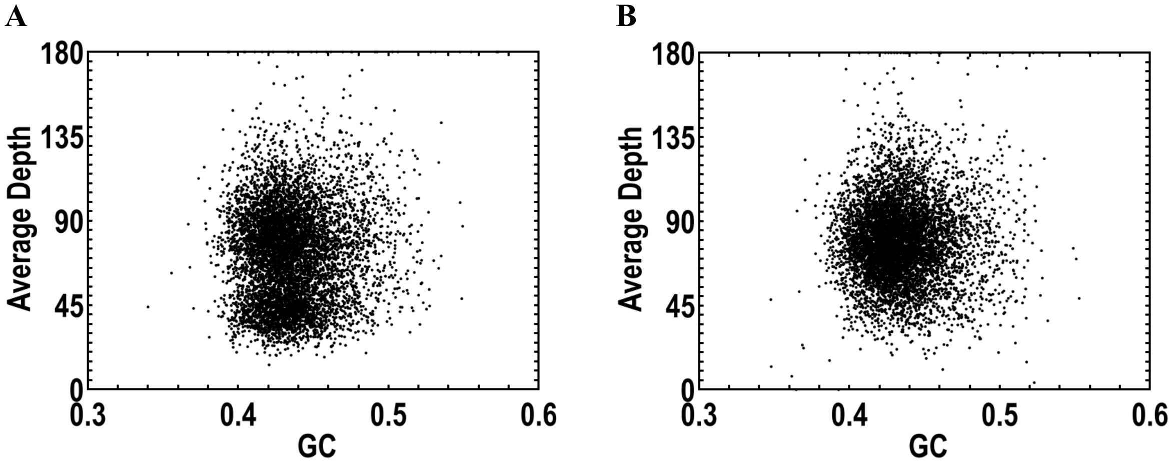
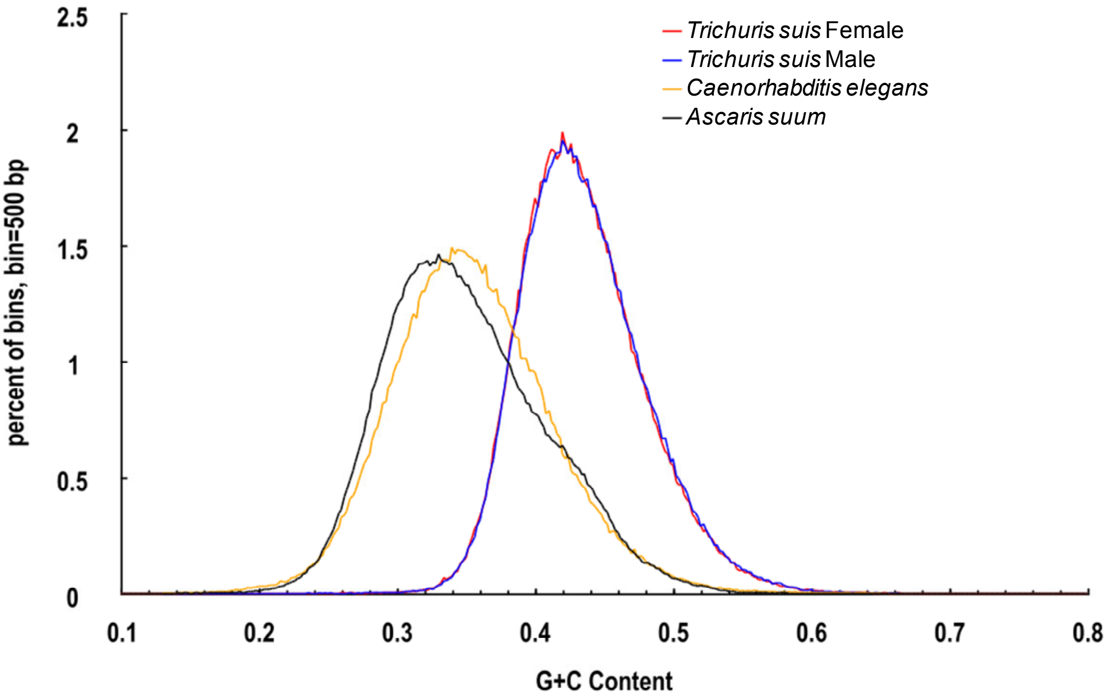
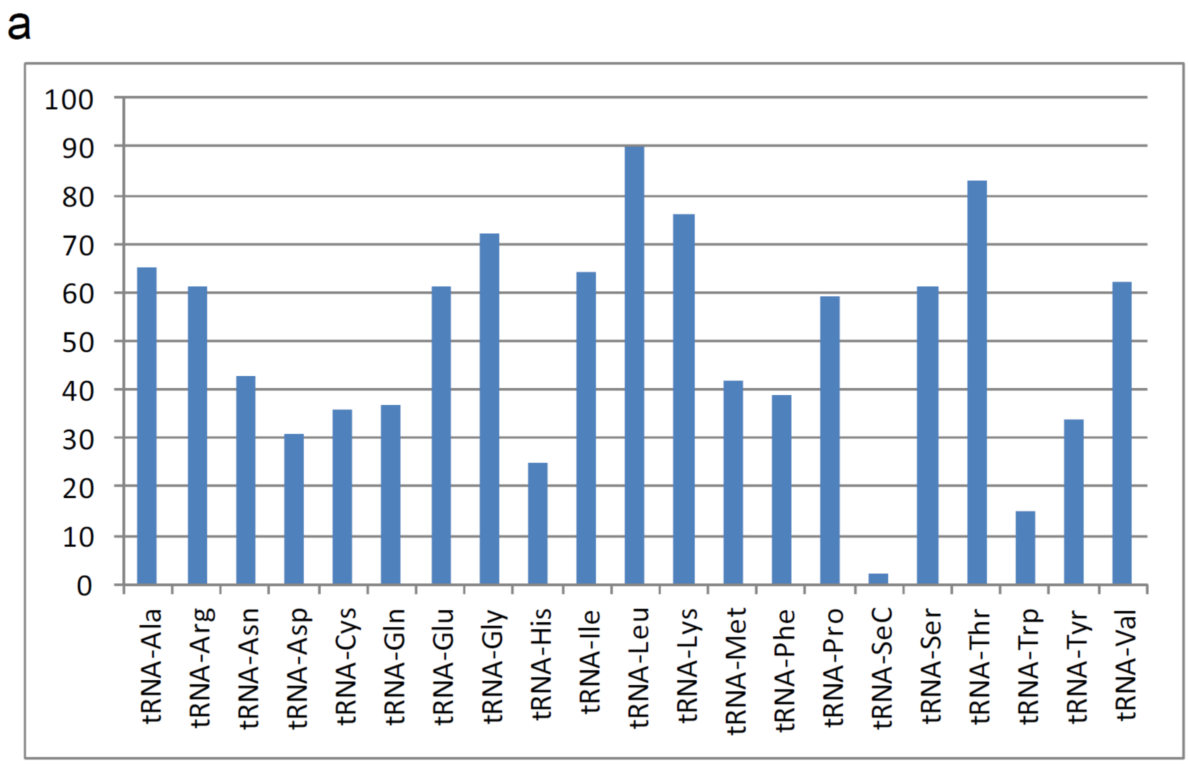
Supplementary Figures

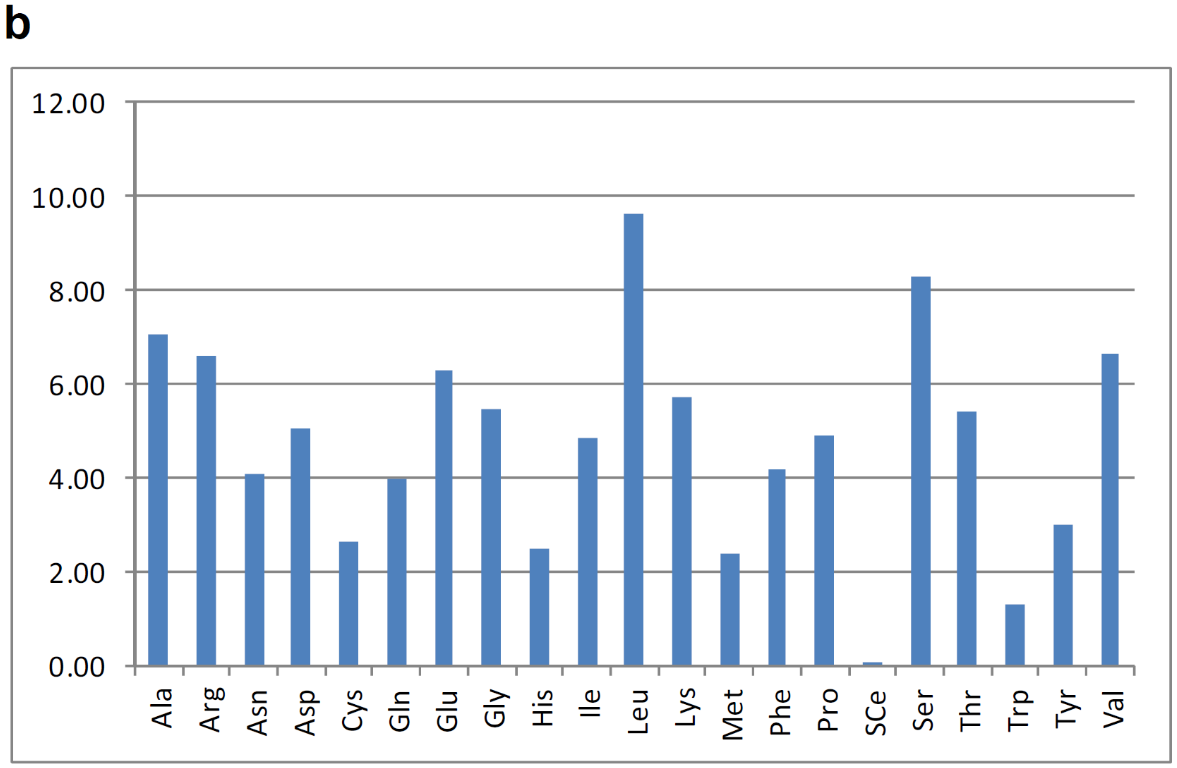


Supplementary Figure 1 | GC content *versus* sequencing depth of the (A) male and (B) female *Trichuris suis* genome. Calculated as GC% per 10 kb non-overlapping sliding window.

