Characterization of the complete mitochondrial genomes of two sibling species of parasitic roundworms, Haemonchus contortus and Teladorsagia circumcincta

Nikola Palevich
AgResearch Limited, Grasslands Research Centre

Paul H Maclean
AgResearch Limited, Grasslands Research Centre

Young-Jun Choi
Washington University School of Medicine in St. Louis

Makedonka Mitreva
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
Palevich, Nikola; Maclean, Paul H; Choi, Young-Jun; and Mitreva, Makedonka, "Characterization of the complete mitochondrial genomes of two sibling species of parasitic roundworms, Haemonchus contortus and Teladorsagia circumcincta." Frontiers in genetics. 11, . (2020). https://digitalcommons.wustl.edu/open_access_pubs/9716

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Characterization of the Complete Mitochondrial Genomes of Two Sibling Species of Parasitic Roundworms, *Haemonchus contortus* and *Teladorsagia circumcincta*

Nikola Palevich*, Paul H. Maclean, Young-Jun Choi and Makedonka Mitreva

*AgResearch Limited, Grasslands Research Centre, Palmerston North, New Zealand, 2 McDonnell Genome Institute and Department of Medicine, Washington University School of Medicine, Saint Louis, MO, United States

*Correspondence: Nikola Palevich nik.palevich@agresearch.co.nz

Haemonchus contortus and Teladorsagia circumcincta are among the two most pathogenic internal parasitic nematodes infecting small ruminants, such as sheep and goats, and are a global animal health issue. Accurate identification and delineation of Haemonchidae species is essential for development of diagnostic and control strategies with high resolution for Trichostrongyloidea infection in ruminants. Here, we describe in detail and compare the complete mitochondrial (mt) genomes of the New Zealand *H. contortus* and *T. circumcincta* field strains to improve our understanding of species- and strain-level evolution in these closely related roundworms. In the present study, we performed extensive comparative bioinformatics analyses on the recently sequenced complete mt genomes of the New Zealand *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP field strains. Amino acid sequences inferred from individual genes of each of the two mt genomes were compared, concatenated and subjected to phylogenetic analysis using Bayesian inference (BI), Maximum Likelihood (ML), and Maximum Parsimony (MP). The AT-rich mt genomes of *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP are 14,011 bp (A+T content of 77.4%) and 14,081 bp (A+T content of 77.3%) in size, respectively. All 36 of the typical nematode mt genes are transcribed in the forward direction in both species and comprise of 12 protein-encoding genes (PCGs), 2 ribosomal RNA (rrn) genes, and 22 transfer RNA (trn) genes. The secondary structures for the 22 trn genes and two rrn genes differ between *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP, however the gene arrangements of both are consistent with other Trichostrongyloidea sequenced to date. Comparative analyses of the complete mitochondrial nucleotide sequences, PCGs, A+T rich and non-coding repeat regions of *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP further reinforces the high levels of diversity and gene flow observed among Trichostrongyloidea, and supports their potential as ideal markers for strain-level identification from different
hosts and geographical regions with high resolution for future studies. The complete mt genomes of *H. contortus NZ_Hco_NP* and *T. circumcincta NZ_Teci_NP* presented here provide useful novel markers for further studies of the meta-population connectivity and the genetic mechanisms driving evolution in nematode species.

**Keywords:** *Haemonchus contortus*, *Teladorsagia circumcincta*, mitochondrial genome, helminth, parasite, anthelmintic-susceptible

**HIGHLIGHTS**

- In-depth comparative and phylogenomic analyses of the recently published complete mitochondrial (mt) genomes of the New Zealand *H. contortus NZ_Hco_NP* and *T. circumcincta NZ_Teci_NP* field strains to improve our understanding of species- and strain-level evolution in roundworms.
- Analysis of the A+T rich and non-coding repeat regions of the complete *Haemonchus* and *T. circumcincta* mt genomes supports their potential as ideal markers for easy species- and possibly strain-level identification with high resolution for future studies.
- Phylogenetic relationships of complete mt genomes for all Trichostrongyloidea species currently available inferred from Bayesian, Maximum Likelihood and Maximum Parsimony analyses of mitochondrial genes, revealed high levels of intra-species diversity and a panmictic structure observed among Trichostrongyloidea.
- Our findings indicate that a three-pronged approach incorporating phylogenetic inertia, pan-genome structure/features and environmental data are needed in order to understand the mitochondrial genome evolution.

**INTRODUCTION**

*H. contortus* (barber’s pole worm) and *T. circumcincta* (brown stomach worm), are the most economically important pathogenic nematodes infecting small ruminants (sheep and goats) worldwide. These blood-feeding strongylid nematodes are orally transmitted via contaminated pasture to the host where they infect the fourth stomach (abomasum) reducing animal production and causing anemia, edema, and associated complications often leading to death (Vlassoff et al., 2001; Sutherland and Scott, 2010). Although these parasites can be managed using existing prophylactic drugs (anthelmintics), their remarkable natural tendency to develop resistance combined with the diminishing efficacy of compounds used, threatens the global livestock industry (Kaplan, 2004; Kaplan and Vidyashankar, 2012). While these monoxenous and obligately sexual species have identical life-cycles (Veglia, 1915), the gene flow and evolutionary relationships observed among the Trichostrongylidae superfamily, and the Strongylida suborder as a whole, remain unclear.

Parasitic roundworms belonging to the Haemonchinae family are able to infect a range of ruminant hosts. For example and on communal pastures in particular, *H. placei* is primarily a cattle parasite but can also infect small ruminants (Amarante et al., 1997; Gasser et al., 2008; Nunes et al., 2013), and vice versa for *H. contortus* (Jacquet et al., 1998). Given the increase in availability of genomic information resources (Palevich et al., 2018), it is crucial for the management of parasitic diseases that readily accessible molecular tools are developed to enable end-users to easily determine which species is infecting which animal host.

