Genomics 113 (2021) 2695–2701

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

Genome sequence of the cardiopulmonary canid nematode *Angiostrongylus vasorum* reveals species-specific genes with potential involvement in coagulopathy

Annageldi Tayyrov [a,*], Nina Gillis-Germitsch [a,b], Lucienne Tritten [a], Manuela Schnyder [a]

[a] Institute of Parasitology, Vetsuisse-Faculty, University of Zurich, Winterthurerstrasse 266a, Zurich 8057, Switzerland
[b] Graduate School for Cellular and Biomedical Sciences, University of Bern, Switzerland

ARTICLE INFO

Keywords:
Angiostrongylus vasorum
Cardiopulmonary nematode
Whole-genome sequencing
Oxford Nanopore
Vaccine candidates

ABSTRACT

Angiostrongylus vasorum is an emerging parasitic nematode of canids and causes respiratory distress, bleeding, and other signs in dogs. Despite its clinical importance, the molecular toolbox allowing the study of the parasite is incomplete. To address this gap, we have sequenced its nuclear genome using Oxford nanopore sequencing, polished with Illumina reads. The size of the final genome is 280 Mb comprising 468 contigs, with an N50 value of 1.68 Mb and a BUSCO score of 93.5%. Ninety-three percent of 13,766 predicted genes were assigned to putative functions. Three folate carriers were found exclusively in *A. vasorum*, with potential involvement in host coagulopathy. A screen for previously identified vaccine candidates, the aminopeptidase H11 and the somatic protein rHc23, revealed homologs in *A. vasorum*. The genome sequence will provide a foundation for the development of new tools against canine angiostrongylosis, supporting the identification of potential drug and vaccine targets.

1. Introduction

Angiostrongylus is a genus of parasitic nematodes in the superfamily Metastrongylidae, which contains two species of relevant zoonotic nature, namely, *Angiostrongylus costaricensis* and *Angiostrongylus cantonensis*. These species normally infect rodents as their definitive host and are also known to incidentally infect humans [1,2]. *Angiostrongylus vasorum* is another clinically important parasitic member of this genus: it is a cardiopulmonary nematode that lives in the pulmonary arteries and the right side of the heart of several canid species, including domestic dogs (*Canis familiaris*), red foxes (*Vulpes vulpes*), and other wild carnivores [3].

Angiostrongylus vasorum, also known as the French heartworm, was described in detail by Baillet in 1866 [4]. It has spread from the region in France where it was originally discovered to many other places in Europe [5] and the Americas [6,7]. The parasite has a heteroxenic lifecycle where snails serve as an intermediate host in which, after ingestion of contaminated canid feces, first-stage larvae develop into infectious third-stage larvae [8,9].

Currently, the treatment of canine angiostrongylosis is limited to the use of anthelmintics such as moxidectin and milbemycin-oxime or fenbendazole. Such treatment is, in case of heavy worm burdens and chronicity, associated with a high risk of side effects such as immune-mediated thrombocytopenia, thrombotic arteritis, shock, or congestive heart failure [10–12]. Besides, anthelmintic drugs are effective at eliminating existing *A. vasorum* infections [13,14], but they do not protect against reinfection. An effective prevention requires regular administration of the drugs. There are also growing concerns about the regular use of anthelmintics in view of reduced efficacy. These issues necessitate the development of alternative preventive and management measures.

The number of sequenced parasite genomes has greatly expanded over the last decade [15]. However, genome sequences of nematodes had been progressing rather slowly due to challenges associated with their relatively big genome size [16] and the high level of polymorphism within species [17,18].

The development of long-read sequencing or third-generation sequencing technologies has opened a new era for whole-genome sequences, including for nematode genomes [19]. While short-read sequencing such as from second-generation sequencing or Sanger platforms are more accurate and supported by a wide range of bioinformatics tools, they alone are not ideal for sequencing of genomes where

* Corresponding author.
E-mail address: annageldi.tayyrov@uzh.ch (A. Tayyrov).

https://doi.org/10.1016/j.ygeno.2021.06.010
Received 2 October 2020; Received in revised form 21 May 2021; Accepted 7 June 2021
Available online 10 June 2021

the repetitive regions are longer than the size of the maximum read lengths [20]. Long read sequencing technologies such as Pacific Biosciences PacBio sequencing [21] and Oxford Nanopore Technologies (ONT) nanopore sequencing technologies [22] routinely generate reads longer than 10 kb, allowing the resolution of many repetitive regions in a genome [20]. One of the drawbacks of long-read sequencing technologies is their higher error rate compared to short-read sequences [23]. The error rate of long reads can be significantly reduced with polishing using more accurate Illumina short reads [24].

The genomic sequence of a parasite provides a foundation for better understanding and control of the pathogen and an invaluable source of potential vaccine candidates [25]. Reverse vaccinology exploits genomic data of an organism to identify novel immunogenic products as putative vaccine candidates [26]. The approach has been successfully used in viral and bacterial vaccine development pipelines and shown to be more efficient than traditional vaccinology on several occasions [27]. In addition, genome sequences provide a new opportunity for the future identification and development of microsatellite markers. These markers are simple repeat sequences consisting of 1–6 nucleotide motifs that can be used as a genetic marker for studying population genetic variation within and between species.

Here we report a high-quality genome of *Angiostrongylus vasorum*, produced by using long-read ONT nanopore sequencing followed by polishment of the exonic regions with Illumina short reads. The genome sequence will serve as a useful resource for a better understanding of the parasite leading to the development of new solutions for the control of canine angiostrongylosis.

