

Article

Making the Most of Its Short Reads: A Bioinformatics Workflow for Analysing the Short-Read-Only Data of *Leishmania orientalis* (Formerly Named *Leishmania siamensis*) Isolate PCM2 in Thailand

Pornchai Anuntasomboon ^{1,2}, Suradej Siripattanapipong ³, Sasimanas Unajak ⁴ , Kiattawee Choowongkamon ⁴ , Richard Burchmore ⁵ , Saovanee Leelayoova ⁶, Mathirut Mungthin ⁶ and Teerasak E-kobon ^{1,2,*} 

- ¹ Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
² Omics Center for Agriculture, Bioresources, Food, and Health, Kasetsart University (OmiKU), Bangkok 10900, Thailand
³ Department of Microbiology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
⁴ Department of Biochemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
⁵ Glasgow Polyomics, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK
⁶ Department of Parasitology, Phramongkutklao College of Medicine, Bangkok 10400, Thailand
* Correspondence: teerasak.e@ku.th; Tel.: +66-812-85-4672



Citation: Anuntasomboon, P.; Siripattanapipong, S.; Unajak, S.; Choowongkamon, K.; Burchmore, R.; Leelayoova, S.; Mungthin, M.; E-kobon, T. Making the Most of Its Short Reads: A Bioinformatics Workflow for Analysing the Short-Read-Only Data of *Leishmania orientalis* (Formerly Named *Leishmania siamensis*) Isolate PCM2 in Thailand. *Biology* **2022**, *11*, 1272. <https://doi.org/10.3390/biology11091272>

Academic Editor: Francesca Mancianti

Received: 26 July 2022

Accepted: 24 August 2022

Published: 26 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Leishmaniasis is a parasitic disease caused by flagellated protozoa of the genus *Leishmania*. Multiple genome sequencing platforms have been employed to complete *Leishmania* genomes at the expense of high cost. This study proposes an integrative bioinformatic workflow for assembling only the short-read data of *Leishmania orientalis* isolate PCM2 from Thailand and produce an acceptable-quality genome for further genomic analysis. This workflow gives extensive information required for identifying strain-specific markers and virulence-associated genes useful for drug and vaccine development before a more exhaustive and expensive investigation.

Abstract: Background: *Leishmania orientalis* (formerly named *Leishmania siamensis*) has been neglected for years in Thailand. The genomic study of *L. orientalis* has gained much attention recently after the release of the first high-quality reference genome of the isolate LSCM4. The integrative approach of multiple sequencing platforms for whole-genome sequencing has proven effective at the expense of considerably expensive costs. This study presents a preliminary bioinformatic workflow including the use of multi-step de novo assembly coupled with the reference-based assembly method to produce high-quality genomic drafts from the short-read Illumina sequence data of *L. orientalis* isolate PCM2. Results: The integrating multi-step de novo assembly by MEGAHIT and SPAdes with the reference-based method using the *L. enriettii* genome and salvaging the unmapped reads resulted in the 30.27 Mb genomic draft of *L. orientalis* isolate PCM2 with 3367 contigs and 8887 predicted genes. The results from the integrated approach showed the best integrity, coverage, and contig alignment when compared to the genome of *L. orientalis* isolate LSCM4 collected from the northern province of Thailand. Similar patterns of gene ratios and frequency were observed from the GO biological process annotation. Fifty GO terms were assigned to the assembled genomes, and 23 of these (accounting for 61.6% of the annotated genes) showed higher gene counts and ratios when results from our workflow were compared to those of the LSCM4 isolate. Conclusions: These results indicated that our proposed bioinformatic workflow produced an acceptable-quality genome of *L. orientalis* strain PCM2 for functional genomic analysis, maximising the usage of the short-read data. This workflow would give extensive information required for identifying strain-specific markers and virulence-associated genes useful for drug and vaccine development before a more exhaustive and expensive investigation.

Keywords: sequence read analysis; *Leishmania orientalis*; leishmaniasis; genomics; bioinformatics

1. Introduction

Leishmaniasis is a significant vector-borne zoonotic disease caused by flagellated protozoans of the order Trypanosomatidae and genus *Leishmania*. The disease occurs in over 98 countries, including countries in Asia, Africa, America, and Europe [1–3]. The number of new patients has increased annually to around 1.7 million people yearly [4]. Human leishmaniasis is classified into three forms: cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL), and visceral leishmaniasis (VL). VL is the most severe form, affecting the liver, lymph node, and spleen. If the patient does not receive appropriate treatment, the fatality rate can be 100% within two years [5,6]. The CL, one of the most common forms of leishmaniasis, shows skin lesions and ulcerates at the site bitten by sandflies [7], while the MCL is a rare form associated with *Leishmania braziliensis* [8]. According to the Centers for Disease Control and Prevention (CDC), approximately 21 of the 30 *Leishmania* species cause human infection.

Leishmaniasis in Thailand was considered as an imported disease before 1999 [9,10]. However, after 1999, a new species of *Leishmania* was identified from a Thai patient with visceral leishmaniasis, named *Leishmania siamensis* [11] and later renamed *Leishmania orientalis* [12]. Several *L. orientalis* isolates were discovered in Thailand, including CU1, PCM1, PCM4, PCM5, and the Trang lineage isolate PCM2 [13]. *L. orientalis* infection was also found in other countries, such as the CL cases in horses in Europe [14,15] and Florida, USA [16]. Livestock, such as donkeys, cows, sheep, goats, and camels, can be a reservoir of leishmaniasis and could spread the infection more easily [17–19]. The sandfly is considered a significant vector of leishmaniasis in Thailand, predominantly *Sergentomyia gemmea* in the northern and southern areas of the country [20]. However, the situation of *Leishmaniasis* in Thailand is also challenging to estimate because the patients often show no symptoms. Monitoring the spread and mutation of *L. orientalis* will prepare for the future outbreak and emergence of new virulent strains. The previous study established the prevalence of coinfection with *Leishmania* among Thai HIV patients who attended the HIV clinic in Trang province. *Leishmania* infection was detected in nearly one-fourth of the 724 individuals using either direct agglutination test (DAT) or ITS1-PCR tests, and the dominant species identified in this investigation were *Leishmania martiniquensis* and *L. orientalis* [21]. The coinfection of *Leishmania* and HIV enhances the host immunological degradation, leading to treatment failure, a high incidence of recurrence, and a high fatality rate [22]. For leishmaniasis treatment, failure has been documented in patients treated with most anti-leishmaniasis drugs [23–30]. Amphotericin B (AmB) remains an effective drug with mitigated toxic effects after liposomal formulation [24].

The *Leishmania* genomes have shown several intriguing aspects, including (1) lack of large subtelomeric regions; (2) absence of C-5 DNA methylation but having a hypermodified nucleobase or base J which is unique to the kinetoplasts; and (3) unusual gene regulation, such as the absence of introns and trans-splicing of mRNAs from ~200 polycistronic transcription units (PTUs), compared with other trypanosomatids [31–34]. Chromosome numbers also vary among the *Leishmania* species: 36 chromosomes in *Leishmania infantum*, *Leishmania donovani*, and *Leishmania major* [35]; 35 chromosomes in *L. braziliensis* [33,36]; and 34 chromosomes in *Leishmania mexicana* [36]. Some cellular components of *Leishmania* are associated with virulence, i.e., glycoinositolphospholipids (GIPLs) [37], lipophosphoglycans (LPGs) [38], proteophosphoglycans (PPGs) [39], and the 11 kDa kinetoplastid membrane protein (KMP-11) [40]. These components contribute to the infection, invasion, and establishment of the mammalian host. Although the precise effect of these *Leishmania* components on clinical symptoms in mammalian hosts is unclear, there is evidence that these components affect *Leishmania*–host immune cell interactions [41].

Despite concerns about the genetic variation of *L. orientalis* and their public health impacts, genomic information of multiple *L. orientalis* isolates has gained attention. To date, three major genome sequencing platforms (Illumina, PacBio, and Oxford Nanopore) have been used to create the complete genomic data of several *Leishmania* species, including the genome of *L. major* strain Friedlin [31,42–44], *L. infantum* strain JPCM5 [33,42],

L. martiniquensis strain LSCM1 [45–47], and a recent *L. orientalis* strain LSCM4 isolated from the northern province of Thailand [46–48], in exchange for considerable cost invested in the genome project. Questions have arisen on whether the new *Leishmania* species shall have their whole genomes decoded using all techniques at first glance. When there are several closely related genomes available, the assistance of bioinformatic analysis on the draft genome derived from only the short-read genome sequencing method would be enough to answer basic genomic questions on virulence, pathogenesis, and drug resistance. Therefore, this research aims to design a bioinformatics workflow for analysing the whole-genome short-read data of *L. orientalis* strain PCM2 isolated from the southern province of Thailand by optimising the assembly methods and maximising the data output. This workflow would give preliminary information required for further plans on the *Leishmania* genome sequencing with other methods.