The development of low cost, high-throughput, next-generation sequencing technologies have led to a recent focus on whole-genome sequencing (WGS) of complete nuclear genomes of parasitic roundworms (Nematodes) (Hu and Gasser, 2006; Hu et al., 2007; Jex et al., 2008; Palevich et al., 2019a,b), instead of earlier markers such as *cox1*, *cox2*, and *nad* genes. Owing to the substantial coverage following WGS (Wit and Gillarde, 2017; Palevich et al., 2019c) and advances in bioinformatics pipelines, further complete mitochondrial genome resources will be available to facilitate detailed comparative phylogenetic analyses across many species complexes and major taxonomic groups of nematodes (Hu et al., 2004; Gasser et al., 2016). In general, complete nematode mt genomes are ∼12–21 kbp circular-DNA molecules (Hu and Gasser, 2006). Nematode mt genomes can provide rich sources of informative markers and typically feature (Wolstenholme, 1992; Boone, 1999; Saccone et al., 1999; Blouin, 2002): 12 protein-coding genes (PCGs) and lack of the *atp8* gene [except for *T. spiralis* (Lavrov and Brown, 2001)]; 2 ribosomal RNA (*rrn*) genes (Gutell et al., 1993), and 22 transfer RNA (*trn*) genes (Wolstenholme et al., 1987); unidirectional transcription (usually in forward direction) with particularly unique initiation codons; and highly variable gene arrangements (Hu and Gasser, 2006). The gene sequences, and in particular the 12 PCGs, are relatively conserved are useful phylogenetic markers that have been used for examining phylogenetic relationships within the phylum Nematoda (Hu et al., 2004). The mitochondrial genetic code of nematodes follows translation (*transl_table = 5*)1 of National Center for Biotechnology Information (NCBI). In addition, nematode mt genomes also possess transposition genotypes of tRNAs that lack either a TΨC or DHU stem in the *trn* secondary structures (Wolstenholme et al., 1994; Yokobori, 1995). In the pursuit of improving the phylogenetic resolution within the phylum Nematoda, mt genomes serve as excellent markers for investigation of species delineation and the genetics of population evolution. Future efforts should focus on the availability of more complete mt genomes across all nematode species, and especially for conspecific strains.

For these reasons, this study interrogates the complete mitochondrial genome sequences of the anthelmintic-susceptible species...
H. contortus NZ_Hco_NP (Palevich et al., 2019a) and T. circumcincta NZ_Teci_NP (Palevich et al., 2019b) derived from pasture-grazed New Zealand sheep. In this study, we performed numerous comparative and phylogenomic analyses to directly compare the molecular characteristics including the nucleotide composition and codon usage profiles of PCGs, as well as the secondary structures of each identified tRNA and rRNA gene within the two species, but also among other closely related trichostrongyloid nematodes. The phylogenetic position of H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP among Trichostrongylidae and of other species of socio-economically important parasites within the phylum Nematoda, are investigated based on mitochondrial nucleotide sequences and PCGs. This research provides insights into the mt genome evolution that may facilitate the development of molecular tools to differentiate between these two sibling species of parasitic roundworms at the strain level.

MATERIALS AND METHODS

Sample Collection and DNA Extraction
The acquisition of parasite samples, DNA extraction, and genome sequencing technologies used to generate the mt genomes used for analysis in this study have recently been outlined in the form of brief reports (Palevich et al., 2019a,b). To fill any knowledge gaps associated with the above-mentioned reports, here we describe in detail the methodologies, sequencing technologies, and annotation software used to generate the H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP mt genomes. H. contortus and T. circumcincta were recovered from sheep infected with pure strains of parasites in Palmerston North, New Zealand, and total high molecular weight genomic DNA was extracted separately from multiple worms using a modified phenol-chloroform protocol, as previously described (Choi et al., 2017; Palevich et al., 2019a,b,c). DNA was deposited, stored and available upon request from AgResearch Ltd., Grasslands Research Center.

Mitochondrial Genome Sequencing and Annotation
The complete mitochondrial genome sequences of H. contortus NZ_Hco_NP (Palevich et al., 2019a,c) and T. circumcincta NZ_Teci_NP (Palevich et al., 2019b) were obtained using next-generation sequencing technologies. The H. contortus NZ_Hco_NP (BioProject ID: PRJNA517503, SRA accession number SRP247265) whole genome shotgun paired-end (PE) and Single-Molecule, Real-Time (SMRT) long-read sequencing (SRA Runs: SRR11022845 and SRR11022846) were generated using the Illumina HiSeq2500 and Pacific Biosciences (PacBio) platforms, as previously described (Palevich et al., 2019a,c). The T. circumcincta NZ_Teci_NP (BioProject ID: PRJNA72569, SRA accession number SRP007648) whole genome shotgun library (SRA Run: SRR3145390) was generated as previously described (Tang et al., 2014; Palevich et al., 2019b), and sequenced by the Genome Center at Washington University (WUGSC, School of Medicine, St. Louis, MO, USA) using the Illumina MiSeq platform with the strategy of 150 bp paired-end sequencing mode. A total of 605 million raw reads were generated and made available in FASTQ format. The quality of the raw sequence reads was evaluated using the software package [http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Andrews, 2010)]. The software Trimmomatic v.0.36 (Bolger et al., 2014) was used for removal of adapter, contaminant, low quality (Phred scores <30), and short (<50 bp) sequencing reads.

The remaining high-quality sequencing reads were assembled de novo using the NOVOPlasty pipeline v.3.1 with default parameters and based on a kmer size of 39 following the developer's suggestions (Dierckxsens et al., 2016). The assembled mt genomes, tRNA and rRNA annotations were checked using the MITOS web server based on translation (transl_table = 5) (Invertebrate Mitochondrial) of NCBI (http://mitos.bioinf.uni-leipzig.de). Repeat-rich subsets of the mt genome assemblies were identified with varying degrees of complexity using the mDUST v.2006.10.17 (Morgulis et al., 2006) software tools with default settings. The sequence regions containing tandem repetitive elements were identified using Tandem Repeat Finder software online server (http://tandem.bu.edu/trf/trf.html). The coordinates of the found repeats were matched with coding sequences and the proportion of repeats overlapping coding regions was calculated. tRNA genes were further corroborated using the software SPOT-RNA (Singh et al., 2019) and the tRNA secondary structures were predicted using the ViennaRNA v.2.4 (Lorenz et al., 2011) and VARNA v.3.93 (Darty et al., 2009) packages. tRNA secondary structures and complete mt genomes were visualized using the Geneious Prime v.2019.1.1 (Kearse et al., 2012). The PCGs were manually curated by searching for ORFs (employing genetic code 5 as above) and alignment against the available H. contortus and T. circumcincta reference genomes in Geneious Prime v.2019.1.1 (Kearse et al., 2012).