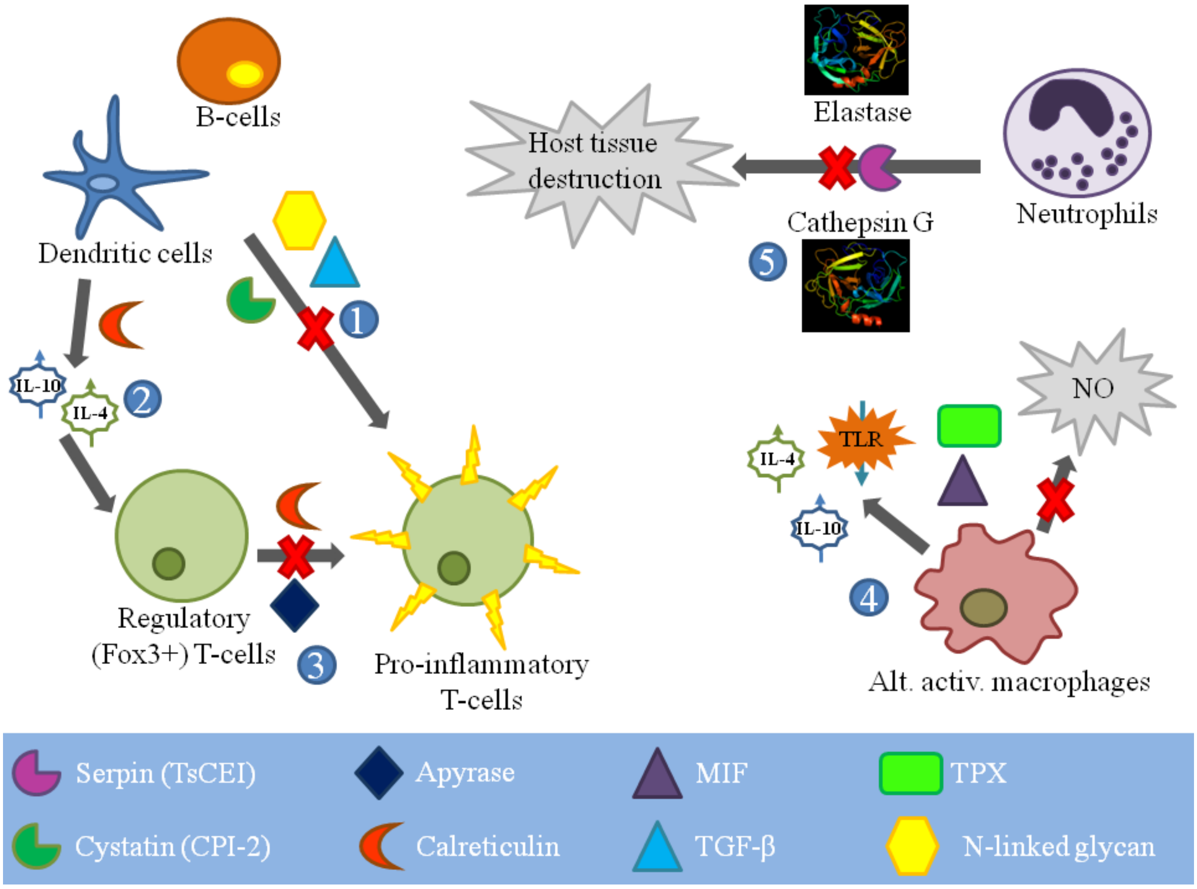


Supplementary Figure 2 | GC-content distribution for the genomes of *Trichuris suis*, *Caenorhabditis elegans* and *Ascaris suum*. Data generated using 500 bp sliding windows using a 250 bp overlap. The x-axis is GC content (%), and the y-axis is the proportion of the windows (=bins) corresponding to each GC measurement.

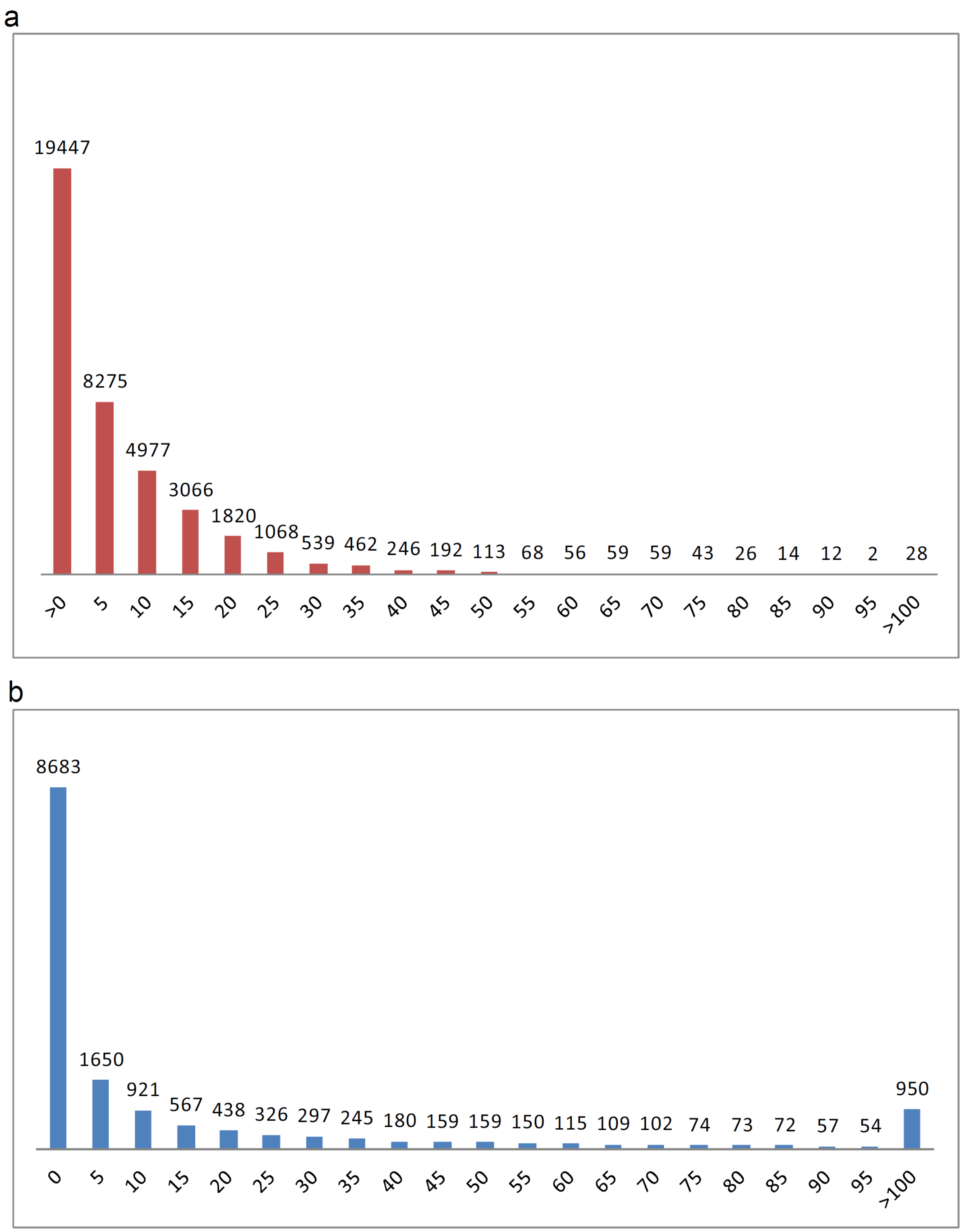




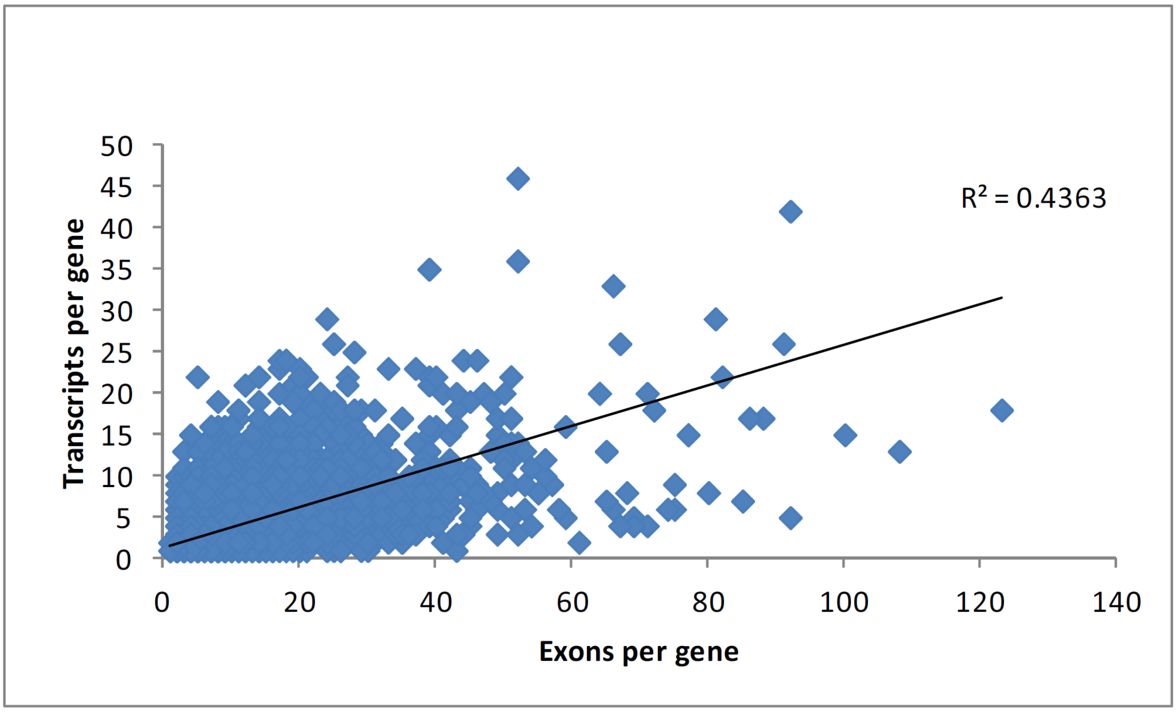
Supplementary Figure 3 | Frequency histograms comparing the tRNA copy number and the predicted amino acid usage for the *T. suis* genome. (a) tRNA copy number; (b) frequency of amino acids encode by the inferred *T. suis* proteome.



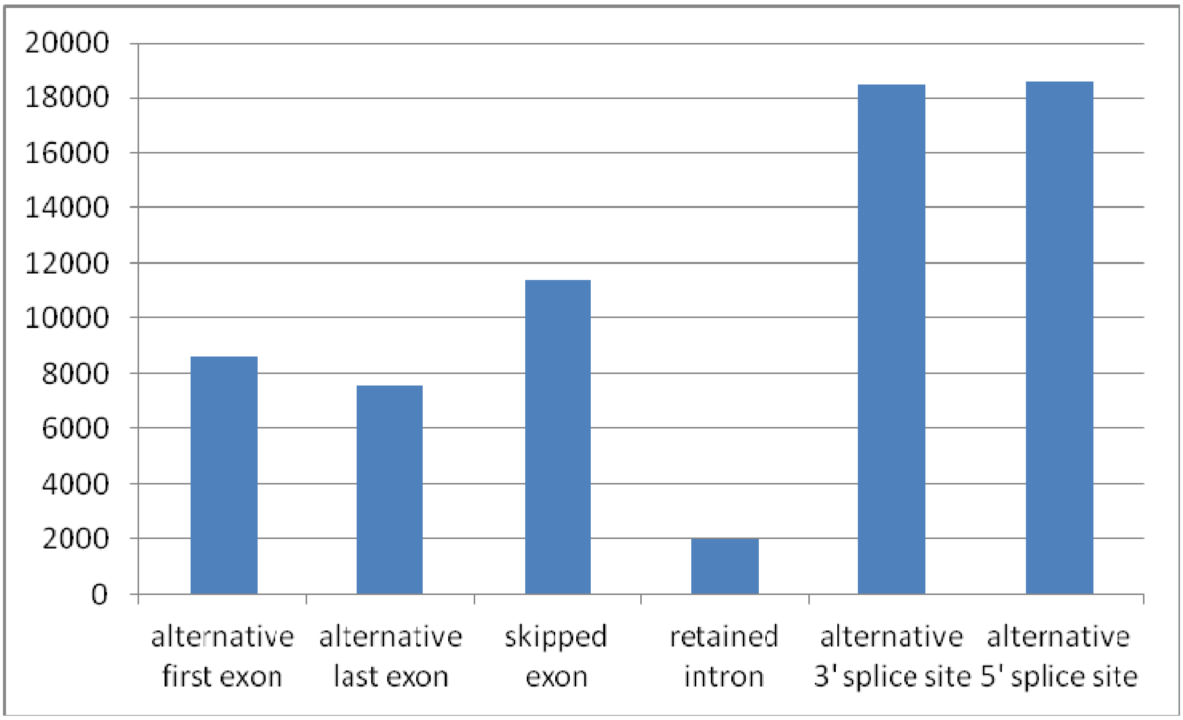
Supplementary Figure 4 | Proposed mechanisms through which *Trichuris suis* modulates inflammation in its vertebrate host. Based on enriched transcripts found in larval stages of *T. suis* embedded in the host gut epithelium and up-regulated in the stichosome relative to non-stichosomal tissues in adult males and females. Mechanisms of action (adapted from [1-4]) include secretion of (1) cysteine protease inhibitors, N-glycosylated proteins[2] and a putative TGF-β mimic to suppress antigen presentation by host dendritic and B-cells and limit prolifation/stimulation of pro-inflammatory T-cells; (2) calreticulins believed to bind to scavenger receptors on dendrtic cells and stimulate production of 'anti-inflammatory' IL-4 and IL-10, leading to proliferation of Fox3+ regulatory T-cells; (3) calreticulins and apyrases that bind free Ca^+ and ATP respectively and prevent conversion of Fox3+ to pro-inflammatory T-cells; (4) thioredoxin peroxidase (TPX) and a putative macrophage initiation factor (MIF) mimic to stimulate alternative activation of host macrophages, inhibiting inflammatory nitrous oxide (NO) production and toll-like receptor (TLR) pathways and stimulating the production of IL-4 and IL-10; and (5) serine protease inhibitors that block neutrophil-secreted cathepsin G and elastases, thereby limiting tissue destruction and reducing inflammation. Red 'X' represents the blocking of a pathways/mechanism due to *Trichuris*-generated immunomodulation.



