The genetic diversity of *Strongyloides papillosus* in Pakistani goats revealed by whole genome sequencing.

Kiran Afshan[1], Yuchen Liu[2], Mark Viney[2*]

1. Department of Zoology, Faculty of Biological Sciences, Quaid-i-Azam University Islamabad, 45320, Pakistan.

2. Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, L69 7ZB, UK.

* To whom correspondence should be addressed; Mark.Viney@liverpool.ac.uk

Brief Report format:

Main text word count: 3,011

Figures: 2

References: 20

Abstract

Strongyloides nematodes are parasites of livestock, and *S. papillosus* infects ruminant livestock that can cause disease. Recent genomic analysis of several *Strongyloides* species is now facilitating the population genomic analyses of natural *Strongyloides* infections, for example finding that *S. ratti* in wild UK rats exists as an assemblage of long-lived, asexual lineages. Here we have initiated an investigation into the population genomics of *S. papillosus* in goats in Pakistan. We find that *S. papillosus*

is common, with a prevalence of 28%; that the population is genetically diverse and that individual goats commonly have mixed-genotype infections; and that there is evidence of only limited admixture. These results now provoke further questions about the host range of different *S. papillosus* genotypes that can in the future be investigated by further population genomic analyses.

Background

Strongyloides is a genus of gut parasitic nematode that infects a wide variety of terrestrial vertebrates. *Strongyloides papillosus* has been reported to parasitise sheep, goats, and cattle and can be experimentally maintained in rabbits (Pinn et al*.,* 2022). However, the species identity of *Strongyloides* infecting these host species has been questioned. Sequence analysis of the rRNA coding gene of *Strongyloides* from cattle and sheep found two different forms that appeared to be substantially (though not completely) host specific (Eberhardt et al*.,* 2008). The names *S. papillosus* and *S. vituli* for parasites from sheep and from cattle, respectively, were originally suggested by Brumpt 1921 (though not widely used), but reinitiated by Eberhardt et al*.,* 2008, with further support for *S. vituli* infection in cattle (Ko et al*.,* 2019). More generally, the host specificity of *Strongyloides* species remains unexplored (Al-Jawabreh et al., 2023). Most *Strongyloides* species have been described morphologically, often from only some stages of the life cycle, which makes it hard to not only define a *Strongyloides* species, but also its host range. However, population genetic analyses of *Strongyloides* has the potential to resolve such questions.

The parasitic phase of *S. papillosus* (in common with other species of *Strongyloides*) consists of adult female worms only that live in the mucosal epithelium of the small intestine and reproduce by mitotic parthenogenesis. Eggs are shed in the host faeces, ultimately resulting in infective, filariform third-stage larvae (iL3s). These iL3s can develop either, directly, via two larval moults, or, indirectly, as the progeny of a facultative adult dioecious free-living generation that reproduces sexually (Viney, 2017). Hosts become infected by percutaneous infection by iL3s (Turner et al*.,* 1960). The parthenogenetic reproduction of the parasitic female worms means that the larvae

passed in faeces are genetic copies of their mother that is resident in the host gut. The only opportunity for sexual reproduction is when free-living males and female worms are present (presumably in faeces or nearby soil). The factors that control the occurrence of this free-living adult generation are not fully understood, but in *S. ratti* in rats this varies among *S. ratti* genotypes, and is affected both by the temperature external to the host and by the host immune response (Viney & Lok 2007). From a population genetic perspective the facultative nature of sexual reproduction may mean that the population genetic structure of *Strongyloides* deviates from that predicted by Hardy-Weinberg expectations.

S. papillosus can be a common infection of ruminants, with a prevalence of up to 58% reported in lambs, though the prevalence differs widely geographically and among different livestock management systems (Thamsborg et al., 2017). *S. papillosus* can cause diarrhoea, dehydration, anorexia, anaemia and malnutrition, especially in young animals (Thamsborg et al., 2017). High intensity infections can cause serious strongyloidiasis, which can be fatal (Pinn et al*.,* 2022). Juvenile animals are most susceptible to heavy infection upon first exposure, but rapidly develop acquired immunity to future *S papillosus* infections (Turner et al*.,* 1960). *S. papillosus* can be treated with anthelminthic drugs such as macrocyclic lactones (*e.g.* Ivermectin) and albendazole (Thamsborg et al., 2017). Rabbits can also be infected with *S. papillosus* and as such can be used as an experimental model for *S. papillosus* (Thamsborg et al., 2017).

The genome of *S. papillosus* has been sequenced, together with 3 other species of *Strongyloides* (Hunt et al., 2016). The *S. papillosus* genome is estimated to be 60Mbp,

though it is less well assembled than the other *Strongyloides* species such that the estimate of its genome size may not be very accurate. Notwithstanding, the *S. papillosus* genome assembly can be used for population genomic analyses of this species.

Population genetic analyses of parasitic nematodes of livestock has found that there is very high gene flow and an absence of parasite population structure (Blouin et al., 1995). This was thought to be due to both the often very large effective population sizes of the nematodes, and because of extensive host movement that occurs as part of commercial farming. For nematode parasites of species of wildlife then there are a diversity of observed parasite population structures, with this dependant on host and parasite specific biology (Cole and Viney 2018). There has been previous population genomic analysis of *Strongyloides*. Specifically, *S. ratti* infecting wild rats in the UK was found to exist as a mixture of mainly asexual lineages parasites that were widely dispersed across the host population, findings that were consistent with earlier, UK-wide, studies (Fisher & Viney, 1998; Cole et al., 2023).

Here we report an initial population genomic analysis of *S. papillosus* in goats in Pakistan where we have used single worm, whole genome sequencing. We describe the patterns of genetic diversity among these worms, find evidence of mixed-genotype infections in hosts, but limited admixture.