Mitochondrial Pangenome Analysis
The entire nucleotide sequences alignments using Mauve (Darling et al., 2004) were performed using Geneious Prime v.2019.1.1 (Kearse et al., 2012). Sequence differences among the Haemonchus and Teladorsagia species as well as between the respective strains was done using the H. contortus NZ_Hco_NP (Palevich et al., 2019a,c) and T. circumcincta NZ_Teci_NP (Choi et al., 2017; Palevich et al., 2019a,b,c) as the reference sequences, respectively. Repeated sequence elements of Haemonchus and Teladorsagia mt genomes were visualized in the form of Dotplots and generated based on pair-wise alignments with chaining coverage using PipMaker (Schwartz et al., 2000). Gene synteny was investigated using BLASTX alignments of sequences and genetic similarity mapping was visualized using the CGView Comparison Tool (http://stothard.afns.ualberta.ca/downloads/CCT) (Grant et al., 2012). For analysis of codon usage and amino acid composition, codon usage counts were extracted from each protein using Geneious Prime version 2019.1.1 (Kearse et al., 2012). These were then compiled and loaded into R version v.3.6.1 (R Core Team, 2013) where a clustered image map was produced using the “cim” function from the mixOmics package v.6.8.5 (Rohart et al., 2017).
Phylogenomics Analysis of Complete mt Genomes

The phylogenetic positions of *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP among other species of nematodes available in the GenBank database were examined. The nucleotide sequences and PCGs of published Trichostrongyloidea mitochondrial genomes were retrieved from the National Center for Biotechnology Information, USA (http://www.ncbi.nlm.nih.gov/). *Caenorhabditis elegans* N2 (JF896456) was used as the outgroup species in this study. All coding sequences (CDS) used in this analysis had correct start and stop codons based on translation (transl_table = 5)\(^1\) of NCBI (Invertebrate Mitochondrial).

Considering the high degree of intraspecific variation in nucleotide sequences of mt genes of nematodes, we used the deduced amino acid sequences of mt proteins for phylogenetic analyses. Amino acid sequences from the 12 mt protein-coding genes of the complete mitochondrial genomes were deduced, aligned individually and concatenated into a single alignment using MAFFT v.7.450 (Katoh and Standley, 2013). Phylogenetic analyses were conducted using three methods: Bayesian inference (BI), Maximum Likelihood (ML), and Maximum Parsimony (MP) methods. BI analysis with the program MrBayes v.3.15265 (Huelsenbeck and Ronquist, 2001) and the MtInv model of amino acid evolution (Le et al., 2017) was selected as the most suitable model of evolution. However, since the MtInv model is not implemented in the current version of MrBayes, an alternative model, MtREV (Adachi and Hasegawa, 1996), was used in BI analysis. Four independent Markov chains were run for 1,000,000 metropolis-coupled MCMC generations, sampling a tree every 100 generations. The first 2,500 trees represented burn-in, and the remaining trees were used to calculate Bayesian posterior probabilities (PP). The evolutionary history was further inferred by using the ML and MP methods. ML analysis was performed with the MtREV General Reversible Mitochondrial + Freq. model (Adachi and Hasegawa, 1996) and MP was performed with PAUP\(^*\) v.4.0b10 (http://paup.csit.fsu.edu/) (Swofford, 2001). Both methods were calculated using 1,000 bootstrap replicates with bootstrapping frequencies (BF) were calculated. Phylograms were drawn using the software MEGA X (Kumar et al., 2018).

RESULTS AND DISCUSSION

mt Genome General Features and Characteristics

The reconstruction of partial and complete mitochondrial genomes is becoming increasingly cost-effective with the advent of next-generation sequencing technology and recent advances in molecular parasitology tools, in particular for species identification and population genetics studies (Hu and Gasser, 2006; Jex et al., 2009). Despite these advances, only 12 species of Trichostrongyloid mt genomes are currently available in GenBank with only half of these represented by multiple strains (dos Santos et al., 2017). For these reasons, the complete mt genomes of New Zealand *H. contortus* NZ_Hco_NP (Palevich et al., 2019a,c) and *T. circumcincta* NZ_Teci_NP (Choi et al., 2017; Palevich et al., 2019b) field strains were sequenced and compared in detail to investigate the strain-level genetic differences of these two sibling Trichostrongyloidea species. Recently, complete mt genomes of NZ_Hco_NP and NZ_Teci_NP were generated using various sequencing platforms and assembled from 582 and 5.6 million (PE) reads with an average coverage depth per nucleotide of 13,293\(\times\) and 2,097\(\times\), respectively.

The smallest Trichostrongyloidea mt genome published to date belongs to *Dictyocaulus* sp. cf. *eckerti* (Gasser et al., 2012) at 13,296 bp with the largest 15,221 bp mt genome belonging to *M. digitatus* (Jex et al., 2009). The NZ_Hco_NP and NZ_Teci_NP mt genomes are 14,001 and 14,081 bp in length, respectively (Figure 1) and are thus within the expected range. The reported mt genomes differ in size by \(\sim\)16–54 bp than other *H. contortus* and *T. circumcincta* characterized strains, except for the 13.7 kb *H. placei* MHpl1 mt genome (dos Santos et al., 2017). The size discrepancies among the different strains primarily relate to A+T-rich control region expansions and longer non-coding regions between numerous transfer RNA genes.

The total GC content of the *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP mt genomes were 21.1 and 22.8\% and contained an A+T bias with an overall base composition of A = 32.9 and 31.2\%, T = 44.5 and 46.1\%, C = 6.3 and 7.3\%, and G = 14.8 and 15.5\%, respectively (Table 1). The overall A+T bias of 77.4 and 77.3\% observed in *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP mt genomes, respectively, are within the typical range reported for nematode mitochondrial genomes (Hu et al., 2004; Hu and Gasser, 2006). Compared with the others, the A+T content of all *T. circumcincta* strains are within 0.1\% of each other with *H. contortus* strains being \(\pm 0.5\) of NZ_Hco_NP. Low-complexity regions (LCRs) are extremely abundant in eukaryotic proteins that are typically represented by amino acid sequences containing repeats of single amino acids or short amino acid motifs, and may facilitate the formation of novel coding sequences. *H. contortus* NZ_Hco_NP had 23 LCRs with 15 (65\%) in coding regions, while *T. circumcincta* NZ_Teci_NP had 26 repeats with 21 (81\%) found in coding regions (Figure 2), however determining the functional role of these repeats is a focus for future work. Comparative analysis of the *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP mtDNAs also revealed five and four tandem repeat (TR) regions, respectively (Figure 2).