Supplementary Figure 5 | Frequency histogram summarizing the number of exons per gene and the number of alternatively spliced isoforms the gene encodes. (a) Exons per gene; (b) Transcripts per gene.

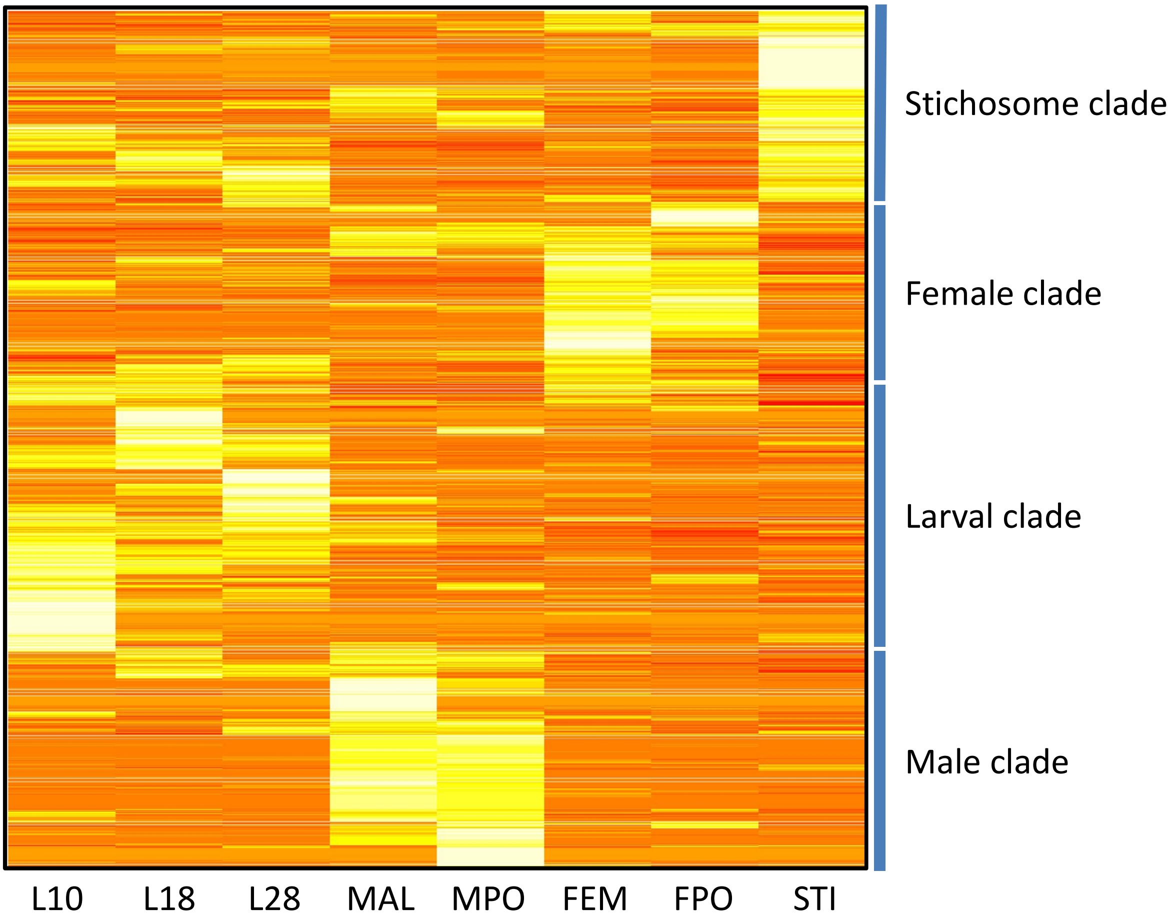


Supplementary Figure 6 | XY scatterplot comparing transcripts vs exons per gene for the *T. suis* transcriptome.

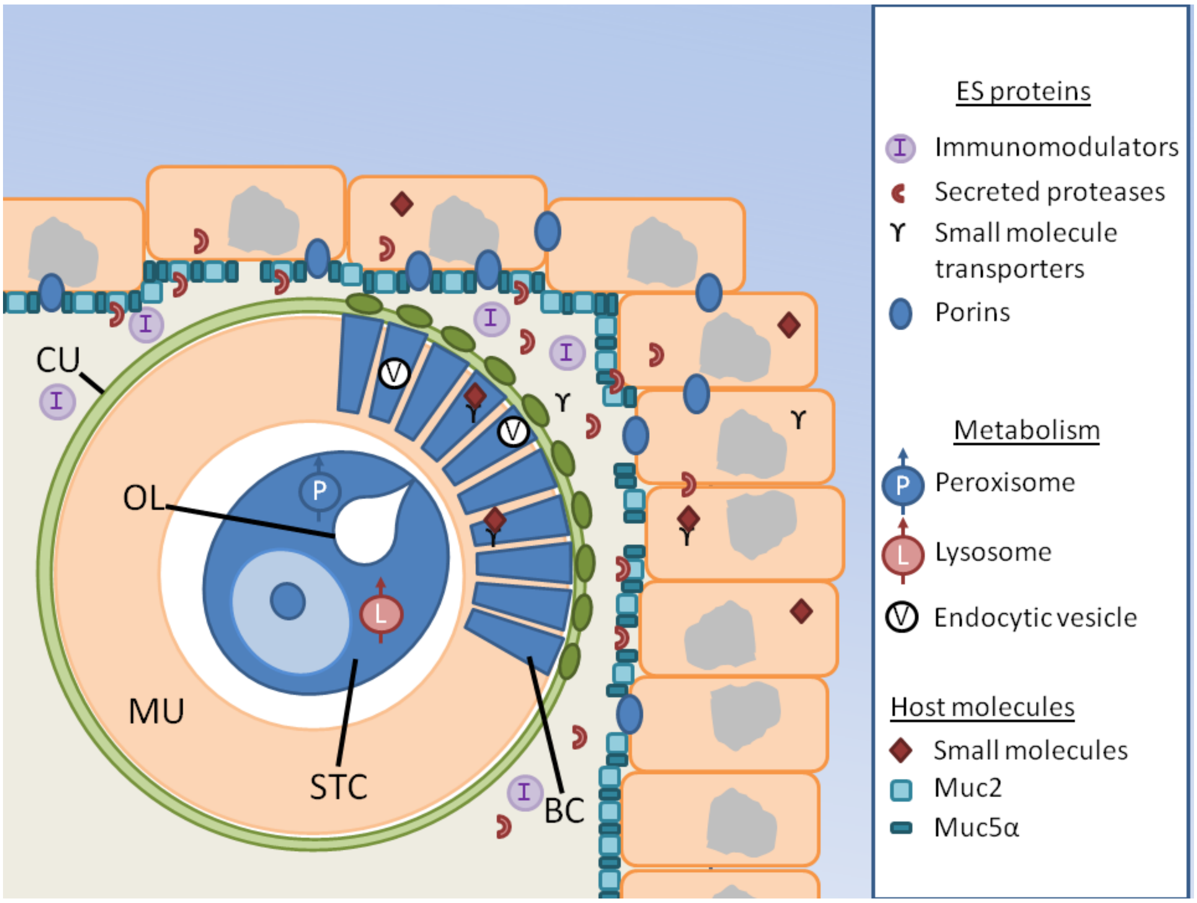


Supplementary Figure 7 | Frequency histogram of alternative splice events characterized for *Trichuris suis* transcripts. Splice events predicted for each transcript are available in Supplementary Data 3.