2. Materials and Methods

2.1. Culture of *Leishmania orientalis* Isolate PCM2

Leishmania orientalis isolate PCM2 was maintained and provided by the Department of Parasitology, Phramongkutklao College of Medicine, Thailand. The promastigotes were grown at 26 °C in RPMI 1640-modified with 13.3 mM glutamine, 2.5 mM arginine, 0.3 mM cysteine, 1.7 mM glutamate, 62.1 mM proline, 0.6 mM ornithine, 3.8 mM glucose, 2.2 mM fructose, 5.1 mM malate, 2.8 mM α -ketoglutarate, 0.5 mM fumarate, 0.5 mM succinate, 25 mM HEPES, 50 μ g/mL gentamicin, 2 \times MEM vitamins (Gibco, Grand Island, NY, USA), and 20% heat-inactivated fetal bovine serum (HIFBS, Gibco, Grand Island, NY, USA).

2.2. Genomic DNA Preparation

Genomic DNAs were prepared from a late logarithmic phase of promastigotes. The promastigote pellet was washed in ultrapure water and suspended in 1 mL of lysis buffer (10 mM Tris, 10 mM KCl, 10 mM MgCl₂, 0.5 M NaCl, 2 mM EDTA, and 0.5% SDS) and 20 μ L of Proteinase K solution (20 mg/mL). The samples were incubated at 56 °C for 30 min; then chloroform:isoamyl alcohol (24:1) was added one time to the sample volume, and the samples were gently shaken vigorously for 10 min. The samples were centrifuged at 10,000 rpm at room temperature for 10 min, and the upper aqueous phase was collected. RNase (20 mg/mL) solution was added for 10 μ L and incubated at room temperature for 3 min. After RNase treatment, one-time chloroform:isoamyl alcohol (24:1) was added, and the samples gently shaken vigorously before centrifugation at 10,000 rpm at room temperature for 10 min. The upper aqueous phase was collected, and DNAs in the upper aqueous phase were precipitated in 200 μ L of 4 M ammonium acetate and 800 μ L absolute ethanol at -70 °C overnight. The precipitated samples were centrifuged at 10,000 rpm at 4 °C for 10 min and washed with 70% ethanol twice. DNA samples were air-dried at room temperature for 30 min and suspended in TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA). The DNA quality and quantity were assayed by measuring absorbance at 260/280 nm using the Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) and at the absorbance of 260 nm. The genomic integrity was analysed by 1% agarose gel electrophoresis. The samples were kept at -70 °C before proceeding to the genome sequencing.

2.3. Quality Check and Processing of the Raw Sequence Reads

A paired-end read library (101 bp) was constructed for the whole-genome sequencing using the Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA). The quality of the raw sequence reads was checked by FastQC Version 0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) (accessed on 7 June 2021) [49], and the raw sequence reads were processed through the filtering and trimming with the cut-off value of 20 using BBDuk in the BBTools pipeline (sourceforge.net/projects/bbmap/) (accessed on 7 June 2021).

2.4. Bioinformatic Workflow for Assembling the Genomic Reads of *Leishmania orientalis* Isolate PCM2 by Hybrid Methods and Salvaging the Unmapped Reads

This study designed a bioinformatics workflow for analysing the short-read genomic data of *Leishmania orientalis* isolate PCM2 as a preliminary overview which could be helpful to further decisions on the incorporation of additional sequencing platforms, as displayed in Figure 1. Initially, the filtered and trimmed reads (Data A in Figure 1) were processed separately through (1) de novo assembly and (2) reference-based assembly and the de novo assembly of the remaining unmapped reads.

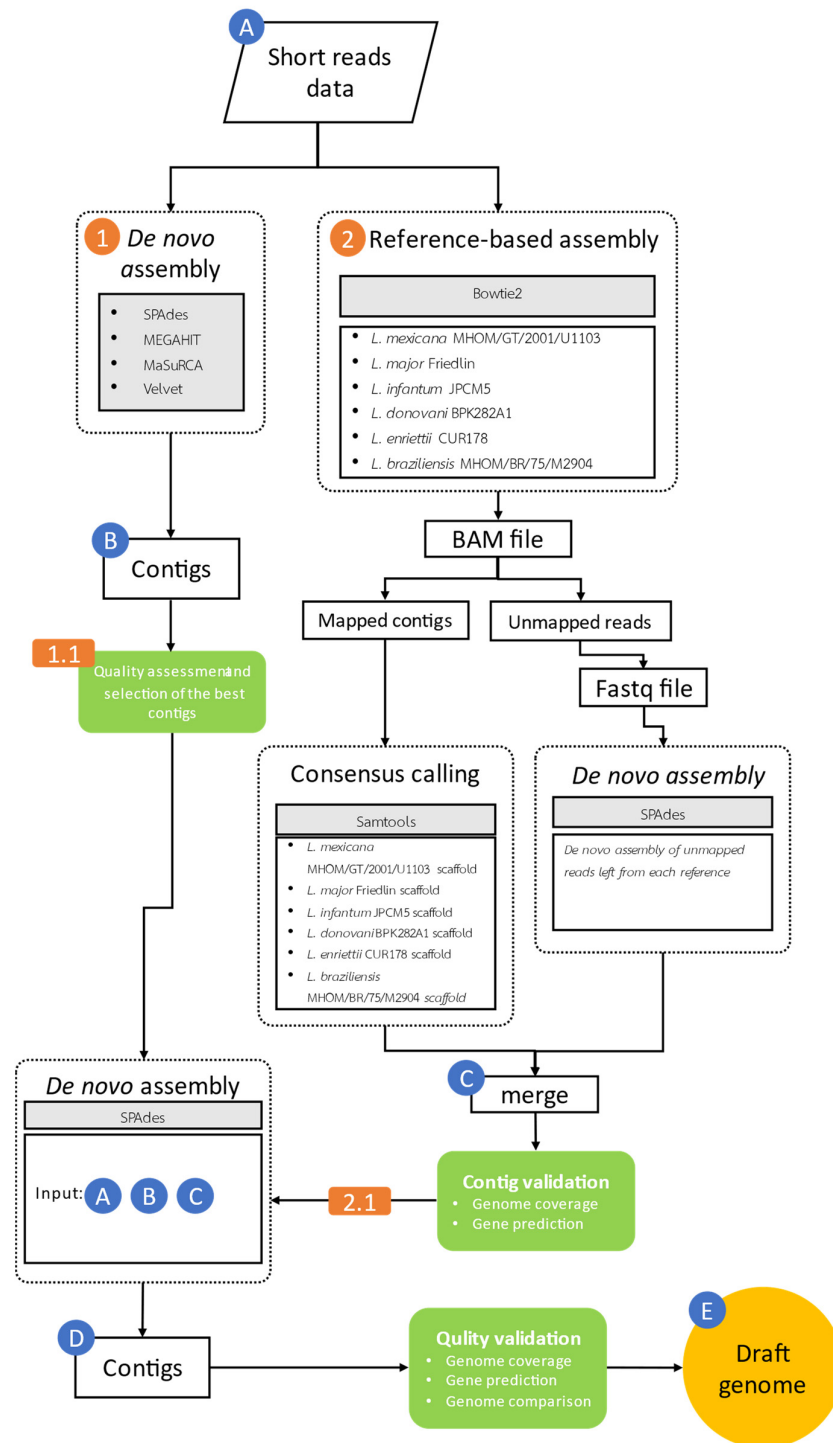


Figure 1. Diagram overview showing the Illumina read data management workflow, comparing the de novo method and the reference-based assembly.

The first de novo assembly was conducted by using four assembly programs, namely SPAdes version 3.14 (Center for Algorithmic Biotechnology, St Petersburg, Russia) [50], MEGAHIT version 1.2.9 (HKU-BGI Bioinformatics Algorithms Research Laboratory & Department of Computer Science, L3 Bioinformatics Limited, Hong Kong, China) (National Institute of Informatics, Tokyo, Japan) [51], MaSuRCA version 4.0.5 (University of Maryland, College Park, MD, USA) [52], and Velvet version 1.2.10 (EMBL-European Bioinformatics Institute, Cambridge, UK) [53]. The assembled contigs (Data B in Figure 1) from each program were assessed by QUAST version 5.0.2 (Center for Algorithmic Biotechnology, St Petersburg, Russia) [54], and the best results were selected based on the contig length, the total number of contigs, and genome size (step 1.1). The Bowtie2 program version 2.4.2 (Johns Hopkins University, Baltimore, MD, USA) [55] performed the latter reference-based assembly using six complete genomes of *Leishmania* as references: *L. mexicana* MHOM/GT/2001/U1103 (GCA_000234665.4), *L. major* Friedlin (GCA_000002725.2), *L. infantum* JPCM5 (GCA_000002875.2), *L. donovani* BPK282A1 (GCA_000227135.2), *L. braziliensis* MHOM/BR/75/M2904 (GCA_000002845.2), and *Leishmania enriettii* CUR178 (GCA_017916305.1). A phylogenetic analysis of 443 partial sequences of heat shock protein 70-coding gene (*hsp70*) of *Leishmania* species and *Trypanosoma equiperdum* (an outgroup) downloaded from the NCBI nucleotide database was performed to select suitable reference genomes for this step. The sequences were trimmed to 1322 bp, multiply aligned by ClustalW version 2.1 (Conway Institute of Biomolecular and Biomedical Research, Belfield, Ireland) [56], and the phylogenetic relationship was reconstructed using neighbor-joining with the *p*-distance model and 10,000 bootstrap iterations using the MEGA X program (Research Center for Genomics and Bioinformatics, Tokyo, Japan) [57]. Although there was a chromosome-scale genome of another isolate LSCM4 of *L. orientalis* from the northern province of Thailand available, this genome was not incorporated into this process to resemble the situation in which no prior genomes of the same species were determined.

