Howard Hughes Medical Institute

Published as: *Nat Genet*. 2014 July ; 46(7): 701–706.

Genome and transcriptome of the porcine whipworm *Trichuris suis*

Aaron R. Jex[1,*], **Peter Nejsum**[2], **Erich M. Schwarz**[1,4], **Li Hu**[3], **Neil D. Young**[1], **Ross S. Hall**[1], **Pasi K. Korhonen**[1], **Shengguang Liao**[3], **Stig Thamsborg**[2], **Jinquan Xia**[3], **Pengwei Xu**[3], **Shaowei Wang**[3], **Jean-Pierre Y. Scheerlinck**[1], **Andreas Hofmann**[5], **Paul W. Sternberg**[6,7], **Jun Wang**[3,8,9,10,11], and **Robin B. Gasser**[1]

[1]Faculty of Veterinary Science, The University of Melbourne, Parkville, Victoria, Australia

[2]University of Copenhagen, Copenhagen, Denmark

[3]BGI-Shenzhen, Shenzen, PR China

[4]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, 14853-2703, USA

[5]Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, Queensland, Australia

[6]Howard Hughes Medical Institute

[7]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA

[8]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark

[9]Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia

[10]Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China

[11]Department of Medicine, University of Hong Kong, Hong Kong

Correspondence and requests for reprints or permissions should be addressed to Aaron R. Jex (ajex@unimelb.edu.au), Jun Wang (wangjun@genomics.org.cn) or Robin B. Gasser (robinbg@unimelb.edu.au).

Accession numbers

All short read data are available via NCBI's sequence read archive (genomic DNA male: SRR1041639-SRR1041644; genomic DNA female: SRR1041645-SRR1041650; messenger RNA: SRR1041651-SRR1041658; small RNA: SRR1041659-SRR1041664, SRR1041669, SRR1041670). Annotated assemblies of each genome are accessible through NCBI via BioProject PRJNA208415 (male) and PRJNA208415 (female).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions P.N. and S.M.T. provided *T. suis*, and N.D.Y. purified nucleic acids for sequencing. L.H. and S.W. coordinated sequencing. A.R.J., S.L., P.K.K. R.S.H., J.X. and P.X. undertook the assembly, annotation and analyses of genomic and transcriptomic data. A.R.J., R.S.H., A.H., P.K.K., and E.M.S. planned and performed additional, detailed bioinformatic analyses. T.W.H. organized and collated archival of the draft assemblies in Wormbase. A.R.J., P.W.S., J-P.S. and R.B.G. drafted and edited the manuscript, tables, figures and supplementary information. A.R.J., N.D.Y., J.W. and R.B.G. conceived and planned the project. A.R.J., N.D.Y. and R.B.G. supervised and coordinated the research.

Abstract

Trichuris (whipworm) infects 1 billion people worldwide, and causes a disease (trichuriasis) that results in major socioeconomic losses in both humans and pigs. Trichuriasis relates to an inflammation of the large intestine manifested in bloody diarrhoea, and chronic disease can cause malnourishment and stunting in children. Paradoxically, *Trichuris* of pigs has shown substantial promise as a treatment for human autoimmune disorders, including inflammatory bowel disease (IBD) and multiple sclerosis (MS). Here, we report ~80 megabase (Mb) draft assemblies of the genomes of adult male and female *T. suis*, and explore stage-, sex- and tissue-specific transcription of messenger and small non-coding RNAs.

Soil-transmitted helminths, including whipworm (*Trichuris*), hookworms (*Necator* and *Ancylostoma*) and large roundworm (*Ascaris*), are among the most prevalent and devastating parasites of humans globally, and predominate in impoverished nations[1]. *Trichuris* infects one billion people, and chronic infection of high intensity can lead to typhlitis, colitis, chronic dysentery, malnutrition through malabsorption as well as reduced physical and cognitive development[2]. Consequently, trichuriasis, which disproportionately affects children, has an estimated global burden of 1–6.5 million Disability Adjusted Life-Years (DALYs), exceeding that of schistosomiasis, trachomiasis, trypanosomiasis or leishmaniasis[1]. Despite this, *Trichuris* species are classified by the World Health Organization (WHO) as neglected parasites, in urgent need of improved control[3].

Contrasting the significant burden of trichuriasis and other neglected helminths is the observation that human populations in endemic countries tend to suffer significantly less from immuno-pathological diseases[4], which are common and increasingly prevalent[5] in countries in which exposure to pathogens is limited. These observations have inspired the "hygiene hypothesis"[6], which proposes that a lack of exposure of humans to common pathogens impairs immune function and leads to increased autoimmune disease. This hypothesis is supported by clinical data, with routine deworming positively[7] and early-childhood helminth negatively[8] correlating with autoimmune disorders. Recent studies have shown that porcine *Trichuris* (*T. suis*) administered to humans suffering from inflammatory bowel disease (IBD; including Crohn's disease and ulcerative colitis) can reduce clinical symptoms[9,10]. Similar observations have been made in patients with multiple sclerosis (MS)[11]. Although helminths can alter immune responses in their hosts via a variety of excretory/secretory (ES) molecules[12], the specific interactions between *T. suis* and its host remain unclear. By sequencing the *T. suis* genome and transcriptomes (messenger and small RNAs), we provide deep insights into the molecular biology of this parasite, and its modulation of host immune responses. These data provide a solid basis for exploring human trichuriasis, developing new anti-parasitic drugs and elucidating how helminths suppress autoimmune disorders.

RESULTS

Sequencing, assembly and synteny

We sequenced the genomes of single adult female and male *T. suis*, each at ~140-fold coverage, producing draft assemblies of 76 and 81 Mb, respectively (Table 1;

Supplementary Tables 1 and 2; Supplementary Figs. 1 and 2). Matches to conserved eukaryotic genes (CEGs) indicated that each assembly is 96% complete, with minimal redundancy (Supplementary Table 3). Alignment of these assemblies showed high similarity, with 68 Mb aligning as direct 1-to-1 matches in blocks of a mean length of 2.5 kb. Overall, sequence identity was 99.2% (38,854 single nucleotide polymorphisms, SNPs). Despite the reported XX/XY karyotypes for female and male *Trichuris*, respectively[13], we found no evidence for a Y chromosome among the male-specific scaffolds, suggesting that this chromosome largely contains repetitive sequences common to both sexes; this finding is consistent with observations made of the *T. suis* karyotype, suggesting that the sex-chromosomes were the smallest chromosomal pair and were morphologically very similar in both sexes[13]. Repetitive sequences comprise ~32% of the genome, including 8% DNA transposable elements, 2.9% long tandem repeats and 3.3% retrotransposons (Supplementary Table 4). Each genome encoded ~1,000 transfer RNA (tRNA) genes, with copy numbers reflecting codon usage in protein-encoding regions (Supplementary Table 5 and 6; Supplementary Fig. 3).

Protein-coding gene-set

The female and male *T. suis* genomes encode at least 14,470 and 14,781 protein-coding genes, respectively, representing ~70% of each genome, including introns and exons. We identified 14,356 and 14,315 female and male genes with an orthologue/homologue in the opposite sex, with 10,403 genes being defined as unambiguous one-to-one orthologues. Evidence for sex- specific genes was limited, with just one and 41 supported as female and male specific, respectively (see Supplementary Results). Of these sex-specific genes, only three male genes have a predicted function, with homology to *C. elegans frk-1* (a receptor tyrosine-kinase), *gpc-1* (a G protein-coupled receptor, GPCR), *his-66* (a histone protein). The sex-specific genes show no clustering among scaffolds, providing little evidence for their association with the sex chromosomes. Most of the remaining differences in the genes of the two genders relate to a higher copy number of some genes in the male. From both assemblies, we defined a unified set of 14,820 genes for *T. suis* (Table 1 and Supplementary Table 7), with 12,910 (87.1%) of them supported by RNA-seq data. The majority (59.8%) of these genes have homologues (BLASTp cut-off: 1×10^{-5}) in other nematodes, including 6,286 (42.4%), 6,340 (42.7%), 6,149 (41.5%) and 8,480 (57.2%) in *Ascaris suum*[14], *Brugia malayi*[15], *Caenorhabditis elegans*[16] and *Trichinella spiralis*[17], respectively (Fig. 1). Functions were assigned to 9,342 (63.0%) protein-encoding genes (Supplementary Tables 8–13). Focusing on key functional or druggable proteins[14], we predicted 653 peptidases and 288 peptidase inhibitors. Peptidase classes S1 (116) and S8 (42) are expanded in *T. suis* compared with those represented in other nematode genomes[14–17]. The *T. suis* genome also encodes 269 phosphatases and 232 kinases. We identified a large complement of receptors and transporters[18]; these molecules include 228 GPCRs, as well as 1,962 channel, pore and transporter proteins. Among the latter are 133 peroxisomal protein importers, being more numerous than in *A. suum* (n = 74)[14], which may suggest an increased importance of fatty-acid digestion and metabolism in *T. suis*.

