Clinical Research Report

Journal of INTERNATIONAL MEDICAL RESEARCH

Journal of International Medical Research
2019, Vol. 47(9) 4353–4364

DOI: 10.1177/0300060519862069
journals.sagepub.com/home/imr
SAGE

Identification of genetic variations in *Necator americanus* through resequencing by whole genome amplification

Yumei Liu[1,2,*], Hongxuan Liu[1,2,*], Yuefeng Xie[1,2,*], Baohuan Zhang[1,2], Xiaoqian Zou[1,2], Meiling Ou[1,2], Xiaohong Ye[1,2], Yajing Han[1,2], Jing Wu[1,2], Xiaojing Chen[1,2], Shirui Dong[1], Kehui Zhu[1,2], Congcong Guo[2], Penghua Wang[3], Hening Zhai[4,*], Chunxia Jing[2,5,6,*] and Guang Yang[1,5,6,*]

Abstract

Objective: To describe a new strategy for the whole genome resequencing of small parasite samples.
Methods: Whole genome resequencing was based on a multiple displacement amplification (MDA) method. Sequencing reads were aligned with the reference genome, and a Bayesian model was used to calculate genotype probabilities. *De novo* genome assembly was conducted, and single nucleotide polymorphisms (SNPs) were detected. Gene ontology (GO) analysis was used to determine connections between SNPs and genes.
Results: In total, 64.12% of the parasite genome sequence was mapped to *Necator americanus*. fa, and 125,553 SNPs were detected. GO analysis revealed that most SNPs in coding regions were probably associated with common drug targets.

[1]Department of Pathogen Biology, School of Medicine, Jinan University, Guangzhou, China
[2]Department of Epidemiology, School of Medicine, Jinan University, Guangzhou, China
[3]Department of Immunology, UConn Health, Farmington, CT, USA
[4]Endoscopy Center, the First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China
[5]Guangdong Key Laboratory of Environmental Pollution and Health, Jinan University, Guangzhou, Guangdong, China
[6]The Key Laboratory for Virology of Guangzhou, College of Life Science and Technology, Jinan University, Guangzhou, China

*These authors contributed equally to this work.

Corresponding author:
Guang Yang, School of Medicine, Jinan University, Guangzhou, Huangpu Road West 601, Room 844, 510632, China.
Email: tyangguang@email.jnu.edu.cn

Conclusion: These results reveal the feasibility of a new strategy to detect genetic variations of small parasites. This study also provides a proof-of-principle for the molecular classification and epidemiological analysis of other parasites.

Keywords
Necator americanus, resequencing, multiple displacement amplification, gene ontology, single nucleotide polymorphism, parasite

Date received: 4 January 2019; accepted: 17 June 2019

Introduction

Parasitic infection is a major global threat to human health.[1,2] Humans can be infected with around 300 species of helminths and more than 70 species of protozoa,[3] and parasites also cause enormous economic losses.[4] Soil-transmitted helminths are a major type of parasitic infection with public health importance.[5] In 2015, the global burden of infections involving soil-transmitted helminths was estimated at 3.4 million disability-adjusted life years.[6] Exploring the genetic information of these parasites is important for understanding the pathogenesis of parasitic diseases. However, the tiny amounts of DNA isolated are usually inadequate for genetic analysis. Therefore, a whole genome amplification (WGA) strategy may help resolve this problem.

Many WGA methods are currently available, including multiple displacement amplification (MDA) which has been widely used in next-generation sequencing.[7] MDA involves the annealing of random primers to denatured DNA, and strand displacement synthesis at a constant temperature using the φ29 enzyme, which has high processivity, strong strand displacement, and exonuclease activity.[8] MDA is an ideal method for WGA because of its high DNA yield, high fidelity, and low amplification bias.[9]

The genome of the helminth *Necator americanus* (*N. americanus*) was first published in 2014 (Genome ID: 770).[10] In this report, Tang and colleagues characterized many important genes involved in invasion, blood feeding, and development, and identified a list of novel potential drug targets against hookworms.

In the present study, we acquired a single small helminth through endoscopic biopsy. Morphologically, its appearance was consistent with *N. americanus* under an inverted microscope. To obtain its genetic information, we conducted whole genome resequencing (WGRS) but failed to construct DNA libraries using traditional DNA extraction methods from such a limited sample. We therefore developed a novel sequencing strategy that directly amplified the entire genome of the helminth using MDA, and aligned the sequencing data to the first version of the genome of *N. americanus* (Genome ID: 770). We successfully identified 125,553 single nucleotide polymorphisms (SNPs), 2,552 insertion–deletions (InDels), and 22,316 structural variations (SVs) in the sample. We believe that this novel method will provide a reference for molecular classification and epidemiological studies of other small parasites.