The quality of the mapped contigs from individual references was assessed by SAMtools version 1.14 (Wellcome Sanger Institute, Cambridge, UK) [58,59]. The unmapped reads from each reference-based assembly were de novo assembled by SPAdes version 3.14 (Center for Algorithmic Biotechnology, St Petersburg, Russia) [50]. These additional contigs were again combined with their corresponding mapped contigs, resulting in Data C in Figure 1. The obtained genomic contigs were subjected to gene prediction by AUGUSTUS Web Server [60–62]. The contigs from the two approaches (Data B and C) were integrated using the de novo assembler, SPAdes version 3.14 [50], using a parameter `-trusted-contigs`. Data A was then re-mapped to the integrated contigs using SPAdes, resulting in six genomic drafts (Data D).

2.5. Quality Examination of *L. orientalis* PCM2 Genomic Contigs by Genomic Comparison with the Reference Genomes

The contigs of *L. orientalis* isolate PCM2 (Data D in Figure 1) were mapped to seven associated reference genomes of *Leishmania* (*L. donovani* BPK282A1 (GCA_000227135.2), *L. braziliensis* MHOM/BR/75/M2904 (GCA_000002845.2), *L. infantum* JPCM5 (GCA_000002875.2), *L. major* Friedlin (GCA_000002725.2), *L. mexicana* MHOM/GT/2001/U1103 (GCA_000234665.4), *L. enriettii* CUR178 (GCA_017916305.1), and *L. orientalis* LSCM4 (GCA_017916335.1)) using Bowtie2 version 2.4.5 (Johns Hopkins University, Baltimore, MD, USA) [55] with their default parameters. Samtools version 1.14 (Wellcome Sanger Institute, Cambridge, UK) [58,59] was used to determine mapping statistics, including the total number of mapped, unmapped, and paired mapped reads, as well as further analysis of the alignment files. Samtools then converted the SAM file from Bowtie2 to the BAM format and sorted it [63]. Pairwise comparison of two *Leishmania* genomes was analysed using pairwise comparison methods using the Smith–Waterman algorithm [64], and the comparative dot plots were produced by the re-DOT-able tools (<https://www.bioinformatics.babraham.ac.uk/projects/redotable/>) (Babraham Institute, Cambridge, UK) (accessed on 7 June 2021). A coverage analysis on the *Leishmania* chromosomes was performed using Bowtie2. Calculation of the

coverage analysis on each alignment was performed using SAMtools (SAMtools coverage). The visualization of our draft genome coverage against the seven *Leishmania* reference genomes was plotted and compared using the R package, karyoploteR library [65]. Genes were also predicted from Data D using the Augustus program [62,66] based on evidence of the protein homology to *L. major* Friedin (GCA_000002725.2). Gene ontology was assigned to the predicted protein-coding genes using Pannzer2 version 2 (University of Helsinki, Helsinki, Finland) [67–72]. The GO sets were simplified by grouping similar terms based on semantic similarity. The accuracy of the predicted GO class was estimated using positive predictive value (PPV). The relationship between PPV and the Argot score was calibrated using a training set of proteins with available correct annotation. The GO enrichment analysis was performed by using the enrichGO program from the enrichplot library for gene ontology over-representation test. Adjusted *p*-values for these multiple comparisons were obtained by using Benjamini and Hochberg methods by setting *p*valueCutoff = 0.05 and *q*valueCutoff = 0.10, and the enrichment results were plotted using the rrvgo library [73]. These parameters were used to select the best final draft genome (Data E in Figure 1).

3. Results

This study sequenced and reconstructed the draft genome of *L. orientalis* isolate PCM2, collected from the southern province in Thailand, exclusively based on the Illumina short-read data. A total of 16,980,871 sequence reads were generated and may frequently be considered insufficient for the current genomic research trend. The authors designed a bioinformatics workflow for use with these genomic reads as depicted in Figure 1 that (1) integrated de novo and reference-based assembly methods to handle novel reads and the reads that shared homology to the reference genomes; (2) de novo assembled by multiple assemblers; (3) recruited multiple reference genomes as choices; (4) employed multiple steps of the de novo assembly to assure that all reads would be involved in the draft genome; and (5) assessed the genome quality by judging the coverage, number of predicted genes, and the associated functional annotation. The first stage of de novo assembly compared the performance of four assemblers, namely SPAdes, MEGAHIT, MaSuRCA, and velvet (Table 1). The four assemblers gave contigs with similar GC contents (an average of 59.07%). MEGAHIT gave the genomic size of 29.94 Mb and contained the largest contigs of 85.32 Kb. Although the MEGAHIT-derived genomic size was slightly less than that of the SPAdes program, MEGAHIT achieved higher N50 and N70 values, representing 50% and 70% of the entire assembly larger than these values. Among these four assemblers, the performance of Velvet was the lowest, and that of MaSuRCA was second-lowest.

Table 1. Summary of the de novo assembly parameters after assembling sequence reads of *L. orientalis* isolate PCM2 using four assemblers.

Features	Sequence Assemblers			
	SPAdes	MEGAHIT	MaSuRCA	Velvet
Total base in the assembly (Mb)	30.15	29.94	28.50	29.05
No. of contigs	5565	6470	11,241	18,409
Largest contigs (kb)	60.19	85.32	33.26	27.81
N50	12,259	13,737	4883	2919
N75	6229	6714	2547	1450
%GC content	59.02	59.07	59.15	59.03

The hsp70 of *Leishmania orientalis* isolate PCM2 was retrieved from the MEGAHIT-derived contigs and was utilised to determine the evolutionary relationship among various *Leishmania* species. The phylogenetic tree clustered *L. orientalis* isolates PCM2 and LSCM4 (MG731233.1) and another PCM2 sample previously named *L. siamensis* (KC202880.1). The phylogenetic result suggested that *L. enriettii* was closely related to the *L. orientalis* PCM2

(Figure 2), consistent with prior analysis [74]. Therefore, the genome of *L. enriettii* was selected and used as the reference along with the other five complete *Leishmania* genomes. The reference-based assembly gave different genome sizes, numbers of contigs, and genes (Table 2). The longest length (33.32 Mb), highest N75 value (709,397 bases), and the highest number of predicted genes (8545 genes) were achieved by using *L. enriettii* CUR178 as the reference (marked as bold on the Scaffold (B) row in Table 2). This method yielded fewer contigs and larger contig sizes than the de novo assembly. The reference genomes of *L. donovani* BPK282A1 and *L. major* Friedlin in the assembly produced the lowest number of contigs (36 contigs), and the N50 value was the highest when using *L. major*. In comparison, the largest contig (3343,498 bases) was obtained when *L. mexicana* MHOM/GT/2001/U1103 was used as a reference. The de novo assembly of the remaining reads from the reference-guided mapping could provide additional contigs: the maximum of 12,929 contigs from the *L. donovani* dataset, the longest contig (33,294 bases) from the *L. infantum* and *L. major* datasets, the highest number of genes (3674 genes) from the *L. mexicana* dataset. To construct genomic scaffolds of *L. orientalis* isolate PCM2, the integrating multi-step de novo assembly with the reference-based method and salvage of the unmapped reads resulted in the lowest contig number (3367 contigs) and the highest number of predicted genes (8887 genes) for the *L. enriettii* dataset, yielding the largest scaffold of 30.27 Mb (marked as the bold red number in Table 2) at the expense of reduced length of the largest contig, N50, and N75 values. The results from the integrated approach were comparable to those of *L. orientalis* isolate LSCM4 (8158 genes) collected from the northern province of Thailand. Changes in the number of predicted genes were the main improvement of our proposed workflow compared to the use of the de novo assembly method or reference-based assembly alone which gave a broad range of the gene number, approximately 2265–7863 genes, with different contig numbers and lengths (Table 2).