2. Materials and Methods

2.1. Sample Collection and Nucleic Acid Extraction

Adult female *A. vasorum* nematodes were collected from the lungs of freshly hunted foxes in Zurich, Switzerland. The worms were washed in Phosphate-buffered saline (PBS) buffer, and subsequently, genomic DNA (gDNA) and total RNA were extracted using E.Z.N.A.® Mollusc DNA Kit (Norcross, OMEGA, USA) and Norgen RNA extraction kit (Norgen Biotek Corporation, Canada), respectively, following manufacturers' protocols. The quantity and quality of the extracted gDNA were assessed with NanoDrop One Microvolume UV–Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and Qubit® (1.0) Fluorometer dsDNA Broad Range assay (Life Technologies, California, USA). The fragment size and quality of DNA samples were assessed using Femto Pulse (Agilent Technologies, Inc., USA). The quantity and quality of extracted RNA samples were assessed using Qubit® (1.0) Fluorometer and TapeStation Instrument (Agilent Technologies, Santa Clara, CA), respectively.

gDNA and RNA libraries were prepared and sequenced at the Functional Genomics Center Zurich (FGCZ), University of Zurich, as described before [28]. Briefly, a gDNA sample with a fragment size of >15 kb was proceeded with ONT (Oxford Nanopore Technologies) library preparation using the ligation sequencing kit (Oxford Nanopore, SQK-LSK109) protocol, without fragmentation [29].

2.2. gDNA Sequencing and Assembly

The ONT library was then sequenced for 48 h using a GridION X5 nanopore sequencer (Oxford Nanopore). ONT adapters were trimmed off using Porechop v0.2.4 [30]. The quality of reads was assessed and filtered using Nanofilt v2.7.1 [31]. Reads with a quality score of >Q7 were assembled using Flye v2.5, an assembler specifically developed to assembly third-generation long-read sequences, with the default parameters except – *iteration 3* [32]. The exonic regions of the draft assembly were further polished using Illumina RNA-seq short reads using 4 rounds of Pilon v1.23 with – *fix all* parameter [23]. The completeness of the assembly was assessed before and after each round of Pilon polishing using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 [33] analyses against nematoda_odb10 (Creation date: 2019-11-20, number of species: 7, number of BUSCOs: 3131) embedded in Quast v5.0.2 *with –conserved-genes-finding and –eukaryote* parameters added [34]. Contigs with smaller sizes (<50 kbp) were filtered out based on BUSCO results without compromising the genome completeness.

2.3. RNA Sequencing

Poly(A) + enriched RNA-seq libraries were prepared at the FGCZ using the TruSeq Stranded RNA Kit v2 (Illumina Inc., California, USA) following the manufacturers' protocols. In brief, RNA samples were enriched for poly-adenylated messenger RNA followed by reverse transcription into cDNA. The cDNA sample was fragmented and end-repaired before adding Illumina adaptors. The quantity and quality of prepared libraries were assessed using Qubit® (1.0) Fluorometer and Bioanalyzer 2100 (Agilent, Waldbronn, Germany), respectively. The transcriptome library was sequenced on the Illumina Novaseq 6000 pair end at 2 × 150 bp using the TruSeq SBS Kit v4-HS (Illumina, Inc., California, USA). Quality control of the generated data was assessed using FastQC v0.11.9 [35]. Trimmomatic v0.39 was used to remove Illumina adaptors [36] which was followed by de nova transcript assembly using Trinity v2.9.1 [37]. Unless stated otherwise, the default parameters were used for each program.

2.4. Genome Annotation

A custom repeat library was generated from the draft genome using RepeatModeler v2.0.1 with - *engine ncbi* option [38]. The custom repeat library along with the Dfam (3.0) database of repetitive DNA families [39] was used for masking on the draft genome using RepeatMasker v4.0.9 [40]. Genes were predicted using the MAKER v2.31.10 pipeline by employing ab initio gene predictors along with transcriptome and proteome datasets [41]. The MAKER pipeline was run for three consecutive rounds. In the first round, species-specific repeat library, transcriptome assembly created in this study, and secretome proteins (n = 1132) along with transcripts (*Caenorhabditis elegans*, BioProject: PRJNA13758) and proteins (*C. elegans*, BioProject: PRJNA13758; *A. costaricensis*, BioProject: PRJEB494; *A. cantonensis*, BioProject: PRJEB493) from related species were used to predict empirical-based gene models. Before running the second MAKER round, ab initio gene predictors Augustus v.3.3.3 [42] and SNAP v2013_11_29 [43] were trained using evidence from the first MAKER round. Augustus was trained using BUSCO with the *–long* parameter. SNAP was trained using the gene models from the first round of MAKER with an AED score of 0.2 or better and a length of minimum 50 amino acids. For the third round, again SNAP and Augustus were trained as before using the predicted models from the second MAKER round. The quality of gene prediction was assessed against nematoda_odb10 dataset using BUSCO v4.1.4. The final generated gene list was annotated using BlastP similarity searches (E-value < 1e-3) against Swiss/Uniprot and NCBI non-redundant (nr) database where for the latter the Blast search was limited to the members of the phylum Nematoda (taxid 6231) using the BLAST+ command-line tool v2.9.0 [44]. Blast results were mapped against the Gene Ontology Annotation database and searched for conserved protein domains with the InterProScan function of Blast2GO v5.2.5 [45].

Two vaccine candidate genes of *Haemonchus contortus*, namely microsomal aminopeptidase H11 (GenBank: ACK57928.1) [46] and somatic protein rHc23 (GenBank: CDJ92660.1) [47] were searched against the predicted proteome of *A. vasorum* using BLASTP v2.6.0. Phylogenetic analysis (MUSCLE alignment, neighbor-joining tree) was performed using MEGA-X v10.1.15 and visualized at the Interactive Tree Of Life (https://itol.embl.de) [48]. Potential *N*- and *O*-Glycosylation sites were predicted using the N-GlyDE tool [49] and the NetOGlyc-4.0 server [50].

Perfect and imperfect SSR (Simple sequence repeats) were detected

using Krait v1.2.2 with default parameters [51]. Primer3 program implemented in the Krait [52] software was used to design primer pairs for the identified perfect SRRs. Primer3 was run with default parameters except where the acceptable GC range was changed from 30 to 80 to 40–70.