We predicted 618 canonical excretory-secretory (ES) proteins in adult *T. suis* (Supplementary Tables 14 and 15), including 165 proteases, many of which might have a

role in disrupting intestinal epithelial cells in the host[19,20] and in the formation of the syncytial tunnel around the *Trichuris* stichosome[21]. Notable among these molecules are 33 chymotrypsin-like serine proteases, which play key roles in helminths associated with host invasion[22], immuno-suppression[23] and tissue destruction[24]. In addition to proteases, non-membrane bound transporters comprise a major component of the secretome. These transporters include 41 pore-forming toxins (= porins), 25 of which have homology to the *Trichuris trichiura* porin TT47, which induces ion-conducting pores in planar lipid bilayers and assists in the formation of the syncytial tunnel in the intestinal epithelium[25]. Helminth-mediated immuno-modulation by ES products is well documented[12]. Among the predicted *T. suis* ES proteins, we found a variety of immuno-modulators (Supplementary Table 16). Based on these findings and available literature for helminths[12,26,27], we propose a *Trichuris*- driven immuno-modulation model (Supplementary Fig. 4), in which the parasite suppresses inflammation by secreting: (a) serpins to inhibit neutrophil cathepsins and elastases; (b) apyrases to prevent conversion of regulatory T-cells to pro-inflammatory T-cells; (c) cystatins to promote anti-inflammatory (IL-4 and IL-10 producing) T-cells by disrupting antigen presentation by dendritic and B-cells; (d) calreticulins that bind to dendritic cells and stimulate IL-4 production, and limit inflammation by binding free calcium ions; and (e) molecular mimics[12] of host galectins, mammalian macrophage inhibitory factor (MIF) and tumor growth factor (TGF-β) that stimulate apoptosis in activated T-cells, promote alternative activation of macrophages and block the stimulation of (pro-inflammatory) Toll-like receptor (TLR) pathways. This model is consistent with the pathophysiology described for *Trichuris* infection[26], and likely operates in tandem with immuno-suppressive processes linked to glycans[27] and lipids (e.g., sphingolipids; see Supplementary Results).

Transcriptome and differential transcription

We explored stage-, sex- and tissue-specific transcription (mRNAs), with a focus on parasite-host interactions. We predicted 36,763 transcripts, with 15,174 of them being perfect matches to the intron/exon chain (excluding UTRs), as annotated in the genome, and 21,589 representing novel splice isoforms (Supplementary Fig. 5 and Supplementary Table 17). In total, 6,293 (43.7%) *T. suis* genes were predicted to encode at least two isoforms. The number of splice-isoforms per gene only moderately correlated with exon number (R^2 = 0.44; Supplementary Figs. 6 and 7). Alternative splicing did correlate with gene function, reflected in an enrichment (p-value ≤0.05) of protein catabolism (i.e. proteases), membrane-bound ion transport and kinase activity among alternatively spliced, and apoptosis among single splice-isoform genes. Multiple and single splice-isoform genes also differed in their conservation and predicted essentiality. Of the 6,307 *T. suis* genes with *C. elegans* homologues, alternatively spliced genes predominated by a 3 to 1 margin (4,510 *vs.* 1,875) and represented 204 (75.8%) of the 269 essential genes predicted. This latter observation needs to be considered when mining helminth genomes for novel drug targets. If a novel inhibitor targeting products of such genes interacts with one of the spliced domains, isoform-switching may be sufficient to overcome its effect.

Some protein domains were significantly associated (p-value ≤ 0.05) with specific, alternative splice events, with exon skipping and the use of alternative first or last exons

appearing to differ in their functional implications for transcription (Fig. 2A). Most notable was an over-representation of substrate-binding motifs (e.g., immunoglobin, EGF-like or DnaJ) for genes biased toward transcripts with skipped exons. Remodelling of binding motif structure through alternative splicing has been shown to affect binding specificity in other organisms[28], and we propose that exon skipping is important in regulating binding specificity of proteins in *T. suis*. Given the varied functions associated with alternative first or last exon splicing events (Fig. 2A), we hypothesise that these specific modifications might be playing a role in regulating protein localisation, another known role for alternative splicing[28].

Trichuris suis undergoes significant developmental changes throughout its direct life-cycle[29]. To understand developmental processes in this parasite, we used RNA-seq to characterise transcription in various stages, sexes or body portions (= tissues) - namely larval stages (i.e. L1/L2, L3 and L4); adult male and female; stichosome; adult male posterior body and female posterior body excluding the stichosome (Supplementary Table 18 and Supplementary Fig. 8). Overall, a number of major functional classes of proteins showed increased representation in the transcriptome of *T. suis* relative to the genome (Fig. 2B). Secretory proteins were notable in this regard, making up ~4% of the *T. suis* gene-set, but representing ~10% of the transcriptional abundance in all libraries. Peptidases, in particular secreted peptidases, were also over-represented in the transcriptome, and notably upregulated during larval development and in the stichosome.

The stichosome is the thin, elongate anterior end of *Trichuris* embedded tightly within a syncytial tunnel[21] in the superficial layer of the large intestinal mucosa. Within this tunnel, the parasite secretes proteins and other molecules, and absorbs nutrients from cell cytoplasm and surrounding tissue fluids, likely through thousands of bacillary cells[30]. Given its central importance in feeding and interaction with the host, we focused on investigating transcription in the stichosome relative to the rest of the worm body (Supplementary Table 18; for detailed comparisons among other stages or tissues, see Supplementary Results). In the stichosome, we observed enriched transcription for 2,210 genes (encoding 3,721 transcripts) relative to both the male and female posterior bodies (Supplementary Table 18). Among these genes are 160 peptidases (encoding 256 transcripts) and 41 porins (85 transcripts), supporting their role in host tissue degradation and syncytial tunnel formation[19,25] (Supplementary Fig. 9). Also notable is the enrichment of a large number of secreted and membrane-bound transporters (222 genes encoding 371 transcripts) of various ions (e.g., sodium, phosphate and calcium) and small molecules (e.g., glucose and nucleosides). Sugar metabolism is enriched in the stichosome, suggesting that absorbed glucose is rapidly metabolised in the stichocytes. Also up-regulated in the stichosome are transcripts associated with endocytosis and vesicle formation, lysozyme and peroxisome pathways as well as fatty acid and amino acid (cysteine/methionine and lysine) degradation. At least one isoform of each putative immuno-modulatory gene encoded by *T. suis* is transcribed in the stichosome, with 22 transcripts encoding galactins, serpins, SCP/TAPS proteins, apyrase or calreticulin being specifically enriched in the stichosome relative to both the male and female posterior bodies.