Table 2. Comparative assembly parameters of *L. orientalis* isolate PCM2 obtained from different assembling strategies: a single de novo step (A), the reference-based assembly alone with six reference genomes of *Leishmania* (B), the de novo assembly of the unmapped reads (C), and the multi-step de novo assembly and referenced-based method (D), compared with the reference genome of another *L. orientalis* isolate LSCM4. Bold numbers represent the maximum numbers for the (B) method. Underlined numbers represent the maximum numbers for the (C) method. Red numbers represent the maximum numbers for the (D) method.

Assembly Methods		Genes	Contigs	Largest Contigs	Total Length (Mb)	N50	N75	L50	L75	GC (%)
MEGAHIT (De novo assembly)	Scaffold (A)	7653	4989	90,086	29.94	11,804	5994	767	1663	59.03
<i>L. braziliensis</i> MHOM/BR/75/M2904	Scaffold (B)	2265	138	2,686,643	32.07	992,961	641,930	11	22	58.68
	Unmapped (C)	3670	12,897	30,803	13.17	967	683	3112	7219	60.02
	Merged (D)	8187	5001	99,540	30.09	17,861	8663	500	1100	59
<i>L. donovani</i> BPK282A1	Scaffold (B)	2977	36	2,713,248	32.44	1,024,085	671,483	11	21	59.22
	Unmapped (C)	3672	<u>12,929</u>	30,803	13.17	968	683	3144	7250	60.02
	Merged (D)	8174	<u>4973</u>	99,540	30.08	18,120	8651	492	1087	59
<i>L. enriettii</i> CUR178	Scaffold (B)	7863	54	2,730,217	33.32	1,075,649	709,397	11	21	59.77
	Unmapped (C)	944	12,331	10,814	9.84	798	627	4427	7922	58.51
	Merged (D)	8887	3367	206,463	30.27	24,513	13,215	359	779	59.05
<i>L. infantum</i> JPCM5	Scaffold (B)	3101	76	2,673,956	32.12	1,043,848	659,512	11	21	59.29
	Unmapped (C)	3668	12,910	33,294	13.17	968	683	3133	7235	60.02
	Merged (D)	8181	4982	99,540	30.08	18,053	8696	498	1095	59
<i>L. major</i> Friedlin	Scaffold (B)	3030	36	2,682,151	32.86	1,091,540	684,829	11	21	59.31
	Unmapped (C)	3670	12,928	<u>33,294</u>	13.17	967	682	3135	7247	60.02
	Merged (D)	8178	4995	99,540	30.09	17,924	8632	495	1095	59

Table 2. Cont.

Assembly Methods		Genes	Contigs	Largest Contigs	Total Length (Mb)	N50	N75	L50	L75	GC (%)
<i>L. mexicana</i> MHOM/GT/2001/ U1103	Scaffold (B)	2930	575	3,343,498	32.10	1,044,075	655,046	10	20	59.41
	Unmapped (C)	3674	12,924	22,679	13.17	968	683	3140	7246	60
	Merged (D)	8184	4984	99,540	30.09	18,053	8615	496	1093	59
<i>L. orientalis</i> LSCM4	Scaffold (E)	8158	98	2,735,713	34.19	1,120,138	682,718	11	22	59.72

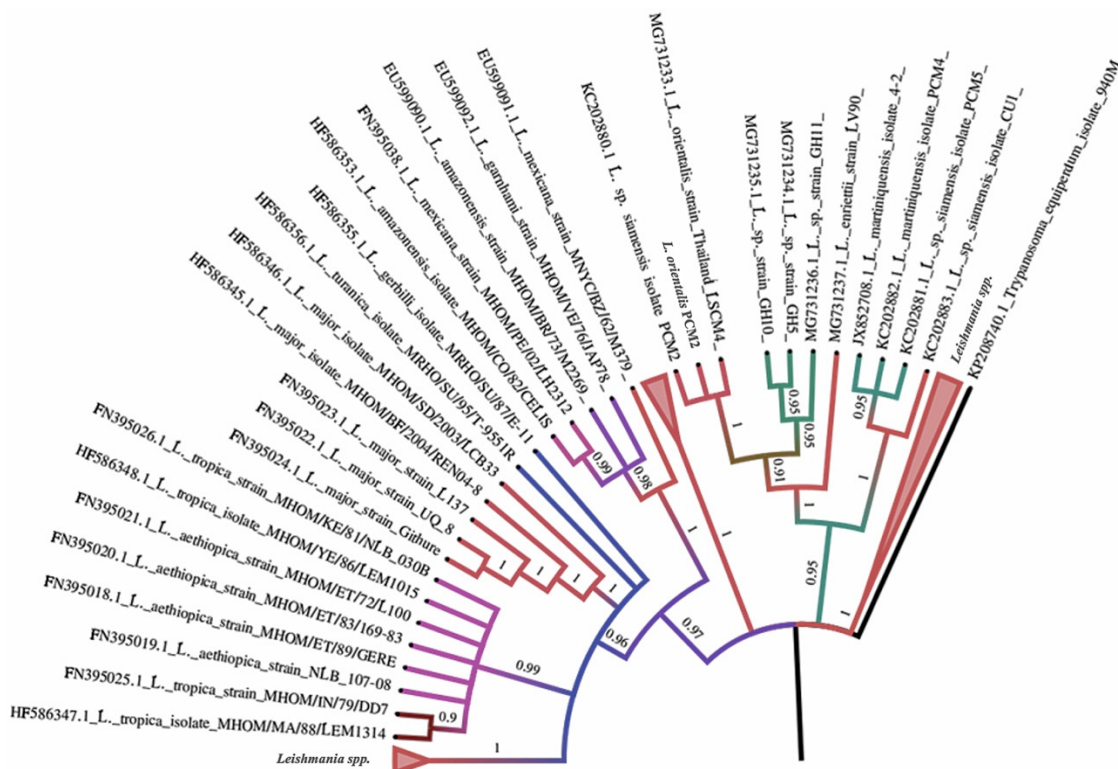


Figure 2. Phylogenetic relationship of *Leishmania* species reconstructed based on the heat shock protein 70 (hsp70) gene using the neighbour-joining method with the *p*-distance model and 10,000 bootstrap iterations. Line colours represent clusters of the samples. The branch number indicates the bootstrap value. The red triangles represent the distant clades which included sequences of *Leishmania infantum* (123 samples), *Leishmania donovani* (91 samples), *Leishmania braziliensis* (90 samples), *Leishmania tropica* (38 samples), *Leishmania major* (31 samples), *Leishmania turanica* (10 samples), *Leishmania naiffi* (9 samples), *Leishmania chagasi* (6 samples), *Leishmania aethiopia* (6 samples), *Leishmania sp.* (5 samples), and *Leishmania panamensis* (4 samples).

The contigs were mapped to the chromosomes of the closely related strain LSCM4 of *L. orientalis* for better visualisation of the assembled contigs (Figure 3a). The contigs from our integrating method with *L. enriettii* as a reference showed the best integrity and coverage of the assembly (Figures 3a and S1) and better contig alignment to the genome of *L. orientalis* isolate LSCM4 (Figure 3b,c), indicating that the de novo assembly method alone may not be able to properly anticipate the contig direction. Our method also solved the problem of gap bridging that occurred during the de novo assembly process and can be seen as multiple gaps (Figure 3d) as examples in chromosomes 3, 4, 6–8, 11, and 13–36 in Figure S1. However, some large gaps remained in chromosomes 1–5, 7–12, 15, 17, 22–23, and 25–26 (in Figure S1), suggesting further low-depth long-read sequencing for closing these remaining gaps if researchers aim to complete these regions.