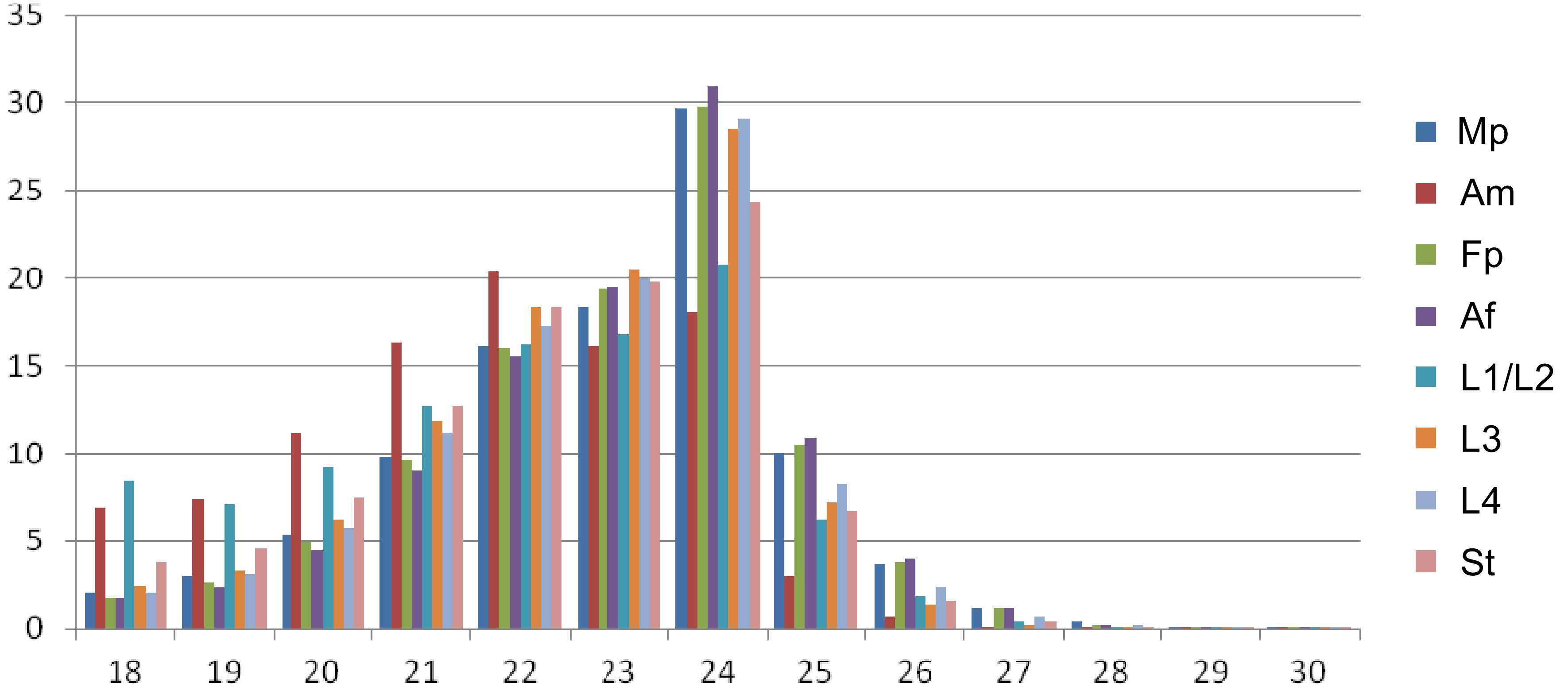
Nature Genetics: doi:10.1038/ng.3012



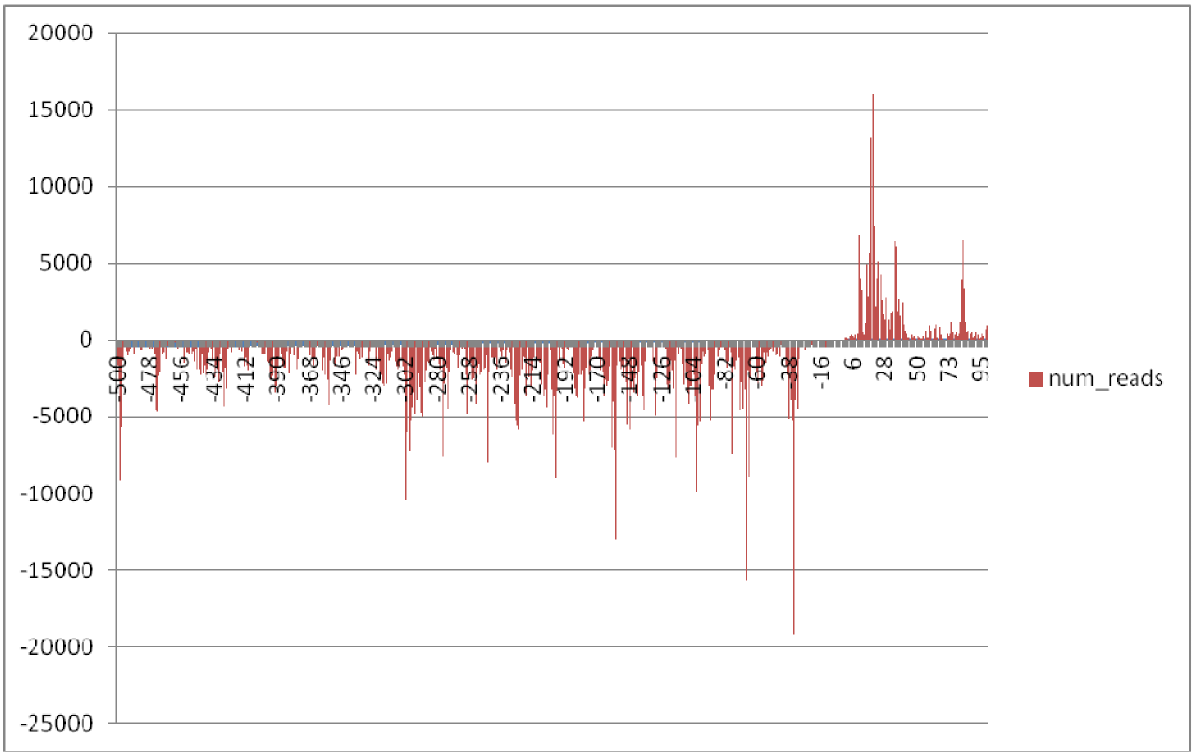
Supplementary Figure 8 | Normalized heatmap showing stage and tissue-specific transcriptional enrichment profiles of each *T. suis* messenger RNA. Blue-banding highlight major clusters based on transcriptional profile for stichosome, larval, made and female enriched transcripts. Each row represents one transcript. Highest transcription represented by white coloration. Lowest transcription represented by red coloration. Colour scale used allows comparison between/among libraries for each transcript, but not between/among distinct transcripts. FPKM values for each transcript available in Supplementary Table 18. Dendrogram clustering cropped for presentation.



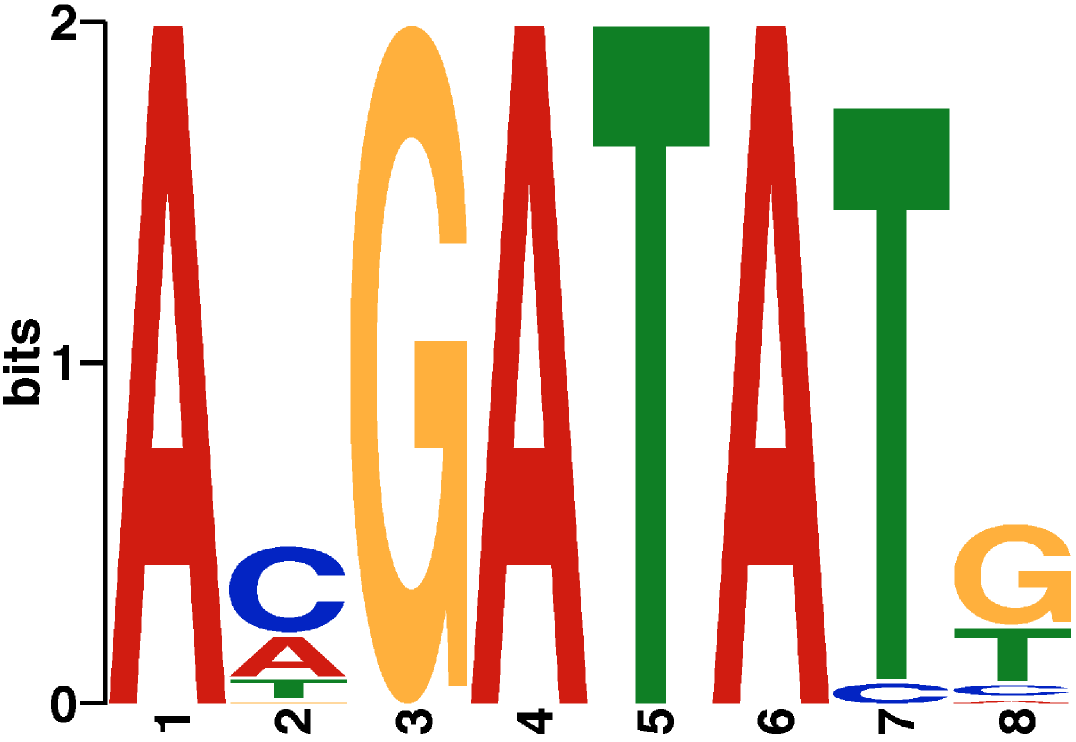
Supplementary Figure 9 | Schematic representation of molecular mechanisms involved in the interaction between the *Trichuris suis* stichosome (St) and surrounding host tissues (cross section). All parasite-derived molecules are enriched in the stichosome tissues relative to non-stichosomal tissues in both male and female adult *T. suis*. In this schematic, we propose that secreted proteases, including S1 (chymotrypsin), S9, M12 (astacin) and C1 (cathepsin) families, play an early role in formation of the syncitial tunnel in the host gut epithelium. Among the secreted S1 proteases are several that degrade Muc2 polymers in the host intestinal mucus barrier, exposing the gut epithelial cell membranes. These membranes are then breached by *T. suis* secreted porins, allowing small molecule transporters to enter the host cell cytoplasm and harvest simple sugars, and nucleic and amino acids, which are then taken back into the bacilliary cells through endocytosis or another unknown mechanism. In addition, S1 proteases degrade long protein polymers found in the host blood/plasma including fibrinogen (thus inhibiting co-agulation and possibly contributing to the gut haemorrhaging) and kallykrien (possibly providing a means to inhibit inflammation by blocking bradykinin formation). While this is occurring a potent mix of immuno-modulatory proteins (see **Supplementary Fig. 4 and Supplementary Table 16**) are secreted and specifically regulate/suppress an inflammatory response. We hypothesize also that larger peptide are taken up into the stichocyte by endocytosis and vesicle transport through the bacillary cells and further degraded in the peroxisomes in the stichocytes themselves. Increased lysosome activity in the stichocytes may also suggest degradation of lipids acquired from the host cells, but we cannot propose a source for these molecules based on our current data. Abbreviations: CU = cuticle, MU = muscle, STC = stichocyte, OL = oesophageal lumen, BC = bacillary cell.



Supplementary Figure 10 | Frequency histograms showing percentage distribution of small RNA reads by length for *Trichurs suis*. Mp = male non-stichosomal (i.e., posterior body) tissue; MALE = adult male; F.POST = female non-stichosomal (i.e., posterior body) tissue; FEMALE = adult female; L10 = larvae at 10 days post infection (p.i.); L18 = larvae at 18 days p.i.; L28 = larvae at 28 days p.i.; STICH = adult (mixed sex) stichosomal tissue.



Supplementary Figure 11 | Frequency distribution of transcription start site associated small RNAs mapping 5' antisense and 3' sense to each transcription start site predicted for the *T. suis* transcriptome.



Supplementary Figure 12 | An 8-nt motif present in 4.5% of 22A-RNAs from *Trichuris suis*. This represents the one significant motif, among possible motifs ranging in size from 6 to 22 nt, that was shared between 1,174 nonredundant 22A-RNA sequences. The logarithmic probability distribution of individual nucleotides in the motif is shown as a WebLogo (see Methods).

Supplementary Table 1 | General data summary

Data	Number of individuals	NCBI BioProject	SRA accession	Fragment size	Library type	Read length (bp)	Total raw reads	Total raw data (bp)	Total clean reads	Total clean data (bp)
Female genomic	1	PRJNA208416	SRR1041650	170	PE	100	33,446,068	3,344,606,800	30,827,784	3,021,122,832
Female genomic	1	PRJNA208416	SRR1041649	500	PE	100	20,971,166	2,097,116,600	18,712,876	1,833,861,848
Female genomic	1	PRJNA208416	SRR1041648	800	MP	100	18,996,962	1,899,696,200	16,771,324	1,643,589,752
Female genomic	1	PRJNA208416	SRR1041647	2000	MP	50	68,951,488	3,378,622,912	57,897,806	2,836,992,494
Female genomic	1	PRJNA208416	SRR1041646	5000	MP	50	54,983,752	2,694,203,848	43,397,704	2,126,487,496
Female genomic	1	PRJNA208416	SRR1041645	10000	MP	50	32,556,620	1,595,274,380	18,414,190	902,295,310
Total female genomic							**229,906,056**	**15,009,520,740**	**186,021,684**	**12,364,349,732**
Male genomic	1	PRJNA208415	SRR1041644	170	PE	100	34,816,402	3,481,640,200	32,525,860	3,187,534,280
Male genomic	1	PRJNA208415	SRR1041643	500	PE	100	26,008,128	2,600,812,800	23,506,604	2,068,581,152
Male genomic	1	PRJNA208415	SRR1041642	800	MP	100	11,200,440	1,120,044,000	9,618,188	942,582,424
Male genomic	1	PRJNA208415	SRR1041641	2000	MP	50	78,253,936	3,834,442,864	63,323,508	3,102,851,892
Male genomic	1	PRJNA208415	SRR1041640	5000	MP	50	37,072,032	1,816,529,568	27,791,474	1,361,782,226
Male genomic	1	PRJNA208415	SRR1041639	10000	MP	50	26,316,276	1,289,497,524	19,410,686	951,123,614
Total male genomic							**213,667,214**	**14,142,966,956**	**176,176,320**	**11,614,455,588**
smallRNA Af	10	PRJNA208415	SRR1041663	-	SE	~20-40	87,716,743	2,087,026,528	86,974,413	2,077,008,801
smallRNA Am	10	PRJNA208415	SRR1041662	-	SE	~20-40	36,335,250	824,579,512	35,622,760	811,386,240
smallRNA Fp	10	PRJNA208415	SRR1041669	-	SE	~20-40	45,524,286	1,082,659,626	45,124,686	1,076,596,732
smallRNA Mp	10	PRJNA208415	SRR1041664	-	SE	~20-40	56,793,761	1,358,177,539	56,308,373	1,351,386,651
smallRNA St	10	PRJNA208415	SRR1041670	-	SE	~20-40	39,407,168	901,792,913	38,974,944	894,308,032
smallRNA L1/L2	0,000 (pooled from 5 pigs)	PRJNA208415	SRR1041659	-	SE	~20-40	52,882,352	1,207,202,505	50,691,760	1,157,224,321
smallRNA L3	5,000 (pooled from 4 pigs)	PRJNA208415	SRR1041660	-	SE	~20-40	68,361,260	1,606,672,788	67,699,233	1,597,514,753
smallRNA L4	3,000 (pooled from 2 pigs)	PRJNA208415	SRR1041661	-	SE	~20-40	54,304,680	1,276,710,244	53,806,103	1,269,943,680
Total small RNA							**441,325,500**	**10,344,821,655**	**435,202,272**	**10,235,369,210**
messengerRNA Female	10	PRJNA208415	SRR1041655	200	PE	100	42,591,650	3,833,248,500	37,942,368	3,414,813,120
messengerRNA Male	10	PRJNA208415	SRR1041654	200	PE	100	41,931,490	3,773,834,100	38,417,134	3,457,542,060
messengerRNA Fp	10	PRJNA208415	SRR1041657	200	PE	100	43,646,026	3,928,142,340	41,248,854	3,712,396,860
messengerRNA Mp	10	PRJNA208415	SRR1041656	200	PE	100	42,386,370	3,813,153,300	39,460,564	3,551,450,760
messengerRNA St	10	PRJNA208415	SRR1041658	200	PE	100	44,345,506	3,991,095,540	39,287,692	3,535,892,280
messengerRNA L1/L2	0,000 (pooled from 5 pigs)	PRJNA208415	SRR1041651	200	PE	100	42,853,824	3,856,844,160	40,355,004	3,631,950,360
messengerRNA L3	5,000 (pooled from 4 pigs)	PRJNA208415	SRR1041652	200	PE	100	44,113,156	3,970,184,040	41,207,030	3,708,632,700
messengerRNA L4	3,000 (pooled from 2 pigs)	PRJNA208415	SRR1041653	200	PE	100	45,374,224	4,083,680,160	41,298,098	3,716,828,820
Total messenger RNA							**347,242,246**	**31,250,182,140**	**319,216,744**	**28,729,506,960**