Materials and Methods

Study area, faecal sampling and culturing. This study was approved by the bioethical committee of Quaid-i-Azam University, Islamabad, Pakistan. Sampling was conducted at four slaughterhouses within Rawalpindi City, with the slaughterhouses ranging from a minimum of 9 to a maximum of 32km between them. Faecal samples were collected directly from the rectum of 150 slaughtered goats using sterile disposable plastic gloves, with fresh gloves used for each goat to avoid cross-contamination. The faeces were cultured by mixing a few grammes of fresh faeces with an approximately equal volume of wet charcoal, which was then kept at room temperature for up to 13 days when individual iL3s were collected, which were identified as *Strongyloides* by morphological examination (Zhou et al*.,* 2019); no free-living adult males or females were observed.

DNA extraction and genome sequencing of single larvae. The DNA extraction protocol is based on the method of Cole et al., 2023 but adapted for ethanol-stored worms after Zhou et al., 2019. Pools of larvae from each host were processed separately, individual larvae in TE were placed into wells of a 96 well plate; lysis buffer was added to give a final concentration of 200 mM NaCl,100 mM Tris-HCl (pH8.5), 50 mM EDTA (pH8), 0.5 % w/v SDS, 0.9 mg /mL Proteinase K, and 45 mM Dithiothreitol (DTT). Plates were then sealed and held at 60°C for 2 hours with gentle agitation, after which they were held 85°C for 15 minutes to inactivate the Proteinase K. Lysates were then stored at -80°C before whole genome sequencing.

125 Whole genome sequencing was performed by the Centre for Genomic Research,
126 University of Liverpool. Illumina libraries were prepared using the NEB FS Ultra 2 DNA
127 kit at half of the manufacturer's specified reaction volumes.
128
129 **Sequence analyses.** We whole genome sequenced 114 *S. papillosus* iL3s, mapped
130 the sequence reads using Bowtie2 to the *S. papillosus* reference genome (PRJEB525,
131 WBPS18 available from WormBase-ParasiteSite; parasite.wormbase.org), and
132 calculated the average depth and coverage across the genome of each alignment
133 using SAMtools. Of these 114 larvae, 37 had a genome coverage of at least 70% and
134 an average depth of at least 10, and these samples were then analysed further.
135
136 Single Nucleotide Polymorphism (SNP) calling was performed using SAMtools. We
137 used VCFtools to filter SNPs using the criteria that (i) they had a mean mapping quality
138 of at least 20, (ii) they had a QUAL score of at least 20, (iii) SNP loci had a minor allele
139 frequency of ≥ 0.02, and (iv) their mean depth was between the 10 and 90%
140 percentiles of the whole data set, which was 6 and 21, respectively. Among the 37 *S.*
141 *papillosus* genome sequences, nucleotides that were identical among all samples (but
142 different from the PRJEB525 reference genome) were removed.
143
144 Using the filtered SNP set of the 37 *S. papillosus* iL3s we performed population
145 structure analysis by calculating the pairwise genetic similarity among the 37 worms
146 and then (i) using principal component analysis (PCA) in PLINK 1.9 and Tassel v.5
147 with plots visualized in R studio v.4.2.1; (ii) constructing a neighbour-joining tree in
148 Tassel, visualising the tree in iTOL; and (iii) using ADMIXTURE version 1.3.0
149 (Alexander and Large, 2011) analysis, with linkage disequilibrium (LD)-based pruning,

150 which was done with window sizes of 50 kb, step sizes of 10 SNPs, and with a r^2
151 threshold of 0.1. ADMIXTURE was run for k = 2 – 15 and results plotted in R studio
152 v.4.2.1. To investigate the genetic distance among worms we calculated the identity-
153 by-distance (IBD) on the filtered SNP set in PLINK. IBD values were categorised into
154 two pairwise comparison groups (i) from within the same host and (ii) and among
155 different hosts.
156
157 The *S. papillosus* reference mitochondrial genome was obtained from NCBI GenBank,
158 NC_028622.1. We generated maximum likelihood trees for the mitochondrial
159 genomes of the 37 larvae, producing consensus fasta sequences for each; where an
160 individual was heterozygous the reference allele was applied. Sequences were
161 aligned with MAFFT (Katoh et al., 2009) using the fast alignment method FF-NST-1.
162 Maximum likelihood tree estimation was carried out utilizing RaxML (Stamatakis,
163 2006), applying a general time reversible gamma model of substitution rate
164 heterogeneity, and rapid bootstrapping with 100 replicates and visualized in ITOL.
165
166 It has been suggested that there are two species of *Strongyloides* that infect
167 ruminants, *S. papillosus* and *S. vituli*, which can be differentiated by their rRNA
168 sequence (Eberhardt et al*.,* 2008; Ko et al., 2019), and a complete mitochondrial
169 genome sequence has also been reported for *S. vituli* (Ko et al*.,* 2023). For the 37
170 samples we analysed the rRNA and mitochondrial sequences to confirm their species
171 identity.