Materials and methods

Sampling

The helminth was collected using endoscopic biopsy forceps during an endoscopic

biopsy conducted with an EC-600WM endoscope (Fujifilm, Tokyo, Japan). It was transferred to the laboratory in a cell culture tube filled with 5 mL saline. After taking photographs under an inverted microscope, the helminth sample was transferred to the Beijing Genomics Institute (BGI) for WGRS.

The study was approved by the Ethics Committee of the School of Medicine at Jinan University, China, and was performed strictly in accordance with the Declaration of Helsinki. The patient's written consent was obtained before the endoscopy.

DNA library construction

Because of the small size of the roundworm, its genomic DNA was directly amplified using MDA by BGI. WGA was performed using the REPLI-g Single Cell Kit (Qiagen Inc., Valencia, CA, USA). The sample was first lysed with lysis buffer from the kit, then the DNA was denatured with denaturation buffer for 10 minutes at 65°C. Denaturation was stopped by the addition of neutralization buffer, then a master mix containing buffer and DNA polymerase was added and the isothermal amplification reaction proceeded for 8 hours at 30°C. Amplified DNA was kept at –20°C for long-term storage.

The library was prepared as follows: 1.5 μg DNA was fragmented by ultrasonication, then its quality was tested by gel electrophoresis. DNA was end-repaired by combining with End Repair Mix (New England Biolabs Inc. Beverly, MA, USA), and incubating at 20°C for 30 minutes. DNA was then purified with the QIA quick PCR Purification Kit (Qiagen), after which A-Tailing Mix was added and incubated at 37°C for 30 minutes. DNA was purified with QIA quick PCR Purification Kit (Qiagen), then combined with 60 nmol adapter and 2 μL Ligation Mix before incubating at 20°C for 15 minutes. Adapter-ligated DNA was recovered from a 2% agarose gel and purified with the QIA quick Gel Extraction kit (Qiagen).

Validation of the library

The final library was quantitated by determining the average molecule length using the Agilent 2100 bioanalyzer instrument (Agilent DNA 1000 Reagents; Agilent Technologies Inc., Santa Clara, CA, USA) and via real-time quantitative PCR using TaqMan probes. This was performed using the Step One Plus Real-Time PCR System Upgrade Kit (Applied Biosystems, Foster City, CA, USA) and Step One Plus Real-Time PCR System (Applied Biosystems) according to the manufacturer's instructions.

Whole genome resequencing and SNP calling

The library was sequenced on an Illumina HiSeq X10 system at the BGI in 2×150 paired-end mode. For quality control, we removed reads with adaptors and reads of low quality where more than half of the base qualities were less than 5 using SOA Pnuke software.[11] The sequencing reads were aligned with the reference genome *N. americanus fa* using SOAP2 software.[12] The reference genome (*N. americanus* fa) and annotation were obtained from the Worm Base Parasite database (http://parasite.wormbase.org/Necator_americanus_prjna72135/Info/Index/; Genome ID: 770).[10] The sequencing depth and coverage compared with the reference genome were calculated based on the alignment. SNPs in the sequenced genome were detected using SOAPsnp software.[13] Based on the alignment results, and considering the analysis of data characteristics, sequencing quality, and other experimental effects, a Bayesian model was used to calculate genotype probabilities. The genotype with the highest probability was selected as the genotype of the sequenced

individual at the specific locus, and a quality value was designated to denote the accuracy of the genotype. We remapped the reads using BWA software and recalled InDels by samtools and bcftools.[14] BreakDancer was used to identify SVs.[15]

De novo assembly and species distribution based on all scaffolds

To calculate the species distribution based on all scaffolds, sequences were assembled using Velvet software, which is the most commonly used genome assembly tool.[16] First, the ErrorCorrection tool was used to correct the data to remove low-frequency sequencing errors after transferring original image data into sequence data stored in FASTQ format through Base Calling software. Next, Kmer analysis was used to estimate the genome size and sequence multiples. Assembled scaffolds were then aligned onto the RefSeq non-redundant protein database, and BLAST results were analyzed via MEGAN software.[17]

Gene ontology analysis

To determine the underlying connections between SNPs and genes, Gene Ontology (GO) analysis was conducted for functional annotations of these SNPs. Enrichment analysis was performed by Fisher's exact test using the Database for Annotation, Visualization and Integrated Discovery (https://david.ncifcrf.gov) by classifying them into three domains: biological process, cellular component, and molecular function.[18] The cut-off P-value was < 0.05, which was considered to be significant.

Results

Patient background and sample description

A 54-year-old woman attended the First Affiliated Hospital of Jinan University on July 20th, 2016, complaining of arthritis for more than 1 year that had worsened in the past 2 months. On physical examination, she showed no obvious symptoms or signs, and her vital signs were stable. Her levels of red blood cells, hemoglobin, hematocrit, mean corpuscular hemoglobin, and ferrum were low, whereas platelets and the blood sedimentation rate were elevated, which suggested iron deficiency (anemia). Eosinophils, the most important indicator for parasitic infection, were also elevated, but a stool examination was negative for parasite eggs.