Chymotrypsin-like (S1) serine proteases (n = 28 of 31 genes, and 51 of 135 transcripts) are also significantly enriched in the stichosome. Many are homologues of vertebrate plasmin, which is thought to regulate blood clotting in the host[31]. A poorly understood consequence of trichuriasis is bloody diarrhoea[32], and there is some evidence to suggest that *Trichuris* ingests blood[33]. It may be that some *T. suis* chymotrypsin-like serine proteases act as anti-coagulants or assist in digesting blood, serum and tissue components (e.g., fibrinogen). Notably, *T. muris* infection alters the mucus barrier in the host's gut epithelium, leading to an increased susceptibility to nematode infections[34] through the degradation of Muc2 polymers[35]. Muc2 depolymerization by *T. muris* is blocked by chymostatin/antipain[35], suggesting a likely role for chymotrypsin-like and other serine proteases. Several of the secreted serine proteases enriched in the stichosome are homologues of *Schistosoma mansoni* SP1 and human kallikrein. The latter molecule regulates the degradation of kininogen to bradykinin, stimulating vasodilation, the cytosolic release of Ca^{2+}, neutrophil recruitment and increased inflammation[36]. SP1 is a potent vasodilator in mice[37], suggesting that it has an ability to convert vertebrate kininogen to bradykinin. Given the anti-inflammatory capacity of *T. suis*, we propose that some of these chymotrypsin-like serine proteases might degrade host kininogen, but do not enable bradykinin production, thereby preventing bradykinin receptor stimulation and, thus, inhibiting inflammation. Notably, bradykinin receptors play a key role in various autoimmune disorders, including IBD[38] and MS[36].

Genetic regulatory networks

The *T. suis* gene-set has complete RNA-interference (RNAi) machinery, suggesting potential for functional genomic studies and indicating a role for small non-coding RNAs in gene regulation. We explored these small RNAs in *T. suis* (Supplementary Tables 19–21 and Supplementary Figs. 10 and 11) and produced ~435 million sequence reads. Approximately 92% of these reads mapped to the *T. suis* genome, with 16%, 23% and 9% classified as micro- (miRNAs), small-interferring (siRNAs) and tiny non-coding RNAs (tncRNAs), respectively. Approximately 4% of the small RNA reads mapped with an anti-sense (> 80% of reads) bias to transposable elements, consistent with PIWI-interacting RNAs (piRNAs)[39]. However, as for small RNAs in *Ascaris suum*[40], < 0.01% had characteristics consistent with 21U-RNAs, which function as piRNAs in *C. elegans*[41].

We identified 319 microRNAs (miRNAs), with 132 having close homologues in other nematodes (Supplementary Table 22). These miRNAs accounted for 16% of all small RNA reads sequenced, with Tsu-let-7 (50% of all miRNA reads), Tsu-miR-1 (17%), Tsu-novel-51 (8%; an homologue of a novel miRNA, Tsp-novel-51, from *Tr. spiralis*) and Tsu-miR-228 (4%) being the most highly transcribed. Approximately two-thirds of the miRNAs were most abundant in larval stages, suggesting a central role in development, with a diminishing number of miRNAs enriched in adults (Fig. 2C). This trend was reversed in the transition from L4 to the adult female. To explore the potential functional implications of differential transcription of these miRNAs, we predicted miRNA-binding sites linked to 3′-UTRs among 22,954 of the 23,824 mRNA isoforms (representing 7,180 genes) for which at least part of the 3′-UTR could be identified based on RNA-seq data. We focused on miRNA:mRNA interactions recognised or proposed for *C. elegans*. Of the 785,143 predicted binding sites

with homology to *C. elegans* (both miRNA and mRNA), 300,042 were supported by information in public databases and 3,238 by experimental findings[42]. Due to differences in gene copy number, these 300,042 binding sites represented 45 and 62 miRNAs as well as 3,205 and 3,877 coding genes in *C. elegans* and *T. suis*, respectively.

For *T. suis*, the shift from L3 to L4 coincides with a universal down-regulation of 24 of these conserved miRNAs, including miR-1, miR-252 and miR-236, being the second, seventh and eighth most abundant miRNAs, respectively, in *T. suis* overall. We identified 69 L4-enriched transcripts (relating to 62 *T. suis* genes and 61 *C. elegans* homologues) with binding sites for each of these miRNAs. Many of the *C. elegans* genes with inferred homology to these transcripts are involved in larval or embryonic development (e.g., *rol-3*, *slt-1* and *sox-3*), growth (e.g., *egl-4*, *unc-44* and *lin-39*) or early sexual determination (e.g., *sex-1*; www.wormbase.org), suggesting similar functional roles in *T. suis*. The maturation of *T. suis* to adulthood coincides with a variety of sex-specific changes in miRNA levels (relative to L4s). In both sexes, miR-228 (the fourth most abundantly transcribed miRNA in *T. suis*) and several isoforms of miR-61 were down-regulated, and miR-34 and two 'minor' miRNAs, namely miR-256 and miR-50, were up-regulated. Many of the coding genes (n = 447) up-regulated in both male and female adults are predicted to be co-regulated by miR-61 and miR-228. Homologues of these coding genes in *C. elegans* are enriched in GO terms (biological process) for embryonic/genital development, reproduction, morphogenesis/growth and metabolism, and include *tbx-2*, *vps-16*, *xnp-1* and *dyci-1* (www.wormbase.org). Notable among predicted miR-34 regulated genes were homologues of *srp-2* (serpin-2; a known anti-inflammatory of helminths[12]), which is down-regulated in the adult worm (with the exception of the stichosome) compared with larval stages.

When we compared male and female adults of *T. suis*, we found that major differences in miRNA transcription also related to miR-61, miR-228, miR-236 and miR-252, highlighting their importance in this nematode. Enriched in males are miR-228 and four copies of miR-61, and in females, miR-236, miR-252 and one copy of miR-61 (with closest homology to Cel-miR-61-5p). Considering the ambiguity associated with the enrichment of different miR-61 isoforms in both males and females, we focussed on miR-228, miR-236 and miR-252. In males, the enrichment of miR-228 coincides with a down-regulation of 412 transcripts (representing 320 *T. suis* genes) with a predicted miR-228 binding site. Based on their function in *C. elegans* homologues, many of these transcripts are inferred to be linked to vulva development (e.g., *exc-4*, *mys-1*, *nekl-2* and *sem-4*), egg production (e.g., *cbd-1*, *nsy-1*, *ppt-1*, *unc-29* and *unc-58*) and embryogenesis/germline development (e.g., *bcat-1*, *lrs-1*, *rnp-4*, *rpt-5*, *slt-1* and *tbp-1*). In females, enriched transcription of miR-236 and miR-252 coincides with a down-regulation of 262 transcripts (representing 205 genes) predicted to be co-regulated by these miRNAs. Homologues of these 'female-suppressed' coding genes in *C. elegans* include genes involved in spermatogenesis (e.g., *cogc*-5 and *cpb-1*), male mating/fertility (e.g., *goa-1* and *odc-1*), the regulation of germline specification or apoptosis (e.g., *glp-1*, *him-1*, *rpt-5*, *let-60*, *vps-16* and *vps-41*) and chemosensation (e.g., *crh-1*, *grk-2* and *lys-2*). Collectively, these data suggest that sexual dimorphism in *T. suis* might relate, at least partially, to post-transcriptional sex-suppression by miRNAs rather than exclusive transcriptional promotion by mRNAs.

In addition to miRNAs, we identified 1,028,808 putative small RNAs mapping to coding regions of the genome. Of these RNAs, 673,355 mapped antisense (≥ 80% of reads at each location) to exons, suggesting a potential role as siRNAs[41]. Most abundant among the siRNAs predicted for *T. suis* (except for those derived from males) were sequences of 24 and 25 nt with a 5′ guanine, compared with 22 and 26G sequences predominating among siRNAs predicted for other nematodes to date[40,41]. Putative siRNAs were predicted for 3,497 protein-coding genes. Many siRNAs have key roles in germline tissues[40,41]. In *T. suis*, we identified transcripts for 508 coding genes, for which putative siRNAs were uniquely transcribed in the adult female and the female posterior body relative to the stichosome, the adult male and the male posterior body (Supplementary Tables 7, 18 and 23). These coding genes were enriched for transposable elements/transposases, histones/histone methytransferases, DNA/RNA-binding, chromatin folding and homeodomain-related proteins. Although less abundant, there was a similar enrichment for these functions in relation to the 69 coding genes associated with siRNAs uniquely transcribed in the adult male and the male posterior body relative to the stichosome and female worm. We hypothesize that these highly transcribed siRNAs protect chromatin in the *T. suis* germline, which is supported by the observation that 162 of the female-enriched siRNAs are absent from the larval stages studied here (Supplementary 7, 18 and 23).