172 **Results**

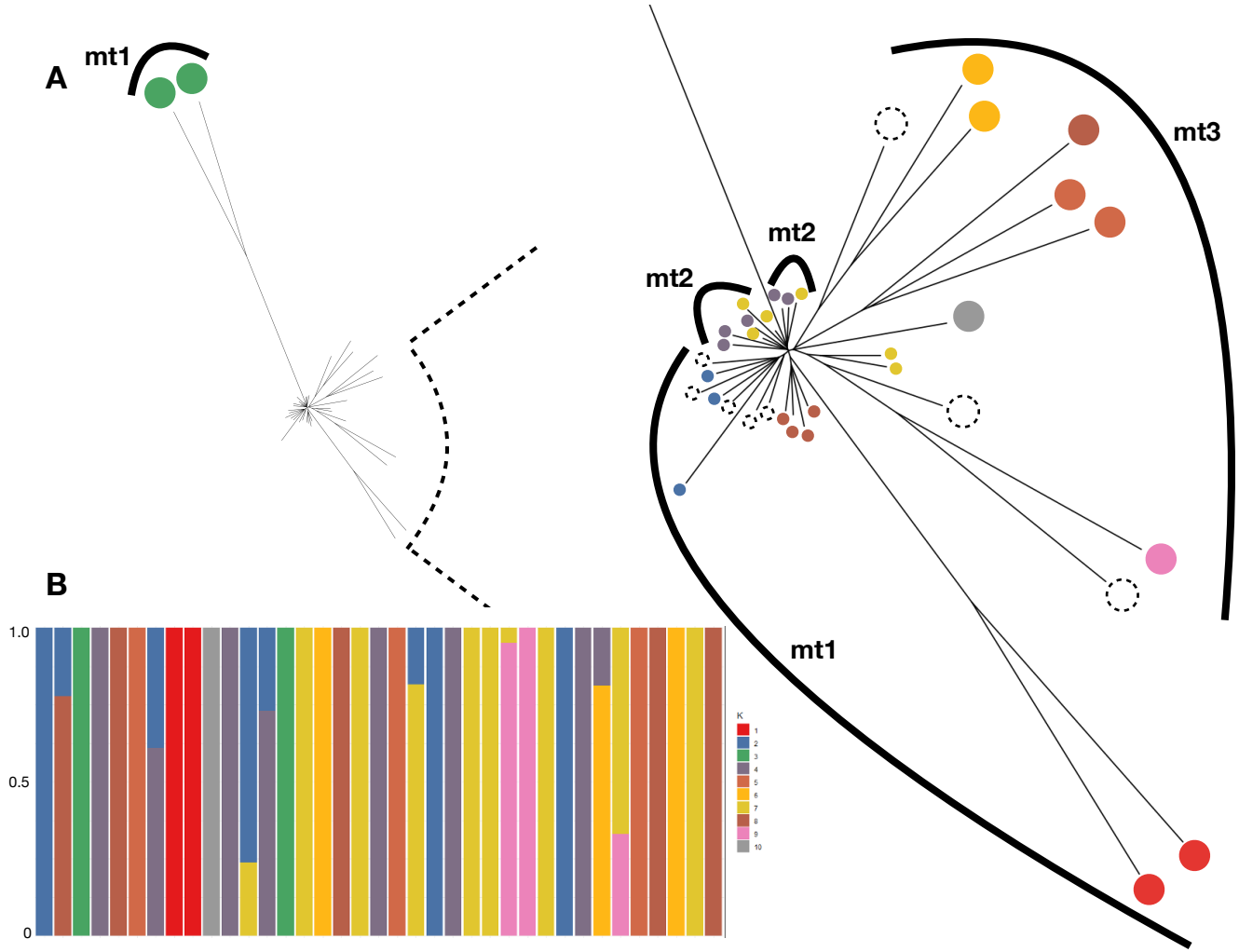
173 We sampled 150 goats of which 42 (28%) were faecal culture positive for
174 *Strongyloides*. In all we obtained 526 iL3s, ranging from 7 – 15 iL3s per infected goat.
175 We sequenced 114 individual iL3s of which 37 generated sufficiently good data and
176 which were taken forward for further analysis. These 37 larvae were obtained from 21
177 goats. For these 37 this resulted in sequence data with an average genome coverage
178 of 89% (range 74 – 94%), and an average read depth of 21 (range 9.7 – 69), giving
179 an average genome coverage size of 53.6 Mb. We detected 11,023 filtered SNPs in
180 the nuclear genome, giving an average SNP density of 0. 2 SNPs per 10 kb. This
181 alignment of reads to the *S. papillosus* genome is consistent with identifying these
182 larvae as *S. papillosus* rather than *S. vituli*. We also analysed the rRNA and
183 mitochondrial genome sequence of the 37 samples, which confirmed their identity as
184 *S. papillosus.*

185

186 To investigate the genetic relationship among the 37 iL3s we calculated the pairwise
187 nuclear genetic similarity among the larvae and used this to construct a neighbour
188 joining tree (**Figure 1A**). The shape of this tree showed that there was extensive
189 genetic heterogeneity among the worms. In this tree most (33 worms) were on short-
190 length branches, showing their relatively close genetic similarity. But two pairs of
191 worms (two of which were from a single host, the other two from two different hosts)
192 were on long, distant branches in the tree, showing their relative genetic dissimilarity
193 from the other 33 worms. A PCA of the same data also showed three clusters of iL3s,
194 with 33 iL3s in a single cluster, and then two smaller clusters each consisting of the
195 same two pairs of divergent iL3s (**Supplementary Figure 1**). We investigated the data
196 for evidence of admixture, with the results for k = 10 (Cross Validation (CV) = 0.13)