2.5. Phylogenomics Analysis

Orthologous protein family analysis of *A. vasorum* was performed using OrthoFinder v2.4.0. [53]. The predicted proteome of *A. vasorum* was analyzed together with protein sequences of the other six clade V nematode species namely: *Ancylostoma caninum* (BioProject: PRJNA72585) [16], *A. cantonensis* (BioProject: PRJEB493) [16], *A. costaricensis* (BioProject: PRJNA350391) [54], *C. elegans* (BioProject: PRJNA13758) [55], *Dictyocaulus viviparus* (BioProject: PRJNA72587) [56] and *H. contortus* (BioProject: PRJEB506) [57]. The diamond algorithm v0.9.24 was used for all-versus-all sequence search and alignment [58]. An unrooted species tree was inferred from concatenated multiple sequence alignment (MSA) of single-copy orthogroups using maximum-likelihood phylogenetic tree maker, FastTree v2.1.10 [59]. The unrooted tree was rooted using the STRIDE algorithm embedded into OrthoFinder [60]. The Newick-format file was visualized using the Interactive Tree Of Life (https://itol.embl.de) [48]. *A. vasorum* genes that are either uniquely assigned to the parasite or did not belong to any orthogroups were considered to be specific for *A. vasorum*. These genes were further analyzed, and biological processes and molecular functions were assigned using Blast2GO v5.2.5 [45].

3. Results

3.1. Assembly

ONT sequencing yielded a total of 8.9 Gb of data. Minimum quality cutoff (>Q7) was applied and resulted in 7.8 Gb of data. The quality of filtered reads ranged from Q7 to Q15 with a median read quantity of 11.6. The mean read length was 7.2 Kb with the N50 value of 14.3 Kb (Table 1).

A total of 1,087,814 reads with >Q7 were used in the genome assembly. The Flye assembler [32] produced a 298 MB genome consisting of 2417 contigs with the largest contig being 7.5 Mb and N50 of 1.6 Mb. Moreover, 95.5% of the draft genome assembly was represented with only 470 contigs that are larger than 50 Kb (Table 2).

The ONT assembly without polishing with short reads showed a Complete BUSCO score of 84.7%. BUSCO analysis after removing the shorter contigs (<50 Kb) did not affect the completeness of the assembly (Table S1). The exonic regions of the ONT-only assembly were polished with Illumina short reads for four consecutive rounds using Pilon. The Illumina data generated in this study contained 149,128,628 pairs of reads with an average quality score of above Q30. The length of each read was 150 bases making the total amount of Illumina-generated data around 45 Gb. The completeness of the genome was assessed after each round of pilon polishing. The Complete BUSCO score was 93.4% after 2 rounds of Pilon polishment. Further rounds of polishment did not improve the assembly (Table 2). After polishing with Illumina short reads, <50 Kb contigs were filtered out based on BUSCO analysis without affecting the completeness of the assembly (Table S1).

Table 1
Summary statistics of ONT sequencing output.

Feature	Raw reads	Pass reads[a]
Mean read length	6847.80	7205.20
Mean read quality	10.4	11.4
Median read length	3548.00	3616.00
Median read quality	11.1	11.6
Number of reads	13,01,673.00	10,87,814.00
Read length N50	13,934.00	14,295.00
Total bases	8,91,35,46,938.00	7,83,79,04,788.00
Number, percentage, and lengths of reads above quality cutoffs		
>Q5	1,208,497 (92.8%) 8541.3 Mb	1,087,814 (100.0%) 7837.9 Mb
>Q7	1,087,814 (83.6%) 7837.9 Mb	1,087,814 (100.0%) 7837.9 Mb
>Q10	821,480 (63.1%) 6061.4 Mb	821,485 (75.5%) 6061.6 Mb
>Q12	464,755 (35.7%) 3514.6 Mb	464,756 (42.7%) 3514.7 Mb
>Q15	11,429 (0.9%) 49.5 Mb	11,429 (1.1%) 49.5 Mb

[a] Represents reads with >Q7.

3.2. Comparison of Angiostrongylus Genomes

Angiostrongylus vasorum genome assembly was compared to the latest genomes of two other *Angiostrongylus* species, namely *Angiostrongylus cantonensis* (BioProject number: PRJNA350391) and *Angiostrongylus costaricensis* (BioProject number: PRJEB494). *Angiostrongylus cantonensis* and *A. vasorum* had the same unfiltered genome size of 293 MB while *A. costaricensis* had a smaller genome assembly of 263 Mb. All three had a GC content between 41 and 42%. While BUSCO scores were comparable for all three species with values >90%, *A. vasorum* had better assembly values for almost all features (contig number, contig sizes, N50, L50; Table 3).

The completed number of BUSCO genes (single copy + duplicated) was 2928/3131 for *A. vasorum*, corresponding to 93.5% completeness. The completed BUSCO numbers are 2888 and 2849 for *A. cantonensis* and *A. costaricensis*, respectively. For all assemblies, predicted BUSCO duplication values were below 1% (Table 4). Out of 3131 BUSCO genes, 157 genes were missing in the *A. vasorum* assembly. Sixty-seven of those missing 157 BUSCO genes were also missing in both *A. cantonensis* and *A. costaricensis*, while 97/157 were missing in either of the other two species (Fig. 1).

3.3. Annotation

Overall, 47.85% (133.9 Mb) of the genome consists of repetitive elements, and these regions were soft-masked for the subsequent analysis. At the end of three rounds of the MAKER pipeline, we predicted 13,766 protein-coding genes with an average gene length of 1.5 kb and an average exon number of 8.4. MAKER mRNA Quality index (QI) and an Annotation Edit Distance (AED) were included for each of the predicted genes as a metric for annotation quality control. The large majority (90.5%) of genes had AED values of <0.5, where an AED score of 1 denoting no support of evidence for annotation while an AED value of 0 indicates perfect support of evidence. The BUSCO completeness score was 74.3% [Single-copy BUSCOs: 73.5%, Duplicated BUSCOs: 0.8%] for annotated transcripts of genes. Among predicted genes, 59.7% (n = 8215) and 92.2% (n = 12,688) were supported by Swiss/UniProt and NCBI nr protein databases, respectively. In total, 10,482 genes could be assigned GO terms (Table S2). The distribution of top Blast2GO hits among nematode species is shown in Fig. S1, where the top two species with the most matches were *Angiostrongylus costaricensis* and *Angiostrongylus cantonensis*, respectively.

