Review by anonymous reviewer 1, 20 Sep 2023 10:59

The paper is an interesting read and makes a very valuable contribution to the equine parasitology research community. The authors are to be commended for taking a very methodical and careful approach to delivering a high-quality genome for a parasite species that has proved challenging in the past. Cyathostomins are not very tractable organisms with which to work and this manuscript reports a first success in provding a chromosomal level genome assembly. The work is very well done and is articulately described in the text. It provides interesting insights in cyathostomin biology and moves on the field significantly. The paper is generally well written and the authors provide convincing evidence for almost all of their conclusions. I have just one query for them to address:

My query centres around the section 'Transcriptomic differences across sexes upon pyrantel exposure'. The authors propose that differences in the expression profiles between males and female may be due to developing eggs in the females or sex-specific differences in their response to pyrantel. However, there could be many reasons for the differences between the sexes, gastrointestinal nematode infections are not equal in terms of the number of male and female worms residing in the gut. Female worms are often more numerous and male worms can be polyandrous by nature. This raises the prospect of males being more motile as they move from mate to mate, which is one other way to explain differences between the sexes. There will likely be many other differences between the sexes. I remain unconvinced that the difference in their data is due to sex-dependent sensitivity as stated in lines 610-611 and would prefer to see some further discussion of these data.

Minor typos and text edits

1. Line 60 'on' rather than 'to' per and livestock species

2. Line 62 'to' the brink of distinction

3. Line 65 'anthelmintic'

The three aforementioned edits were made.

4. Lines 70 and 72 'S. vulgaris' and 'C. nassatus' should be in italics, if fact there are many examples throughout the text where parasite names are used and not italicised. They all need to be corrected.

Species names were italicised throughout the manuscript and references accordingly.

5. Lines 71-73 use a comma rather than a hyphen

6. Line 75 'and' zebras

7. Line 108 'decreased'

8. 129 'Ukraine to Kentucky'

The four aforementioned suggestions have been applied.

9. Line 130-132 sentence needs reworking – what does ‘in that species’ refer to?

The sentence was changed to: “*Finally, comparison of the genetic diversity of pyrantel-resistant and -susceptible isolates identified novel candidates for pyrantel resistance in C. nassatus. Orthologs of these candidate genes that species that were found differentially expressed in evolved C. elegans lines.*“

10. Line 317 ‘end-repaired is repeated’

This was modified accordingly.

11. Line 423 ‘Cylicostephanus’ spelling

This was modified accordingly.

12. Line 504 ‘The last century affected allele frequencies over known anthelmintic drug targets’ sounds a little clumsy. Do the authors mean ‘Allele frequencies of known anthelmintic targets has altered over the last century’? Or ‘Alteration in allele frequencies over the last century are found within genomic regions encoding anthelmintic drug targets? I think this title needs rewording.

The title has been changed.

13. Line 562 Figure 3 legend a.’ The number of private SNPs for a given isolate’

14. Line 585 (reference isolate – this mean STR?)

15. Table 1 legend ‘Bold names were covered by at least one? Decisive SNP….’

Comments 13 to 15 have been modified as suggested.

16. Figure 5 The legend and citation of this figure in the text is somewhat confusing and requires clarification. Some further explanation of the relationship between FECRT and allele frequency needs to be provided.

This has been changed accordingly. The legend was rephrased as:

The three left panels (a,d,f) represent the statistical support of the association (measured as a Bayes Factor, BF) between SNP allele frequency with pyrantel resistance on three chromosomes of interest (the red dashed line stands for the significance cut-off) for decisive associations) after correction for population structure. The analysis was focused on contemporary isolates with known pyrantel sensitivity, i.e. four isolates from Normandy, France (two drug-resistant: STB and STL, and two drug-susceptible: SBA and STG), the reference drug-susceptible isolate (STR) and the drug-resistant isolate from Kentucky, USA (SKE). For each chromosome, the association between the raw allele frequency of the most likely candidate gene is plotted against the pyrantel sensitivity for every considered isolate (panels b & c for chromosome 2, e & g for chromosomes 1 and 4 respectively).

h. Differential gene expression profile (between male and female worms exposed to pyrantel) for the candidate genes laying within identified QTL regions (colours match the respective

chromosomes as in panels a, d, f). i. Significance and expression fold change for the genes differentially expressed between control and pyrantel-selected Caenorhabditis elegans *lines.*

17. Line numbers now absent – First paragraph page 21. 'can be reproduced in another organism' – this sentence needs clarification as to what organisms this is referring to.