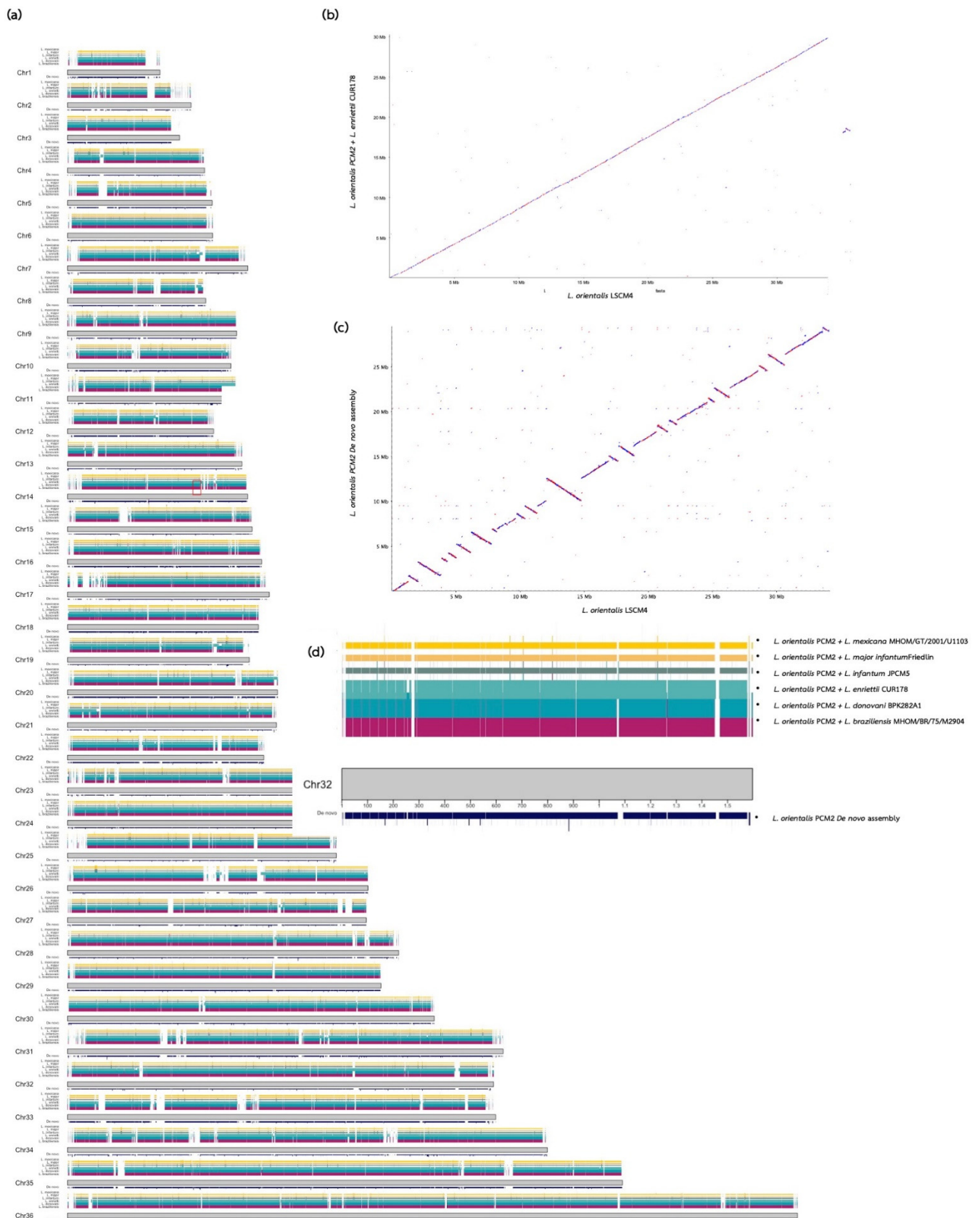


Figure 3. Comparative read mapping of the contigs of *L. orientalis* isolate PCM2 from different assembly methods: (1) a single de novo step; (2–7) six reference-based assemblies integrated with multi-step de novo assembly, and the salvage of the unmapped reads, with the chromosome-scale

genome of *L. orientalis*, isolate LSCM4. The contigs assembled by six *Leishmania* reference genomes are shown in different colour strips. (a) Overview of genome alignments between these seven datasets with the LSCM4 chromosomes. Dot plots represent the high levels of synteny when comparing the *L. enriettii*-based contigs of the isolate PCM2 (b) and the single de novo assembled contigs of the PCM2 isolate (c) with the LSCM4 genome. (d) Example of regions on chromosome 32 with better genomic coverage of *L. enriettii*-based assembled contigs.

Analysis of the annotated genes from *L. orientalis* strain PCM2 obtained by using our proposed method (Figure 4b,e) compared with the single de novo assembly (Figure 4a,d) and those of the reference *L. orientalis* strain LSCM4 (Figure 4c,f) showed similar patterns of gene ratios and frequency of each functional category. Fifty GO terms in the biological process category were assigned to the assembled genomes, and 23 of these (accounting for 61.6% of the annotated proteins) showed higher gene counts and ratios when the results from our workflow were compared to those of the LSCM4 isolate, i.e., biological regulation (GO:0065007), regulation of biological process (GO:0050789), cellular component organisation or biogenesis (GO:0071840), regulation of cellular process (GO:0050794), localisation (GO:0051179), transport (GO:0006810), establishment of localisation (GO:0051234), cellular component organisation (GO:0016043), nucleobase-containing compound biosynthetic process (GO:0034654), response to stimulus (GO:0050896), organelle organisation (GO:0006996), etc. Similarly, 21 GO terms from our workflow had greater gene counts and ratios compared to the single assembly method. Very few GO terms assigned to the results from the combined workflow showed a lower number of gene counts and ratios compared to the other two methods, including regulation of transcription (GO:0006355), nitrogen compound transport (GO:0071705), and transcription (GO:0006351). Moreover, the orthologous gene comparison between the PCM2 and LSCM4 strains revealed 7626 shared functional clusters, 29 clusters unique to the LSCM4, and 19 clusters unique to the PCM2 strain, confirming the difference between the two *L. orientalis* strains. These results indicated that our proposed bioinformatic workflow produced an acceptable-quality genome of *L. orientalis* strain PCM2 for functional genomic analysis.

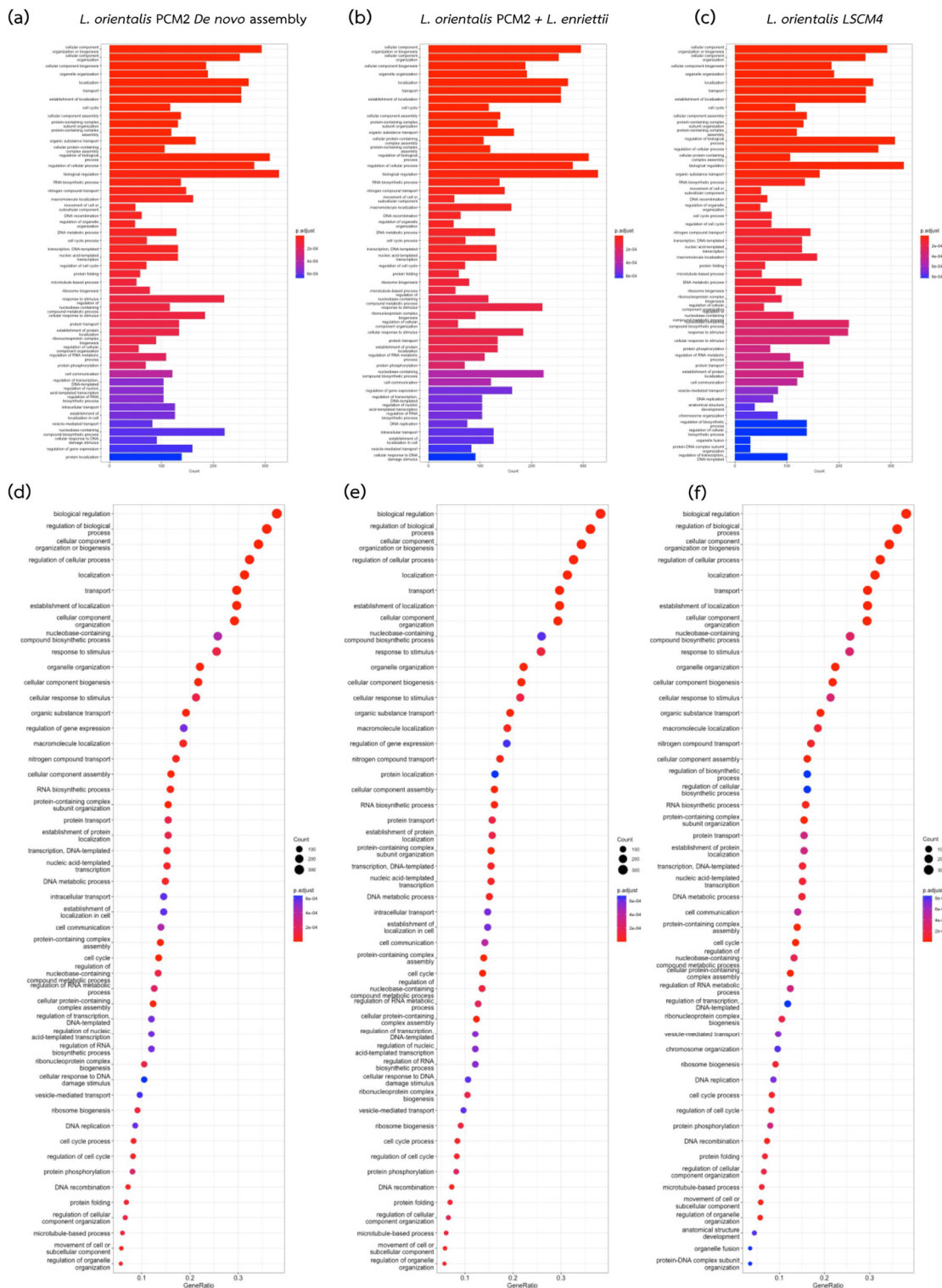


Figure 4. Distribution of the GO-annotated genes of the assembled genome of *L. orientalis* isolate PCM2 obtained from the single de novo assembly (a,d) and the integrated workflow (b,e) compared with the reference genome of *L. orientalis* isolate LSCM4 (c,f). (a–c) Plots between the GO term categories and gene count with their associated *p* values. (d–f) Plots between the GO terms and the gene ratios (proportion of all assigned GO functions in each genome) with their associated *p* values. The balloon sizes indicate the gene count.