Genomic reads for each sex all sequenced from the same individual worm

Small and messenger RNA libraries for each stage, sex and tissue sequenced from the same original pool of individual worms

All adults used for whole and tissue specific small and messenger RNA library construction isolated from the same experimentally infected pig

Supplementary Table 2 | Genome assembly metrics.

	Male assembly				Female Assembly			
	Scaffold		Contig		Scaffold		Contig	
	Size(bp)	Number	Size(bp)	Number	Size(bp)	Number	Size(bp)	Number
N90	819	1,011	481	3,510	20,572	250	4,081	1,535
N80	95,829	165	13,187	1,071	146,581	143	20,461	798
N70	188,455	107	28,194	669	232,222	103	38,602	535
N60	307,826	73	44,892	449	282,838	74	55,778	376
N50	452,488	51	65,844	304	424,593	51	76,312	262
Longest	1,594,463	-	468,935	-	1,448,326	-	577,627	-
Total Size	81,300,690	-	78,876,884	-	76,029,588	-	74,207,564	-
Total Number (>= 100bp)	81,300,690	60,856	78,876,884	63,057	76,029,588	42,663	74,207,095	44,412
Total Number (>=2kb)	72,700,674	628	69,712,838	2,199	69,887,603	506	67,694,526	1,849

Supplementary Table 3 | Core Eukaryotic Genes Mapping Approach (CEGMA) data.

Match	Value	***T. suis*** **male genome**	***T. suis*** **female genome**	***C. elegans*** **WS240 genome**	***C. briggsae*** **WS240 genome**	***T. suis*** **male gene predictions**	***T. suis*** **female gene predictions**
Complete	Unique proteins	231	231	244	246	214	215
Complete	%Completeness	93.15	93.15	98.39	99.19	86.29	86.69
Complete	Total proteins	273	258	271	274	411	398
Complete	Protein redundancy	1.18	1.12	1.11	1.11	1.92	1.85
Complete	%Ortho	14.72	9.96	10.25	11.38	60.75	55.81
Partial	Unique proteins	238	238	248	247	236	238
Partial	%Completeness	95.97	95.97	100	99.6	95.16	95.97
Partial	Total proteins	298	286	298	295	469	461
Partial	Protein redundancy	1.25	1.2	1.2	1.19	1.99	1.94
Partial	%Ortho	21.85	16.81	17.74	18.62	62.71	58.82

Supplementary Table 4 | Repeat Content.

		Male			**Female**		
Total Repeat content (bp; % of genome):		23,498,734; 31.65			22,926,863; 32.26		
Number of scaffolds		4,293			3,288		
Total assembly length (bp) excluding Ns		71,811,605			69,238,595		
GC content (%)		43.57			43.48		
Class	Elements	Number	Length (bp)	% of genome	Number	Length (bp)	% of genome
SINEs:		2,166	400,850	0.54	2,229	516,257	0.73
	ALUs	6	88	0	49	14,613	0.02
	MIRs	26	1,815	0	32	2,276	0
LINEs:		5,190	1,745,994	2.35	6,632	1,812,986	2.55
	LINE1	147	7,690	0.01	194	27,732	0.04
	LINE2	62	3,624	0	70	3,737	0.01
	L3/CR1	59	3,134	0	61	3,399	0
LTR	elements:	5,304	2,498,133	3.37	4,671	2,028,152	2.85
	ERVL	15	865	0	16	804	0
	ERVL-MaLRs	0	0	0	0	0	0
	ERV_classI	364	100,858	0.14	102	5,364	0.01
	ERV_classII	64	3,407	0	75	3,872	0.01
DNA	elements:	22,177	6,900,281	9.3	17,523	5,721,863	8.05
	hAT-Charlie	2,221	865,108	1.17	1,515	727,849	1.02
	TcMar-Tigger	8,218	2,564,104	3.45	6,854	2,087,420	2.94
Unclassified:		35,968	7,377,876	9.94	32,468	6,595,924	9.28
Total interspersed reads			18,923,134	25.49		16,675,182	23.47
tRNAs		991	72,188	0.11	1,021	76,442	0.11
Small RNA		737	84,792	0.11	788	83,004	0.12
Satellites:		98	40,678	0.05	104	10,707	0.02
Simple repeats:		2,991	248,417	0.33	2,504	185,093	0.26
Low complexity:		1,793	88,471	0.12	1,789	92,693	0.13

Supplementary Table 7 | Coding gene annotation summary.

	Total	**% of total gene set**
Total number coding genes	14,820	-
Supported by transcriptomic evidence	12,979	87.6
Genes with IPR domains (mean per annotated gene)	7,938 (9.2)	53.6
Genes with TM domains (mean per annotated protein)	2,583 (4.0)	17.4
Genes with SP domains	1,568	10.6
Genes with GO classification	6,147	41.5
Cellular Component	2,100	14.2
Molecular Function	5,696	38.4
Biological Process	3,825	25.8
Hits UniProt/SwissProt	8,123	54.8
Hits KEGG (unique terms)	4,037 (2,664)	27.2
A. suum homologues	6,286	42.4
B. malayi homologues	6,340	42.7
C. elegans homologues	6,149	41.5
T. spirallis homologues	8,480	57.2
MEROPs hits	1,671	11.3
ks-sarfari hits	232	1.6
IUPHAR:gpcr-sarfari hits	228	1.5
TCDB hits	1,962	13.2
SPD hits	4,956	33.4
Key functional classes		
Proteases	623	4.2
Protease Inhibitors	287	1.9
Kinases	231	1.6
Phosphatases	262	1.8
GPCRs	225	1.5
Transporters and Channels	1,928	13.0
1.Channels/Pores	529	3.6
2. Electrochemical Potential-driven transporters	305	2.1
3. Primary Active Transporters	513	3.5
4. Group Translocators	14	0.1
5. Transport Electron Carriers	14	0.1
8. Accessory Factors Involved in Transport	128	0.9
9. Incompletely characterized transport systems	425	2.9

Supplementary Table 19| Small RNA summary.

small RNA category	Total	L1/L2	L3	L4
Total sRNA reads sequenced	435,202,272	50,691,760	67,699,233	53,806,103
Mean total read size (range)	23.5 (18-44)	22.8 (18-44)	23.6 (18-44)	23.6 (18-44)
Unique sRNA reads	58,038,955	7,651,785	8,267,136	7,246,902
Mean unique read size (range)	23.6 (18-44)	23 (18-44)	23.6 (18-44)	23.8 (18-44)
Redundancy (total/unique sRNA reads)	7.5	6.6	8.2	7.4
Total sRNA reads aligning to genome	400,033,829	38,114,747	63,235,915	50,864,941
% aligning to genome	92	75	93	95
Clusters				
Total small RNA clusters (n: mean reads per cluster)	23,013	16,781	16,781	16,545
Mean cluster size (range)	484.2	629.3	629.3	635.5
Protein-coding clusters	11,749	7,827	7,827	7,581
Noncoding clusters	11,264	8,954	8,954	8,964
Consensus reads (min cov n ≥ 20)	2,004,317	245,861	450,730	341,437
Coding reads	1,028,808	129,619	247,847	187,744
Exonic	900,133	112,756	218,898	166,607
sense strand bias (≥ 80%)	176,853	12,681	26,130	19,190
anti-sense strand bias (≥ 80%)	673,355 (91,640,543: 22.9)	99,058	188,773	144,491
Coding genes inhibited (n)	3,497	1,993	2,299	2,141
Map to first exon	434,124	62,284	120,657	93,540
Map to second exon	120,632	19,827	35,103	26,248
Map to third exon	58,640	8,776	17,256	12,988
Map to another exon	59,959	8,171	15,757	11,715
22G-RNAs (total reads)	21,491 (2,574,054)	3,487 (247,853)	6,173 (445,054)	4,401 (285,472)
26G-RNAs (total reads)	9,269 (734,887)	516 (29,796)	1,147 (59,819)	1,157 (53,466)
Most abundant (total reads)	25G (11,056,328)	25G (667,580)	25G (1,836,068)	25G (1,443,468)
Intronic	128,675	16,863	28,949	21,137
sense strand bias (≥ 80%)	45,919	5,629	9,658	6,989
anti-sense strand bias (≥ 80%)	75,535	10,986	18,684	13,748
Exonic and Intronic	0	0	0	0
Non-coding reads	939,360	116,242	202,883	153,693
miRNAs (total miRNA reads: % all sRNA reads)	307 (63,140,093: 15.8)	276 (2,319,929: 4.6)	282 (5,878,451: 8.6)	278 (9,225,232: 17.1)
mean mature miRNA size (range)	22.9 (18-25)	22.9 (18-25)	22.9 (18-25)	22.9 (18-25)
mean star miRNA size (range)	22.7 (18-28)	22.7 (18-28)	22.7 (18-28)	22.7 (18-28)
mean precursor miRNA size (range)	66 (44-106)	65.9 (47-106)	65.5 (46-106)	65.8 (47-106)
Antisense to TE (unique sequences)	187,811	23,445	47,882	37,047
unique 21U-RNA (total reads)	2,102 (190,455)	257 (17,006)	1,237 (80,581)	833 (49,618)
unique 24U-RNA (total reads)	10,123 (1,217,781)	1,278 (72,695)	4,801 (469,887)	3,862 (321,648)
22nt, 5'A small-RNAs (total reads: % all mapped reads)	58,307 (36,954,546: 9.2)	7,760 (1,307,173: 3.4)	13,103 (2,966,295: 4.7)	8,846 (4,467,479: 8.8)
Coding (total reads: % all mapped reads)	29,883 (3,067,868: 0.8)	3,987 (312,764: 0.8)	7,043 (471,863: 0.7)	4,683 (276,753: 0.5)
Noncoding (total reads: % all mapped reads)	28,424 (33,886,678: 8.5)	3,773 (994,409: 2.6)	6,060 (2,494,432: 4.0)	4,163 (4,190,726: 8.2)
Annotated space (total reads: % all mapped reads)	8,438 (1,166,419: 0.3)	1,350 (181,170: 0.5)	1,960 (184,205: 0.3)	1,362 (98,260: 0.2)
Unannotated space (total reads: % all mapped reads)	19986 (32,720,259: 8.2)	2,423 (813,239: 2.1)	4,100 (2,310,227: 3.7)	2,801 (4,092,466: 8.0)
Strand-biased (total reads: % all mapped reads)	17,736 (32,602,599: 8.2)	2,398 (812,067: 2.1)	4,028 (2,307,422: 3.6)	2,766 (4,091,267: 8.0)
"Sense"-bias (total reads: % all mapped reads)	10,864 (1,856,699: 0.5)	1,308 (138,678: 0.4)	2,400 (245,702: 0.4)	1,642 (162,142: 0.3)
"Antisense"-bias (total reads: % all mapped reads)	6,872 (30,745,900: 7.7)	1,090 (673,389: 1.8)	1,628 (2,061,720: 3.3)	1,123 (3,929,122: 7.7)