Mining the *A. vasorum* genome assembly for vaccine candidates identified 20 and two significantly similar (E-value < 1e-5) homologs of aminopeptidase H11 [46] and somatic protein rHc23 [47] proteins, respectively. Blast results of these two proteins along with their annotations are shown in Table S3. Phylogenetic trees of *H11* (Fig. S2) and *rHc23* (Fig. S3) homologous genes from *A. vasorum* and *H. contortus* show scattered distribution of homolog genes between two species rather than species-specific clustering. In silico glycosylation site analysis of identified vaccine candidates revealed up to six potential N-Glycosylation sites and up to 11 O-Glycosylation sites in the sequences of aminopeptidase H11 homologs.

Table 2
ONT-only and Illumina short-read polished assembly features.

Assembly	Unpolished	Pilon polishment with Illumina reads[a]			
	ONT-only	Round 1	Round 2	Round 3	Round 4
# contigs (≥ 0 bp)	2417	2417	2417	2417	2417
# contigs (≥ 50 Kb)	470	468	468	468	468
Total length (≥ 0 bp)	29,81,84,885	29,61,16,400	29,33,35,652	29,30,21,587	29,29,89,225
Total length (≥ 50 Kb)	28,48,93,995	28,27,40,098	27,99,69,565	27,96,56,640	27,96,24,380
# contigs (≥ 3 Kb)[b]	1238	1238	1238	1238	1238
Largest contig	75,39,874	74,77,795	73,43,437	73,34,913	73,34,962
Total length	29,66,00,403	29,45,33,234	29,17,52,611	29,14,38,583	29,14,06,234
GC (%)	41.67	41.71	41.73	41.73	41.73
N50	16,24,931	16,08,840	15,91,569	15,91,478	15,91,478
N75	6,52,672	6,09,104	6,09,030	6,08,966	6,08,966
L50	52	52	52	52	52
L75	125	126	126	126	126
# N's per 100 kbp	1.52	1.5	1.52	1.52	1.52
Complete BUSCO (%)	84.73	93.39	93.39	93.36	93.39
Partial BUSCO (%)	2.46	1.56	1.63	1.47	1.44
Complete + Partial BUSCO (%)	87.19	94.95	95.02	94.83	94.83

[a] Exonic regions were polished.
[b] All statistics are based on contigs of size ≥ 3000 bp, unless otherwise noted (e.g., "#contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Table 3
Statistics and comparison of the current genome assembly with other *Angiostrongylus* species.

Assembly	*A. cantonensis*	*A. costaricensis*	*A. vasorum*[a]	*A. vasorum*[b]
# contigs (≥ 0 bp)	26,577	6384	2417	468
# contigs (≥ 50 Kb)	452	1677	468	468
# contigs (≥ 100 Kb)	385	827	304	304
Total length (≥ 0 bp)	29,33,06,699	26,27,82,696	29,33,35,652	27,99,69,565
Total length (≥ 50 Kb)	26,06,21,061	20,50,59,529	27,99,69,565	27,99,69,565
Total length (≥ 100 Kb)	25,59,63,369	14,39,87,101	26,87,17,000	26,87,17,000
# contigs (≥ 3 Kb)[c]	1577	4601	1238	468
Largest contig	46,79,122	8,23,818	73,43,437	73,43,437
Total length	27,95,76,253	26,06,48,492	29,17,52,611	27,99,69,565
GC (%)	41.5	41.22	41.73	41.66
N50	9,39,093	1,13,261	15,91,569	16,73,247
N75	4,38,224	56,680	6,09,030	7,28,034
L50	84	700	52	48
L75	193	1498	126	113
# N's per 100 kbp	6985.50	280.02	1.52	1.58
Complete BUSCO (%)	92.24	91.63	93.39	93.52
Partial BUSCO (%)	1.63	2.3	1.63	1.47
Complete + Partial BUSCO (%)	93.87	93.93	95.02	94.99

[a] Unfiltered draft assembly.
[b] Filtered assembly (≥ 50 Kb).
[c] All statistics are based on contigs of size ≥ 3000 bp, unless otherwise noted (e.g., "#contigs (≥ 0 bp)" and "Total length (≥ 0 bp)" include all contigs).

Table 4
BUSCO numbers and percentage for the three species of *Angiostrongylus*.

	A. cantonensis		*A. costaricensis*		*A. vasorum*	
	N	%	N	%	N	%
Total BUSCO groups searched	3131	100	3131	100	3131	100
Complete BUSCOs (S + D)	2888	92.2	2869	91.6	2928	93.5
Complete and single-copy BUSCOs (S)	2871	91.7	2849	91	2911	93
Complete and duplicated BUSCOs (D)	17	0.5	20	0.6	17	0.5
Fragmented BUSCOs (F)	51	1.6	72	2.3	46	1.5
Missing BUSCOs (M)	192	6.2	190	6.1	157	5

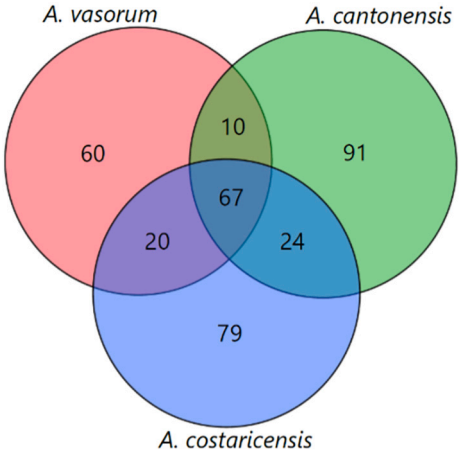

Fig. 1. Missing BUSCO genes comparison of three Angiostrongylus species.