4. Discussion

Several studies have shown that reconstructing a complete *Leishmania* genome requires multiple whole-genome sequencing methods, such as combining data from second- and third-generation sequencing technologies [75]. However, the expense of employing the third-generation sequencing platform might make it inaccessible to everyone. Dividing the genome sequencing project into phases would maximise data usage and be more cost-effective. The first phase could begin with either the short- or long-read sequencing method depending on relevant reference genomes and research objectives. This study started with the short-read sequencing method because of (1) the availability and quality of reference genomes from the related *Leishmania* species and (2) the focus on the gene function. In a similar genomic study on *Leishmania naiffi* and *Leishmania guyanensis*, the researchers reconstructed the draft genomes based on the short sequence reads and found additional genes compared to the previous related reference of *Leishmania braziliensis* strain M2904 [76]. Their short-read genomic draft allowed identification of novel genes, alteration of the TATE transposon, and the change in the gene copy number, consistent with different numbers of annotated genes in Figure 4 of the present study. If there were no appropriate references, multiple short-read sequencing with varying library sizes or the Oxford Nanopore long-read sequencing would be the choices for estimating the genomic landscape.

Analysis of the short-read outputs with the combination of reference-based and multi-step de novo assembly methods ensured that almost all reads would be incorporated into the draft genome of the isolate PCM2. Using multiple assemblers and reference genomes could maximise the chance of obtaining good genomic drafts using the best available assembling algorithms (MEGAHIT and SPAdes) and genomic guides using the genome of the closely related species *L. enriettii* in this study. This study also found that the de Bruijn graph-based MEGAHIT program was another good assembler for the *Leishmania* genome compared to SPAdes which was used several times in our workflow, consistent with the report that the MEGAHIT program could handle large and complex next-generation sequencing datasets [51]. In contrast, several recent chromosome-scale *Leishmania* genomes, including *L. orientalis* isolate LSCM4, were assembled using de novo assembly of the long-read MinION data as guidance for mapping the Illumina short reads [45]. In this study, our assembled genome of *L. orientalis* isolates PCM2 from the proposed analytic workflow on the short-read data shared highly similar genomic synteny with those of the LSCM4, as shown in Figure 3, implying the genome structural similarity between the selected reference, *L. enriettii*, and the LSCM4 genome. The critical success of this phase would also depend on the selection of the reference genome to obtain results comparable to the use of both long and short reads. The reference in this study also enhanced the investigation of genome structure compared between the PCM2 and LSCM4 isolates, as displayed in Figure 3b.

Although our workflow produced genomic drafts qualified for functional analysis, limitations were also addressed. Certain structures of the *Leishmania* genome could be complex to resolve by the short-read technology, such as duplicated genes and long tandem repeats, which may play a role in gene expression mechanisms [42,43]. Repetition and aneuploidy have challenged the assembly of *Leishmania* genomes [43,77–80]. These would be estimated by counting the read coverage as explained in [75], which estimated the aneuploidy of each chromosome (S) by $2 \times di/dm$ from the short-read data, where di is the median depth of each chromosome and dm is the median depth of the whole genome. The gene copy number was calculated by $dFG = S \times dHG$, where dFG is the full cell depth with some effect and dHG was the average haploid depth per gene without the some effect. To estimate the ploidy in this study, the long-read data would be essential for preparing the chromosomal outline onto which the short reads are mapped.

With the first phase results, researchers could manage the second phase of the genome project, whether accepting the genomic draft quality for the downstream analysis or applying additional sequencing platforms for the genome refinement and filling the genomic gaps. As shown in this study, the proposed workflow (Figures 1 and S2) generated an acceptable

genomic draft and a similar number of annotated genes compared to the reference genome of *L. orientalis* isolate LSCM4 (Figures 3 and 4). These are useful for further functional comparative genomic analysis to identify core and accessory genes involved in virulence and pathogenesis, such as genes involved in establishing localisation. Our workflow could also be extended to explore more genomes of related *Leishmania* variants and species with cost-effective expenditure, particularly in low-income and developing countries.

5. Conclusions

The proposed bioinformatics workflow, which employs the usage of multiple assemblers and genomic references and combines multi-step de novo assembly with the reference-based method and the salvage of the unmapped reads, could maximise the use of the short-read data of *L. orientalis* strain PCM2 with the genome structure and number of annotated genes comparable to those of the strain LSCM4. This workflow would give preliminary information required for further plans to expand the *Leishmania* genome sequencing project to cover more isolates and incorporate other sequencing methods.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology11091272/s1>, Figure S1: Mapping the contigs of *Leishmania orientalis* strain PCM2 from the integrating assembly method based on six reference genomes of *Leishmania* and the single de novo method to the individual chromosomal sequences of *L. orientalis* isolate LSCM4. Figure S2: Graphical summary of the proposed analytic workflow based on the short-read data. The reads were assembled by the integration reference-based and multi-step de novo assembly.

Author Contributions: Conceptualisation, T.E.-k., M.M., S.L. and R.B.; methodology, T.E.-k., P.A. and S.S.; software, T.E.-k. and P.A.; validation, T.E.-k. and P.A.; investigation, T.E.-k. and P.A.; resources, K.C. and S.U.; data curation, T.E.-k.; writing—original draft preparation, T.E.-k. and P.A.; writing—review and editing, M.M., S.L. and T.E.-k.; supervision, M.M., S.L., R.B. and T.E.-k.; project administration, T.E.-k.; funding acquisition, T.E.-k. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Kasetsart University Research and Development Institute (KURDI), Kasetsart University, grant number FF(KU)6.64.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: None.

Acknowledgments: The authors would like to thank the Department of Parasitology, Phramongkutkhalao College of Medicine; the Department of Microbiology, Faculty of Science, Mahidol University; and the Department of Genetics, Faculty of Science, Kasetsart University, for supporting this project and the Faculty of Science, Kasetsart University, for providing high-performance computing facilities.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Alvar, J.; Vélez, I.D.; Bern, C.; Herrero, M.; Desjeux, P.; Cano, J.; Jannin, J.; den Boer, M.; Team, W.L.C. Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* **2012**, *7*, e35671. [[CrossRef](#)] [[PubMed](#)]
2. Lainson, R.; Ward, R.; Shaw, J. Leishmania in phlebotomid sandflies: VI. Importance of hindgut development in distinguishing between parasites of the *Leishmania mexicana* and *L. braziliensis* complexes. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1977**, *199*, 309–320.
3. Murray, H.W.; Berman, J.D.; Davies, C.R.; Saravia, N.G. Advances in leishmaniasis. *Lancet* **2005**, *366*, 1561–1577. [[CrossRef](#)]
4. Steverding, D. The history of leishmaniasis. *Parasites Vectors* **2017**, *10*, 82. [[CrossRef](#)] [[PubMed](#)]
5. Belo, V.S.; Struchiner, C.J.; Barbosa, D.S.; Nascimento, B.W.L.; Horta, M.A.P.; da Silva, E.S.; Werneck, G.L. Risk factors for adverse prognosis and death in American visceral leishmaniasis: A meta-analysis. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2982. [[CrossRef](#)] [[PubMed](#)]