Supplementary Table 19| Small RNA summary (cont'd).

small RNA category	Am	Af	Mp	Fp	St
Total sRNA reads sequenced	35,622,760	86,974,413	56,308,373	45,124,686	38,974,944
Mean total read size (range)	22.8 (18-44)	23.9 (18-44)	24.0 (18-44)	23.9 (18-44)	22.9 (18-44)
Unique sRNA reads	5,146,745	10,425,332	8,047,656	6,788,911	4,464,488
Mean unique read size (range)	22.7 (18-44)	24.1 (18-44)	23.9 (18-44)	24 (18-44)	23.4 (18-44)
Redundancy (total/unique sRNA reads)	6.9	8.3	7.0	6.6	8.7
Total sRNA reads aligning to genome	34,069,157	80,932,479	53,896,484	42,296,461	36,623,645
% aligning to genome	96	93	96	94	94
Clusters					
Total small RNA clusters (n: mean reads per cluster)	16,597	17,794	17,010	16,798	15,672
Mean cluster size (range)	633.6	598.3	621.6	628.4	665.1
Protein-coding clusters	7,639	8,305	7,787	7,691	7,108
Noncoding clusters	8,958	9,480	9,223	9,107	8,564
Consensus reads (min cov n ≥ 20)	188,427	530,341	348,242	266,264	155,442
Coding reads	66,264	277,976	128,425	138,659	79,568
Exonic	52,354	243,155	105,222	120,750	79,568
sense strand bias (≥ 80%)	11,463	45,236	16,085	21,837	17,612
anti-sense strand bias (≥ 80%)	40,011	190,846	86,804	96,461	53,956
Coding genes inhibited (n)	1,749	2,749	2,214	2,272	1,530
Map to first exon	27,428	124,344	59,695	62,678	35,039
Map to second exon	6,615	33,553	14,033	17,451	10,131
Map to third exon	2,953	16,732	6,466	8,232	4,768
Map to another exon	3,015	16,217	6,610	8,100	4,018
22G-RNAs (total reads)	1,887 (116,533)	4,462 (291,106)	1,851 (107,811)	2,462 (148,476)	2,257 (136,771)
26G-RNAs (total reads)	98 (5,832)	2,766 (152,308)	1,002 (53,826)	1,164 (58,437)	270 (12,009)
Most abundant (total reads)	25U (254,080)	25G (2,419,425)	25G (1,008,204)	25G (1,010,613)	25G (394,102)
Intronic	13,910	34,821	23,203	17,909	11,360
sense strand bias (≥ 80%)	6,344	11,995	10,286	5,878	4,502
anti-sense strand bias (≥ 80%)	7,061	21,823	12,078	11,675	6,637
Exonic and Intronic	0	0	0	0	0
Non-coding reads	122,163	252,365	219,817	127,605	75,874
miRNAs (total miRNA reads: % all sRNA reads)	270 (7,215,803: 20.2)	284 (12,285,795: 14.1)	263 (4,540,483: 8.1)	274 (6,315,467: 14.0)	281 (15,358,933: 39.4)
mean mature miRNA size (range)	22.9 (18-25)	23.0 (18-25)	23.0 (18-25)	22.9 (18-25)	22.9 (18-25)
mean star miRNA size (range)	22.7 (18-28)	22.8 (18-28)	22.7 (18-28)	22.7 (18-28)	22.7 (18-28)
mean precursor miRNA size (range)	65.7 (47-106)	66.0 (47-106)	66.0 (47-106)	65.9 (47-106)	65.8 (44-106)
Antisense to TE (unique sequences)	11,683	53,885	24,571	27,766	15,922
unique 21U-RNA (total reads)	250 (19,836)	246 (13,900)	169 (10,155)	475 (35,285)	143 (7,274)
unique 24U-RNA (total reads)	536 (34,064)	2,922 (211,322)	1,256 (76,263)	3,205 (231,939)	856 (48,930)
22nt, 5'A small-RNAs (total reads: % all mapped reads)	8,099 (4,482,799: 13.1)	9,784 (7,037,538: 8.7)	6,514 (2,671,525: 5.0)	5,055 (3,819,964: 9.0)	4,797 (8,138,025: 22.2)
Coding (total reads: % all mapped reads)	2,826 (185,843: 0.5)	4,896 (302,971: 0.4)	2,311 (147,746: 0.3)	2,511 (148,010: 0.3)	2,327 (133,111: 0.3)
Noncoding (total reads: % all mapped reads)	5,273 (4,296,956: 12.6)	4,888 (6,734,567: 8.3)	4,203 (2,523,779: 4.7)	2,544 (3,671,954: 8.7)	2,470 (8,004,914: 21.9)
Annotated space (total reads: % all mapped reads)	984 (91,418: 0.3)	1,589 (119,132: 0.1)	811 (68,309: 0.1)	843 (60,622: 0.1)	819 (58,523: 0.2)
Unannotated space (total reads: % all mapped reads)	4,289 (4,205,538: 12.3)	3,299 (6,615,435: 8.2)	3,392 (2,455,470: 4.6)	1,701 (3,611,332: 8.5)	1,651 (7,946,391: 21.7)
Strand-biased (total reads: % all mapped reads)	4,118 (4,199,318: 12.3)	3,226 (6,612,505: 8.2)	3,276 (2,451,440: 4.5)	1,683 (3,610,713: 8.5)	1,625 (7,945,217: 21.7)
"Sense"-bias (total reads: % all mapped reads)	2,629 (279,584: 0.8)	1.949 (230,736: 0.3)	2,066 (194,073: 0.3)	999 (124,664: 0.3)	958 (126,048: 0.3)
"Antisense"-bias (total reads: % all mapped reads)	1,489 (3,919,734: 11.5)	1,277 (6,381,769: 7.9)	1,210 (2,257,367: 4.2)	684 (3,486,049: 8.2)	667 (7,819,169: 21.4)

Supplementary Table 20| Frequency distribution of consensus and total small RNAs of *Trichuis suis.*

Read length	Consensus G	Total G	Consensus A	Total A	Consensus U	Total U	Consensus C	Total C
18	2048	261364	2495	431583	1824	356183	2886	719071
19	5051	675010	5859	963021	4655	1156032	7183	1053792
20	11633	1357976	11980	1900188	10368	1735912	14563	2857699
21	25371	3264089	26197	5701204	25629	3535730	29153	4574815
22	56589	8838799	58307	36954526	59408	17342211	61274	7150725
23	86871	12380064	96281	16230111	104279	13876185	92238	11478801
24	106758	15858803	107163	16789294	110550	13976614	88454	11132008
25	163774	30329161	175201	30506478	167791	29308221	147129	25211794
26	26878	2614952	25200	2604983	35012	3348772	23242	3312923
27	4346	443878	5074	624806	6657	745239	3949	515546
28	889	660733	956	169292	942	301863	785	100930
29	261	47120	278	41476	240	28580	259	57779
30	76	28149	73	4979	42	21092	78	8336
31	28	61337	27	58959	10	865	15	517
>31	32	2428	4	1159	0	0	0	0

Supplementary Note

Differential transcription during larval development

Following ingestion of *T. suis* eggs by the host, the L1 stage undergoes histotropic development in the mucosa of the large intestine[5,6]. From L1 to L3 stage, the parasite grows substantially and must withstand the host's early immune response. At the L3 stage, the posterior end begins to protrude from the mucosa. As the nematode grows and develops to the L4, it increases significantly in size and begins sexual differentiation. Exploring the mRNA transcription of *T. suis* during the invasion and establishment phases in its host is central to understanding the biology of the parasite and the host-parasite interaction, underpinning new drugs and control strategies, and yielding new insights into how *T. suis* might suppress autoimmune disorders. To this end, we characterised the transcriptome of larval *T. suis*, investigating key genetic changes associated with the development of L1/L2s and the transition to L3 and L4 stages (**Supplementary Table 18**).

In the transition from L1/L2 to L3, *T. suis* undertakes a significant shift in its transcriptional activity, with 2,195 (representing 1,544 genes) and 2,507 (representing 1,607 genes) transcripts significantly enriched in these stages, respectively. Switching among alternatively-spliced transcripts accounts for ~20% of these differences, representing 503 L1/L2- and 529 L3-enriched transcripts (relating to 373 genes), respectively (**Supplementary Table 18**). From a metabolic perspective, L1/L2 *T. suis* have enriched transcription associated with the pentose phosphate pathway. There is also significant transcription linked to the reductive citrate cycle; however, compared with L3, the major transcriptional changes associated with this pathway relate to splice-isoform switching. Nonetheless, these findings suggest that the early metabolic activities of larval *T. suis* emphasize anabolism rather than energy production. At the L1/L2 stages, there is a disproportionate enrichment among transporters and channel proteins for porins, including 22 transcripts with homology to TT47[7]. Also enriched are porters, including various ionic, small molecule- and glucose transporters, and 15 RND superfamily transporters, including homologs of Niemann-Pick C1 and C1-like proteins and several homologues of the PATCH hedgehog receptor. Considering the immunoregulatory role that larval *T. suis* play during TSO (*Trichuris suis* ova) therapy of IBD[8,9] and other autoimmune disorders in humans[10], an observed up-regulation specifically in the L1/L2 phase of lactosylceramide biosynthesis is of significant interest. An intermediary molecule between ceramide and lactosyclceramide is the sphingolipid β-glucosylceramide; studies in mouse models for IBD have shown that the intraperitoneal injection of β-glucosylceramide results in decreased inflammation and the stimulation of a Th2-mediated immune response[11,12]. Glycoslyceramides have been identified previously in at least one species of *Trichuris* (from goats[13]) and although their potential as immunomodulators has not been explored for *T. suis*, such molecules have been shown to have important immunomodulatory roles for schistosomes[14] and *Ascaris*[15].We have predicted also a variety of proteins (representing 44 genes) with homology to known helminth-derived immunomodulators (**Supplementary Table 16**). Nearly all (103) of the 116 transcripts predicted for these genes are transcribed during the L1/L2 stage; most abundantly transcribed are several homologues of *Strongyloides ratti sra-hsp-17*[4] and galectins with significant peptide homology to host-derived galectin-9 as well as a thioredoxin peroxidase, cystatin (nearest to *bmal-cpi-2*), two SCP/TAP and two calreticulin homologues, and several serpins. Relative to the L3 stage, 17 transcripts encoding proteins with putative immunomodulatory roles are statistically significantly enriched in L1/L2; these transcripts represent homologues of *sra-hsp-17*, various serpins and SCP/TAPS proteins, a galectin and a putative TGF-β mimic.