3.4. Detection and Distribution of Microsatellites

The assembled genome was scanned for perfect and imperfect microsatellite markers with a repeat length of 1–6 nucleotides. A total of 24,307 and 147,554 perfect and imperfect microsatellite sequences were detected, respectively. The total length of perfect SSRs was 1138 Mb, while this number was 5618 Mb for imperfect SSRs making the relative density of perfect and imperfect SSRs 86.82 and 527.04 (bp/Mb), respectively (Table 5).

The most common perfect SSR type was dinucleotide repeats with a total length of 552,038 nucleotides (Table S4). Among those, the three most frequent dinucleotide microsatellite motifs were (AT)n, (AC)n, and (AG)n. Trinucleotide repeats were the most abundant repeats (2,240,813 bp) among imperfect SSRs (Table S5) where (AAG)n, (AAC)n, and (AAT)n were the three most frequent trinucleotide microsatellite patterns. Primer pairs can be designed for 15,062 perfect microsatellite sequences by bioinformatic analysis. Of those 24,306 identified perfect

Table 5
Summary of perfect and imperfect SSRs (Simple Sequence Repeats) detected in the *Angiostrongylus vasorum* genome.

Item	Description	Number
Total number of perfect SSRs	Counts	24,307
Total length of perfect SSRs	bp	11,37,676
Relative abundance of perfect SSRs	total SSRs/total valid length (loci/Mb)	86.82
Relative density of perfect SSRs	total SSR length/total valid length (bp/Mb)	4063.63
Total number of imperfect SSRs	Counts	1,47,554
Total length of imperfect SSRs	bp	56,17,689
Relative abundance of imperfect SSRs	total iSSR counts/total valid length (loci/Mb)	527.04
Relative density of imperfect SSRs	total iSSR length/total valid length (bp/Mb)	20,065.70

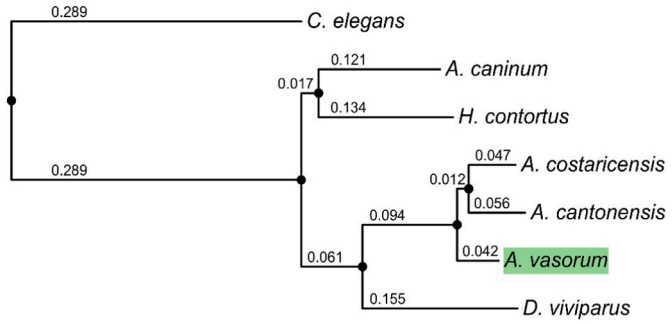

Fig. 2. Rooted phylogenetic tree analysis of seven related clade V nematode species based on 2630 single-copy genes. The length of branches represents a number of amino acid substitutions per site. Every node is supported by the best node values of 100/100.

microsatellite sequences, 15,062 had a suitable sequence pattern for the design of primer pairs.

3.5. Phylogenomics Analysis

Phylogenomics analysis was performed using 139,371 genes from seven nematode species. OrthoFinder assigned 125,517 of those genes (90.1% of total) to 17,467 orthogroups. *A. vasorum* had 12,622 (91.7%) of its predicted genes assigned to 9793 orthogroups and 89 of those gene families encompassing 437 genes were unique to the parasite. These genes were combined and analyzed together with the unassigned 1144 genes (Table S6). Fifty-one percent (813/1581) of the *A. vasorum* unique genes had significant hits when blasted against the NCBI database (Fig. S4A). Among the blast results, there were genes encoding for aminopeptidase-N (GO: 0008233), metallopeptidases (GO: 0008237), and three (GO: 0090482) folate carriers (Table S2). Visualization of GO-terms showed that the majority of unique genes were assigned to either "Cellular process" (n = 161) or "Metabolic process" (n = 99) (Fig. S4B) while the "Hydrolase activity" was the most common molecular function (Fig. S4C). Among the analyzed seven nematode species, *A. vasorum* had most of its orthogroups shared with *A. cantonensis* (n = 8510) and *A. costaricensis* (n = 8595) while the least overlap was with *C. elegans* (n = 7294) and *H. contortus* (n = 7708) (Table S7). There were 5549 orthogroups with all species present and 725 of these consisted entirely of single-copy genes. The species tree was constructed using 2630 orthogroups with a minimum of 85.7% of species having single-copy genes in any orthogroup. *C. elegans* was detected as the best outgroup for species tree where it was supported by 39 out of 41 non-terminal duplications (Fig. 2).

4. Discussion

In this study, we generated a draft genome assembly of *A. vasorum* using the combination of ONT and Illumina data. The final polished draft resulted in a highly contiguous genome (N50: 1.68 Mb, #contig: 468) with a BUSCO completeness score of 93.5%. We predicted 13,766 protein-coding genes, of which 93% were functionally annotated using several databases.

The genome size of *A. vasorum* was considerably larger (~280 Mb) than that of *A. costaricensis* (~263 Mb) and *A. cantonensis* (253 Mb) [16]. This difference is attributed to the fact that we used long read derived sequences while the other *Angiostrongylus* genomes were sequenced using Illumina short-read sequences. Long sequences support a more accurate determination of repetitive regions, resulting in highly complete and longer genomes. Coghlan et al. (add year?) estimated the actual genome size of *A. costaricensis* to be 284.5 Mb and for *A. cantonensis* to be around 279 Mb when sequences gained by "uncollapsing" repeats are taken into account [16]. For instance, the reference genome of *C. elegans* was expanded by more than 2 Mb when MinION long-read derived sequences were used, even though *C. elegans* had one of the most complete genomes [61]. For the latest genome of *A. cantonensis* a combination of Illumina short-read and PacBio long-read data were used and this assembly resulted in 282 Mb in size [54], in agreement with the present *A. vasorum* genome size. However, ONT-only genomes, despite their contingency, are known to produce many small SNPs and indel errors. Thus, to produce a more accurate genome, it is highly recommended to polish ONT-only assemblies with sequences from another technology with a low error rate, such as Illumina [28]. Indeed, the BUSCO completeness score increased from 87% to 95% when we polished exonic regions of the ONT-only draft genome with Illumina reads. Polishing the non-coding regions of the genome might be needed for future studies. Comparing the missing 5% of BUSCO genes with the other two *Angiostrongylus* species revealed that 61.2% of those missing genes were also missing in either of *A. cantonensis* or *A. costaricensis*, suggesting the absence of these orthologues in this group of species rather than an assembly issue. Similar to the other *Angiostrongylus* species, duplicated BUSCO value was very low for *A. vasorum* suggesting low heterozygosity and genome duplication in this nematode genus. Sampling worms from rural foxes might have helped us, as they are known to be genetically less diverse compared to urban fox isolates [62].