6. Banuls, A.-L.; Bastien, P.; Pomares, C.; Arevalo, J.; Fisa, R.; Hide, M. Clinical pleiomorphism in human leishmaniasis, with special mention of asymptomatic infection. *Clin. Microbiol. Infect.* **2011**, *17*, 1451–1461. [[CrossRef](#)]
7. Ameen, M. Cutaneous leishmaniasis: Advances in disease pathogenesis, diagnostics and therapeutics. *Clin. Exp. Dermatol. Clin. Dermatol.* **2010**, *35*, 699–705. [[CrossRef](#)]
8. de Oliveira Guerra, J.A.; Prestes, S.R.; Silveira, H.; Câmara, L.I.d.A.R.; Gama, P.; Moura, A.; Amato, V.; Barbosa, M.d.G.V.; de Lima Ferreira, L.C. Mucosal leishmaniasis caused by *Leishmania (Viannia) braziliensis* and *Leishmania (Viannia) guyanensis* in the Brazilian Amazon. *PLoS Negl. Trop. Dis.* **2011**, *5*, e980.
9. Viriyavejakul, P.; Viravan, C.; Riganti, M.; Punpoowong, B. Imported cutaneous leishmaniasis in Thailand. *Southeast Asian J. Trop. Med. Public Health* **1997**, *28*, 558–562.
10. Suttinont, P.; Thammanichanont, C.; Chantarakul, N. Visceral leishmaniasis: A case report. *Southeast Asian J. Trop. Med. Public Health* **1987**, *18*, 103–106.
11. Sukmee, T.; Siripattanapipong, S.; Mungthin, M.; Worapong, J.; Rangsin, R.; Samung, Y.; Kongkaew, W.; Bumrungsana, K.; Chanachai, K.; Apiwathanasorn, C. A suspected new species of *Leishmania*, the causative agent of visceral leishmaniasis in a Thai patient. *Int. J. Parasitol.* **2008**, *38*, 617–622. [[CrossRef](#)] [[PubMed](#)]
12. Jariyapan, N.; Daroontum, T.; Jaiwong, K.; Chanmol, W.; Intakhan, N.; Sor-Suwan, S.; Siriyasatien, P.; Somboon, P.; Bates, M.D.; Bates, P.A. *Leishmania (Mundinia) orientalis* n. sp. (Trypanosomatidae), a parasite from Thailand responsible for localised cutaneous leishmaniasis. *Parasites Vectors* **2018**, *11*, 351. [[CrossRef](#)] [[PubMed](#)]
13. Leelayoova, S.; Siripattanapipong, S.; Hitakarun, A.; Kato, H.; Tan-ariya, P.; Siriyasatien, P.; Osatakul, S.; Mungthin, M. Multilocus characterization and phylogenetic analysis of *Leishmania siamensis* isolated from autochthonous visceral leishmaniasis cases, southern Thailand. *BMC Microbiol.* **2013**, *13*, 1–7. [[CrossRef](#)] [[PubMed](#)]
14. Müller, N.; Welle, M.; Lobsiger, L.; Stoffel, M.H.; Boghenbor, K.K.; Hilbe, M.; Gottstein, B.; Frey, C.F.; Geyer, C.; von Bomhard, W. Occurrence of *Leishmania* sp. in cutaneous lesions of horses in Central Europe. *Vet. Parasitol.* **2009**, *166*, 346–351. [[CrossRef](#)] [[PubMed](#)]
15. Lobsiger, L.; Müller, N.; Schweizer, T.; Frey, C.; Wiederkehr, D.; Zumkehr, B.; Gottstein, B. An autochthonous case of cutaneous bovine leishmaniasis in Switzerland. *Vet. Parasitol.* **2010**, *169*, 408–414. [[CrossRef](#)] [[PubMed](#)]
16. Reuss, S.M.; Dunbar, M.D.; Mays, M.B.C.; Owen, J.L.; Mallicote, M.F.; Archer, L.L.; Wellehan, J.F., Jr. Autochthonous *Leishmania siamensis* in horse, Florida, USA. *Emerg. Infect. Dis.* **2012**, *18*, 1545. [[CrossRef](#)]
17. Chusri, S.; Thammapalo, S.; Silpapojakul, K.; Siriyasatien, P. Animal reservoirs and potential vectors of *Leishmania siamensis* in southern Thailand. *Southeast Asian J. Trop. Med. Public Health* **2014**, *45*, 13.
18. Mukhtar, M.M.; Sharief, A.; El Saffi, S.; Harith, A.; Higazzi, T.; Adam, A.; Abdalla, H.S. Detection of antibodies to *Leishmania donovani* in animals in a kala-azar endemic region in eastern Sudan: A preliminary report. *Trans. R. Soc. Trop. Med. Hyg.* **2000**, *94*, 33–36. [[CrossRef](#)]
19. Rohousova, I.; Talmi-Frank, D.; Kostalova, T.; Polanska, N.; Lestinova, T.; Kassahun, A.; Yasur-Landau, D.; Maia, C.; King, R.; Votycka, J. Exposure to *Leishmania* spp. and sand flies in domestic animals in northwestern Ethiopia. *Parasites Vectors* **2015**, *8*, 360. [[CrossRef](#)]
20. Polseela, R.; Vitta, A.; Nateeworanart, S.; Apiwathanasorn, C. Distribution of cave-dwelling phlebotomine sand flies and their nocturnal and diurnal activity in Phitsanulok Province, Thailand. *Southeast Asian J. Trop. Med. Public Health* **2011**, *42*, 1395–1404.
21. Manomat, J.; Leelayoova, S.; Bualert, L.; Tan-Ariya, P.; Siripattanapipong, S.; Mungthin, M.; Naaglor, T.; Piyaraj, P. Prevalence and risk factors associated with *Leishmania* infection in Trang Province, southern Thailand. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0006095. [[CrossRef](#)] [[PubMed](#)]
22. Lindoso, J.A.L.; Cunha, M.A.; Queiroz, I.T.; Moreira, C.H.V. Leishmaniasis–HIV coinfection: Current challenges. *HIV/AIDS* **2016**, *8*, 147.
23. Faraut-Gambarelli, F.; Piarroux, R.; Deniau, M.; Giusiano, B.; Marty, P.; Michel, G.; Faugère, B.; Dumon, H. In vitro and in vivo resistance of *Leishmania infantum* to meglumine antimoniate: A study of 37 strains collected from patients with visceral leishmaniasis. *Antimicrob. Agents Chemother.* **1997**, *41*, 827–830. [[CrossRef](#)] [[PubMed](#)]
24. Croft, S.L.; Sundar, S.; Fairlamb, A.H. Drug resistance in leishmaniasis. *Clin. Microbiol. Rev.* **2006**, *19*, 111–126. [[CrossRef](#)]
25. Sundar, S.; Sinha, P.K.; Rai, M.; Verma, D.K.; Nawin, K.; Alam, S.; Chakravarty, J.; Vaillant, M.; Verma, N.; Pandey, K. Comparison of short-course multidrug treatment with standard therapy for visceral leishmaniasis in India: An open-label, non-inferiority, randomised controlled trial. *Lancet* **2011**, *377*, 477–486. [[CrossRef](#)]
26. Musa, A.; Khalil, E.; Hailu, A.; Olobo, J.; Balasegaram, M.; Omollo, R.; Edwards, T.; Rashid, J.; Mbui, J.; Musa, B. Sodium stibogluconate (SSG) & paromomycin combination compared to SSG for visceral leishmaniasis in East Africa: A randomised controlled trial. *PLoS Negl. Trop. Dis.* **2012**, *6*, e1674.
27. van Griensven, J.; Gadisa, E.; Aseffa, A.; Hailu, A.; Beshah, A.M.; Diro, E. Treatment of cutaneous leishmaniasis caused by *Leishmania aethiopica*: A systematic review. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004495. [[CrossRef](#)]
28. Lindoso, J.A.; Cota, G.F.; da Cruz, A.M.; Goto, H.; Maia-Elkhoury, A.N.S.; Romero, G.A.S.; de Sousa-Gomes, M.L.; Santos-Oliveira, J.R.; Rabello, A. Visceral leishmaniasis and HIV coinfection in Latin America. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3136. [[CrossRef](#)]
29. Sundar, S. Drug resistance in Indian visceral leishmaniasis. *Trop. Med. Int. Health* **2001**, *6*, 849–854. [[CrossRef](#)]
30. Vanaerschot, M.; Dumetz, F.; Roy, S.; Ponte-Sucre, A.; Arevalo, J.; Dujardin, J.-C. Treatment failure in leishmaniasis: Drug-resistance or another (epi-) phenotype? *Expert Rev. Anti-Infect. Ther.* **2014**, *12*, 937–946. [[CrossRef](#)]