Novel class of tncRNAs

Conspicuous among the *T. suis* small RNAs is an abundance of 22 nt sequences with a 5′ adenine cap. Although representing just 2.9% (n = 58,307) of consensus small RNAs, these sequences represent 9.2% of all small RNA transcription. By location, 22-nt 5′-adenylated sequences are evenly distributed between coding and non-coding regions, and within non-coding regions, between annotated (e.g., transposable elements, tRNAs or other known, non-coding RNAs) and un-annotated space. However, 89% of transcription attributed to these sequences relates to un-annotated, non-coding space in the genome. Based on their size and abundance, these sequences are consistent with tncRNAs[43]; however, they have characteristics not previously attributed to this class. For instance, in *T. suis*, they have a significant strand-bias, with 83% of their transcription occurring on the Watson (i.e., antisense) strand. These ‘antisense’-biased tncRNAs (henceforth called 22A-RNAs) form 1,208 clusters (ranging from 22 to 11,831 nt) among 238 assembly scaffolds, with a median of three 22A-RNA sequences per cluster at a median spacing of 97 bp. Although 40% of multi-copy 22A-RNA sequences are found in the same cluster, clusters comprising one repeated 22A-RNA sequence are rare, and sequences are often shared among clusters and genomic scaffolds, indicating that tandem duplication is not the only mechanism associated with cluster formation. Few transposable elements are found within 25 kb of these clusters, suggesting that their insertion/translocation within the genome is not recent.

At this stage, we can only speculate about the function(s) and mechanism(s) of action of these sequences. Their 100-nt genomic neighbourhoods are a mixture of some regions resembling (but not overlapping with) known protein-coding sequences, others resembling known ncRNAs such as tRNAs and still others that resemble neither. Eleven of these neighbourhoods show partial similarities to cryptic tRNAs, ~100 nt in size, discovered in *C. elegans* by the modENCODE Consortium[44]. The sequences of the 22A-RNAs themselves

are also heterogeneous, with only one significant 8-nt sequence motif (5′-A[CA]GATAT[GT]-3′) occurring in 4.5% (245/5,457) of 22A-RNA sequences (Supplementary Fig. 12). Given these findings, we propose that 22A-RNAs may be processed from larger ncRNAs of diverse types, some of which are highly conserved and familiar, others of which are both hypothetical and unfamiliar. Despite having no obvious promoter motif (such as that proposed for 21U-RNAs[41]), these sequences do appear to be transcriptionally regulated, with their abundance varying significantly among stages, sexes and tissues in *T. suis*; including, notably, a significant enrichment in the adult male body and male posterior body relative to all other stages/tissues (Fig. 2D), which may suggest a role in the male germline. As a proportion of overall sRNA transcription, 22A-RNAs are most abundant in the stichosome, wherein they comprise ~22% of all sRNA reads determined. Indeed, the stichosome is notably restricted in its classes of sRNAs, with miRNAs (39% of all sRNA reads from the stichosome) and 22A-RNAs dominating the sRNA population in this organ. Whether this finding points to an involvement of these novel *T. suis* ncRNAs in host interactions deserves detailed investigation.

DISCUSSION

Globally, helminthiases are seriously neglected causes of morbidity and mortality. Genomic and transcriptomic explorations of *T. suis* should enable the design of urgently needed therapeutics against human trichuriasis, one of the world's most significant and neglected helminthiases. An intriguing feature of *T. suis* is its possible use as a therapy for human autoimmune disorders[9–11]. A detailed characterisation of how this parasite modulates the host immune response is thus a key priority. Secreted proteins (including cystatins/serpins, thioredoxin peroxidase and various putative mimics of host proteins) appear to have a central role in this process, primarily through inhibiting inflammation. Our findings indicate a role for parasite-derived lipids, including the inferred synthesis of β-glucosylceramide, a known anti-inflammatory and putative therapy for IBD[45], during the early developmental phase of *T. suis* (see Supplementary Results). It is likely that both proteins and lipids work in concert with N-linked glycans, which are known immuno-modulators produced by *T. suis*[27], particularly L4 and adult stages, in which pathways associated with their synthesis are transcriptionally enriched (see Supplementary Results). The detailed characterization of these molecules *in vitro* and *in vivo*, using existing models of IBD and other autoimmune disorders, might pave the way for parasite-derived therapies[5]. Indeed, a better understanding of the *T. suis*-host interactions might shed new light on why helminth exposure seems crucial for the development of a healthy immune system in humans. This is the first study to characterize the genomes of male and female individuals of a dioecious nematode. We found little evidence for sex-specific genes or assembly contigs, despite the reported XY-karyotype of this species. However, intriguingly, miRNAs appear to play a major role in regulating sexual development in this species, with tsu-miR-228 in male, and tsu-miR-236 and tsu-miR-252 in female worms predicted to regulate and suppress key feminizing and masculinizing developmental genes, respectively. This is the first time that this has been observed for a metazoan.

URLs

Salient data files are accessible via http://gasser-research.vet.unimelb.edu.au/Trichuris_suis/ and ftp://ftp.wormbase.org/pub/wormbase/species/t_suis. Browsable male and female genomes are accessible via http://gasser-research.vet.unimelb.edu.au/jbrowse/JBrowse-1.11.2/index.html?data=TsuisMale and http://gasser-research.vet.unimelb.edu.au/jbrowse/JBrowse-1.11.2/index.html?data=TsuisFemale, respectively, or through Wormbase via ftp://ftp.wormbase.org/pub/wormbase/species/t_suis/.

ONLINE METHODS

Sample preparation and storage

Trichuris suis were isolated from experimentally infected pigs inoculated orally with a single dose of 5,000–50,000 embryonated eggs (Animal Ethics Permission No. 2010/561-1914; University of Copenhagen). Individuals of *T. suis* were isolated at 10 (L1/L2 larvae), 18 (L3s), 28 (L4s) and 49 (adulthood) days post-inoculation (p.i.)[29,46] and washed in physiological saline (37 °C) and RPMI1640 (GIBCO) with Antibiotic-Antimycotic (GIBCO). Adult male and female *T. suis* were separated. Stichosomes were excised from whole adult worms (n = 10, irrespective of sex), pooled and frozen, as were the posterior portions of the worms (n = 10 of each sex). All stages/tissues were snap frozen in liquid nitrogen and stored at −80 °C.

Genomic sequencing and assembly

Total genomic DNA was isolated each from a single adult male or female *T. suis*[47,48]. Paired-end (insert sizes: 170 bp and 500 bp) and mate-paired (800 bp, 2 kb, 5 kb and 10 kb) libraries were constructed from total and whole genomic amplified (WGA) genomic DNA, respectively[14,49] and sequenced using a HiSeq 2000 machine (Illumina). Low quality sequences, base-calling duplicates and adapters were removed using standard approaches. Sequence quality and heterozygosity were assessed by 17-mer frequency distribution[50], and genome sizes estimated[51]. Corrected and filtered data were assembled into contigs using SOAPdenovo v2.0[50] and assessed for accuracy using SOAP2aligner[52]. Assembly completeness and redundancy were assessed using CEGMA[53] and RNA-seq data using Bowtie2[54].

RNA isolation and RNA-seq

Total RNAs from L1/L2 (n = 50,000 from 5 pigs), L3 (n = 15,000 from 4 pigs) and L4 (n = 3,000 from 2 pigs), adult male (n = 10), adult female (n = 10), and stichosomal (mixed sex; n = 10) and non-stichosomal portions of adult females (n = 10) and males (n = 10) were individually purified using TriPure reagent (Roche). Polyadenylated (polyA+) RNA was purified from 10 μg of total RNA for each library using Sera-mag oligo(dT) beads and fragmented, purified and sequenced using HiSeq 2000[14,49]. Small non-coding RNAs (~18–30 nt) were isolated from 10 μg of total RNA for each library by size fractionation on polyacrylamide gels, purified, adapter-ligated, reverse transcribed, amplified by PCR and sequenced using HiSeq 2000. All RNA-seq data were adaptor trimmed, and length- and quality-filtered using standard approaches.