Line numbers were added and the sentence was removed.

18. Line 9 page 22 'about twice the size' or twice as large'

Modified accordingly.

19. Paragraph 3 for 'complexifies' substitute 'becomes more complex, as it remains to be determined….'?

Modified accordingly.

Review by anonymous reviewer 2, 01 Oct 2023 12:08

The manuscript "*Spatio-temporal diversity and genetic architecture of pyrantel resistance in Cylicocyclus nassatus, the most abundant horse parasite*" by Sallé et al. presents a detailed study of the genome and genetic diversity of *Cylicocyclus nassatus*, the most prevalent species from the cyathostomin complex, a group of parasitic nematodes that infect horses and wild equids. The manuscript is the result of an impressive set of work: the authors have taken an important parasite from a clade of nematodes with practically no genomic data and have generated a very high-quality genome, a resequencing dataset, and sex-specific transcriptomes. They use these datasets to explore the genetic basis of pyrental resistance. The results are interesting and highlight several candidate loci that may play a role in pyrantel sensitivity. The manuscript will be of interest to those interested in parasite genetics and drug resistance in nematodes and flatworms. I commend the authors for making all data available on the relevant INSDC databases, which facilitated my review of their manuscript.

My only major comment is that the manuscript, in some places, suffers from an overly simplistic or one-sided interpretation of the results (which I detail below). In most cases, this doesn't invalidate the major claims of the manuscript, but providing a more nuanced discussion of the results would really help readers understand the strengths and limitations of the study.

General comments

Although the reference genome is evidently of high quality, I found the presentation of the genome assembly results lacking detail in places, which led me to download the genome and annotation and calculate my own metrics. The following additions would help:

- "666,884 Kbp" would be better as 666.8 Mbp.

This was modified accordingly.

- The authors only quote scaffold N50 of the final assembly, but it would be good to quote the contig N50 as well (which looks to be 671 kb).

The contig N50 was 674,848 bp.

- The post-purging assembly size was 666.8 Mb, but the final assembly size was 514.7 Mb - what happened to the other ~150 Mb during Hi-C scaffolding and curation?

The hifiasm assembly was harbouring numerous haplotypic contig copies meaning that both haplotypes were present in the assembly. These duplicates were located one against the other after Hi-C scaffolding and presented a specific counter diagonal pattern which enables to remove one of them. This explains the reduction in size between contigs and scaffolds. This was already specified in main text as "*Following Hi-C data curation*".

- The authors list BUSCO completeness which, based on the text, is based on running BUSCO on the proteome. It would be useful to also present the BUSCO completeness of the genome because gene sets may have lower completeness due to missing genes. It would also be useful to contextualise both values by comparing them to e.g. the *H.*

contortus reference genome and annotation (my analysis suggests that the C. nassatus genome is more complete, at least at the genome level, than *H. contortus*).

Thank you for this suggestion. To address this comment, we have inserted the following under the corresponding result section: *Following Hi-C data curation, a chromosome-scale assembly of 514.7 Mbp was built (scaffold N50 = 91,661,033 bp, scaffold L50 = 3; contig N50 = 674,848 bp; Table 1, Fig. 1a, Fig. S1) with high degree of completeness as supported by the 85.8% complete BUSCOs found that slightly outperformed the H. contortus genome statistics of 82.7%. The difference between the two genome assemblies mostly owed to higher missing BUSCOs in the H. contortus assembly (9.6%, n = 300) relative to C. nassatus (6.4%, n = 203).*

- Adding the coverage of the HiFi data (along with whether it was derived from PacBio LI or ULI or both) would be also useful.

To address this comment, the following was added at the end of the materials and methods section (*PacBio sequencing of a single C. nassatus worm*): "*The low- and ultra-low input protocols yielded 6 and 26 Gbp respectively, corresponding to 615,427 and 2,780,682 CSS reads respectively and 10 and 44x depth of coverage.*"

In my opinion, the authors' analysis of gene family evolution adds very little to the manuscript:

We broadly agree with this view although we decided to describe as many genomic features as possible. To account for this comment and the suboptimal analyses performed, this section has been moved into the supplementary data while we kept figure 1d unchanged to illustrate the phylogenomic position of *C. nassatus*. Comments on the limitations of our approach has been added under the discussion section as follows:

- When deciding which nematode species to analyse, the authors used only those with scaffold N50s above 1 Mb (presumably to avoid their analysis being influenced by poor-quality assemblies). However, this necessarily led to a very spare sampling of the tree, including just three species of strongylid. The lack of dense sampling within Strongylida means that it's hard to know if the gene families they highlight are the result of expansion in *C. nassatus* specifically, the result of expansion in the suborder Strongylina generally, or the result of contraction in *A. ceylanicum* and/or *H. contortus*. A more dense sampling, even if it included lower quality assemblies, would have provided far greater resolution (and likely far more informative results).