A 10-mm-long slender hyperemic round-cylindrical helminth was found via gastroduodenoscopy in the duodenal bulb (Figure 1a). To avoid further suffering to the patient, we did not obtain more helminths by further endoscopy. After being collected by the endoscopic biopsy forceps, the sample was put in a 24-well plate (Figure 1b). The helminth was relatively slender in the front section and slightly curved in the dorsal section. Under the inverted microscope, a well-developed buccal cavity was seen to contain a pair of cutting plates (Figure 1d). The end of the helminth was conical, indicating that it was a female worm (Figure 1c). Its appearance was consistent with that of *N. americanus*. Because it was isolated from the First Affiliated Hospital of Jinan University, we named the strain JNU-1.

Sequencing, mapping, and SNP annotation

A total of 20.38 G bases were sequenced of the raw data, and ultimately 19.3 G of clean data were obtained. Q20 and Q30 reflect the accuracy of the sequenced bases, and denote base error rates of 1% and 0.1%, respectively. Our Q20 rate of clean data was 97.41%, indicating that the data quality was very high. However, mapping of the reads to the human genome (hg19) resulted in only 0.85% (552,556/64,757,726) being

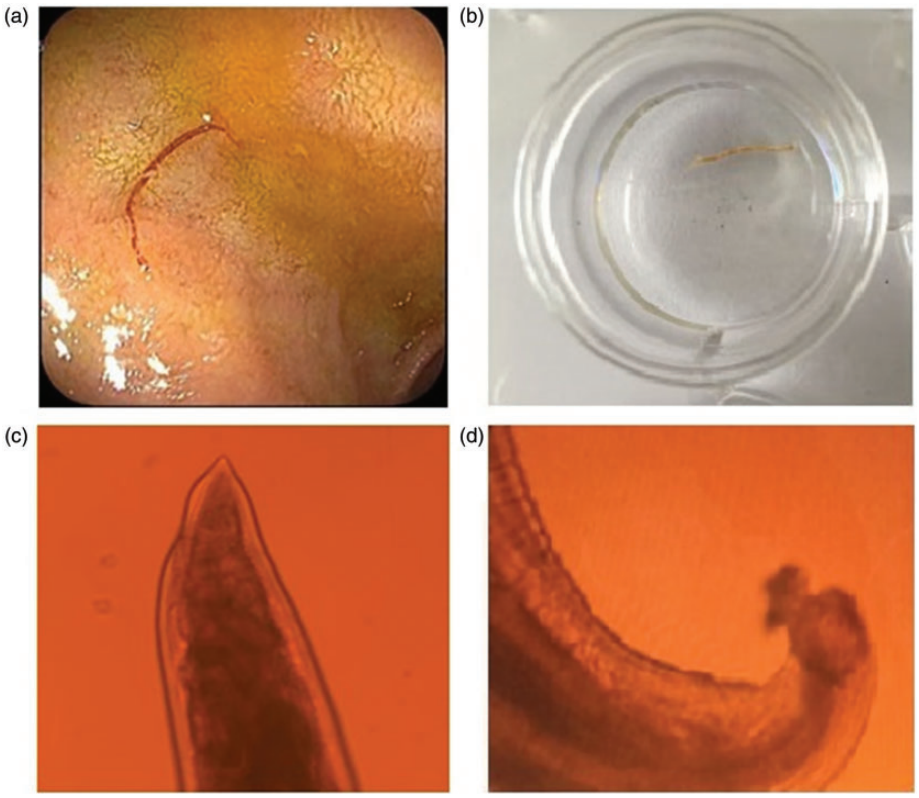

Figure 1. Morphological characteristics of strain JNU-1. a: JNU-1 strain in the duodenal bulb. b: JNU-1 strain in the 24-well plate. c: End of the JNU-1 strain under the inverted microscope. d: Cutting plates of the JNU-1 strain under the inverted microscope.

aligned (reads with 80% of the bases matching the human genome are considered to be human contamination reads). The proportion of contaminations of the host genome was relatively low.