To date, mt genomes are available for many species of Trichostrongyloidea and the mt genomes of the trichostrongyloid *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP are in agreement, as they both comprise of 12 protein-encoding genes (PCGs), 22 transfer RNA (trn) genes, and 2 ribosomal RNA genes (rrnS or 12S ribosomal RNA and rrnL or 16S ribosomal RNA). The 12 mt PCGs of both *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP were encoded on the heavy strand and transcribed in the same direction. The order of genes observed in *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP are identical to the previously reported *H. contortus* (Palevich et al., 2018) and *T. circumcincta* (Jex et al., 2009), respectively (Figure 1). In accordance to the mt genomes of
<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Position</th>
<th>Length (bp)</th>
<th>Strand</th>
<th>Start</th>
<th>Stop</th>
<th>T (%)</th>
<th>C (%)</th>
<th>A (%)</th>
<th>G (%)</th>
<th>AT (%)</th>
<th>GC (%)</th>
<th>GT (%)</th>
<th>AT skew</th>
<th>GC skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>atp6</td>
<td>Coding</td>
<td>8,809</td>
<td>9,001</td>
<td>600</td>
<td>600</td>
<td>636.7</td>
<td>636.7</td>
<td>29.503</td>
<td>16.3</td>
<td>15.8</td>
<td>75.785</td>
<td>22.6</td>
<td>22.5</td>
<td>62.6</td>
<td>62.6</td>
</tr>
<tr>
<td>cox1</td>
<td>Coding</td>
<td>1/7</td>
<td>1,582</td>
<td>1,578</td>
<td>1,582</td>
<td>1,572</td>
<td>64.6</td>
<td>46.7</td>
<td>29.528</td>
<td>19.6</td>
<td>20.3</td>
<td>69.268</td>
<td>29.331</td>
<td>62.364</td>
<td>-0.234</td>
</tr>
<tr>
<td>cox2</td>
<td>Coding</td>
<td>1,862</td>
<td>1,896</td>
<td>2,554</td>
<td>2,593</td>
<td>693</td>
<td>696</td>
<td>30.428</td>
<td>17.5</td>
<td>18.5</td>
<td>73.017</td>
<td>25.328</td>
<td>60.162</td>
<td>-0.167</td>
<td>-0.219</td>
</tr>
<tr>
<td>cox3</td>
<td>Coding</td>
<td>11,880</td>
<td>1,929</td>
<td>12,481</td>
<td>1,759</td>
<td>769</td>
<td>801</td>
<td>24.371</td>
<td>17.4</td>
<td>15.5</td>
<td>71.573</td>
<td>24.822</td>
<td>64.216</td>
<td>-0.309</td>
<td>-0.193</td>
</tr>
<tr>
<td>cytB</td>
<td>Coding</td>
<td>10,711</td>
<td>1,109</td>
<td>11,825</td>
<td>1,109</td>
<td>1,113</td>
<td>1,113</td>
<td>29.525</td>
<td>19.8</td>
<td>17.8</td>
<td>71.873</td>
<td>27.763</td>
<td>65.673</td>
<td>-0.272</td>
<td>-0.321</td>
</tr>
<tr>
<td>nad1</td>
<td>Coding</td>
<td>9,589</td>
<td>9,705</td>
<td>10,443</td>
<td>10,544</td>
<td>846</td>
<td>860</td>
<td>30.454</td>
<td>18.3</td>
<td>18.6</td>
<td>78.390</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
</tr>
<tr>
<td>nad2</td>
<td>Coding</td>
<td>3,559</td>
<td>3,928</td>
<td>3,894</td>
<td>3,933</td>
<td>336</td>
<td>336</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
</tr>
<tr>
<td>nad3</td>
<td>Coding</td>
<td>12,702</td>
<td>1,726</td>
<td>13,930</td>
<td>1,726</td>
<td>1,730</td>
<td>1,730</td>
<td>29.525</td>
<td>19.8</td>
<td>17.8</td>
<td>71.873</td>
<td>27.763</td>
<td>65.673</td>
<td>-0.272</td>
<td>-0.321</td>
</tr>
<tr>
<td>nad4</td>
<td>Coding</td>
<td>6,868</td>
<td>6,672</td>
<td>9,916</td>
<td>9,603</td>
<td>232</td>
<td>232</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
</tr>
<tr>
<td>nad5</td>
<td>Coding</td>
<td>3,863</td>
<td>3,942</td>
<td>5,477</td>
<td>5,520</td>
<td>1,582</td>
<td>1,579</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
</tr>
<tr>
<td>nad6</td>
<td>Coding</td>
<td>2,641</td>
<td>1,635</td>
<td>6,681</td>
<td>6,572</td>
<td>441</td>
<td>438</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
</tr>
<tr>
<td>Transfer RNA</td>
<td>tRNA-Asp</td>
<td>1,728</td>
<td></td>
<td>1,728</td>
<td>1,728</td>
<td>1,728</td>
<td>1,728</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
</tr>
<tr>
<td>tRNA-Cys</td>
<td>1,583</td>
<td>1,579</td>
<td>1,636</td>
<td>1,633</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Glu</td>
<td>10,570</td>
<td>1,067</td>
<td>10,624</td>
<td>1,073</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Gly</td>
<td>9,868</td>
<td>6,990</td>
<td>7,042</td>
<td>7,047</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Ile</td>
<td>10,455</td>
<td>10,552</td>
<td>10,513</td>
<td>10,612</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Met</td>
<td>6,085</td>
<td>5,932</td>
<td>6,140</td>
<td>5,988</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Phe</td>
<td>10,657</td>
<td>10,736</td>
<td>10,710</td>
<td>10,790</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Thr</td>
<td>12,647</td>
<td>12,724</td>
<td>12,702</td>
<td>12,777</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Tyr</td>
<td>7,881</td>
<td>7,975</td>
<td>7,934</td>
<td>8,030</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>tRNA-Val</td>
<td>6,180</td>
<td>6,081</td>
<td>6,234</td>
<td>6,135</td>
<td>13.912</td>
<td>13.9</td>
<td>13.9</td>
<td>78.377</td>
<td>16.3</td>
<td>17.8</td>
<td>59.763</td>
<td>-0.145</td>
<td>-0.242</td>
<td>0.460</td>
<td>0.382</td>
</tr>
<tr>
<td>Ribosomal RNA</td>
<td>rRNA-18S</td>
<td>7,041</td>
<td>7,050</td>
<td>7,743</td>
<td>7,749</td>
<td>703</td>
<td>700</td>
<td>50.874</td>
<td>16.3</td>
<td>15.6</td>
<td>79.770</td>
<td>19.3</td>
<td>23.0</td>
<td>51.566</td>
<td>0.048</td>
</tr>
<tr>
<td>rRNA-28S</td>
<td>2,611</td>
<td>2,638</td>
<td>3,561</td>
<td>3,599</td>
<td>951</td>
<td>962</td>
<td>50.874</td>
<td>16.3</td>
<td>15.6</td>
<td>79.770</td>
<td>19.3</td>
<td>23.0</td>
<td>51.566</td>
<td>0.048</td>
<td>-0.065</td>
</tr>
</tbody>
</table>

