90)
CHA25	2206 (0.1)	1 (CY37)
CHA70	2200 (0.1)	1 (CY136)
CHA72	1994 (0.1)	1 (CY145)
CHA29	1888 (0.1)	1 (CY39)
CHA58	2186 (0.1)	1 (CY90)
CHA59	2985 (0.1)	1 (CY90)
CHA60	2185 (0.1)	1 (CY90)
CHA22	1674 (0.1)	1 (CY34)
CHA61	1873 (0.1)	1 (CY90)
CHA73	2162 (0.1)	1 (CY145)
CHA62	2459 (0.1)	1 (CY90)
CHA63	2647 (0.1)	1 (CY90)
CHA42	2143 (0.1)	2 CY83, CY84)
CHA36	2539 (0.1)	1 (CY42)
CHA30	1835 (0.1)	1 (CY39)
CHA31	1933 (0.1)	1 (CY39)
CHA6	2132 (0.1)	1 (CY6)
CHA71	2627 (0.1)	1 (CY136)
CHA79	2526 (0.1)	1 (CY177)
CHA32	2845 (0.1)	1 (CY39)
CHA46	1515 (0.1)	1 (CY87)
CHA64	1408 (0.1)	1 (CY90)
CHA21	1805 (0.1)	1 (CY33)

(B) Buffalo





Genotypes	Sequence reads	Number of populations
BHA1	278937 (39.1)	26 (CY91, CY92, CY96, CY98, CY103, CY105, CY107, CY110, CY111, CY112, CY113, CY117, CY123, CY126, CY128,CY129, CY133, CY138, CY139, CY149, CY157, CY7, CY8, CY15, CY16, CY19)
BHA2	77171 (10.8)	15 (CY19, CY98, CY113, CY117, CY128, CY129, CY133, CY138, CY148, CY149, CY7, CY8, CY15, CY16, CY19)
BHA16	39389 (5.5)	3 (CY109, CY128, CY142)
BHA31	14779 (2.0)	1 (CY129)
BHA23	19109 (2.6)	1 (CY123)
BHA17	18232 (2.5)	1 (CY109)
BHA34	19535 (2.7)	3 (CY138, CY8, CY19)
BHA35	19437 (2.7)	1 (CY138)
BHA24	20260 (2.8)	1 (CY123)
BHA25	10998 (1.5)	1 (CY123)
BHA4	12882 (1.8)	6 (CY91, CY133, CY138, CY141, CY148, CY19)
BHA36	14721 (2.0)	1 (CY138)
BHA11	13670 (1.9)	2 (CY98, CY133)
BHA18	15501 (2.1)	1 (CY109)
BHA38	12485 (1.7)	1 (CY19)
BHA33	14428 (2.0)	1 (CY133)
BHA37	12365 (1.7)	1 (CY148)
BHA15	8310 (1.1)	3 (CY107, CY117, CY126)
BHA41	7283 (1.0)	1 (CY94)
BHA3	8268 (1.1)	2 (CY91, CY138)
BHA28	6217 (0.8)	2 (CY128, CY129)
BHA39	7212 (1.0)	1 (CY19)
BHA40	5199 (0.7)	1 (CY19)
BHA12	4192 (0.5)	1 (CY98)
BHA13	5192 (0.7)	1 (CY98)
BHA20	4169 (0.5)	1 (CY110)
BHA9	5151 (0.7)	1 (CY92)
BHA27	3925 (0.5)	1 (CY128)
BHA19	3024 (0.4)	1 (CY109)
BHA32	3117 (0.4)	2 (CY129, CY157)
BHA29	4110 (0.5)	1 (CY128)
BHA5	3102 (0.4)	1 (CY91)
BHA30	2790 (0.3)	1 (CY128)
BHA10	2078 (0.2)	1 (CY96)
BHA6	2378 (0.3)	1 (CY91)
BHA7	2477 (0.3)	1 (CY91)
BHA21	2869 (0.4)	1 (CY110)
BHA14	1264 (0.1)	1 (CY98)
BHA26	1864 (0.2)	2 (CY123),
BHA22	1657 (0.2)	1 (CY128)
BHA8	1954 (0.2)	1 (CY91)