According to basic morphological features, we used *N. americanus* fa as the reference genome for this project. The genome size of *N. americanus* fa is 245,853,160 bp and the effective size is 208,173,610 bp (not including N base in the reference). The mapping rate was 64.12%, and the final effective mapping depth was 59.61. Paired-end sequencing libraries with an insert size of approximately 500 bp were constructed for the sample. We plotted the distribution of the insert size based on the alignment

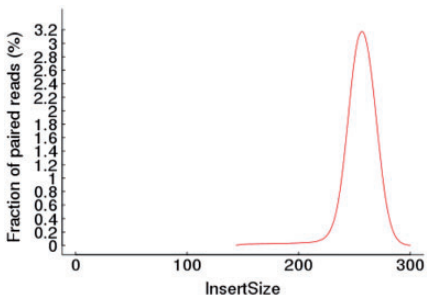

Figure 2. The distribution of insert size based on the result of alignment.

results (Figure 2). To determine if the WGA had a sequence bias, we assessed the distribution of reads mapped to the reference genome and the distribution of

coverage of 10 chromosomes (Figure 3). There were no significant differences in the mapped reads distributions on each chromosome.

In total, 125,553 SNPs in 42,131 genes were found in the *N. americanus* genome. Of these, 6289 (5.01%) were in coding regions. Among these, 3933 synonymous and 2356 nonsynonymous SNPs were annotated. Based on these annotations, we identified some SNPs that could potentially influence gene functions which we called potential effect SNPs (Table 1). Of these, 36 SNPs were predicted to induce premature stop codons, 21 were predicted to change the start codon to a non-start codon, 11 were predicted to remove the annotated stop codons resulting in longer open reading frames, and 35 were predicted to disrupt splicing donor or acceptor sites. Some of these potential SNPs might have major impacts on biological processes. For instance, some SNPs predicted to induce premature

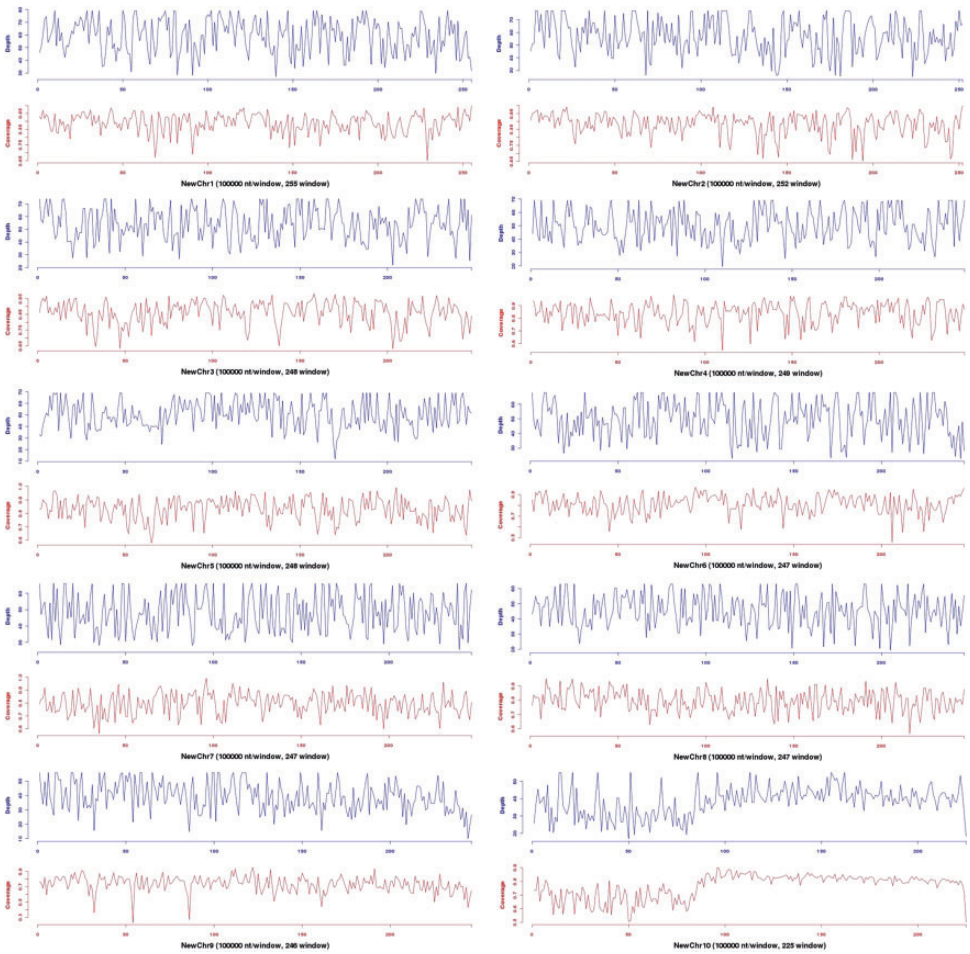

Figure 3. The distribution of the reads mapped to reference genome. X-axis represents the window numbers. Y-axis represents the coverage or depth. Coverage represents the ratio of the area covered by reads to the length of each window. A coverage close to 1 indicates that the reads are evenly distributed. Depth means the average sequence depth in each window. Nt, nucleotide.

Table 1. The list of potential effect SNPs.