197

Figure 1



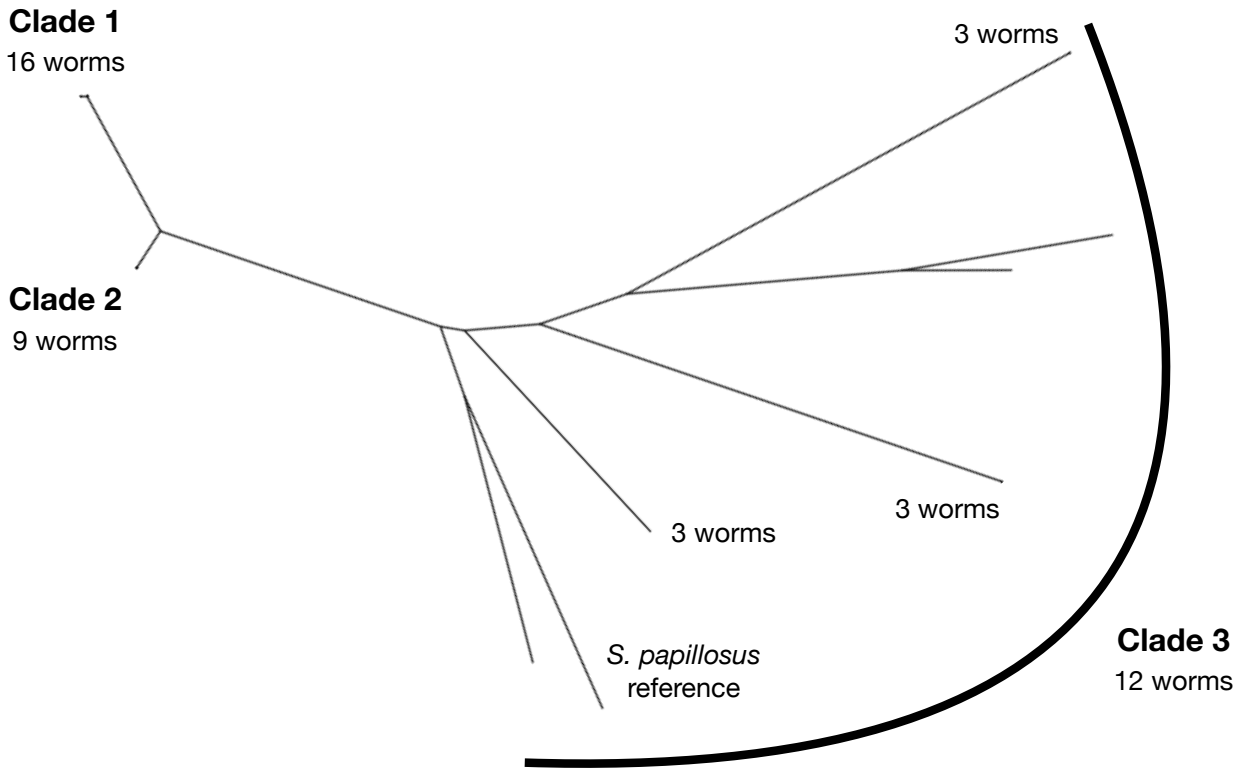
198

199

200 showing that only eight worms (22%) had evidence of admixture (**Figure 1B**).
201 Examining admixture results for other values of k, also showed low levels of admixture;
202 specifically for k = 3, 7, 15 then six, three, and six worms showed admixture,
203 respectively (**Supplementary Figure 2**)
204
205 The mitotic parthenogenetic reproduction of *Strongyloides* parasitic females means
206 that offspring of a single parasitic female worm will be genetically identical, and so that
207 genetically identical iL3s would be observed in host faeces. Conversely, if hosts are
208 infected with genetically different parasitic females, then genetically diverse iL3s will
209 be present in hosts faeces. With this in mind, we considered hosts from whom multiple
210 iL3s were sequenced. Among the 21 hosts, 11 each had more than one worm
211 sequenced, giving a total of 27 worms (five hosts with three worms; six hosts with two
212 worms). We compared the average genetic distance of worms from within hosts, with
213 worms from among different hosts, finding that there was little difference: mean (SD)
214 within hosts 0.063 (0.068), but 0.059 (0.059) among hosts. This suggests that hosts
215 may harbour genetically different worms. We asked whether worms from a single host
216 were in one or more k = 10 ADMIXTURE-defined groups and / or were admixtured.
217 We found that 10 of the 11 hosts contained worms from more than one ADMIXTURE-
218 defined group and / or were admixtured. This suggests that multi-genotype *S.*
219 *papillosus* infections are common in goats in Pakistan.
220
221 For the mitochondrial genome we detected 357 SNPs and used these to construct a
222 maximum likelihood tree of the 37 iL3 and the reference *S. papillosus* mitochondrial
223 genome. This showed that there were three *S. papillosus* clades (**Figure 2**). Two
224 clades contained most (16 and 9 iL3s) iL3s, with the third clade containing 12 iL3s

225

Figure 2



226

227 that were more divergent. This third clade also contained the *S. papillosus* reference
228 mitochondrial genome (**Figure 2**).
229
230 As for the nuclear genome, we considered the 11 hosts that each had more than one
231 sequenced worm. We found that nine hosts contained worms that were in more than
232 one mitochondrial clade. This is further evidence of mixed *S. papillosus* genotype
233 infections in these hosts. Interestingly, the four iL3s with divergent nuclear genomes
234 (**Figure 1A**) were within mitochondrial clade 1 (**Figure 2**). This result therefore shows
235 some divergence in the genetic relationship among iL3s as judged by nuclear and
236 mitochondrial data.
237
238 We compared the nuclear and mitochondrial trees. This showed that, with the
239 exception of one nuclear clade, the three mitochondrial clades grouped related nuclear
240 clades (**Figure 1A**). The exception is the highly divergent nuclear clade, which belongs
241 to mitochondrial clade 1, whereas its closest neighbours belong to mitochondrial clade
242 2.

243 **Discussion**

244 This work has used whole genome sequencing to describe the genetic diversity of *S.*
245 *papillosus* in goats in Pakistan. We have found that *S. papillosus* is a common infection
246 of goats in Pakistan, with a prevalence of 28%. *S. papillosus* can infect a number of
247 species of ruminants, and it remains to be discovered what its prevalence is in these
248 other species. Given that *S. papillosus* is able to infect a number of different potential
249 host species, there is the possibility for among-host species transmission of *S.*
250 *papillosus*. Such sharing of parasites among different host species has the potential
251 to genetically structure the parasite population, with this possibility being dependant
252 on the amount of within, verses between, host species transmission.