The transition to the L3 stage sees an increase in transcription associated with metabolic activity, particularly fatty acid, ceramide/sphingosine, and some components of glycan (i.e, 'mannose' and 'complex' N-glycan) biosynthesis, and of purine/pyrimidine (conversion of IMP to adenine/guanine ribonucleotides) or cysteine/methionine (biosynthesis and degradation) metabolism. The transcription of porin proteins appears to increase in richness during the transition from L1/L2 to L3, with 57 transcripts encoding these proteins up-regulated (p-value < 0.05). Also chymotrypsin-like serine proteases are prominently represented among transcripts with most enriched in L3s relative to L1/L2. This latter finding might suggest a significant shift and/or up-regulation in the digestion of host protein polymers as the worm matures. Notably, in acute *T. muris* infection in mice, host-derived serpins secreted into the intestinal mucosa, which are proposed to block the degradation of the mucus barrier by preventing the depolymerization of Muc2 by parasite secreted serine proteases, peak between 14 and 21 days after infection[16], coinciding with the development of the L3 stage of this nematode. Although the effect of *T. suis* on the host intestinal mucus barrier has not been explored, increased transcription of Muc5α in goblet cells is observed in pigs with this infection,[16] suggesting a similar degradation of the mucus barrier by chymotrypsin- or trypsin-like serine proteases. Also notable among the highly differentially transcribed sequences are several serpins, including an isoform of Tsu_03130, which appears to encode *Ts*-CEI, a demonstrated inhibitor of cathepsin G and elastases of host neutrophils and, thus, a known immunomodulator of *T. suis*.[17,18]

The next major transition in the *T. suis* life-cycle is the moult from L3 to L4, which occurs ~3 to 4 weeks post-infection[5,6]. In the present RNA-seq study, this developmental transition coincided with 1,737 (1,247 genes) down- and 1,749 (1,221 genes) up-regulated transcripts. As observed in the change from L1/L2 to L3, many of the differentially transcribed sequences between L3 and L4 stages relate to isoform-switching (287 genes representing 384 and 392 transcripts, respectively). From a metabolic perspective, the major differences in the transition from L3 to L4 are an up-regulation of fatty acid biosynthesis (initiation and elongation) and an enrichment of β-oxidation. Also notable is a shift in glycosylation pathways, namely from N-linked ('mannose' and 'complex') to O-linked (i.e., glycosaminoglycan biosynthesis) glycans. Whether these changes are indicative of a shift in the glycan profile on the surface of the parasite is unknown, but, considering the potential role of helminth glycans (particularly O-linked) as immunostimulatory[19] and/or antigen-masking agents,[20] and the finding that N-linked (mannose) glycans play a key role in modulating antigen presentation in dendritic cells in *T. suis* infection[2], this finding is worth additional investigation. Common among the genes under-going the greatest quantitative increases in transcription from L3 to L4 are a variety associated with oxidative stress (e.g., homologues of *C. elegans sod-1*, *pah-1* and *fah-1*). Their occurrence in the present dataset may be indicative of an increased immunological attack on the parasite by the host, reflected in a neutrophil oxidative burst, for example. This is consistent with immunohistological findings for *T. suis* infection of pigs,[21] which indicate that neutrophil activity peaks ~3-5 weeks post-innoculation (in the present study, L4s were harvested precisely 4 weeks p.i.). Notably, putative inhibitors of neutrophil cathepsin G/elastase[17,18], including 3 splice-isoforms of Tsu_03130 ("TsCEI"),, are all among the genes showing the greatest enrichment in transcription in L4 compared with L3. Considering that L4 is the stage at which sexual dimorphism becomes apparent in dioecious nematodes, we note an up-regulation of various transcripts involved in vulval development/morphogenesis and/or expressed in vulval precursor cells (e.g., *acn-1*, *ced-10*, C15B12.7b, *eif-3*, *lin-1 and -39*, *rab-9* and *rsp-9*), gonad development (e.g., *acin-1*, *baf-1*, *hda-1* and *mig-6* and *-17*), sex determination (e.g., *fox-1*, *sex-1*, *rsp-6* and *ufd-2*) and male gonadal/tail development (e.g., *adt-1*, *col-34*, *evl-20*, *lin-29*, *ptr-2* and *sur-2*) based on homology with *C. elegans*.

Differential transcription associated with maturation to adulthood

Following larval development, maturation to adulthood represents a major morphological and biological transformation in *T. suis*. To explore the early phases of sexual development and maturation, we compared transcription of L4s with adult male and female worms (**Supplementary Table 18**). This transition reflects the most substantial transcriptional changes observed here in the life cycle of *T. suis.* We identified 4,467 down- (2,789 genes) and 5,026 up-regulated (2,949 genes) transcripts between L4s and adult males, with 28% and 23% of these changes relating to isoform-switching for 776 genes between L4 and the adult male, respectively. Among all sequences differentially transcribed between L4 and the adult male, 3,837 (2,420 genes) and 3,736 (2,283 genes) are similarly down- and up-regulated, respectively, between L4 and male posterior body. Maturation to the adult male stage appears to coincide with increased energy production, with glycolysis, gluconeogenesis and the TCA cycle all being enriched, and a transition away from anabolism, with the pentose phosphate pathway continuing to be down-regulated. This increased energy production related mainly to glucose metabolism, with fatty-acid degradation being down-regulated compared with L4s. In contrast, fatty acid elongation and β-oxidation are enriched in male worms, with β-oxidation possibly supplying acetyl-CoA to the TCA cycle, allowing fatty acids to be used for energy production. In addition to these metabolic changes, 1,434 male-enriched transcripts with homology to 797 *C. elegans* genes are enriched in males relative to the L4 stage, with no evidence for isoform-switching. Among the *C. elegans* homologues representing these differentially transcribed genes are a variety associated with sperm/spermatogenesis (e.g., *cbp-1*, *cpb-1*, *msp-3*, *-33*, *-57*, *-63* and *-79*), male mating behaviour (e.g., *aex-3*, *arl-3*, *rab-3* and *sax-2*) and masculinization/male sex-determination (e.g., *fem-2* and Y54E10A.9c). Also notable among male-enriched sequences are 35 *C. elegans* homologues associated with development/maintenance and regulation of germline tissue, including *dbr-1*, *glp-1*, *rpt-3* and *spk-1*. By comparing L4s to the adult female, we identified 4,217 down- (linked to 2,629 genes) and 3,313 up-regulated (2,260 genes) transcripts associated with maturation; 3,795 (2,366 genes) and 1,984 (1,363 genes) of these transcripts are also down- and up-regulated, respectively, between L4 and the female posterior body. As observed in the comparison between L4s and the adult male, approximately one-quarter (22 and 25% for down- and up-regulated, respectively) of these differentially transcribed sequences are predicted to relate to isoform-switching among genes (n = 596 genes). Among the female-enriched transcripts, 1,253 (958 genes) are also enriched in males relative to L4, and are likely associated with maturation to adulthood rather than sexual differentiation. Indeed, the vast majority (72.4%) of these male- and female-enriched sequences have an orthologue in the KEGG database, including genes associated with glycolysis, and some components of amino (methionine), fatty and nucleic acid metabolism. Among the 2,060 transcripts (1,448 genes) specifically up-regulated in the transition from L4 to the adult female, 1,158 (806 genes) were also up-regulated in the female posterior body and not in the adult male or male posterior body relative to L4. These female-enriched transcripts represented fatty acid and N- and O-linked glycan biosynthesis pathways, and homologues of 565 *C. elegans* genes. Among the *C. elegans* homologues are genes associated with egg-laying/oogenesis (e.g., *car*-1, *cbd-1*, *cej-1*, emo-*1*, *mau-2* and *spas-1*), embyrogenesis (e.g., *ags-3*, *hnd-1*, *let-767*, *ptp-3*, *sym-5* and *unc-112*) and vulval development (*die-1*, *hda-1*, *let-341*, *met-1* and *tag-185*).

Of the transcripts down-regulated in the adult female relative to L4, 90% are also down-regulated in Fp. Indeed, there was a high level of consistency among the L4 sequences down-regulated during the maturation to adulthood, with 65% of the transcripts down-regulated in The adult female compared with L4, also down-regulated in the adult male, and 59% (n = 2,472) down-regulated in The adult female, the female posterior body, the adult male and male posterior body. Among the 2,472 transcripts (1,600 genes) enriched in L4s relative to each adult library (The adult female, the female posterior body, the adult male and male posterior body), 521 (relating to 363 genes) related to isoform-switching. Of the remaining 1,951 transcripts (1,237 genes), only a relatively small proportion (representing 186 genes) have an orthologue in the KEGG database. Conspicuous among the KEGG pathways found to be down regulated during maturation is the synthesis of lactosylceramide, which appears to continually diminish as *T. suis* develops throughout the life-cycle. Among the transcripts shown to be most down-regulated in the adult stage/sexes are several porin homologues, 21 chymotrypsin-like serine proteases, one serpin (thought to be involved in inhibiting neutrophil secreted cathepsin G/elastases)[17,18], and a CPI-2 cystatin homologue (thought to disrupt antigen presentation in dendritic cells). Also notable among the transcripts down-regulated with *T. suis* maturation are 79 *C. elegans* homologues, including some involved in larval development/morphogenesis and growth rate (e.g., *dao-5*, *daf-21*, *grl-25*, *ifb-1*, *mig-6*, *noah-2* and *nhr-95*).

Differential transcription between adult sexes

Sexual dimorphism in adult *T. suis* is of significant interest in relation to reproductive biology and potentially for the development of novel drug targets, considering the high likelihood that genes differentially transcribed between male and female worms are involved in critical reproductive processes and, thus, essential for the survival and transmission of this nematode[22-24]. By comparing the adult male and the adult female, we identified 3,411 female- (2,320 genes) and 5,189 male-enriched transcripts (3,153 genes). Considering reproductive processes, we explored differential transcription in the female posterior body compared with male posterior body, and identified 1,976 female- (1,340 genes) and 3,666 male-enriched transcripts (2,252 genes), respectively. As for other stages, isoform-switching was inferred to play a role in the differential transcription between adult females and males of *T. suis*, with 450 and 478 transcripts (representing 336 shared genes) relating to such switching events. Considering only the 1,004 genes (1,324 transcripts) enriched in the adult female and the female posterior body, and the 1,916 genes (2,477 transcripts) enriched in the adult male and male posterior body, with no evidence of isoform-switching compared with the opposite gender, we identified 845 and 328 KEGG orthologues respectively, suggesting that many of the differences between the genders of adult *T. suis* relate to changes in activity of fundamental biological pathways, particularly in adult females. In females, these differences include an up-regulation of the reductive citrate cycle and fatty acid synthesis, suggesting an increased need for anabolic pathways. By contrast, male-enriched metabolic activity appears to relate to a conversion among metabolites, with an enrichment of sugar/starch and nucleic acid metabolism being most conspicuous. As would be expected, many of the differentially transcribed genes between male and female worms relate to sex-determination, gonadal development and gamete production. Among the genes encoding female-enriched transcripts are *C. elegans* homologues linked to vulval development/function (e.g., *ced-10*, C15B12.7, *die-1*, *nekl-2*, *rnp-4*, *sos-1*, *sqv-1*, *sqv-4* and *syg-2*), oogenesis (e.g., *cdk-1*, *cpb-1*, *cogc-2*, *ppk-1* and *unc-31*), ovulation (e.g., *ceh-18, lin-3*, *par-3* and *vab-1*), egg-laying (*asd-2*, *gbp-2*, *nhr-85*, *mau-2*, *sup-9*, *unc-2* and *unc-58*) and embryogenesis (e.g., *cht-1*, *hnd-1*, *let-767*, *uba-1*, *uba-2* and *xpo-1*). Of the genes encoding male-enriched transcripts are *C. elegans* homologues with functions including sperm/spermatogenesis (e.g., *alg-3*, *cbp-1*, *cogc-5*, and *msp-3, -33*, *-57*, *63*, *78* and *-79*), masculinisation (e.g., Y54E10A.9c), male mating behaviour (e.g., *arl-3*, *crt-1*, *goa-1* and *lin-39*) and male development (e.g., *alg-3* and *-4*, *mab-5*, *dbl-1*, *lin-29* and *tag-170*).