Synteny and polymorphism analysis, and annotation of repeat content

For comparative analysis, the assemblies for adult male or female *T. suis* were aligned using MUMmer3[55]. Repetitive sequences in each assembly were identified[14], using Tandem Repeats Finder (TRF)[56], RepeatMasker[57], LTR_FINDER[58], PILER[59] and RepeatScout[60], with a consensus population of predicted repetitive elements constructed in RepeatScout using fit-preferred alignment scores. Transfer RNAs were predicted using tRNA-SCAN[61]. The male assembly was explored for scaffolds likely representing the male-specific Y-chromosome[13,62]. Reads from all genomic sequence libraries each for male and female *T. suis* were aligned to their own and the opposite sex (both repeat unmasked and hard-masked) assembly (i.e., male-to-male, male-to-female, female-to-male and female-to-female) using Bowtie2[54]. Contigs with >80% coverage in same-sex, but <20% coverage in opposite-sex read alignments were deemed ‘sex-specific’.

Prediction and functional annotation of the protein-encoding gene-set

The male and female protein-coding gene-set of *T. suis* was inferred in MAKER2[63]. Briefly (1) the non-redundant *T. suis* transcriptome was aligned each assembly using BLAT[64] and filtered for full-length open reading frames (ORFs), which (2) were used to train Hidden Markov Models (HMM) for *de novo* gene prediction using SNAP[65] and AUGUSTUS[66]; with these models supplemented using (3) homologous genes from *Tr. spiralis*[17] and *C. elegans*[16]; and (4) all *T. suis* RNA-seq data from all libraries used to infer each transcript using Tophat2[67] and Cufflinks2[68,69]; (5) all HMM-predicted, homology and evidence-based information was combined into a single consensus gene-set and (6) genes overlapping with predicted repetitive regions of the genome and/or having significant homology to known repetitive sequences (i.e., transposable elements) in RepBase[57] and no close homology to *C. elegans* or *T. spiralis* protein-coding genes were removed. The male and female *T. suis* gene-sets were unified by orthology prediction using InParanoid[70], with *Trichinella spiralis*[17] as an out-group.

Each gene was assessed for encoding conserved protein domains using InterProScan[71], with these data used to infer Gene Ontology[72]. Using Reciprocal BLASTp and OrthoMCL[73], the *T. suis* inferred proteome was clustered with predicted homologues/orthologues for other nematodes, including *Ascaris suum*[14], *Brugia malayi*[15], *C. elegans*[16], and *T. spiralis*[17]. Each contig was assessed for a known functional orthologue in the Kyoto Encylopedia of Genes and Genomes (KEGG) using the KEGG orthology bases annotation system (kobas)[74]. In addition, *T. suis* inferred proteins were compared by BLASTx/BLASTp with protein sequences available for *A. suum*, *B. malayi*, *C. elegans* and *T. spiralis*, and in the databases UniProt[75], SwissProt and TREMBL[76], as well as specialist databases for key protein groups represented in MEROPS[77], WormBase[78], KS-Sarfari and GPCR-Sarfari, and the Transporter Classification database (TCDB)[79]. Excretory/secretory (ES) proteins were predicted using Phobius[80] and by BLASTp comparison with the validated signal peptide database (SPD)[81] and proteomic data for the nematodes *B. malayi*[82,83] and *Meloidogyne incognita*[84] and the trematode *Schistosoma mansoni*[85].

Differential transcription analysis of messenger RNA

Reconstruction and quantification [in Fragments per Kilobase per Million Reads (FPKMs)] of the *T. suis* transcriptome was conducted using TopHat2[67] and Cufflinks2[68,69]. Predicted alternative splice events were classified[86]. Comparisons of splice events and gene function (based on encoded Pfam domains) were conducted by pair-wise Pearson Chi-squared analysis (p-value ≤ 0.05). We also compared the relationship between gene essentiality and being a single or multi-isoform gene, with essentiality predicted[14]. Differential transcription was assessed using NOISeq[87], with 20% of the evaluated reads for each library used in 5 iterations to simulate technical replicates.

Annotation and differential transcription analysis of small non-coding RNAs

Canonical micro-RNAs (miRNAs) were identified and quantified in miRDeep2[88–90], and supported using miRNAs published for *A. suum*[40], *C. elegans*[91], *Haemonchus contortus*[92], *Brugia pahangi*[92] and *T. spiralis*[93]. The 3′ untranslated region (3′-UTR) for each Cufflinks-predicted *T. suis* transcript was identified by comparison with the *T. suis* genome annotation. Each 3′-UTR was screened for miRNA binding sites using the program PITA[94]. These binding sites were filtered based on homologous miRNA/transcript binding interactions predicted for *C. elegans* in curated databases (microRNA.org, RNA22, and TargetScan) or demonstrated empirically[42]. All non-miRNA reads from each small RNA library for *T. suis*, were then aligned to the male *T. suis* genome using Bowtie2[54] and clustered using ShortStack[95], with a minimum cluster depth cut-off of 10. Small RNA reads having perfect alignment overlap (i.e., the same start and stop position) were defined as homologous and condensed into a consensus sequence by majority-rule. Each consensus small RNA was classified[41] and nucleotide diversity within homologous sRNA reads was assessed using custom Perl scripts. Specific small RNA sequences (e.g., 21U-RNAs[41]) and their 5′ and 3′ flanking regions were explored for sequence motifs using MEME[96]. Differential transcription among stage/tissue specific libraries was assessed (in reads per million mapped reads, RPKM) for miRNAs, siRNAs and 22A-RNAs using NOISeq[87], and clustered by stage/tissue specific transcription pattern using R.

Analysis of the genomic neighborhoods and primary sequences of 22A-RNAs

We extracted 22A-RNAs with 100-nt flanks lacking scaffolding (‘N’) residues, merged those with spatial overlaps along the genome, and further condensed them to 80% sequence identity with CD-HIT-EST[107]. We probed resultant nonredundant 22A-RNA regions for protein-coding exons with BlastX[79], and known ncRNAs with INFERNAL 1.1/cmscan[108]. BlastX was run against predicted proteomes from all published nematode genomes in WormBase WS240[87], as well as the *T. suis* male proteome from this study; cmscan was run against the ncRNA database RFAM 11.0[109]. Regions passing these filters were tested for similarity to (1) genomic DNA from other nematode species, (2) to other 22A-RNA neighborhoods, and (3) to novel ncRNAs from *C. elegans*. The first was assayed by BlastN against published nematode genomic sequences in WormBase WS240[87]. The second was assayed by BlastN against 22A-RNA regions spatially merged for genomic overlaps but not condensed with CD-HIT-EST. The third was assayed by BlastN against a set of 8,126 *C.*

elegans ncRNAs taken from the WS240 release of WormBase[87]. All searches used E-value thresholds of $\leq 10^{-3}$.

A set of 25,259 *C. elegans* ncRNAs were obtained from WormBase WS240. We filtered out those named ‘asRNA’, ‘rRNA’, ‘scRNA’, ‘snoRNA’, ‘snRNA’, or ‘tRNA’, leaving 8,126 ncRNAs with no official similarity to well-known structures. Most of these ncRNAs had been discovered by modENCODE[45]; 176 others were long non-coding RNAs[110]. We then checked for previously undescribed motifs via RFAM 11.0 and cmscan.