More than 90% of transcripts have an AED score (the quality indicator for MAKER) of less than 0.5 and 60% of predicted genes were assigned to UniProtKB/SwissProt database, indicating overall good annotation quality. However, the BUSCO completeness score for predicted genes was considerably lower than the one resulting from the whole genome assembly. Thus, it will be important to improve gene prediction in the future.

Angiostrongylus vasorum has experienced successful expansion throughout Europe in the last decades, supported by foxes acting as wildlife reservoirs [63]. Anthelmintics are the current mainstay for disease treatment and prevention. Vaccines would represent a sustainable control alternative, possibly also applicable to the wildlife reservoir [64]. High-quality genomes can be a valuable resource to mine for potential vaccine candidates based on gene homology to related species. Thus, we looked for the presence of homologs of the two well-known vaccine candidate genes of *Haemonchus contortus*, namely H11 [46] and rHc23 [47], in the predicted proteome of *A. vasorum*. Challenges with the production of recombinant forms of the protein can be expected due to the high number of putative glycosylation sites of *A. vasorum* H11 homologs. Indeed, *H. contortus* H11 is highly protective in its native form, whereas recombinant H11 has failed to induce the same level of protection [65], which is at least partially attributed to the incorrect folding and inappropriate glycosylation of recombinant proteins along with the presence of several H11 isoforms. However, immunization with another recombinant vaccine candidate, rHc23, was found to be

partially effective against *H. contortus* infection [47]. Thus, homologs of *A. vasorum* rHc23 make them a promising vaccine candidate. Whether immunization with recombinant rHc23 proteins could induce protection against *A. vasorum* remains to be determined.

The genome sequence also offers a new opportunity to identify and develop new genomic markers. In particular, markers originating from the nuclear genome are needed for *A. vasorum*, as the currently most widely used genomic marker ITS-2 (second internal transcribed spacer) often produces errors when sequenced directly, making cloning a prerequisite [66]. To this end, we mined the genome and identified around 15,000 perfect SSRs for which primer pairs can be designed for gene amplification. Further experimental work is needed to validate those microsatellite regions for use as genetic markers. Computer predictions offer advantages over lab-based screening of genomic markers. For example, out of 12,160 microsatellite repeats identified computationally, 509 new genomic markers for *Tribolium castaneum* could be validated experimentally [67]. It was shown that microsatellite density and coverage are independent of the genome size and differs greatly among different nematodes [68]. For example, *A. vasorum* had two-fold higher microsatellite coverage than *C. elegans*, while three-fold lower coverage than *B. malayi* [68]. Similar to many other organisms [69], dinucleotides were the most dominant microsatellite for *A. vasorum*, too. We also found that *A. vasorum* microsatellites are AT-rich, which is in line with previous findings, where the analysis of 26 eukaryotic genomes of 26 identified predominantly GC-low repeats [70].

The phylogenomic analysis based on orthologous gene families has shown *A. cantonensis* and *A. costaricensis* as mostly related species, thus validating the expected phylogeny. Interestingly, among the *A. vasorum* exclusive genes, three genes were found to be encoding for folate carriers. These membrane proteins are essential to keep blood homocysteine level at a minimum to prevent accidental blood clotting [71]. *Angiostrongylus vasorum* infection is known to be associated with coagulopathy in dogs [72]. Thus, the folate transport proteins may affect blood clotting in the host to facilitate feeding and survival of this blood-borne parasite.

Mitochondrial genome [73] and the adult stage de novo transcriptome assembly [74] were the only available large-scale genomic data for *A. vasorum*. Thus, the current whole-genome assembly will be an important complement. The exploitation of genomic data is pivotal for the advancement of the ongoing research on *A. vasorum* infections towards the development of more sustainable control measures. It will be equally important to complement the genome assembly with transcriptome and proteome data from different developmental stages of the parasite to considerably improve our understanding of the parasite's biology and host-parasite interactions.

5. Conclusion

Angiostrongylus vasorum is an emerging parasite responsible for serious disease with potentially fatal outcomes in dogs if left untreated. Preventive measures such as vaccines are much needed. However, the lack of a deep understanding of the parasite's biology has impaired the development of novel control strategies. To boost the research in the field, we sequenced a high-quality genome of *A. vasorum*. We hope the genome sequence will be an important contribution to the advancement of the ongoing research on *A. vasorum* infections towards the development of more sustainable control measures.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.06.010.

Authors' contribution

Annageldi Tayyrov: Conceptualization, Methodology, Data curation and analysis, Writing - Original draft preparation.

Nina Gillis-Germitsch: Fox dissection, Nematode collection, Reviewing & Editing.

Lucienne Tritten: Conceptualization, Reviewing & Editing.

Manuela Schnyder: Funding acquisition, Supervision, Conceptualization, Writing - Review & Editing.

Data availability

All sequenced sequences were deposited in the GenBank database and are available under Bioproject number PRJNA663250.

Funding

Bayer Animal Health GmbH, Leverkusen, Germany, and the Swiss National Science Foundation (Grant No: 310030_201045/1) funded the study. The funder had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

We would like to thank members of the Functional Genomics Center Zurich for their technical support.