Chromosome	Position	Gene ID	Name	AA Mutation
NW_013562409.1	65040	742	XP_013308799.1	CAG<->TAG
NW_013562027.1	8965	2664	XP_013306871.1	TGG<->TAG
NW_013561991.1	12602	2856	XP_013306683.1	TAC<->TAA
NW_013561981.1	73695	2961	XP_013306563.1	CAA<->TAA
NW_013561750.1	37908	3768	XP_013305757.1	CGA<->TGA
NW_013561523.1	892311	4699	XP_013304922.1	CAG<->TAG
NW_013561477.1	109313	4818	XP_013304712.1	TGT<->TGA
NW_013561310.1	788985	5542	XP_013304039.1	GGA<->TGA
NW_013561123.1	77954	6080	XP_013303452.1	CGA<->TGA
NW_013561101.1	163698	6135	XP_013303371.1	TGG<->TGA
NW_013560795.1	1050	6989	XP_013302549.1	CGA<->TGA
NW_013560786.1	354745	7021	XP_013302505.1	TGC<->TGA
NW_013560572.1	139025	7615	XP_013301923.1	CAG<->TAG
NW_013560466.1	126603	7790	XP_013301701.1	GGA<->TGA
NW_013559795.1	27264	9340	XP_013300184.1	CAG<->TAG
NW_013559785.1	42716	9493	XP_013300046.1	GAA<->TAA
NW_013559784.1	146561	9512	XP_013300037.1	GAG<->TAG
NW_013559709.1	5094	10851	XP_013298669.1	TCA<->TAA
NW_013559689.1	575652	11164	XP_013298405.1	CGA<->TGA
NW_013559653.1	87276	11650	XP_013297892.1	AAG<->TAG
NW_013559590.1	154897	12652	XP_013296866.1	GAG<->TAG
NW_013559531.1	103834	13406	XP_013296146.1	AAA<->TAA
NW_013559528.1	104542	13424	XP_013296113.1	CGA<->TGA
NW_013558779.1	149543	13971	XP_013295537.1	GAA<->TAA
NW_013558779.1	205375	13978	XP_013295544.1	TAT<->TAG
NW_013554582.1	71	15129	XP_013294409.1	TTA<->TAA
NW_013554317.1	710	15190	XP_013294348.1	TAC<->TAG
NW_013552452.1	33832	15654	XP_013293882.1	TGG<->TGA
NW_013551777.1	2561	15830	XP_013293697.1	TAC<->TAA
NW_013551390.1	849	15917	XP_013293621.1	CAA<->TAA
NW_013551101.1	29418	16041	XP_013293493.1	CAA<->TAA
NW_013551101.1	30221	16041	XP_013293493.1	TGG<->TGA
NW_013550967.1	221434	16927	XP_013292606.1	GGA<->TGA
NW_013550919.1	38067	17191	XP_013292230.1	TGG<->TGA
NW_013550772.1	1898	18154	XP_013291357.1	TTA<->TAA
NW_013562348.1	55633	1032	XP_013308510.1	ATG<->GTG
NW_013562335.1	40527	1176	XP_013308367.1	ATG<->ATT
NW_013562312.1	13932	1318	XP_013308220.1	ATG<->ATA
NW_013561997.1	5358	2808	XP_013306729.1	ATG<->ATT
NW_013561842.1	9873	3513	XP_013306024.1	ATG<->ATC
NW_013561689.1	17800	4029	XP_013305510.1	ATG<->GTG
NW_013561551.1	166514	4522	XP_013305005.1	ATG<->ATC
NW_013561437.1	8390	4927	XP_013304611.1	ATG<->ACG
NW_013561041.1	151223	6338	XP_013303207.1	ATG<->GTG
NW_013560947.1	21475	6590	XP_013302937.1	ATG<->TTG
NW_013559734.1	156583	10308	XP_013299150.1	ATG<->ACG

(continued)

Table 1. Continued.

Chromosome	Position	Gene ID	Name	AA Mutation
NW_013559703.1	154427	10929	XP_013298617.1	ATG<->ATA
NW_013559680.1	12453	11243	XP_013298287.1	ATG<->AAG
NW_013559652.1	130195	11663	XP_013297880.1	ATG<->CTG
NW_013559623.1	780755	12105	XP_013297443.1	ATG<->ATC
NW_013556537.1	419	14632	XP_013294906.1	ATG<->ATA
NW_013555186.1	1123	14962	XP_013294576.1	ATG<->AGG
NW_013553526.1	31	15398	XP_013294140.1	ATG<->ATA
NW_013550952.1	290570	17045	XP_013292512.1	ATG<->ATC
NW_013550860.1	4391	17683	XP_013291813.1	GTA<->GCA
NW_013550693.1	36717	18751	XP_013290789.1	ATG<->ATA
NW_013562267.1	189514	1514	XP_013308023.1	TAG<->TGG
NW_013561551.1	287713	4541	XP_013305024.1	TAG<->TGG
NW_013561299.1	225108	5591	XP_013303951.1	TGA<->CGA
NW_013560708.1	746	7229	XP_013302309.1	TAA<->TAC
NW_013559914.1	98071	9121	XP_013300416.1	TAA<->TTA
NW_013559511.1	515304	13740	XP_013295835.1	TAG<->TAC
NW_013558462.1	1471	14098	XP_013295440.1	TGA<->CGA
NW_013552930.1	92	15558	XP_013293971.1	TAA<->CAA
NW_013550973.1	38612	16888	XP_013292648.1	TGA<->AGA
NW_013550910.1	30541	17368	XP_013292168.1	TAA<->AAA
NW_013550673.1	28643	18850	XP_013290684.1	TAA<->TCA