To discover whether 22A-RNA sequences contained novel motifs, we extracted 22A-RNA sequences without flanking protein-coding or ncRNA similarities, merged them spatially and for 80% identity, and scanned with MEME[106], using a first-order Markov model from the adult male *T. suis* genome (via MEME’s fasta-get-markov). We ran MEME with arguments: ‘-dna -revcomp -nsites 100 -bfile TS_M_200bpormore.1markov.txt -nmotifs 10 -evt 0.05 -minw 6 -maxw 22 -mod anr’. For one resulting 8-nt motif, we used FIMO[111] to determine where it occurred in original, unmerged 22A-RNA sequences, with arguments: ‘--bgfile TS_M_200bpormore.1markov.txt --output-pthresh 0.001’. The motif was displayed as a logarithmic WebLogo[112].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was funded by the Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC) of Australia and BGI-Shenzhen. This research was supported by a Victorian Life Sciences Computation Initiative (VLSCI) grant number VR0007 on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government. Other support to R.B.G. from the Alexander von Humboldt Foundation, Australian Academy of Science, the Australian-American Fulbright Commission, Melbourne Water Corporation and the IBM Research Collaboratory for Life Sciences – Melbourne is gratefully acknowledged. P.N. was supported by Danish Agency for Science, Technology and Innovation. N.D.Y. holds an NHMRC Early Career Research Fellowship. E.M.S. was supported by startup funds from Cornell University. P.W.S. is an investigator with the Howard Hughes Medical Institute (HHMI) and acknowledges support from the National Institutes of Health (NIH). We are indebted to the staff of the BGI-Shenzhen who assisted this study but whose names are not included in the authorship. We also acknowledge the continued contributions of staff at WormBase.

References

1. Hotez PJ, Fenwick A, Savioli L, Molyneux DH. Rescuing the bottom billion through control of neglected tropical diseases. Lancet. 2009; 373:1570–1575. [PubMed: 19410718]
2. Stephenson LS, Holland CV, Cooper ES. The public health significance of Trichuris trichiura. Parasitology. 2000; 121 (Suppl):S73–95. [PubMed: 11386693]
3. Schistosomiasis and soil-transmitted helminth infections (World Health Assembly Resolution WHA54.19). 2001. (http://www.who.int/wormcontrol/about_us/en/ea54r19.pdf)
4. Okada H, Kuhn C, Feillet H, Bach JF. The ‘hygiene hypothesis’ for autoimmune and allergic diseases: an update. Clin Exp Immunol. 2010; 160:1–9. [PubMed: 20415844]
5. Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med. 2002; 347:911–920. [PubMed: 12239261]
6. Elliott DE, Summers RW, Weinstock JV. Helminths as governors of immune-mediated inflammation. Int J Parasitol. 2007; 37:457–464. [PubMed: 17313951]

7. Endara P, Vaca M, Chico ME, et al. Long-term periodic anthelmintic treatments are associated with increased allergen skin reactivity. Clin Exp Allergy. 2010; 40:1669–1677. [PubMed: 21039971]
8. Rodrigues LC, Newcombe PJ, Cunha SS, et al. Early infection with *Trichuris trichiura* and allergen skin test reactivity in later childhood. Clin Exp Allergy. 2008; 38:1769–1777. [PubMed: 18547322]
9. Summers RW, Elliott DE, Urban JF Jr, Thompson RA, Weinstock JV. *Trichuris suis* therapy for active ulcerative colitis: a randomized controlled trial. Gastroenterology. 2005; 128:825–832. [PubMed: 15825065]
10. Summers RW, Elliott DE, Qadir K, et al. *Trichuris suis* seems to be safe and possibly effective in the treatment of inflammatory bowel disease. Am J Gastroenterol. 2003; 98:2034–2041. [PubMed: 14499784]
11. Fleming JO, Isaak A, Lee JE, et al. Probiotic helminth administration in relapsing-remitting multiple sclerosis: a phase 1 study. Mult Scler. 2011; 17:743–754. [PubMed: 21372112]
12. Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Mol Biochem Parasitol. 2009; 167:1–11. [PubMed: 19406170]
13. Spakulova M, Kralova I, Cutillas C. Studies on the karyotype and gametogenesis in *Trichuris muris*. J Helminthol. 1994; 68:67–72. [PubMed: 8006389]
14. Jex AR, Liu S, Li B, et al. *Ascaris suum* draft genome. Nature. 2011; 479:529–533. [PubMed: 22031327]
15. Ghedin E, Wang S, Spiro D, et al. Draft genome of the filarial nematode parasite *Brugia malayi*. Science. 2007; 317:1756–1760. [PubMed: 17885136]
16. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 1998; 282:2012–2018. [PubMed: 9851916]
17. Mitreva M, Jasmer DP, Zarlenga DS, et al. The draft genome of the parasitic nematode *Trichinella spiralis*. Nat Genet. 2011; 43:228–235. [PubMed: 21336279]
18. Kaminsky R, Ducray P, Jung M, et al. A new class of anthelmintics effective against drug-resistant nematodes. Nature. 2008; 452:176–180. [PubMed: 18337814]
19. Cooper ES, Whyte-Alleng CA, Finzi-Smith JS, MacDonald TT. Intestinal nematode infections in children: the pathophysiological price paid. Parasitology. 1992; 104 (Suppl):S91–103. [PubMed: 1589304]
20. Drake LJ, Bianco AE, Bundy DA, Ashall F. Characterization of peptidases of adult *Trichuris muris*. Parasitology. 1994; 109 (Pt 5):623–630. [PubMed: 7831097]
21. Tilney LG, Connelly PS, Guild GM, Vranich KA, Artis D. Adaptation of a nematode parasite to living within the mammalian epithelium. J Exp Zoolog A Comp Exp Biol. 2005; 303:927–945.
22. Toubarro D, Lucena-Robles M, Nascimento G, et al. Serine protease-mediated host invasion by the parasitic nematode *Steinernema carpocapsae*. J Biol Chem. 2010; 285:30666–30675. [PubMed: 20656686]
23. Balasubramanian N, Toubarro D, Simoes N. Biochemical study and in vitro insect immune suppression by a trypsin-like secreted protease from the nematode Steinernema carpocapsae. Parasite Immunol. 2010; 32:165–175. [PubMed: 20398179]
24. Toubarro D, Lucena-Robles M, Nascimento G, et al. An apoptosis-inducing serine protease secreted by the entomopathogenic nematode *Steinernema carpocapsae*. Int J Parasitol. 2009; 39:1319–1330. [PubMed: 19481087]
25. Drake L, Korchev Y, Bashford L, et al. The major secreted product of the whipworm, *Trichuris*, is a pore-forming protein. Proc Biol Sci. 1994; 257:255–261. [PubMed: 7991635]
26. Kringel H, Iburg T, Dawson H, Aasted B, Roepstorff A. A time course study of immunological responses in *Trichuris suis* infected pigs demonstrates induction of a local type 2 response associated with worm burden. Int J Parasitol. 2006; 36:915–924. [PubMed: 16750534]
27. Klaver EJ, Kuijk LM, Laan LC, et al. *Trichuris suis*-induced modulation of human dendritic cell function is glycan-mediated. Int J Parasitol. 2013; 43:191–200. [PubMed: 23220043]
28. Stamm S, Ben-Ari S, Rafalska I, et al. Function of alternative splicing. Gene. 2005; 344:1–20. [PubMed: 15656968]