31. Myler, P.J.; Audleman, L.; DeVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.; Rickel, E.; Sisk, E.; Sunkin, S. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2902–2906. [[CrossRef](#)] [[PubMed](#)]
32. Berriman, M.; Ghedin, E.; Hertz-Fowler, C.; Blandin, G.; Renauld, H.; Bartholomeu, D.C.; Lennard, N.J.; Caler, E.; Hamlin, N.E.; Haas, B. The genome of the African trypanosome *Trypanosoma brucei*. *Science* **2005**, *309*, 416–422. [[CrossRef](#)] [[PubMed](#)]
33. Peacock, C.S.; Seeger, K.; Harris, D.; Murphy, L.; Ruiz, J.C.; Quail, M.A.; Peters, N.; Adlem, E.; Tivey, A.; Aslett, M. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat. Genet.* **2007**, *39*, 839–847. [[CrossRef](#)]
34. Cuypers, B.; Dumetz, F.; Meysman, P.; Laukens, K.; De Muyllder, G.; Dujardin, J.-C.; Domagalska, M.A. The absence of C-5 DNA methylation in *Leishmania donovani* allows DNA enrichment from complex samples. *Microorganisms* **2020**, *8*, 1252. [[CrossRef](#)] [[PubMed](#)]
35. Wincker, P.; Ravel, C.; Blaineau, C.; Pages, M.; Jauffret, Y.; Dedet, J.-P.; Bastien, P. The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res.* **1996**, *24*, 1688–1694. [[CrossRef](#)]
36. Britto, C.; Ravel, C.; Bastien, P.; Blaineau, C.; Pagès, M.; Dedet, J.-P.; Wincker, P. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene* **1998**, *222*, 107–117. [[CrossRef](#)]
37. Proudfoot, L.; O'Donnell, C.A.; Liew, F.Y. Glycoinositolphospholipids of *Leishmania major* inhibit nitric oxide synthesis and reduce leishmanicidal activity in murine macrophages. *Eur. J. Immunol.* **1995**, *25*, 745–750. [[CrossRef](#)]
38. Argueta-Donohué, J.; Carrillo, N.; Valdés-Reyes, L.; Zentella, A.; Aguirre-García, M.; Becker, I.; Gutiérrez-Kobeh, L. *Leishmania mexicana*: Participation of NF- κ B in the differential production of IL-12 in dendritic cells and monocytes induced by lipophosphoglycan (LPG). *Exp. Parasitol.* **2008**, *120*, 1–9. [[CrossRef](#)]
39. Peters, C.; Stierhof, Y.-D.; Ilg, n.T. Proteophosphoglycan secreted by *Leishmania mexicana* amastigotes causes vacuole formation in macrophages. *Infect. Immun.* **1997**, *65*, 783–786. [[CrossRef](#)]
40. Carvalho, L.P.; Passos, S.; Dutra, W.O.; Soto, M.; Alonso, C.; Gollob, K.; Carvalho, E.; Ribeiro de Jesus, A. Effect of LACK and KMP11 on IFN- γ Production by Peripheral Blood Mononuclear Cells from Cutaneous and Mucosal Leishmaniasis Patients. *Scand. J. Immunol.* **2005**, *61*, 337–342. [[CrossRef](#)]
41. Almeida, M.S.; Pereira, B.A.S.; Guimarães, M.L.R.; Alves, C.R. Proteinases as virulence factors in *Leishmania* spp. infection in mammals. *Parasites Vectors* **2012**, *5*, 160. [[CrossRef](#)] [[PubMed](#)]
42. Rogers, M.B.; Hilley, J.D.; Dickens, N.J.; Wilkes, J.; Bates, P.A.; Depledge, D.P.; Harris, D.; Her, Y.; Herzyk, P.; Imamura, H. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res.* **2011**, *21*, 2129–2142. [[CrossRef](#)] [[PubMed](#)]
43. Ivens, A.C.; Peacock, C.S.; Worthey, E.A.; Murphy, L.; Aggarwal, G.; Berriman, M.; Sisk, E.; Rajandream, M.-A.; Adlem, E.; Aert, R. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **2005**, *309*, 436–442. [[CrossRef](#)] [[PubMed](#)]
44. Worthey, E.; Martinez-Calvillo, S.; Schnauffer, A.; Aggarwal, G.; Cawthra, J.; Fazelinia, G.; Fong, C.; Fu, G.; Hassebrock, M.; Hixson, G. *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Res.* **2003**, *31*, 4201–4210. [[CrossRef](#)]
45. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Jariyapan, N.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia) martiniquensis*, Isolate LSCM1, Strain LV760. *Microbiol. Resour. Announc.* **2021**, *10*, e00058-21. [[CrossRef](#)]
46. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Jariyapan, N.; Kwakye-Nuako, G.; Thomaz Soccol, V.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. Chromosome-scale genome sequencing, assembly and annotation of six genomes from subfamily *Leishmaniinae*. *Sci. Data* **2021**, *8*, 1–9. [[CrossRef](#)]
47. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Jariyapan, N.; Kwakye-Nuako, G.; Thomaz-Soccol, V.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. LGAAP: *Leishmaniinae* genome assembly and annotation pipeline. *Microbiol. Resour. Announc.* **2021**, *10*, e00439-21. [[CrossRef](#)]
48. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Jariyapan, N.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia) orientalis*, Isolate LSCM4, Strain LV768. *Microbiol. Resour. Announc.* **2021**, *10*, e00574-21. [[CrossRef](#)]
49. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*; Babraham Bioinformatics; Babraham Institute: Cambridge, UK, 2010.
50. Pribelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. Using SPAdes de novo assembler. *Curr. Protoc. Bioinform.* **2020**, *70*, e102. [[CrossRef](#)]
51. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [[CrossRef](#)]
52. Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA genome assembler. *Bioinformatics* **2013**, *29*, 2669–2677. [[CrossRef](#)]
53. Zerbino, D.R.; Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [[CrossRef](#)] [[PubMed](#)]
54. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUASt: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)] [[PubMed](#)]

55. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
56. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)]
57. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547. [[CrossRef](#)]
58. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
59. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [[CrossRef](#)]
60. Keller, O.; Kollmar, M.; Stanke, M.; Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **2011**, *27*, 757–763. [[CrossRef](#)]
61. Stanke, M.; Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **2003**, *19*, ii215–ii225. [[CrossRef](#)]
62. Stanke, M.; Diekhans, M.; Baertsch, R.; Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **2008**, *24*, 637–644. [[CrossRef](#)] [[PubMed](#)]
63. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M. Twelve years of SAMtools and BCftools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)] [[PubMed](#)]
64. Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.* **1981**, *147*, 195–197. [[CrossRef](#)]
65. Gel, B.; Serra, E. karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **2017**, *33*, 3088–3090. [[CrossRef](#)]
66. Stanke, M.; Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **2005**, *33*, W465–W467. [[CrossRef](#)]
67. Koskinen, J.P.; Holm, L. SANS: High-throughput retrieval of protein sequences allowing 50% mismatches. *Bioinformatics* **2012**, *28*, i438–i443. [[CrossRef](#)]
68. Radivojac, P.; Clark, W.T.; Oron, T.R.; Schnoes, A.M.; Wittkop, T.; Sokolov, A.; Graim, K.; Funk, C.; Verspoor, K.; Ben-Hur, A. A large-scale evaluation of computational protein function prediction. *Nat. Methods* **2013**, *10*, 221–227. [[CrossRef](#)]
69. Koskinen, P.; Törönen, P.; Nokso-Koivisto, J.; Holm, L. PANNZER: High-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics* **2015**, *31*, 1544–1552. [[CrossRef](#)]
70. Somervuo, P.; Holm, L. SANSparallel: Interactive homology search against Uniprot. *Nucleic Acids Res.* **2015**, *43*, W24–W29. [[CrossRef](#)]
71. Törönen, P.; Medlar, A.; Holm, L. PANNZER2: A rapid functional annotation web server. *Nucleic Acids Res.* **2018**, *46*, W84–W88. [[CrossRef](#)]
72. Törönen, P.; Holm, L. PANNZER—A practical tool for protein function prediction. *Protein Sci.* **2021**, *31*, 118–128. [[CrossRef](#)] [[PubMed](#)]
73. Sayols, S. rrvgo: A Bioconductor Package to Reduce and Visualize Gene Ontology Terms. 2020. Available online: <https://bioconductor.org/packages/release/bioc/html/rrvgo.html> (accessed on 20 July 2022).
74. Leelayoova, S.; Siripattanapipong, S.; Manomat, J.; Piyaraj, P.; Tan-Ariya, P.; Bualert, L.; Mungthin, M. Leishmaniasis in Thailand: A review of causative agents and situations. *Am. J. Trop. Med. Hyg.* **2017**, *96*, 534–542. [[CrossRef](#)] [[PubMed](#)]
75. Patiño, L.H.; Muñoz, M.; Pavia, P.; Muskus, C.; Shaban, M.; Paniz-Mondolfi, A.; Ramírez, J.D. Filling the gaps in *Leishmania naiffi* and *Leishmania guyanensis* genome plasticity. *G3* **2022**, *12*, jkab377. [[CrossRef](#)]
76. Coughlan, S.; Taylor, A.S.; Feane, E.; Sanders, M.; Schonian, G.; Cotton, J.A.; Downing, T. *Leishmania naiffi* and *Leishmania guyanensis* reference genomes highlight genome structure and gene evolution in the *Viannia* subgenus. *R. Soc. Open Sci.* **2018**, *5*, 172212. [[CrossRef](#)]
77. Ravel, C.; Dubessay, P.; Bastien, P.; Blackwell, J.M.; Ivens, A.C. The complete chromosomal organization of the reference strain of the *Leishmania* genome project, *L. major* Friedlin'. *Parasitol. Today* **1998**, *14*, 301–303. [[CrossRef](#)]
78. Dubessay, P.; Ravel, C.; Bastien, P.; Stuart, K.; Dedet, J.-P.; Blaineau, C.; Pagès, M. Mitotic stability of a coding DNA sequence-free version of *Leishmania major* chromosome 1 generated by targeted chromosome fragmentation. *Gene* **2002**, *289*, 151–159. [[CrossRef](#)]
79. Cruz, A.K.; Titus, R.; Beverley, S.M. Plasticity in chromosome number and testing of essential genes in *Leishmania* by targeting. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1599–1603. [[CrossRef](#)]
80. Martínez-Calvillo, S.; Stuart, K.; Myler, P.J. Ploidy changes associated with disruption of two adjacent genes on *Leishmania major* chromosome 1. *Int. J. Parasitol.* **2005**, *35*, 419–429. [[CrossRef](#)]