H. contortus NZ_Hco_NP T. circumcincta NZ_Teci_NP
other Trichostrongyloidea, the reported mt genomes also all lack the atp8 gene. All of the 12 PCGs in both mt genomes are predicted to use unique translation initiation codons such as ATT (atp6, cyt b, and nad4-6), TTG (nad2-3), and ATA (nad1 and cox1-3) in H. contortus NZ_Hco_NP, with T. circumcincta NZ_Teci_NP predominantly using ATT and ATA (cyt b and cox1) (Table 1). Both mt genomes are predicted to use the complete termination codons TAA. Nevertheless, cox2 had a TAG translation termination codon similar to nad3 of H. placei where not all termination codons were complete stop codons, typical of nematodes (Jex et al., 2008; dos Santos et al., 2017).

Of the 22 predicted trn genes, size ranged from 54 to 63 bp in length (Table 1), in H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP mt genome sequences and are similar to those of other nematodes studied to date (Supplementary Figure 1). The characteristics associated with the secondary structures predicted for most tRNA genes include: a 7–8 bp acceptor stem (amino-acyl arm), a 5 bp anticodon stem and a T/U residue preceding. With the exception of T. spiralis (Lavrov and Brown, 2001), these characteristics are consistent with all previously published mt genomes for the Trichostrongyloidea (dos Santos et al., 2017) and Rhabditida [C. elegans (Blaxter et al., 1998)].

The rrnS (12S) and rrnL (16S) genes identified in the mt genomes of H. contortus NZ_Hco_NP (703 and 951 bp) and T. circumcincta NZ_Teci_NP (700 and 962 bp) were A+T biased (Table 1). Secondary structure models for the rrn genes (Supplementary Figure 2) of H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP were constructed based on models of nematodes that have been previously reported (Hu et al., 2004, 2007; Hu and Gasser, 2006). The rrn genes were particularly A+T rich with 79.8 and 77.0% for rrnS and 82.9 and 83.0% for rrnL in H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP, respectively. The overall base compositions of the rrnS genes were; A = 41.8 and 36.0%, T = 38.0 and 41.0%, C = 5.8 and 7.4%, and G = 13.5 and 15.6%, and that of the rrnL genes were; A = 42.2 and 39.7%, T = 40.7 and 43.3%, C = 5.3 and 5.5%, and G = 10.7 and 11.4%. Therefore, the A+T content of the rrnS and rrnL genes were comparable to the overall A+T content of both whole mt genomes. The rrnL gene in both H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP, is located between trnH and nad3. Similarly, the
The rrnS gene is located between trnE and trnS1 almost 3.5 Kb from the rrnL and between nad4L and nad1 in both mt genomes (Figures 1, 2). The locations, sizes and secondary structure characteristics of the rrnL and rrnS genes are relatively conserved among all Trichostrongyloidea mt genomes analyzed, especially the stem regions of the rrnL genes between H. contortus and T. circumcincta strains.

Comparative mt Pangenome Analysis of Trichostrongyloidea

Comparisons of the gene synteny as well as the nucleotide, codon usage and amino acid compositions revealed a highly conserved GA2 gene arrangement (nad6, nad4L, rrnS, nad1, atp6, nad2, cyrb, cox3, nad4, cox1, cox2, rrnL, nad3, nad5), and mt genome structure as previously reported for trichostrongyloids (Blouin et al., 1997; Blouin, 2002; Hu et al., 2004, 2007). Also, in accordance with previously published mt genomes the absence of the presumed non-essential ATP synthase F0 subunit 8 (atp8) gene reported only in Trichinella spiralis (Lavrov and Brown, 2001; Jex et al., 2008, 2009) and Trichuris species such as Trichuris ovis and Trichuris discolor (Nematoda: Trichuridae), was absent in our data.

The predicted lengths of all the genes in H. contortus NZ_Hco_NP are very similar to those of T. circumcincta NZ_Teci_NP (≤1 amino acid difference). The sequence variability between Haemonchus species/strains and T. circumcincta strains of the Trichostrongyloidea superfamily was assessed by performing pairwise comparisons for each PCG of the mt genomes (Figure 2). Among the Haemonchus species/strains, the most variable genes were nad6 and nad2 with 13.8 and 10.2% of sequence divergence, respectively, and the most conserved genes were cox1, atp6, and cox2 with 93.6, 92.2, and 92.5% of sequence identity, respectively. For the T. circumcincta strains, we found that the most variable gene was nad4 with 2.2% sequence divergence, with most genes being conserved at around 98.6% sequence identity. Based on our results, the most variable genes among Trichostrongyloidea nematodes were nad2, nad6, nad3, and nad5 with 56.1, 38.4, 23.1, and 21.3% of sequence divergence, respectively, with the most conserved genes were cox1, atp6, cox2, and cox3 with 87.9, 83.5, 85.3, and 86.9% of sequence identity, respectively. Pairwise comparisons of all the PCGs among Trichostrongyloidea nematodes and it was observed that the nad4 gene was the most variable in terms of both the gene length and the sequence identity (Figure 2), where the predicted lengths differ by exactly 6 amino acids between Haemonchus (1,230 bp for H. contortus and H. placei) and T. circumcincta strains (1,212 bp). In comparison, the nad4 gene appears to be conserved among different Trichostrongyloidea species, for example Trichostrongylus vitrinus and T. axei (1,227 bp), while M. digitatus, C. oncophora, and C. elegans N2 all share 1,230 bp nad4 genes (Blouin et al., 1997; Blouin, 2002; Hu et al., 2004; Jex et al., 2009; Gasser et al., 2012; Xu et al., 2015).