stop codons were distributed in the coding region of oxidoreductase proteins, ligand-binding proteins, and nucleotide–sugar transporters. Some SNPs were located in the coding region of endonucleases, calpain family cysteine proteases, and lectin C-type domain proteins, changing the start codon to a non-start codon. Other SNPs were related to 50S ribosomal protein, brix domain protein, and triacylglycerol lipase. We also found 35 SNPs located at splicing sites.

Gene ontology analysis based on SNPs

To better understand the genetic function of mutations in the coding regions of JNU-1, GO analysis was performed of SNP-containing genes in coding regions. In sequential order of enrichment, the top five categories with the most SNPs of all GO terms were: adenosine triphosphate (ATP) binding, zinc ion binding, collagen trimers, G-protein coupled receptor (GPCR) signaling, and structural constituents of the cuticle (Figure 4), which are major molecular drug targets. Other SNPs were also associated with intracellular signal transduction, transporter activity, aspartic-type endopeptidase activity, and microtubule-based movement.

GO terms were divided into three parts: biological process, cellular component, and molecular function. The number of SNPs in each part is shown in Figure 4. Approximately 78% of SNP-containing genes were associated with molecular function. The greatest number of SNPs was associated with intracellular signal transduction, collagen trimers, and ATP binding.

Assembly result

We attempted *de novo* assembly of the JNU-1 genome to identify scaffolds with potential function, but lacked sufficient reads for this. Functional analysis based on

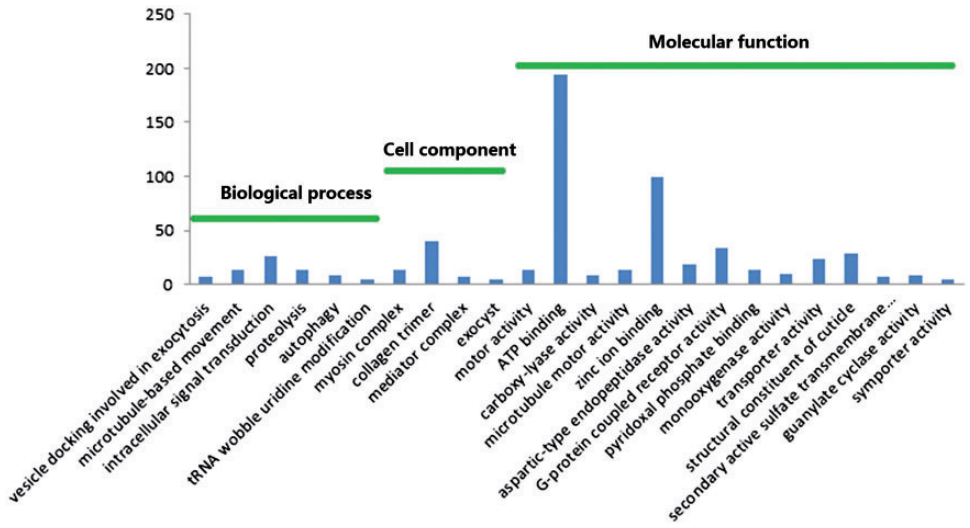

Figure 4. Gene ontology (GO) analysis on SNP-containing genes in coding regions. GO terms were divided into three parts: biological process, cellular component, and molecular function.

scaffolds assembled a total of 290,018 sequences of JNU-1 containing 386,897,339 bp, with an N rate of 0.01519%. The GC content was 39.89%. The maximum length of the sequence was 30,923 bp, and the median was 940 bp with an average length of 1334 bp. Regarding length, 50% of the genome was contained within sequences with a minimum length of 1758 bp, and 90% of the genome was in sequences with a minimum length of 636 bp.

InDels and SVs

We remapped the reads to identify InDels and SVs. We found 2552 InDels, including 726 frameshift deletions, 1011 frameshift insertions, 409 nonframeshift deletions, and 405 nonframeshift insertions. We also found 22,316 SVs, including 8393 inter-chromosomal translocations, 11,711 inversions, and 2212 intra-chromosomal translocations.