We acknowledge that the sampling was restricted to a limited set of species. This restriction aimed to minimize biases linked to assembly fragmentation that may be associated with poorer genome annotations.

- The authors also comment on the expansion of a family of transposons and how that may explain the larger genome size of C. nassatus. While I have no doubt that transposon proliferation has led to an increase in genome size in *C. nassatus*, analysing individual gene families via orthology clustering is not the most appropriate analysis to do this. If the authors do want to understand genome size evolution in *C. nassatus*/strongylids (and I'm not convinced they need to, given all the other work they present), it would be far better to use repeat annotations for a variety of species and

analyse which classes appear to have proliferated (or been lost) and what their relative contribution to genome size is.

Thank you for this suggestion. As indicated above, the whole section has been transferred into the supplementary material. Reports on transposons families have been removed.

I am not an expert on population genetic analyses, so other reviewers may be able to critically assess those sections better than I could, but I do have the following comments/questions:

- I was surprised not to see more discussion of the genome-wide patterns of genetic diversity in the main text, especially given the authors generated a chromosome-level reference genome. How was genetic diversity distributed through the genome? Was the pattern qualitatively similar to that reported for other nematodes (e.g. *H. contortus*)?

Genetic diversity was indeed presented as summary statistics as no particular pattern was found. We instead discussed variations in pi either between chromosomes or candidates of interest.

- As far as I can tell, the author's only evidence that present-day *C. nassatus* populations have less genetic diversity than those that existed ~200 years ago was that the single ancient Egyptian isolate has more genetic diversity than any of the present-day isolates. Without having species-wide sampling of *C. nassatus*, surely the authors cannot say whether this is due to a loss of genetic diversity in modern populations over the last century, or simply because e.g. North African populations have higher genetic diversity (both in the present-day and in the past). The authors should consider this possibility in the results and discussion and note the under/un-sampled regions of the world.

Thank you for raising this issue. This is indeed a major limitation owing to the lack of worm samples from (Northern) Africa or other major regions of the world, e.g. eastern Asia, southern America and Oceania. To deal with this comment, the abstract was modified and the following was added:

- under the results section: "*The old Egyptian worms had both the highest nucleotide diversity estimates (0.087% difference with modern isolates, t = 37.19, P < 10-4) and the highest count of private variants, which would be compatible with a global loss of diversity in C. nassatus populations in the sampled isolates over the last century or higher diversity within North African worms. Investigation of that latter hypothesis would require contemporary Egyptian worms that were not available for this study.*"
- Under the discussion section:"*Similar to observations made between ancient and modern T. trichuria (Doyle, Søe, et al. 2022), lowersignificant loss of nucleotide diversity was found in contemporary isolates relative to old North African C. nassatus individuals. In that respect, additional sampling effort from contemporary north African worms would help resolve the temporal and geographical contributions to the observed contrast in nucleotide diversity.*"

The authors used a range of techniques and approaches to sequence the *C. nassatus* genome, but it was often unclear why they used the methods they did, and how the data were used during assembly. Given their genome sequencing/assembly methods are likely to be of interest to many readers, it would be great to be clearer in the methods and results about what they did:

- The authors isolated DNA from a single worm and an aliquot of this DNA was used for REPLI-G amplification, while the rest was left unamplified. What was the amplified DNA used for?
  The authors used both the PacBio LI (which doesn't involve amplification) and ULI protocol (which does use amplification). What was the reasoning for this? How were the data used during assembly? How did they compare?

To produce as many data as possible out of the same individual while minimising putative biases associated with amplification protocols, we reasoned that raw DNA could be subjected to amplification and PacBio LI on one hand, or ULI protocol on the other hand. The amplified DNA was hence subjected to PacBio LI and the unamplified raw DNA left was used for the ULI PacBio protocol. Both datasets were subsequently combined for the first assembly round.

- The authors mention using 1 µg of DNA for PacBio LI - how much DNA did the authors isolate in total from their worm? 1 µg is a surprisingly large amount of DNA from a single worm (although it's possible I'm underestimating how large these nematodes are).