References

[1] R.N. Incani, et al., Human infection by *Angiostrongylus costaricensis* in Venezuela: first report of a confirmed case, Rev. Inst. Med. Trop. Sao Paulo 49 (3) (2007) 197–200.
[2] Q.P. Wang, et al., Human *Angiostrongylus cantonensis*: an update, Eur. J. Clin. Microbiol. Infect. Dis. 31 (4) (2012) 389–395.
[3] G. Bolt, et al., Canine Angiostrongylosis - a review, Vet. Rec. 135 (19) (1994) 447–452.
[4] C. Baillet, Strongyle des vaisseaux et du coeur du chien. Strongylus vasorum (Nobis) 8, Nouveau Dictionnaire Pratique de Médecine, de Chirurgie et d'Hygiène Vétérinaires, 1866, pp. 587–588.
[5] J.R. Helm, et al., Canine angiostrongylosis: an emerging disease in Europe, J. Vet. Emerg. Crit. Care (San Antonio) 20 (1) (2010) 98–109.
[6] G.A. Conboy, Canine angiostrongylosis: the French heartworm: an emerging threat in North America, Vet. Parasitol. 176 (4) (2011) 382–389.
[7] F. Penagos-Tabares, et al., *Angiostrongylus vasorum* and *Aelurostrongylus abstrusus*: neglected and underestimated parasites in South America, Parasit. Vectors 11 (1) (2018) 208.
[8] J. Guilhon, B. Cens, Migrations and evolution of Angiostrongylus vasorum (Baillet, 1866) in dogs, C. R. Acad. Hebd. Seances Acad. Sci. D. 269 (24) (1969) 2377–2380.
[9] R. Sauerlander, J. Eckert, The African giant snail (Achatina fulica) as experimental intermediate host of Angiostrongylus vasorum (Nematoda) (author's transl), Zeitschrift fur Parasitenkunde. 44 (1) (1974) 59–72.
[10] J. Soland, G. Bolt, Hypovolaemic shock after anthelmintic treatment of canine angiostrongylosis, J. Small Anim. Pract. 37 (12) (1996) 594–596.
[11] S.M. Gould, E.L. McInnes, Immune-mediated thrombocytopenia associated with *Angiostrongylus vasorum* infection in a dog, J. Small Anim. Pract. 40 (5) (1999) 227–232.
[12] A. Corda, et al., Pulmonary arterial response to *Angiostrongylus vasorum* in naturally infected dogs: echocardiographic findings in two cases, Parasit. Vectors 12 (1) (2019) 286.
[13] M. Schnyder, et al., Larvicidal effect of imidacloprid/moxidectin spot-on solution in dogs experimentally inoculated with *Angiostrongylus vasorum*, Vet. Parasitol. 166 (3–4) (2009) 326–332.
[14] J.L. Willesen, et al., Efficacy and safety of imidacloprid/moxidectin spot-on solution and fenbendazole in the treatment of dogs naturally infected with *Angiostrongylus vasorum* (Baillet, 1866), Vet. Parasitol. 147 (3–4) (2007) 258–264.
[15] C. Talavera-Lopez, B. Andersson, Parasite genomics-time to think bigger, PLoS Negl. Trop. Dis. 11 (4) (2017), e0005463.
[16] International Helminth Genomes Consortium, Comparative genomics of the major parasitic worms, Nat. Genet. 51 (1) (2019) 163–174.
[17] P.D. Keightley, B. Charlesworth, Genetic instability of *C. elegans* comes naturally, Trends Genet. 21 (2) (2005) 67–70.
[18] A.D. Cutter, A. Dey, R.L. Murray, Evolution of the *Caenorhabditis elegans* genome, Mol. Biol. Evol. 26 (6) (2009) 1199–1234.
[19] A.M. Giani, et al., Long walk to genomics: history and current approaches to genome sequencing and assembly, Comput. Struct. Biotechnol. J. 18 (2020) 9–19.
[20] S.L. Amarasinghe, et al., Opportunities and challenges in long-read sequencing data analysis, Genome Biol. 21 (1) (2020) 30.
[21] A. Rhoads, K.F. Au, PacBio sequencing and its applications, Genomics Proteomics Bioinforma. 13 (5) (2015) 278–289.
[22] N. Kono, K. Arakawa, Nanopore sequencing: review of potential applications in functional genomics, Develop. Growth Differ. 61 (5) (2019) 316–326.