Discussion

Because of the limited size of small parasites, constructing libraries by extracting DNA using traditional methods often leads to failure. In the present study, we acquired a helminth sample but construction of the DNA library for WGRS was unsuccessful. Therefore, we developed a novel amplification technique to probe the genome of this strain.

The MDA method was previously shown to be feasible for accurately amplifying a nematode parasite genome in an unbiased manner for the genomic characterization of filarial parasites.[19] The technique was also efficient in genotyping *Schistosoma mansoni miracidia*.[20] Therefore, MDA is a reliable, efficient, and accurate procedure for WGA. Here, based on the MDA method, we successfully acquired a large amount of genomic information about the JNU-1 strain, which provided materials for identifying numerous SNPs.

The top five categories with the most SNPs of all GO terms were ATP binding, zinc ion binding, collagen trimers, GPCR signaling, and structural constituents of the cuticle. ATP provides energy and phosphate for the helminth,[21] while the collagen

trimer is a protein complex consisting of three collagen chains assembled into a left-handed triple helix. These trimers typically assemble into higher order structures in the helminth.[22] GPCR senses an extracellular signal and transmits the signal across the cell membrane by activating an associated G-protein. It also promotes the exchange of guanosine diphosphate for guanosine triphosphate on the alpha subunit of a heterotrimeric G-protein complex, which is important for the cuticle structure and cellular integrity.[23] Together, these genes are important for the development and biological process of the helminth.

Parasite drug resistance has been documented in many studies.[24] Benzimidazole derivatives such as albendazole and mebendazole are the most common drugs used to control hookworm infection. They irreversibly disrupt glucose absorption, leading to kinetic disorder and eventually death.[25] We detected a large number of SNPs in coding regions associated with ATP binding. This indicated that these SNPs influence proteins that bind to ATP, such as ribonucleotide adenosine which carries three phosphate groups esterified to the sugar moiety. Because ATP is the major cellular source of energy and phosphate, we propose that long-term use of benzimidazole derivatives has induced the development of SNPs that improve the ATP binding ability of the hookworm. This may be genetic evidence of drug resistance to benzimidazole derivatives.

Other SNPs were associated with zinc ion binding, GPCRs, intracellular signal transduction, and collagen trimers. According to a previous study, more than half (53%) of all current drug targets consist of GPCRs, nuclear receptors, ligand-gated ion channels, kinases, and voltage-gated ion channels.[26] Among them, it is estimated that one-third of currently approved drug targets are GPCRs because these are highly amenable to modulation by pharmaceuticals.[27] Although SNPs are only small variations, they have a large impact on genomes and biological traits, especially SNPs located in coding regions. Although the impact of the SNPs on gene functions was not investigated in this study, it will be interesting to pursue this in future endeavors because splicing SNPs may influence pleiotropic biological processes.[28,29] We are aware that more samples are needed to confirm the relationship between the identified SNPs with pathogenicity or other effects, and additional sample collection has been instigated. We also note that because the published *N. americanus* genome sequence is of low quality, JNU-1 may differ from the published *N. americanus* strain. Therefore, the SNPs and other variants may be falsely positive. Additionally, some sequences with potential functions may have been missed.

In summary, we uncovered a distinct strain of *N. americanus* and innovatively explored potential mutations with the help of MDA technology. Our results demonstrate the feasibility of MDA technology for genome resequencing. Additionally, they shed light on possible future investigations of many aspects of small parasites, such as species evolution and epidemiology. This will be beneficial not only to the study of multicellular helminths but also to that of unicellular protozoans. We believe that with the reduction of sequencing costs, this strategy will be widely accepted and applicable to species identification and the sifting of genetic variations.

Acknowledgements

We would like to thank Qihan Wu and Minjie Xu from the Key Laboratory of Contraceptives and Devices, Shanghai Institute of Planned Parenthood Research, Institute of Reproduction and Development, Fudan University for their help analyzing sequencing data.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

This work was supported in part by the Training Program of the Major Research Plan of the National Natural Science Foundation of China (Grant numbers: 91543132, 81630025, 81541070, 81101267, and 30901249), Guangdong Natural Science Foundation (Grant numbers: 10151063201000036, S2011010002526, and 2016A030313089), Guangdong Province Medical Research Foundation (Grant numbers: A2014374 and A2015310), the Project from Jinan University (Grant numbers: 21612426, 21615426, JNUPHPM2016001, and JNUPHPM2016002), and the Traditional Chinese Medicine Bureau Of Guangdong Province (20181071).