This is an overlooked mistake in main text. In our hands, extractions from single cyathostomins yielded 10.7 ± 4.83 ng per worm on average (n = 130 worms sampled for the population study; variation owing to sample sex, sampling data and likely bacterial contamination). For the very reference genome sample, 5.39 ng of raw DNA was available before amplification.

The bias in sex ratios after pyrantel treatment is interesting. However, I am not fully convinced that the transcriptomic differences the authors identify are the result of sex-specific differences in response to pyrantel treatment:

- The authors implicitly assume that the sex ratio in the infecting population is 50/50, meaning that the bias they observe after pyrantel treatment must be due to sex-specific differences in response. Although that seems likely, is there any information on sex ratios in an adult (in vivo population) of cyathostomins or other strongyles? Is it possible that there are normally more females than males? Do females live longer than males over the course of the infection?

To our knowledge, worm life expectancy is not known and limited evidence is available on cyathostomin sex ratio. Upon deworming with ivermectin, a 0.58 male to female ratio was observed in Polish horses (Sallé et al. 2018) while that ratio was 40% (Anjos & Rodrigues 2006) and 54% (Silva et al. 1999) from horses necropsied in Brazil. In that respect, our observations (male to female ratio of 20%) after pyrantel treatment remains particularly low and would be compatible with differential pyrantel sensitivity between sexes. However, pyrantel resistance was not sex-biased in A. caninum, another clade V parasitic nematode (Kopp et al. 2007).

To deal with this comment, the following was added under the corresponding paragraph: "*This 20% male to female ratio was lower than previous reports ranging from 0.4 to 0.6 (Silva et al. 1999; Anjos and Rodrigues, 2006; Sallé et al, 2018).*"

- The authors suggest that differences between the male and female transcriptomes transcriptome must either be the result of egg development in utero in females or transcriptomic differences due to pyrantel treatment and therefore, after removing genes that are believed to be expressed in the oocyte, germ line or embryonic lineages, all remaining differentially expressed genes are the result of pyrantel treatment. This seems overly simplistic and ignores the fact that there are likely general differences in male and female transcriptomes. Indeed, a quick scan of the literature suggests that sex-specific transcriptome analyses consistently find hundreds of genes that show sex-specific differences in expression (in the absence of an external stressor, like pyrantel).

We agree that the current wording is overly enthusiastic about the putative sex-biased pyrantel effect and our design cannot conclude on that without the addition of unexposed male and female worms. To better reflect this limitation, the corresponding paragraph has been reworded to start as follows: "*To delineate somatic transcriptional differences between males and females not owing to the presence of eggs in utero, we selected C. nassatus DE genes with one-to-one ortholog in C. elegans that were not known to be expressed in any of the oocyte, germ line or embryonic stage of that model species."*

The conclusion of the corresponding paragraph was also tone down.

- I think it's also important to acknowledge the lack of replication in the dataset. I realise how difficult this would be to achieve but simply adding more detail to the results about what samples were used for transcriptome sequencing would facilitate the interpretation of the results.

To deal with this comment, the number of replicates and individuals was added under the results section.

Minor comments

The term 'telomere-to-telomere' refers to gap-free genome assembly (e.g. the new human reference genome, or the C. elegans reference genome). There are no telomere-to-telomere strongylid genomes. Better to use 'chromosome-level' or similar.

This was changed accordingly.

The Hi-C plot in Figure S1 appears to be truncated.

The plot was modified to provide a complete view of the Hi-C contact map.

There doesn't seem to be any information in the methods about how the *C. nassatus* genome was aligned to the *H. contortus* genome (although promer is noted in the Figure S3 caption). Also, using one-to-one orthologues, rather than promer, might reduce some of the noise seen in Figure S3.

To deal with this comment, the following was added under the materials and methods section : "*The synteny between the C. nassatus and H. contortus genomes was inferred after aligning both references against one another, using the PROMER software (Delcher et al. 2002) with*

the -mum option to recover the only exact matches being unique between the query and reference sequences. Hits were subsequently filtered with the delta-filter tool, setting the minimal length of a match at 500 amino acids. A custom perl script (Cotton et al. 2017) was subsequently used to draw the circus plot, showing links with 70% similarity and a minimal length of 10 Kbp."

It might just be me but I think it's better to write 19th century rather than XIXth.

This was changed accordingly.

The authors might want to define what they mean by 'decisive SNP'.

This was changed to" pyrantel-associated SNPs" and the considered Bayes Factor was indicated in the legend.