A recent investigation comparing the divergences observed between ITS-1 and ITS-2 genetic nuclear markers with cox1 or nad4 genes (Blouin et al., 1997; Blouin, 2002), showed that mt DNA accumulates substitutions more quickly than nuclear genome markers and that nad4 genes in particular are a superior tool for prospecting species. For example, regarding H. placei and H. contortus mt genomes, a divergence in nuclear ITS-2 rDNA between these two species has previously been reported as 1.3% (Stevenson et al., 1995). Whereas, differences between nucleotide sequences of H. placei and H. contortus mt genome PCGs ranged between 10 and 20% (Gasser et al., 2006; dos Santos et al., 2017). Taken together, our findings are in agreement with previous reports suggesting the use of the most conserved mt cox genes for investigating systematics and speciation in nematodes (Hu et al., 2004; Hu and Gasser, 2006; Zarowiecki et al., 2007), while the variation observed in nad mt genes would be more appropriate for population genetics studies (Jex et al., 2009).

Analysis of codon usage patterns in Trichostrongyloidea and C. elegans N2 revealed that certain codons are predicted to be utilized almost ubiquitously across all mitochondrial PCGs (Figure 3). As expected, the A+T bias reported in the Trichostrongyloidea mt genome sequences also affects the amino acid sequence composition of the predicted proteins. Within each codon family, we found that the non-polar Leu (TTG, 5.0%), basic His (CAC, 0.4%), polar Asn (AAC, 0.5%), and acidic Asp (GAC, 0.3% but except nad3) were highly utilized across all PCGs. Among each codon family, the T-ending (49.9%) codons are used most frequently than codons ending in any of the three other nucleotides, with C-ending (2.7%) codons least used. Similarly, for both Haemonchidae and Trichostrongylidae the average nucleotide content of their mt PCGs are 26.0% (A) and 43.7% (T), compared to 25.4% (A), and 44.2% (T), respectively. In particular the nad4L was the most T-rich at 53.5 and 54.9%, in each family, respectively. Also the most commonly codon observed was TAA as the termination codon with the majority of initiation codons also being AT-rich.

The observed nucleotide biases also have substantial effects on codon usage patterns for both H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP mt genomes (Table 2). The most frequently used codons among the PCGs were: TGT (Cys, N = 122 and 101 times used, 3.73% and 2.94% of the total), TTT (Phe, N = 435 and 428, 13.32% and 12.47%), ATT (Ile, N = 201 and 215, 6.15% and 6.26%), TTA (Leu, N = 199 and 271, 6.09% and 7.90%), TTT (Asn, N = 179 and 123, 5.48% and 3.60%), GTT (Val, N = 143 and 119, 4.40% and 3.50%), and TAT (Tyr, N = 181 and 198, 5.54% and 5.77%), respectively. In addition, T-ending codons are far more frequent than codons ending in any of the other three nucleotides, with C-ending (2.7%) codons least used. For example, the least used codons included CCA (Pro, N = 3 and 14, 0.09% and 0.41%), GCG (Ala, N = 3 and 7, 0.09% and 0.20%), along with the C-ending codons GAC (Asp, N = 8 and 10, 0.25% and 0.29%) and CTC (Leu, N = 3 and 10, 0.09% and 0.29%).

Our pairwise alignments and synteny analysis of the complete mt genome nucleotide sequences for the Haemonchus and Teladorsagia species and strains revealed a common and highly variable A+T-rich non-coding region between 5,500 and 6,100 bp in each mt genome (Figure 2 and Supplementary Figure 3). The A+T content of this region is unusually high in both Haemonchus (89.0%) and Teladorsagia (88.2%) mt genomes, and is located between trnA and trnP (or between genes nad5 and nad6). Among Haemonchus, the A+T-rich region is ~600 bp in H. contortus (89.7%) and 155 bp in H. placei (87.1%), whereas the
FIGURE 2 | Gene arrangement and distribution of *Haemonchus* (A) and *Teladorsagia* (B) mt genomes. (A) Gene map of the complete mitochondrial genomes for *H. contortus* Hco_McMaster (light red, EU346694), *H. contortus* Hco_ISE (light green, LS997568) and *H. placei* Hpl_MHpl1 (light purple, NC_029736). (B) Gene map of the complete mitochondrial genomes for *T. circumcincta* AP017699 (light red) and GQ888720 (light green). *H. contortus* NZ_Hco_NP (A) and *T. circumcincta* NZ_Teci_NP (B) were used as the reference sequences, respectively. The outermost ring indicates gene arrangement and distribution based on the primary sequence GenBank file. Tandem repeat regions are shown with green blocks (full length) for *H. contortus* NZ_Hco_NP (circle 2) and light red blocks for all other *Haemonchus* mt genomes (circle 3). Low complexity repeat regions for *H. contortus* NZ_Hco_NP are also shown with green blocks (half length) on circle 2. The next two (A) or three (B) rings represent BLAST hits obtained from BLASTX searches performed using a threshold of 1e−5 for protein sequences of genomes described above, in which the query sequences were translated in reading frames 1 and 2. The inner-most rings indicate the GC content information. The contents of the A+T-rich regions have been shown in more detail in expanded 10× zoomed views. nad1-6, NADH dehydrogenase subunits 1–6; cox1-3, cytochrome c oxidase subunits 1–3; atp6, ATPase subunit 6; cyt b, cytochrome b.
region is 340 bp in *Teladorsagia* (88.2%). A pairwise alignment of the nucleotide sequences of the A+T-rich regions of *H. contortus* NZ_Hco_NP (611 bp) and closely related *H. contortus* McMaster (EU346694, 578 bp) resulted in 91.5% identity. Interestingly, the region in *H. contortus* is among the largest among other nematodes studied to date with the exception of *As. suum* (886 bp) (Okimoto et al., 1992; Keddie et al., 1998; Hu et al., 2003).