253

254 This work revealed evidence of extensive genetic diversity among the parasites
255 sampled, for both the nuclear and mitochondrial genomes. The mitotic
256 parthenogenetic reproduction of *Strongyloides* parasitic females means that their
257 offspring will be genetically identical, though sequencing errors will not result in
258 identical gene sequences. Our data show that mixed-genotype infections are common
259 in the goats that were sampled, which is likely due to goats being repeatedly exposed
260 to infection from a genetically diverse *S. papillosus* population. This pattern is broadly
261 consistent with patterns of population genetics seen in other parasitic nematodes
262 (Blouin et al., 1995). Among the parasites sampled, a minority were genetically
263 relatively more divergent. Specifically, by analysis of the nuclear genome 4 of the 37
264 worms were more distinct. This might point to the existence of even higher levels of
265 genetic diversity in the *S. papillosus* population, which might be revealed by more
266 extensive sampling in goats. Further, because *S. papillosus* is able to infect different
267 ruminant species, the more-divergent genotypes we observed may be because these

268 are parasites whose host preference (but not restriction) is for species other than
269 goats. However, further sampling and genetic analysis is needed to understand the
270 distribution of *S. papillosus* among different ruminant host species and the relationship
271 of *S. papillosus* and *S. vituli* (Eberhardt et al., 2008). The results of the analysis of the
272 mitochondrial genome are broadly consistent with the analysis of the nuclear genome,
273 showing 3 distinct clades with the third clade containing mitochondrially diverse worms
274 (and the *S. papillosus* reference genome). Again, these results may potentially indicate
275 a diversity of parasites that infect a range of host species, of which only a small portion
276 has been sampled in the current study. However, it is notable that the grouping of
277 parasites by their nuclear and mitochondrial genomes is not fully concordant. This is
278 indicative of different histories of these two genomes in some of the parasites that we
279 have sampled, which has also been observed in *Ascaris* (Anderson et al., 1997).
280
281 The *Strongyloides* life cycle is obligatorily parthenogenetic with facultative sexual
282 reproduction, where the degree of sexual reproduction varies between species,
283 geographical location, and genotype (Viney & Lok 2007). The amount of sexual
284 reproduction in *S. papillosus* that we have sampled is unknown, but during the culture
285 of faeces we did not observe any free-living adult stages, suggestive of rare or no
286 sexual reproduction. A previous analysis of wild *S. ratti* in the UK found that it consisted
287 of a mixture of long-lived asexual lineages (Cole et al., 2023), and a similar pattern
288 may exist among the *S. papillosus* that we have sampled, though further study is
289 required to confirm this.
290
291 To extend and develop this work and to better understand the population genomics of
292 *S. papillosus* it would be desirable to more extensively sample larvae from a diversity

293 of sympatric host ruminants. This will allow an analysis of the degree of partitioning, if
294 any, of *S. papillosus* genetic diversity among different host species. Future work would
295 also be facilitated by a better, more complete assembly of the *S. papillosus* genome,
296 which could be achieved by long-read DNA sequencing. Analysis of diverse, wild
297 genotypes of *S. ratti* has discovered clusters of parasitism genes (Hunt et al., 2016)
298 that are hyperdiverse compared with the rest of the *S. ratti* genome (Cole et al., 2023).
299 It would be interesting to study this in *S. papillosus* too, but this will also require a more
300 complete genome assembly.

Acknowledgments

This work was funded by an International Veterinary Vaccinology Network Laboratory Exchange Award and Quaid-i-Azam University, Islamabad Pakistan. YL is supported by the University of Liverpool and the China Scholarship Council.

305 **References**

306 Alexander, D. H. and Large, K. 2011. Enhancements to the ADMIXTURE algorithm
307 for individual ancestry estimation. BMC Bioinf 12, 246.
308
309 Al-Jawabreh, R., Anderson, R.M., Atkinson, L., Bickford-Smith, J., Bradbury, R.,
310 Breloer, M., *et al.* 2023. *Strongyloides* questions – a research agenda for the future.
311 Phil Trans Roy Soc B *379, 20230004.*
312
313 Anderson, T. J., and Jaenike, J. 1997 Host specificity, evolutionary relationships and
314 macrogeographic differentiation among *Ascaris* populations from humans and pigs.
315 Parasitol 115, 325-342.
316
317 Blouin, M.S., Yowell, C.A., Courtney, C.H. and Dame, J.B. 1995. Host movement and
318 the genetic structure of populations of parasitic nematodes. Genetics 141, 1007-1014.
319
320 Brumpt, E. 1921 Recherches sur le determinisme des sexes et de l 'evolution des
321 *Anguillules* parasites (*Strongyloides*). Compt Rend Séances Mém Soc Biol Fil 85, 149-
322 152.
323
324 Cole, R., Holroyd, N., Tracey, A., Berriman M. and Viney, M. 2023. The parasitic
325 nematode *Strongyloides ratti* exists predominantly as populations of long-lived
326 asexual lineages. Nat Comm 14, 6427.
327
328 Cole, R. and Viney, M.E. 2018. The population genetics of parasitic nematodes of wild
329 animals. Parasit Vect 11, 590.

330

331 Eberhardt, A.G., Mayer, W.E., Bonfoh, B. and Streit, A., 2008. The *Strongyloides*
332 (Nematoda) of sheep and the predominant *Strongyloides* of cattle form at least two
333 different, genetically isolated populations. Vet Parasitol 157, 89-99.

334

335 Fisher, M.C. and Viney, M.E. 1998. The population genetic structure of the facultatively
336 sexual parasitic nematode *Strongyloides ratti* in wild rats. Proc Roy Soc Lond B 265,
337 703-709.