ORCID iD

Guang Yang https://orcid.org/0000-0003-1688-515X

References

1. Siddiqui ZA. An overview of parasitic infections of the gastro-intestinal tract in developed countries affecting immunocompromised individuals. *J Parasit Dis* 2017; 41: 621–626.
2. Donnelly S, Huston WM, Johnson M, et al. Targeting the master regulator mTOR: a new approach to prevent the neurological of consequences of parasitic infections? *Parasit Vector* 2017; 10: 581.
3. Cruz AA, Cooper PJ, Figueiredo CA, et al. Global issues in allergy and immunology: parasitic infections and allergy. *J Allergy Clin Immunol* 2017; 140: 1217–1228.
4. O'Sullivan JDB, Behnsen J, Starborg T, et al. X-ray micro-computed tomography (muCT): an emerging opportunity in parasite imaging. *Parasitology* 2018; 145: 848–854.
5. Bharti B, Bharti S and Khurana S. Worm infestation: diagnosis, treatment and prevention. *Indian J Pediatr* 2017; 85: 1017–1024.
6. DALYs GBD and Collaborators H. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; 388: 1603–1658.
7. Binga EK, Lasken RS and Neufeld JD. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* 2008; 2: 233–241.
8. Spits C, Le Caignec C, De Rycke M, et al. Optimization and evaluation of single-cell whole-genome multiple displacement amplification. *Hum Mutat* 2006; 27: 496–503.
9. Wang W, Ren Y, Lu Y, et al. Template-dependent multiple displacement amplification for profiling human circulating RNA. *Biotechniques* 2017; 63: 21–27.
10. Tang YT, Gao X, Rosa BA, et al. Genome of the human hookworm Necator americanus. *Nat Genet* 2014; 46: 261–269.
11. Chen Y, Chen Y, Shi C, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 2018; 7: 1–6.
12. Luo R, Liu B, Xie Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 2012; 1: 18.
13. Azam S, Rathore A, Shah TM, et al. An integrated SNP mining and utilization (ISMU) pipeline for next generation sequencing data. *PLoS One* 2014; 9: e101754.
14. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754–1760.
15. Chen K, Wallis JW, McLellan MD, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009; 6: 677–681.
16. Zerbino DR and Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; 18: 821–829.
17. Huson DH, Tappu R, Bazinet AL, et al. Fast and simple protein-alignment-guided assembly of orthologous gene families from microbiome sequencing reads. *Microbiome* 2017; 5: 11.

18. Jiao X, Sherman BT, Huang da W, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 2012; 28: 1805–1806.
19. McNulty SN, Weil GJ, Heinz M, et al. Brugia malayi: whole genome amplification for genomic characterization of filarial parasites. *Exp Parasitol* 2008; 119: 256–263.
20. Valentim CL, LoVerde PT, Anderson TJ, et al. Efficient genotyping of Schistosoma mansoni miracidia following whole genome amplification. *Mol Biochem Parasitol* 2009; 166: 81–84.
21. Ferreira-Neto HC, Yao ST and Antunes VR. Purinergic and glutamatergic interactions in the hypothalamic paraventricular nucleus modulate sympathetic outflow. *Purinergic Signal* 2013; 9: 337–349.
22. Lohi J, Korhonen M, Leivo I, et al. Expression of type IV collagen alpha1(IV)-alpha6(IV) polypeptides in normal and developing human kidney and in renal cell carcinomas and oncocytomas. *Int J Cancer* 1997; 72: 43–49.
23. Woodard GE, Jardin I, Berna-Erro A, et al. Regulators of G-protein-signaling proteins: negative modulators of G-protein-coupled receptor signaling. *Int Rev Cell Mol Biol* 2015; 317: 97–183.
24. Diawara A, Halpenny CM, Churcher TS, et al. Association between response to albendazole treatment and beta-tubulin genotype frequencies in soil-transmitted helminths. *PLoS Negl Trop Dis* 2013; 7: e2247.
25. Vinaud MC, Ferreira CS, Lino Junior Rde S, et al. Taenia crassiceps: energetic and respiratory metabolism from cysticerci exposed to praziquantel and albendazole in vitro. *Exp Parasitol* 2008; 120: 221–226.
26. Overington JP, Al-Lazikani B and Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* 2006; 5: 993–996.
27. Purcell RH and Hall RA. Adhesion G protein-coupled receptors as drug targets. *Annu Rev Pharmacol Toxicol* 2018; 58: 429–449.
28. Mesrian Tanha H, Rahgozar S and Mojtabavi Naeini M. ABCC4 functional SNP in the 3' splice acceptor site of exon 8 (G912T) is associated with unfavorable clinical outcome in children with acute lymphoblastic leukemia. *Cancer Chemother Pharmacol* 2017; 80: 109–117.
29. Matsuura K, Sawai H, Ikeo K, et al. Genome-wide association study identifies TLL1 variant associated with development of hepatocellular carcinoma after eradication of hepatitis C virus infection. *Gastroenterology* 2017; 152: 1383–1394.