29. Beer RJ. Studies on the biology of the life-cycle of *Trichuris suis* Schrank, 1788. Parasitology. 1973; 67:253–262. [PubMed: 4761766]
30. Sheffield HG. Electron Microscopy of the Bacillary Band and Stichosome of *Trichuris muris* and *T. vulpis*. J Parasitol. 1963; 49:998–1009. [PubMed: 14084216]
31. Hoover-Plow J. Does plasmin have anticoagulant activity? Vasc Health Risk Manag. 2010; 6:199–205. [PubMed: 20407627]
32. MacDonald TT, Choy MY, Spencer J, et al. Histopathology and immunohistochemistry of the caecum in children with the *Trichuris* dysentery syndrome. J Clin Pathol. 1991; 44:194–199. [PubMed: 2013619]
33. Burrows RB, Lillis WG. The Whipworm as a blood sucker. J Parasitol. 1964; 50:675–680. [PubMed: 14215490]
34. Hasnain SZ, Wang H, Ghia JE, et al. Mucin gene deficiency in mice impairs host resistance to an enteric parasitic infection. Gastroenterology. 2010; 138:1763–1771. [PubMed: 20138044]
35. Hasnain SZ, McGuckin MA, Grencis RK, Thornton DJ. Serine protease(s) secreted by the nematode *Trichuris muris* degrade the mucus barrier. PLoS Negl Trop Dis. 2012; 6:e1856. [PubMed: 23071854]
36. Golias C, Charalabopoulos A, Stagikas D, Charalabopoulos K, Batistatou A. The kinin system--bradykinin: biological effects and clinical implications. Hippokratia. 2007; 11:124–128. [PubMed: 19582206]
37. Da'dara A, Skelly PJ. Manipulation of vascular function by blood flukes? Blood Rev. 2011; 25:175–179. [PubMed: 21543145]
38. Marceau F, Regoli D. Therapeutic options in inflammatory bowel disease: experimental evidence of a beneficial effect of kinin B1 receptor blockade. Br J Pharmacol. 2008; 154:1163–1165. [PubMed: 18536746]
39. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009; 10:94–108. [PubMed: 19148191]
40. Wang J, Czech B, Crunk A, et al. Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. Genome Res. 2011; 21:1462–1477. [PubMed: 21685128]
41. Ruby JG, Jan C, Player C, et al. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell. 2006; 127:1193–1207. [PubMed: 17174894]
42. Zisoulis DG, Lovci MT, Wilbert ML, et al. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. Nat Struct Mol Biol. 2010; 17:173–179. [PubMed: 20062054]
43. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr Biol. 2003; 13:807–818. [PubMed: 12747828]
44. Lu ZJ, Yip KY, Wang G, et al. Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. Genome Res. 2011; 21:276–285. [PubMed: 21177971]
45. Zigmond E, Preston S, Pappo O, et al. Beta-glucosylceramide: a novel method for enhancement of natural killer T lymphoycte plasticity in murine models of immune-mediated disorders. Gut. 2007; 56:82–89. [PubMed: 17172586]
46. Kringel H, Roepstorff A, Murrell KD. A method for the recovery of immature *Trichuris suis* from pig intestine. Acta Vet Scand. 2002; 43:185–189. [PubMed: 12564548]
47. Gasser R. Molecular tools - advances, opportunities and prospects. Veterinary Parasitology. 2006; 136:69–89. [PubMed: 16457951]
48. Sambrook, J.; Russell, DW. Molecular cloning: a laboratory manual. 3. Cold Spring Harbor Laboratory; 2001.
49. Young ND, Jex AR, Li B, et al. Whole-genome sequence of *Schistosoma haematobium*. Nat Genet. 2012; 44:221–225. [PubMed: 22246508]
50. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]

51. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics. 1988; 2:231–239. [PubMed: 3294162]
52. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009; 25:1966–1967. [PubMed: 19497933]
53. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. Nucleic Acids Res. 2009; 37:289–297. [PubMed: 19042974]
54. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]
55. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics. 2003; Chapter 10(Unit 10):13.
56. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999; 27:573–580. [PubMed: 9862982]
57. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009; Chapter 4(Unit 4):10. [PubMed: 19274634]
58. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007; 35:W265–268. [PubMed: 17485477]
59. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005; 21 (Suppl 1):i152–158. [PubMed: 15961452]
60. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005; 21 (Suppl 1):i351–358. [PubMed: 15961478]
61. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25:955–964. [PubMed: 9023104]
62. Carvalho AB, Clark AG. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. Genome Res. 2013; 23:1894–1907. [PubMed: 23921660]
63. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011; 12:491. [PubMed: 22192575]
64. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002; 12:656–664. [PubMed: 11932250]
65. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004; 5:59. [PubMed: 15144565]
66. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005; 33:W465–467. [PubMed: 15980513]
67. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]
68. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. [PubMed: 20436464]
69. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7:562–578. [PubMed: 22383036]
70. O'Brien KP, Remm M, Sonnhammer EL. Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 2005; 33:D476–480. [PubMed: 15608241]
71. Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier. Nucleic Acids Res. 2005; 33:W116–120. [PubMed: 15980438]
72. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004; 32:D258–261. [PubMed: 14681407]
73. Fischer S, Brunk BP, Chen F, et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics. 2011; Chapter 6(Unit 6):12, 11–19. [PubMed: 21901743]
74. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res. 2006; 34:W720–724. [PubMed: 16845106]
75. Wu CH, Apweiler R, Bairoch A, et al. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res. 2006; 34:D187–191. [PubMed: 16381842]

76. Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003; 31:365–370. [PubMed: 12520024]
77. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. Nucleic Acids Res. 2010; 38:D227–233. [PubMed: 19892822]
78. Harris TW, Baran J, Bieri T, et al. WormBase 2014: new views of curated biology. Nucleic Acids Res. 2014; 42:D786–93.
79. Saier MH Jr, Yen MR, Noto K, Tamang DG, Elkan C. The Transporter Classification Database: recent advances. Nucleic Acids Res. 2009; 37:D274–278. [PubMed: 19022853]
80. Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 2007; 35:W429–432. [PubMed: 17483518]
81. Chen Y, Zhang Y, Yin Y, et al. SPD--a web-based secreted protein database. Nucleic Acids Res. 2005; 33:D169–173. [PubMed: 15608170]
82. Bennuru S, Semnani R, Meng Z, et al. *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. PLoS Negl Trop Dis. 2009; 3:e410. [PubMed: 19352421]
83. Hewitson JP, Harcus YM, Curwen RS, et al. The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. Mol Biochem Parasitol. 2008; 160:8–21. [PubMed: 18439691]
84. Bellafiore S, Shen Z, Rosso MN, et al. Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. PLoS Pathog. 2008; 4:e1000192. [PubMed: 18974830]
85. Cass CL, Johnson JR, Califf LL, et al. Proteomic analysis of *Schistosoma mansoni* egg secretions. Mol Biochem Parasitol. 2007; 155:84–93. [PubMed: 17644200]
86. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456:470–476. [PubMed: 18978772]
87. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011; 21:2213–2223. [PubMed: 21903743]
88. Friedlander MR, Chen W, Adamidi C, et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol. 2008; 26:407–415. [PubMed: 18392026]
89. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012; 40:37–52. [PubMed: 21911355]
90. Mackowiak SD. Identification of novel and known miRNAs in deep-sequencing data with miRDeep2. Curr Protoc Bioinformatics. 2011; Chapter 12(Unit 12):10. [PubMed: 22161567]
91. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006; 34:D140–144. [PubMed: 16381832]
92. Winter AD, Weir W, Hunt M, et al. Diversity in parasitic nematode genomes: the microRNAs of *Brugia pahangi* and *Haemonchus contortus* are largely novel. BMC Genomics. 2012; 13:4. [PubMed: 22216965]
93. Chen MX, Ai L, Xu MJ, et al. Identification and characterization of microRNAs in *Trichinella spiralis* by comparison with *Brugia malayi* and Caenorhabditis elegans. Parasitol Res. 2011; 109:553–558. [PubMed: 21327987]
94. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007; 39:1278–1284. [PubMed: 17893677]
95. Axtell MJ. ShortStack: comprehensive annotation and quantification of small RNA genes. RNA. 2013; 19:740–751. [PubMed: 23610128]
96. Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol. 1995; 3:21–29. [PubMed: 7584439]