338

339 Hunt V.L., Asai I.J., Coghlan A., Reid A.J., Holroyd N., Foth B.J., et al. 2016. The
340 genomic basis of parasitism in the *Strongyloides* clade of nematodes. Nat Genet 48,
341 299-307.

342

343 Katoh, K., Asimenos, G. & Toh, H. 2009. Multiple alignment of DNA sequences with
344 MAFFT. Meth Mol Biol 537, 39-64.

345

346 Ko, P.P., Sakaguchi, K., Yoshida, A., Maruyama, H., Nonaka, N. and Nagayasu, E.
347 2019. First molecular identification of *Strongyloides vituli* in cattle in Japan and insights
348 into the evolutionary history of *Strongyloides* parasites of ruminants. Parasit Int 72,
349 101937.

350

351 Ko, P.P., Haraguchi, M., Hara, T., Hieu, D.D., Ito, A., Tanaka, R., et al. 2023.
352 Population genetics study of *Strongyloides fuelleborni* and phylogenetic
353 considerations on primate-infecting species of *Strongyloides* based on their
354 mitochondrial genome sequences. Parasit Int 92, 102663.

Pinn, T.L., Forrestal, A.M., Duhamel, G.E., Crouch, E.E., Thompson, B.S. and Lejeune, M., 2022. *Strongyloides papillosus* causes sudden death in weaned calves on New York dairies. J Am Vet Med Ass 260, 244-250.

Stamatakis, A. 2006. RaxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinf 22, 2688-2690.

Thamsborg, S.M., Ketzis, J., Horii, Y. and Matthews, J.B., 2017. *Strongyloides* spp. infections of veterinary importance. Parasitol 144, 274-284.

Turner, J.H., Shalkop, W.T. and Wilson, G.I., 1960. Experimental strongyloidiasis in sheep and goats. IV. Migration of *Strongyloides papillosus* in lambs and accompanying pathologic changes following percutaneous infection. Am J Vet Res 21, 536-546.

Viney, M., 2017. *Strongyloides*. Parasitol 144, 259-262.

Viney, M.E. and Lok, J.B. 2007. *Strongyloides* spp. WormBook, ed. The C. elegans Research Community, *WormBook*, doi/10.1895/wormbook.1.7.1, http://www.wormbook.org

Zhou, S., Harbecke, D. and Streit, A. 2019. From the feces to the genome: a guideline for the isolation and preservation of *Strongyloides stercoralis* in the field for genetic and genomic analysis of individual worms. Parasit Vect 12, 496.

380 **Figure legends**

381

382 **Figure 1.** Population genomic analyses of 37 *S. papillosus* larvae from 21 goat hosts.
383 (A) Neighbour-joining tree of the nuclear genome (full view to the left, and the central
384 region enlarged to the right), with individual worms coloured according to the
385 ADMIXTURE groups, with the eight admixtured worms shown as dotted circles and
386 the mitochondrial clades shown as numbered (mt1, mt2, mt3) arcs and (B)
387 ADMIXTURE analysis showing the degree of ancestry on the y-axis, with K = 10. In
388 (A), the two worms from ADMIXTURE group 3 (green) are from different hosts; two
389 worms in ADMIXTURE group 1 (red) are from the same host.

390

391 **Figure 2.** A maximum likelihood tree of the mitochondrial genome of the 37 samples
392 and the *S. papillosus* reference mitochondrial genome.

393

394 **Supplementary Figures**

395

396 **Supplementary Figure 1.** Principal component analysis showing the first two
397 principal components, which accounts for 38% of the variance.

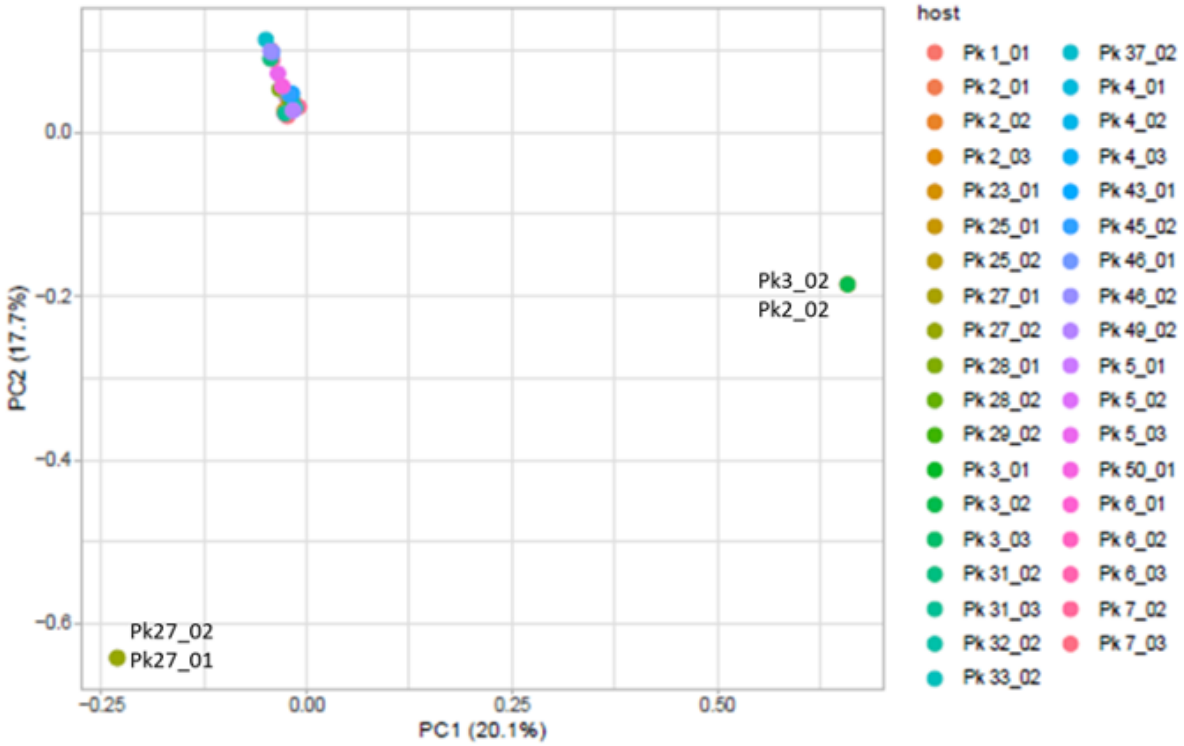
398

399 **Supplementary Figure 2.** ADMIXTURE analyses for k = 3, 7 and 15.