Putative gender-specific genes and their transcription

The karyotyping of *Trichuris* spp. indicates members of this genus are XX (female) and XY (male) [25], suggesting the potential that male-specific genes exist in *T. suis*. A surprising result in our comparative analyses of the male and female genome assemblies was that we could identify few examples of gender-specific genes in either assembly. In initial comparisons between the sexes, we have identified 14,115 and 14,240 female and male genes with an orthologue/paralogue in the opposite sex respectively (10,403 genes are defined as unambiguous one-to-one orthologues). Based on these comparisons we identified 281 and 341 genes 'specific' to the female and male assemblies respectively. Subsequent alignment of the gene sequences and their flanking regions using BLAT [26] identified 247 of the female 'specific' genes were present in the male assembly and simply absent from our male gene models. Of the remaining 34 genes predicted from the female assembly lacking a homolog in the male, just one is transcribed in a female and absent from all male RNA-seq libraries. Thus, as expected, we find no evidence for female-specific genes in *T. suis*. Only 75 of the male 'specific' genes were located in the female assembly and determined to absent only from the initial female gene models. Of the remaining 266 genes annotated in the male and unaccounted for in the female, just 41 are transcribed in a male, but absent from all female RNA-seq libraries, supporting their being sex-specific genes. Of these 41 male-specific sequences, only 3 have a homologue of known function in *C. elegans* [*frk-1* (a receptor tyrosine-kinase), *gpc-1* (a g-protein coupled receptor), *his-66* (a histone protein)]. As these latter sequences are hypothetical and scattered among numerous scaffolds, we do not speculate about their function or relation to the Y-chromosome. We found no evidence of the clustering of the male-specific genes among assembly scaffolds, which might provide evidence of the Y chromosome. Indeed a 4-way assessment (per Carvalho and Clark [27]) of the coverage of the male and female assembly using all male-derived or female-derived genomic reads revealed little evidence of sex-specific assembly scaffolds (data not shown), suggesting that much of the Y-chromosome of male *T. suis* may relate to repetitive sequence and few sex-specific genes. Noting this, we also can find no evidence of these male-specific genes or their flanking regions (up to 1,000 bp to the 5' or 3' of each gene) in the female assembly, suggesting these sequences are the best candidates we have for contigs contributing to the male chromosome in *T. suis*

Composition of 22A-RNA sequences and their 100-nt genomic neighborhoods

We began our analysis of 22A-RNA genomic neighborhoods with 5,517 sequences from the male *T. suis* genome. For our purposes, we defined a genomic neighborhood (prior to such operations as merging overlapping neighborhood sequences) as a genomic site encoding a 22A-RNA itself, plus 100 nt of flanking DNA on that site's 5' and 3' ends. The 5,517 sequences fit this basic criterion (i.e., they did not come from 22A-RNA sites that were less than 100 nt from either end of a genomic scaffold). We then probed their possible nature by filtering out neighborhoods by various criteria: (1)

We chose to consider only those neighborhoods with complete determined DNA sequences (i.e., with no scaffolding 'N' residues). This criterion modestly reduced the number of sequences to 5,457. (2) We tested these 5,457 genomic neighborhoods for potential protein-coding or ncRNA sequences with BlastX[28] and INFERNAL[29]. This yielded 2,657 neighbourhoods with protein-coding potential, 171 with the potential to encode a familiar ncRNA (typically tRNA), and 470 with both potentials. (3) Removing these sequences left a set of 2,159 neighborhoods without such obvious traits. These 2,159 neighbourhoods, in many cases, overlapped one another on the genome. (4) To avoid having further analyses confused by spurious redundancy, we merged overlapping sequences, which yielded 671 spatially nonredundant 22A-RNA neighbourhood sequences. These 671 sequences still had redundancy at the DNA sequence level. When (5) merged further with CD-HIT-EST[30] at a threshold of 99% identity, they decreased to 646 sequences; at a threshold of 80% identity, they decreased to 563. Because 80% identity is still a quite conservative threshold for redundancy (requiring 178/222 nucleotide identities before merging two sequences), we used the latter threshold for further analysis. We (6) used BlastN of the 563 neighborhoods to test them for similarity to the 671 spatially nonredundant genomic neighborhoods in the male *T. suis* genome, to genomic DNA in other nematode species, and to uncharacterized ncRNAs in *C. elegans*. This showed that intragenomic similarities between neighborhoods were common: 42% (236/563) gave a (non-identical) BlastN hit to one or more sequences in the set of 671 genomic neighborhoods. Matches to other nematode genomes were much less common, but did occur in 3.9% (22/563) of cases. Still fewer (2.1%; 12/563) gave a match to a *C. elegans* ncRNA. (7) Examining 11 of the genomic neighborhoods that gave matches both to *C. elegans* ncRNAs and to non-*T. suis* genomic DNA, we found that all of them identified ncRNAs with cryptic, previously undescribed similarities to tRNAs (which we detected by searching the ncRNAs with INFERNAL).

We concluded, from all of the above data, that the 22A-RNAs of *T. suis* may be nucleolytic products of larger RNAs, and that these larger RNAs might be heterogeneous: some could be current/remnant mRNAs (annotated or unannotated in our current genomic analysis), others could be familiar ncRNAs (likewise, either annotated or unannotated), and yet others might identify unfamiliar ncRNAs in *T. suis* that we have not yet sufficiently characterized to define.

Even if 22A-RNAs arose from diverse sources, they might have common signals in their mature sequences that were responsible for their being generated. To detect such possible signals, we searched a nonredundant collection of 1,174 22A-RNA sequences with MEME for recurrent, statistically significant motifs. This non-redundant set came from 2,159 22A-RNAs embedded in the 2,159 neighbourhoods that lacked obvious protein-coding or ncRNA characteristics. We extracted 22A-RNA sequences from these 2,159 neighbourhoods, and merged those 22A-RNAs that overlapped spatially in the *T. suis* genome to yield a set of 1,344 22A-RNAs; we then merged them again with CD-HIT-EST to a threshold of 80% identity, to generate a final set of 1,174 22A-RNAs. The possible size of motifs was allowed to range from 6 to 22 nt (6 nt being the smallest motif we felt likely to be informative, and 22 nt being the full size of a mature 22A-RNA). We found one significant 8-nt motif (E-value = $1.6 \cdot 10^{-40}$; **Supplementary Figure 11**). The motif's consensus sequence was 5'-A[CA]GATAT[GT]-3', and its highest-scoring individual sequence was 5'-ACGATATG-3'. Among 5,457 original, unmerged 22A-RNA sequences, this motif occured at a frequency of 4.5% (in 245 sequences) with a significance of $p \leq 10^{-3}$. This distribution was not entirely random: among 2,159 unmerged 22A-RNAs whose neighborhoods lacked obvious protein or ncRNA similarities (from which whose 22A-RNA sequences we had generated the 8-nt motif), there were 180 sequences containing the 8-nt motif (8.3%, roughly twice the overall average), whereas among 3,298 nonmerged 22A-RNAs whose neighborhoods did exhibit associated protein or ncRNA similarities (and whose 22A-RNA sequences had not been used to generate the 8-nt motif), there were only 65 unmerged sequences containing the 8-nt motif (2.0%, less than half the overall average).

References

1 Hewitson JP, Grainger JR & Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol. Biochem. Parasitol.* **167**, 1-11 (2009).

2 Klaver EJ, Kuijk LM, Laan LC *et al. Trichuris suis*-induced modulation of human dendritic cell function is glycan-mediated. *Int. J. Parasitol.* **43**, 191-200 (2013).

3 McSorley HJ, Hewitson JP & Maizels RM. Immunomodulation by helminth parasites: Defining mechanisms and mediators. *Int. J. Parasitol.* **43**, 301-310 (2013).

4 McSorley HJ & Maizels RM. Helminth infections and host immune regulation. *Clin. Microbiol. Rev.* **25**, 585-608 (2012).

5 Beer RJ. Morphological descriptions of the egg and larval stages of *Trichuris suis* Schrank, 1788. *Parasitology* **67**, 263-278 (1973).

6 Beer RJ. Studies on the biology of the life-cycle of *Trichuris suis* Schrank, 1788. *Parasitology* **67**, 253-262 (1973).

7 Drake L, Korchev Y, Bashford L *et al.* The major secreted product of the whipworm, *Trichuris*, is a pore-forming protein. *Proc. Biol. Sci.* **257**, 255-261 (1994).

8 Summers RW, Elliott DE, Urban JF, Jr., Thompson RA & Weinstock JV. *Trichuris suis* therapy for active ulcerative colitis: a randomized controlled trial. *Gastroenterology* **128**, 825-832 (2005).

9 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.* **98**, 2034-2041 (2003).

10 Fleming JO, Isaak A, Lee JE *et al.* Probiotic helminth administration in relapsing-remitting multiple sclerosis: a phase 1 study. *Mult. Scler.* **17**, 743-754 (2011).

11 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* **56**, 82-89 (2007).

12 Lalazar G, Preston S, Zigmond E, Ben Yaacov A & Ilan Y. Glycolipids as immune modulatory tools. *Mini. Rev. Med. Chem.* **6**, 1249-1253 (2006).

13 Sarwal R, Sanyal SN & Khera S. Lipid metabolism in *Trichuris globulosa* (Nematoda). *J. Helminthol.* **63**, 287-297 (1989).

14 Nagayama Y, Watanabe K, Niwa M, McLachlan SM & Rapoport B. *Schistosoma mansoni* and alpha-galactosylceramide: prophylactic effect of Th1 Immune suppression in a mouse model of Graves' hyperthyroidism. *J. Immunol.* **173**, 2167-2173 (2004).

15 Deehan MR, Goodridge HS, Blair D *et al.* Immunomodulatory properties of *Ascaris suum* glycosphingolipids - phosphorylcholine and non-phosphorylcholine-dependent effects. *Parasite Immunol.* **24**, 463-469 (2002).

16 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.* **6**, e1856 (2012).

17 Rhoads ML, Fetterer RH, Hill DE & Urban JF, Jr. *Trichuris suis*: a secretory chymotrypsin/elastase inhibitor with potential as an immunomodulator. *Exp. Parasitol.* **95**, 36-44 (2000).

18 Rhoads ML, Fetterer RH & Hill DE. *Trichuris suis*: A secretory serine protease inhibitor. *Exp. Parasitol.* **94**, 1-7 (2000).

19 Johnston MJ, MacDonald JA & McKay DM. Parasitic helminths: a pharmacopeia of anti-inflammatory molecules. *Parasitology* **136**, 125-147 (2009).

20 van Die I & Cummings RD. Glycan gimmickry by parasitic helminths: a strategy for modulating the host immune response? *Glycobiology* **20**, 2-12 (2010).

21 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.* **36**, 915-924 (2006).

22 Nisbet AJ, Cottee PA & Gasser RB. Genomics of reproduction in nematodes: prospects for parasite intervention? *Trends Parasitol.* **24**, 89-95 (2008).

23 Nisbet AJ, Cottee P & Gasser RB. Molecular biology of reproduction and development in parasitic nematodes: progress and opportunities. *Int. J. Parasitol.* **34**, 125-138 (2004).

24 Boag PR, Gasser RB, Nisbet AJ & Newton SE. Genomics of reproduction in parasitic nematodes-fundamental and biotechnological implications. *Biotechnol. Adv.* **21**, 103-108 (2003).

25 Spakulova M, Kralova I & Cutillas C. Studies on the karyotype and gametogenesis in *Trichuris muris*. *J. Helminthol.* **68**, 67-72 (1994).

26 Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).

27 Carvalho AB & Clark AG. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* **23**, 1894-1907 (2013).

28 Altschul SF, Madden TL, Schaffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).

29 Nawrocki EP & Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935 (2013).

30 Fu L, Niu B, Zhu Z, Wu S & Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).