Comparative analysis of the control (A+T-rich) region of *Haemonchus* mtDNAs identified three tandem repeat (TR) units (69, 120, and 62 bp) compared to only one (126 bp) found in *T. circumcincta* strains mt genomes (Figure 2). The number of repeat units within the A+T-rich region was variable depending on the mtDNAs. For example, the copy numbers of the A+T-rich region TRs for *T. circumcincta* NZ_Teci_NP corresponding to TTATAATTATTATAAATTA (23 bp), TATATATATAAATTATAATTATTATAAAT (37 bp), and ATATTATAATATTATTATAATTATAAAT (20 bp) consensus units were 4, 2, and 4. In general, TRs of *Haemonchus* mt genomes were much more abundant and distributed randomly throughout the mtDNAs, whereas TRs were mostly in the intergenic regions of *T. circumcincta* mtDNA. Some TRs were found within the mitochondrial genes, such as those in *cox1*, *nad5*, *nad6*, and *nad4*, and also *nad4L* in *T. circumcincta* mtDNA. Overall, our analyses show that the number of tandem repeats was highly variable among different species and strains of *Haemonchus* and *Teladorsagia* and largely account for the variations among the mtDNA sequences lengths.

Based on our comparative genomics of A+T-rich regions of complete *Haemonchus* species/strains and *T. circumcincta* strains mt genomes, we were able to differentiate individual strains with high confidence due to the high mutation rate and sequencing coverage for these regions. We propose that such regions can serve as ideal markers for future studies investigating nematode population structure (Jex et al., 2009; Gasser et al., 2012), population genetics (Blouin, 2002; Hu et al., 2004; Gasser et al., 2012), and phylogeny (Blouin et al., 1997; Blouin, 2002; Xu et al., 2015). For future work, we aim to design and test numerous primer pairs targeting the A+T-rich region between the *nad5* and *nad6* genes to develop accurate PCR-based molecular techniques to differentiate Trichostrongyloidea species.
Mitochondrial Phylogenomics

The super-family Trichostrongyloidea includes genera such as Ostertagia, Teladorsagia, Trichostrongyulus, Haemonchus, Cooperia, Nematodirus, Dictyocaulus. To determine the phylogenetic relationship of the H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP with other H. contortus and T. circumcincta strains and members of the Trichostrongyloidea nematodes, the concatenated amino acid sequences predicted from 12 mtDNA PCGs were analyzed using BI, ML, and MP methods. Topologies of all trees inferred by the three different distance models and methods were identical (Figure 4), with phylogenetic relationships among the different Trichostrongyloidea species well-resolved with very high nodal support throughout. The presented phylogenomic tree of Trichostrongyloidea mt genomes, is consistent in topology with other published data (Chilton et al., 2006; Jex et al., 2009), and further support the hypothesis that the primary driver of early divergence is actually the site of infection in the host. Overall, our results corroborate this hypothesis given that each of different species and strains of the Haemonchidae compared with Cooperidae and Trichostrongylidae families, are genetically closer to abomasal than to non-abomasal (intestine) Trichostrongyloidea, respectively. The anomaly to this hypothesis was T. axei that occupies the abomasum as the site of infection oppose to the small intestine of T. vitrinus, which implies the presence of other and yet uncharacterized drivers of genetic species diversity of these closely related parasites.

The phylogenomic tree based on the concatenated PCGs of the complete mt genomes (Figure 4), placed H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP in their respective clades with other H. contortus and T. circumcincta strains, which are in agreement with previous studies (Jacquiet et al., 1998; Nunes et al., 2013; Rohart et al., 2017; Palevich et al., 2018). However, our tree is not directly comparable to members. Such molecular tools would enable easy species- and possibly strain-level identification with sufficient resolution and confidence from fecal samples that would also help prevent animal sacrifice for such studies.

TABLE 2 | Codon usage analysis of PCGs in the mitochondrial genomes of H. contortus NZ_Hco_NP and T. circumcincta NZ_Teci_NP.

<table>
<thead>
<tr>
<th>AA</th>
<th>Codon</th>
<th>N /1,000a</th>
<th>Freq</th>
<th>AA</th>
<th>Codon</th>
<th>N /1,000a</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>GCG</td>
<td>37</td>
<td>0.922</td>
<td>0.04</td>
<td>Pro</td>
<td>CCG</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>GCA</td>
<td>12</td>
<td>10</td>
<td>3.672</td>
<td>0.91</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GCT</td>
<td>13</td>
<td>32</td>
<td>3.989</td>
<td>0.33</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td>4</td>
<td>8</td>
<td>1.222</td>
<td>0.33</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Cys</td>
<td>TGT</td>
<td>12</td>
<td>21</td>
<td>37.35</td>
<td>2.44</td>
<td>Gin</td>
<td>CAG</td>
</tr>
<tr>
<td></td>
<td>TGC</td>
<td>2</td>
<td>4</td>
<td>7.35</td>
<td>5.54</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GAT</td>
<td>56</td>
<td>11</td>
<td>17.35</td>
<td>2.44</td>
<td>Arg</td>
<td>CGG</td>
</tr>
<tr>
<td></td>
<td>GAC</td>
<td>8</td>
<td>10</td>
<td>2.45</td>
<td>2.91</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>17</td>
<td>32</td>
<td>5.21</td>
<td>3.33</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>GAA</td>
<td>4</td>
<td>0</td>
<td>1.22</td>
<td>0.33</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Phe</td>
<td>TTT</td>
<td>435</td>
<td>428</td>
<td>133.19</td>
<td>124.74</td>
<td></td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>TTC</td>
<td>36</td>
<td>66</td>
<td>11.02</td>
<td>13.41</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>GGG</td>
<td>25</td>
<td>16</td>
<td>7.65</td>
<td>4.66</td>
<td>Ser</td>
<td>AGT</td>
</tr>
<tr>
<td></td>
<td>GGA</td>
<td>25</td>
<td>16</td>
<td>7.65</td>
<td>4.66</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>GCT</td>
<td>38</td>
<td>72</td>
<td>11.64</td>
<td>20.99</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td>8</td>
<td>13</td>
<td>2.45</td>
<td>3.79</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>CAT</td>
<td>19</td>
<td>28</td>
<td>5.82</td>
<td>8.16</td>
<td>Thr</td>
<td>ACG</td>
</tr>
<tr>
<td></td>
<td>CAC</td>
<td>5</td>
<td>16</td>
<td>1.53</td>
<td>4.66</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Ile</td>
<td>ATA</td>
<td>133</td>
<td>103</td>
<td>40.72</td>
<td>30.02</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>ATT</td>
<td>20</td>
<td>125</td>
<td>61.54</td>
<td>62.66</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>ATC</td>
<td>27</td>
<td>19</td>
<td>8.27</td>
<td>5.54</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>AAC</td>
<td>17</td>
<td>32</td>
<td>5.21</td>
<td>3.33</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Lys</td>
<td>AAG</td>
<td>49</td>
<td>37</td>
<td>15</td>
<td>10.78</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>ACA</td>
<td>24</td>
<td>20</td>
<td>7.36</td>
<td>5.83</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>AAC</td>
<td>17</td>
<td>32</td>
<td>5.21</td>
<td>3.33</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>Met</td>
<td>ATG</td>
<td>52</td>
<td>66</td>
<td>15.92</td>
<td>16.32</td>
<td></td>
<td>142</td>
</tr>
</tbody>
</table>