Supplementary Material

Supplementary Figure 1. Principal component analysis showing the first two principal components, which accounts for 38% of the variance.

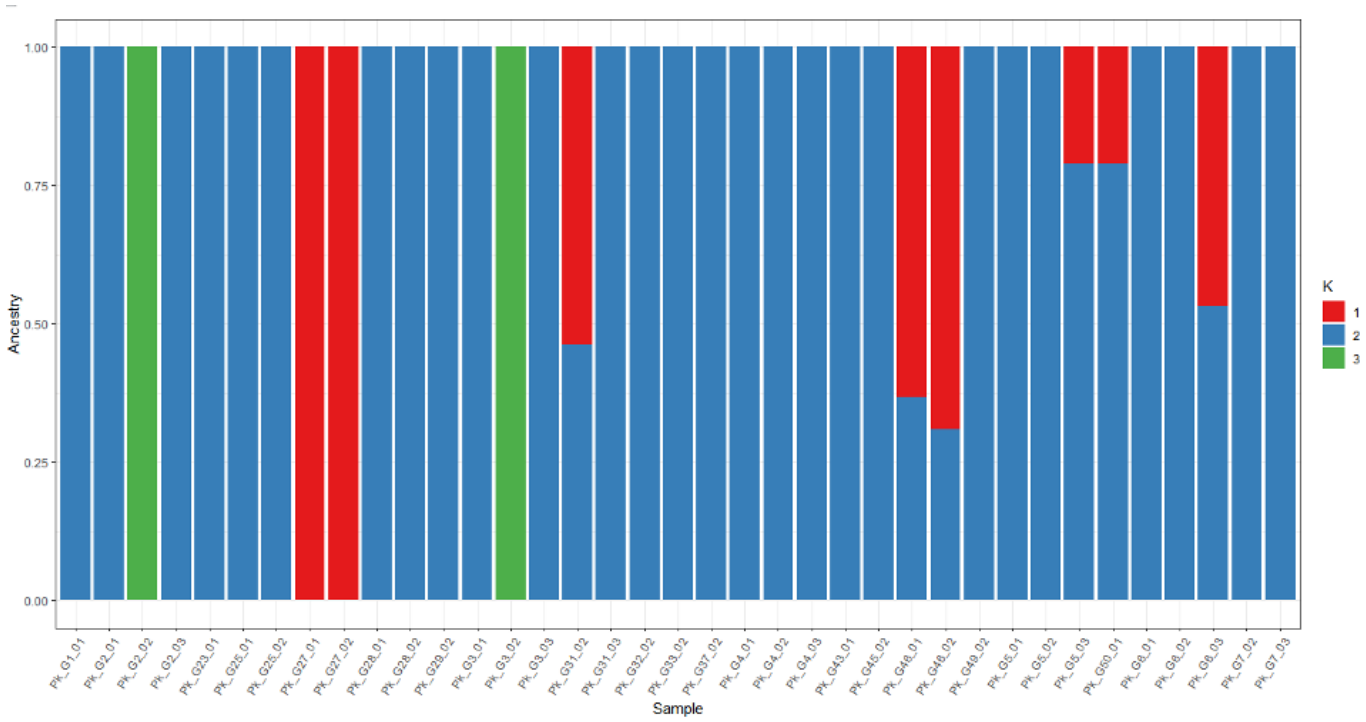
Supplementary Figure 1



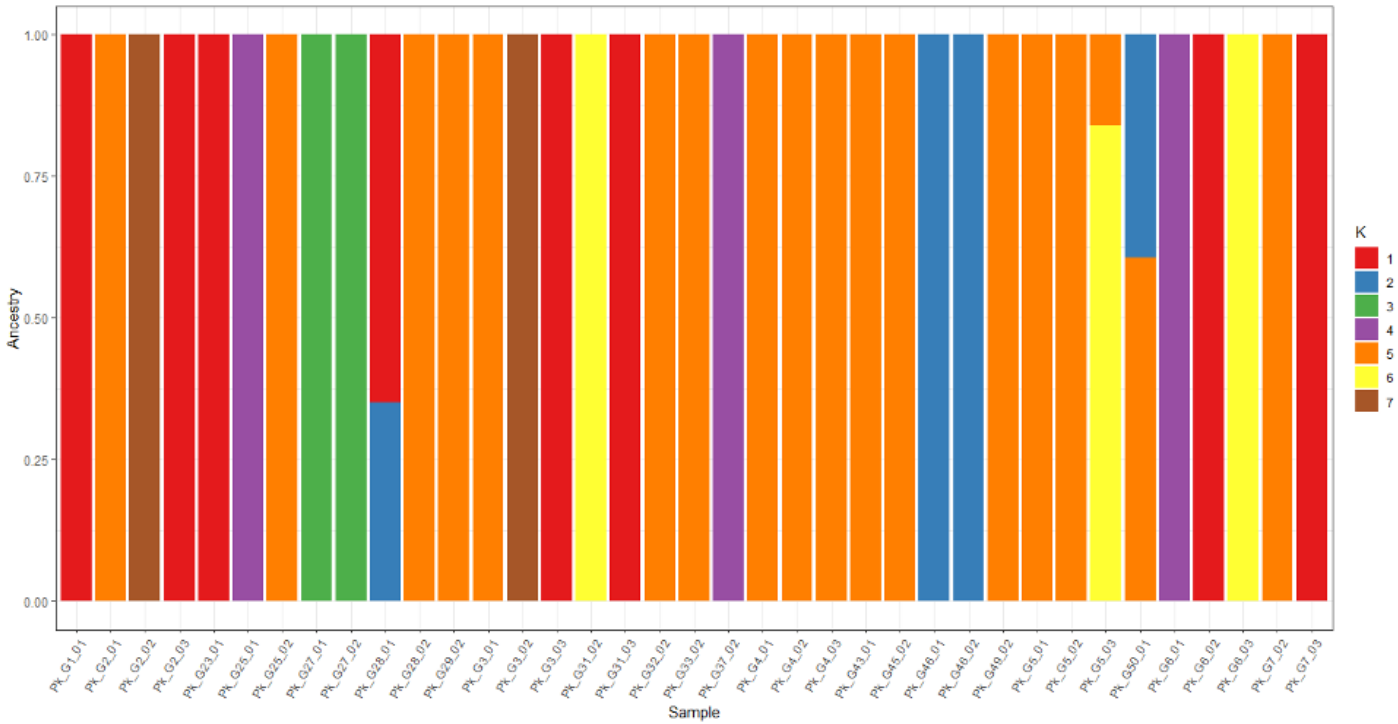