[23] F.J. Rang, W.P. Kloosterman, J. de Ridder, From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy, Genome Biol. 19 (1) (2018) 90.
[24] S.C. Shin, et al., Nanopore sequencing reads improve assembly and gene annotation of the Parochlus steinenii genome, Sci. Rep. 9 (1) (2019) 5095.
[25] D. Serruto, R. Rappuoli, Post-genomic vaccine development, FEBS Lett. 580 (12) (2006) 2985–2992.
[26] R. Rappuoli, Reverse vaccinology, a genome-based approach to vaccine development, Vaccine 19 (17–19) (2001) 2688–2691.
[27] K.L. Seib, X. Zhao, R. Rappuoli, Developing vaccines in the era of genomics: a decade of reverse vaccinology, Clin. Microbiol. Infect. 18 (2012) 109–116.
[28] W. Qi, et al., New insights on Pseudoalteromonas haloplanktis TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies, Sci. Rep. 9 (1) (2019) 16444.
[29] I. Elliott, et al., Oxford Nanopore MinION sequencing enables rapid whole genome assembly of rickettsia typhi in a resource-limited setting, Am. J. Trop. Med. Hyg. 102 (2) (2020) 408–414.
[30] R.R. Wick, et al., Completing bacterial genome assemblies with multiplex MinION sequencing, Microb. Genom. 3 (10) (2017), e000132.
[31] W. De Coster, et al., NanoPack: visualizing and processing long-read sequencing data, Bioinformatics 34 (15) (2018) 2666–2669.
[32] M. Kolmogorov, et al., Assembly of long, error-prone reads using repeat graphs, Nat. Biotechnol. 37 (5) (2019) 540–546.
[33] M. Seppey, M. Manni, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness, Methods Mol. Biol. 1962 (2019) 227–245.
[34] A. Gurevich, et al., QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (8) (2013) 1072–1075.
[35] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2010.
[36] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120.
[37] B.J. Haas, et al., De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis, Nat. Protoc. 8 (8) (2013) 1494–1512.
[38] A. Smit, R. Hubley, RepeatModeler Open-1.0.. http://www.repeatmasker.org, 2008-2015.
[39] R. Hubley, et al., The Dfam database of repetitive DNA families, Nucleic Acids Res. 44 (D1) (2016) D81–D89.
[40] A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0.. http://www.repeatmasker.org, 2013-2015.
[41] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinforma. 12 (2011) 491.
[42] M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics 19 (Suppl. 2) (2003) ii215–25.
[43] I. Korf, Gene finding in novel genomes, BMC Bioinforma. 5 (2004) 59.
[44] C. Camacho, et al., BLAST+: architecture and applications, BMC Bioinforma. 10 (2009) 421.
[45] S. Gotz, et al., High-throughput functional annotation and data mining with the Blast2GO suite, Nucleic Acids Res. 36 (10) (2008) 3420–3435.
[46] T.S. Smith, et al., Purification and evaluation of the integral membrane-protein H11 as a protective antigen against Haemonchus-Contortus, Int. J. Parasitol. 23 (2) (1993) 271–280.
[47] M.E. Gonzalez-Sanchez, et al., Immunization with recombinant rHc23 partially protects lambs against trickle infections by Haemonchus contortus, BMC Vet. Res. 15 (1) (2019) 333.
[48] I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: recent updates and new developments, Nucleic Acids Res. 47 (W1) (2019) W256–W259.
[49] T. Pitti, et al., N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding, Sci. Rep. 9 (1) (2019) 15975.
[50] C. Steentoft, et al., Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology, EMBO J. 32 (10) (2013) 1478–1488.
[51] L. Du, et al., Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design, Bioinformatics 34 (4) (2018) 681–683.
[52] A. Untergasser, et al., Primer3–new capabilities and interfaces, Nucleic Acids Res. 40 (15) (2012) e115.
[53] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (1) (2019) 238.
[54] L. Xu, et al., The genetic basis of adaptive evolution in parasitic environment from the *Angiostrongylus cantonensis* genome, PLoS Negl. Trop. Dis. 13 (11) (2019), e0007846.
[55] R. Li, et al., Illumina synthetic long read sequencing allows recovery of missing sequences even in the "finished" *C. elegans* genome, Sci. Rep. 5 (2015) 10814.
[56] S.N. McNulty, et al., Dictyocaulus viviparus genome, variome and transcriptome elucidate lungworm biology and support future intervention, Sci. Rep. 6 (2016) 20316.
[57] R. Laing, et al., The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery, Genome Biol. 14 (8) (2013) R88.
[58] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (1) (2015) 59–60.
[59] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2–approximately maximum-likelihood trees for large alignments, PLoS One 5 (3) (2010), e9490.
[60] D.M. Emms, S. Kelly, STRIDE: species tree root inference from gene duplication events, Mol. Biol. Evol. 34 (12) (2017) 3267–3278.
[61] J.R. Tyson, et al., MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome, Genome Res. 28 (2) (2018) 266–274.
[62] A. Tayyrov, et al., Genetic diversity of the cardiopulmonary canid nematode *Angiostrongylus vasorum* within and between rural and urban fox populations, Infect. Genet. Evol. 87 (2021) 104618.
[63] N. Gillis-Germitsch, et al., Conquering Switzerland: the emergence of *Angiostrongylus vasorum* in foxes over three decades and its rapid regional increase in prevalence contrast with the stable occurrence of lungworms, Parasitology 147 (10) (2020) 1071–1079.
[64] U. Breitenmoser, et al., The final phase of the rabies epizootic in Switzerland, Schweizer Archiv. Fur. Tierheilkunde 142 (8) (2000) 447–454.
[65] D.P. Knox, et al., The nature and prospects for gut membrane proteins as vaccine candidates for Haemonchus contortus and other ruminant trichostrongyloids, Int. J. Parasitol. 33 (11) (2003) 1129–1137.
[66] R. Jefferies, et al., Elucidating the spread of the emerging canid nematode *Angiostrongylus vasorum* between Palaearctic and Nearctic ecozones, Infect. Genet. Evol. 10 (4) (2010) 561–568.
[67] J.P. Demuth, et al., Genome-wide survey of Tribolium castaneum microsatellites and description of 509 polymorphic markers, Mol. Ecol. Notes 7 (6) (2007) 1189–1195.
[68] P. Castagnone-Sereno, et al., Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species *Meloidogyne incognita*, BMC Genomics 11 (2010) 598.
[69] Y.C. Li, et al., Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, Mol. Ecol. 11 (12) (2002) 2453–2465.
[70] P.C. Sharma, A. Grover, G. Kahl, Mining microsatellites in eukaryotic genomes, Trends Biotechnol. 25 (11) (2007) 490–498.
[71] D.L. Sauls, A.S. Wolberg, M. Hoffman, Elevated plasma homocysteine leads to alterations in fibrin clot structure and stability: implications for the mechanism of thrombosis in hyperhomocysteinemia, J. Thromb. Haemost. 1 (2) (2003) 300–306.
[72] M.C. Cury, et al., Hematological and coagulation profiles in dogs experimentally infected with *Angiostrongylus vasorum* (Baillet, 1866), Vet. Parasitol. 104 (2) (2002) 139–149.
[73] R.B. Gasser, et al., Mitochondrial genome of *Angiostrongylus vasorum*: comparison with congeners and implications for studying the population genetics and epidemiology of this parasite, Infect. Genet. Evol. 12 (8) (2012) 1884–1891.
[74] B.R. Ansell, et al., Insights into the immuno-molecular biology of *Angiostrongylus vasorum* through transcriptomics–prospects for new interventions, Biotechnol. Adv. 31 (8) (2013) 1486–1500.