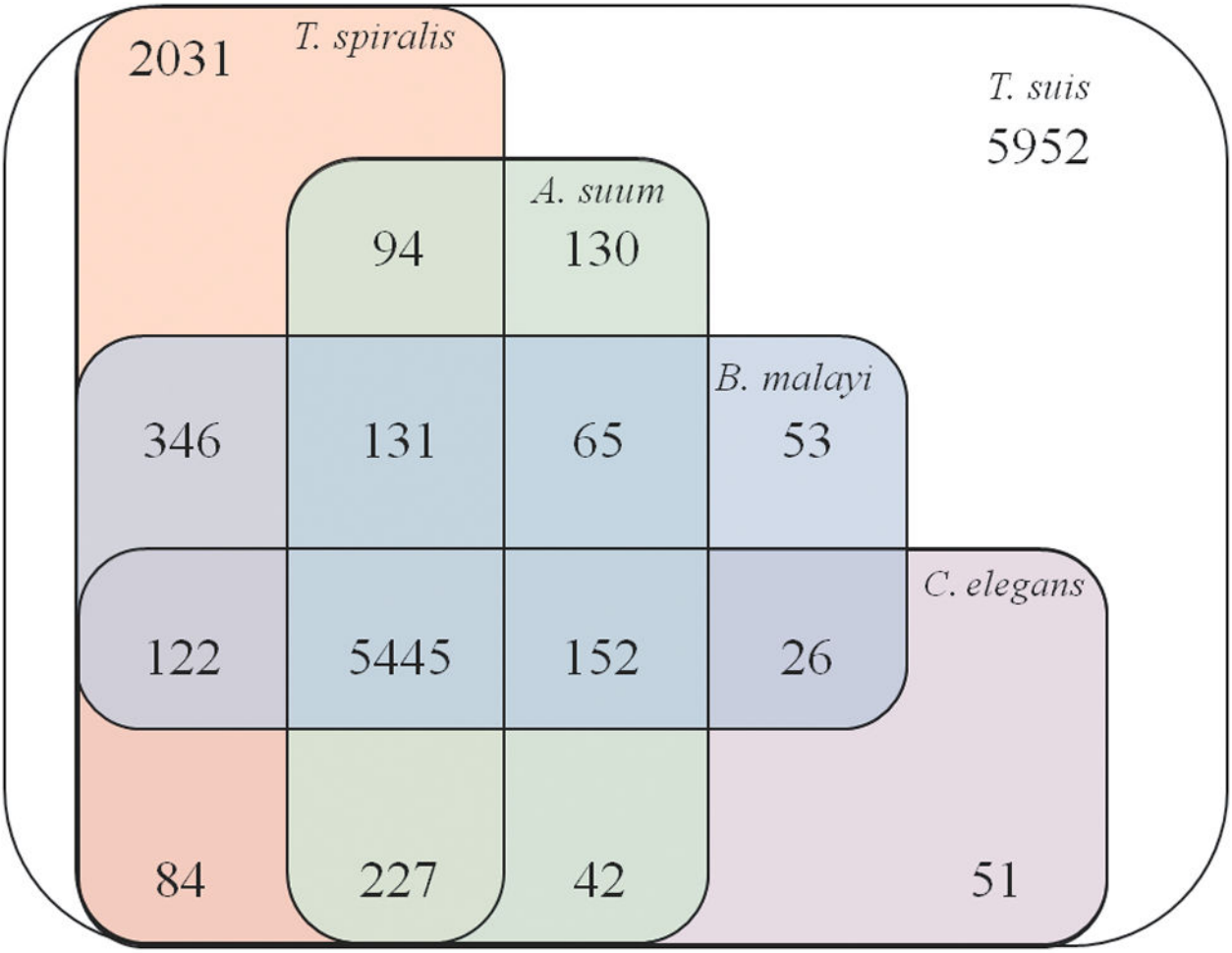

Figure 1.

Homologues shared between *Trichuris suis* (Class Enoplea, Order Trichocephalida) and related nematode species.

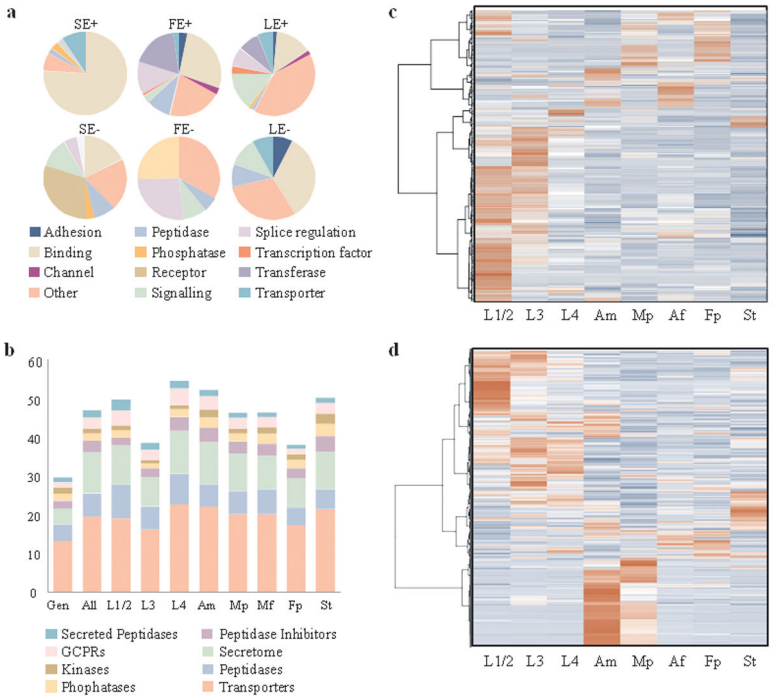

Figure 2.

Stage and tissue-specific small and messenger RNA transcriptome of *Trichuris suis*. (a) Association between gene function and alternative splice variation. Charts show the inferred function of protein domains encoded by genes showing a statistically significant ($p \leq 0.05$) positive (+) or negative (−) bias toward skipped exon (SE), alternative first (FE) or last exon (LE) splice events. Only genes encoding 10 or more transcripts are included in this analysis. (b) Proportional representation of major protein classes/groups encoded by the genome (Gen), and their proportional abundance in all transcriptomic data (All) and in larval (L1/2, L3 and L4), adult male (Am), female (Af) and tissue specific libraries, including in the male (Mp) and female posterior-body (Fp) and the stichosome (St). (c) Self-organizing heatmap (TPM values normalized by gene) clustering microRNAs by their transcription abundance (represented as $\log_2$-transformed RPKM values) in each larval, adult and tissue-specific library. (d) Self-organizing heatmap (TPM values normalized by gene) clustering 22A-RNAs by their transcription abundance (represented as $\log_2$-transformed RPKM values) in each larval, adult and tissue-specific library.

Table 1

Features of the scaffolded assembly of the adult male and female *Trichuris suis* genomes

	Male genome	Female genome
k-mer (17) estimated genome size (in megabases, Mb)	83.6	87.2
Total read data; estimated coverage	11.73 Gb; 140x	12.36 Gb; 142x
Total scaffolded assembly size (Mb); total scaffolds	81.3; 60,856	76.0; 42,663
Total scaffolds of > 200 bp: length (Mb); no. of scaffolds	74.2; 4,293	71.0; 3,288
Largest scaffold (Mb)	1.59	1.44
N50 in kb (scaffolds of > 200 bp)	500	440
N90 in kb (scaffolds of > 200 bp); total number for > N90	81.0; 185	104; 168
% GC content: whole genome; scaffolds > 200 bp)	43.9; 43.6	43.6; 43.5
% repetitive sequence (scaffolds > 200 bp)	31.7	32.3
Proportion of the genome that is coding: exonic; including introns	26.2; 66.2	29.2; 73.4
Number of protein-coding genes	14,781	14,470
Mean gene length (kb)	3.7	3.9
Mean number of exons per gene	5.4	5.7
Mean exon length (bp)	271	270
Mean number of introns per gene	4.4	4.7
Mean intron length (bp)	511	509
Number of transfer RNAs	991	1,021

Author/s:
Jex, AR; Nejsum, P; Schwarz, EM; Hu, L; Young, ND; Hall, RS; Korhonen, PK; Liao, S; Thamsborg, S; Xia, J; Xu, P; Wang, S; Scheerlinck, J-PY; Hofmann, A; Sternberg, PW; Wang, J; Gasser, RB

Title:
Genome and transcriptome of the porcine whipworm Trichuris suis

Date:
2014-07-01

Citation:
Jex, A. R., Nejsum, P., Schwarz, E. M., Hu, L., Young, N. D., Hall, R. S., Korhonen, P. K., Liao, S., Thamsborg, S., Xia, J., Xu, P., Wang, S., Scheerlinck, J. -P. Y., Hofmann, A., Sternberg, P. W., Wang, J. & Gasser, R. B. (2014). Genome and transcriptome of the porcine whipworm Trichuris suis. NATURE GENETICS, 46 (7), pp.701-706. https://doi.org/10.1038/ng.3012.