H. contortus NZ_Hco_NP, T. circumcincta NZ_Teci_NP. aRepresents number of codons per 1,000 codons.
FIGURE 4 | Inferred phylogenetic relationship among the nematode superfamily Trichostrongyloidea. Phylogenetic analysis of the complete mitochondrial genomes of different nematode species or conspecific strains is based on the concatenated amino acid sequences of 12 protein-coding genes by Bayesian inference (BI), maximum likelihood (ML), and maximum parsimony (MP) analyses. *species with different site of infection to other members of family. The numbers along branches indicate posterior probabilities and bootstrap values resulting from different analyses in the order: BI/ML/MP. Members of different families were shown in colored boxes with the associated main site of infection in parenthesis. GenBank accession numbers are provided (in parentheses) for all reference sequences. Scale bar represents Posterior Probability.

Previously described phylogenetic trees that performed different analyses and compared a broader set of taxa, but mainly because the trees were based on the ITS rDNA nuclear genetic markers (Chilton et al., 2001; Hoberg et al., 2004; Jex et al., 2008). While both *Haemoncus* and *Teladorsagia* strains grouped together in their respective families, *T. circumcincta* NZ_Teci_NP was clearly separated from the other strains. Interestingly, *H. contortus* NZ_Hco_NP and McMaster (Australia) strains grouped together and originate from nearby geographic locations, but separate from *H. contortus* ISE (East Africa). This observation corroborates our hypothesis on the impact of environmental constraints and preferences acting as secondary drivers of within-species diversity. Overall, we are in agreement with previous phylogenetic studies suggesting that ITS rDNA nuclear genetic markers alone may not provide sufficient information to reveal higher taxonomic level relationships within the Trichostrongyloidea. Our findings indicate that a three-pronged approach that incorporates phylogenetic inertia,
pangenome structure/features and environmental data in order to understand the mitochondrial genome evolution.

CONCLUSIONS

This study compares the recently sequenced complete mitochondrial genomes of sibling species of parasitic roundworms, *H. contortus* and *T. circumcincta*, two of the most economically important and common pathogenic nematodes infecting small ruminants worldwide. We explored the mitochondrial pangenome features and phylogenomic relationships to assess species- and strain-level diversity among Trichostrongyloidea nematodes and with available mt genome in public databases. Our analyses corroborate previous studies showing that our *H. contortus* NZ_Hco_NP and *T. circumcincta* NZ_Teci_NP strains position in their respective clades. Future work should focus on utilizing our insights on the highly variable regions of *Haemoncus* and *Teladorsagia* conspecific strains to develop cost-effective DNA-based approaches for novel parasite management and control strategies. The complete mt genomes of the New Zealand *H. contortus* and *T. circumcincta* field strains are important contributions to our understanding of meta-population connectivity as well as species- and strain-level evolution in nematodes. With the continuing improvements in sequencing technology combined with a community effort we may be able to reconstruct the true origins of *Haemoncus* and *Teladorsagia*, as well as other parasitic nematodes that are of a global interest.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

NP conceived and designed the project, analyzed and interpreted the data, and wrote the manuscript. PM provided bioinformatics support for the project. MM and Y-JC provided resources and critically revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was supported by the Agricultural and Marketing Research and Development Trust (AGMARDT) Postdoctoral Fellowship Programme, grant number P17001.

ACKNOWLEDGMENTS

Special thanks to PM, AgResearch Limited, Grasslands Research Centre, Palmerston North, New Zealand, for his contribution to data analysis. This manuscript has been released as a pre-print at Research Square (Palevich et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.573395/full#supplementary-material

Supplementary Figure 1 | Predicted secondary structures of the 22 tRNAs in the *H. contortus* NZ_Hco_NP (left) and *T. circumcincta* NZ_Teci_NP (right) mitochondrial genomes. The two-dimensional predicted layout predicted RNA secondary structures is shown with canonical base-pair (BP) in blue, and non-canonical, lone-pairs and triplets in green.

Supplementary Figure 2 | Predicted secondary structures of the rrnS (A) and rrnL (B) genes in the *H. contortus* NZ_Hco_NP (left) and *T. circumcincta* NZ_Teci_NP (right) mitochondrial genomes. The two-dimensional predicted layout predicted RNA secondary structures were predicted based on the energy model of Mathews et al. (2004). The minimum free energy (MFE) structure of harpins is colored according to the base-pairing probabilities (red, high; green, mid; blue, low). Blue and red circles around nucleotides represent the beginning and the end of molecules, respectively.

Supplementary Figure 3 | Dot plots showing *Haemonchus* and *Teladorsagia* species and strain-level synten. The horizontal and vertical axes represent the entire translated mt genome nucleotide sequences, respectively. Each aligning gap-free segment with more than 50% identity is plotted as a black line or dot. Analysis was performed using chaining coverage with PipMaker (Schwartz et al., 2000).

REFERENCES


roundworms, *Haemonchus contortus* and *Teladorsagia circumcincta*. *Res. Square.* 1, 1–37. doi: 10.21203/rs.3.rs-48588/v1


R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Bellingham, WA.


**Conflict of Interest**: NP was employed by the company AgResearch Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Palevich, Maclean, Choi and Mitreva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*