Supplementary Figure 2. ADMIXTURE analyses for k = 3, 7 and 15.

Supplementary Figure 2

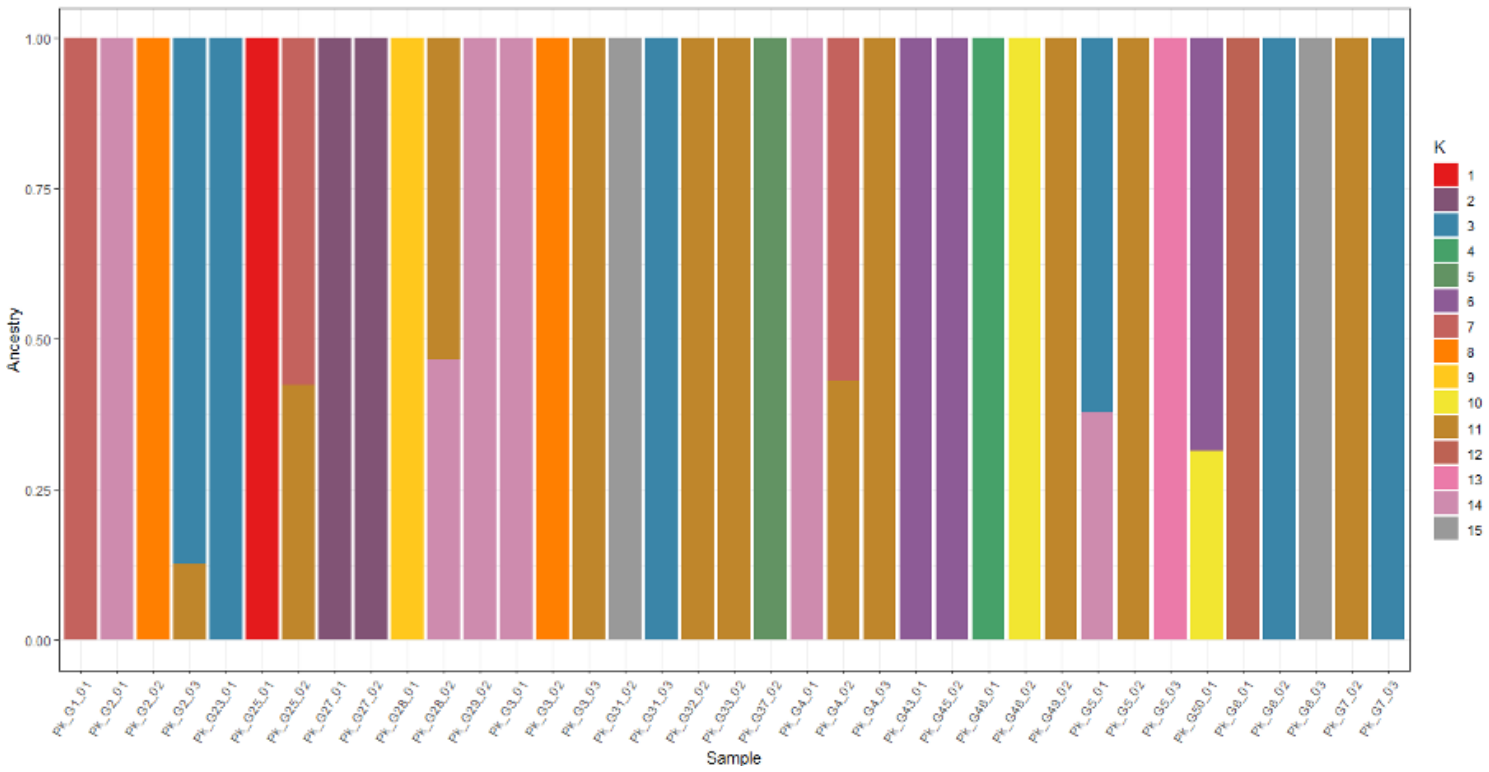
k = 3



k = 7



k = 15


1.00
0.75
0.50
0.25
0.00
Ancestry
Sample
K
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15