Zoological Studies **64:**44 (2025)

Complete Mitochondrial Genome with Phylogenetic Analyses of a New Caenrhabditis Species from Ta Kou Nature Reserve, Vietnam

Viet Dai Dang[1,2,*], Jung-Chen Hsu[1], and John Wang[1,*]

[1]*Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan. *Correspondence: E-mail: dvdai@iat.vast.vn & vietdaidang@gmail.com (Dang); johnwang@gate.sinica.edu.tw (John)*

[2]*Institute of Advanced Technology, Vietnam Academy of Science and Technology, Ho Chi Minh city, Viet Nam. E-mail: ronjan0923@gmail.com (Hsu)*

(Received 3 May 2025 / Accepted 16 July 2025 / Published -- 2025)
Communicated by Jen-Pan Huang

ORCID:
Viet Dai Dang: https://orcid.org/0000-0002-3462-9419
Jung Chen Hsu: https://orcid.org/0000-0002-1080-3689
John Wang: https://orcid.org/0000-0002-6179-775X

Caenorhabditis species diversity has been surveyed in several places around the world and tends to be higher in subtropical and tropical regions. Vietnam is a long country with complex topography and a climate largely influenced by the northeast monsoon from the Siberian plateau. Here, we report the first *Caenorhabditis* survey in Ta Kou Nature Reserve located in south-central Vietnam, where we have found two isolates of a new species, *Caenorhabditis* sp. 71. Whole genome skimming of one isolate using a combination of the Oxford Nanopore Technologies MinION long read and Illumina short read sequencing platforms allowed us to assemble a complete mitochondrial genome that is 13,654 bp in length. Annotation of the mitochondrial genome revealed 12 protein coding genes, 22 transfer RNA genes, 2 ribosomal RNA genes, and 1 control region. Phylogenetic analysis, based on the concatenated sequences of the 12 mitochondrial protein coding genes, suggests that *Caenorhabditis* sp. 71 is the sister species to *C. imperialis*. Our findings not only contribute to the understanding of *Caenorhabditis* diversity in Vietnam but also highlight the ecological significance of isolated habitats, such as Ta Kou Mountain. These fragile environments are essential for biodiversity conservation and require dedicated protection to preserve their unique and diverse ecosystems.

Keywords: Nanopore, *Caenorhabditis* sp. 71, Nematode, Binh Thuan, Diversity

Citation: Dang VD, Hsu JC, Wang J. 2025. Complete mitochondrial genome with phylogenetic analyses of a new Caenrhabditis species from Ta Kou Nature Reserve, Vietnam. Zool Stud **64:**44.

BACKGROUND

The nematode *Caenorhabditis elegans* is a major model organism in biological research. For the first ~100 years after its discovery in 1899 (Sudhaus 2011), only a limited number of species were recognized within the genus. An initial difficulty in finding new *C. elegans* isolates and other species was partly due to the common misconception that *Caenorhabditis* species were soil nematodes. However, the search for new *Caenorhabditis* species gained momentum upon the realization that they are more frequently found on decomposing vegetation—such as rotting fruits and stems—rather than in soil (Félix and Braendle 2010). Following this insight, global sampling has increased (Kiontke et al. 2011; Felix and Duveau 2012; Felix et al. 2013 2014) and intensive sampling efforts have been carried out in certain regions such as Hawaii and French Guiana, resulting in the discovery and description of many new *Caenorhabditis* species and significantly expanding our understanding of their diversity and the ecological niches they occupy (Felix et al. 2013; Crombie et al. 2019 2022; Sun et al. 2022). In less than 10 years, the number of recorded *Caenorhabditis* species has increased from just over 50 in 2019 to more than 80 today (Marie-Annie Félix and others, pers. comm.) (Stevens et al. 2019). These surveys have shown that *Caenorhabditis* diversity seems to be highest in the subtropics and tropics (Ferrari et al. 2017; Stevens et al. 2019; Sloat et al. 2022). Nevertheless, there have been relatively few reports of *Caenorhabditis* species in Southeast Asia due to low sampling intensity.

Vietnam is a medium-sized country with a subtropical and tropical climate, featuring a considerable elevation range from sea level to 3,143 m. As of 2021, there are 181 biodiversity protected areas (including national parks, nature reserves, and biosphere reserves) in the country. Due to its geographical location, diverse topography, and climatic variation, Vietnam continues to be recognized as a global biodiversity hotspot with numerous new species, including insects, plant, and fish, discovered in recent years (Bresseel and Constant 2018; Cumming et al. 2021; Luu et al. 2023; Bresseel and Constant 2024; Chen et al. 2024; Constant and Pham 2024). Despite this biodiversity richness, only six described *Caenorhabditis* species have been reported from three province in Vietnam (Dong Nai, Lam Dong, and Ninh Binh), and therefore *Caenorhabditis* diversity remains largely unexplored (Le et al. 2023a, Le et al. 2023b, Lê et al. 2024).

To address this gap, we initiated a survey at Ta Kou Nature Reserve (10°41'–10°53' N latitude and 107°52'–108°01' E longitude), located in the south-central coastal region of Vietnam.

Characterized by a dry coastal monsoon climate, the nature reserve has a protected core zone of 118 square km^2 including the 697-meter-high Ta Kou Mountain and 10.8 km^2 of coastal sandy flatlands. Surrounding this core zone is a 60 km^2 buffer zone comprising anthropogenic ecosystems (mainly dragon fruit plantations and rice paddies) and coastal sandy land.

Ta Kou Nature Reserve is rich in both animal and plant biodiversity. The nature reserve is home to endangered primates listed by the IUCN (International Union for Conservation of Nature), such as the long-tailed macaque (*Macaca fascicularis*) and the pygmy slow loris (*Nycticebus pygmaeus*) (Hoang et al. 2010; Blair et al. 2021; Hansen et al. 2022). Evaluations of its native forests have identified four main ecosystems: subtropical evergreen forest, tropical semi-deciduous forest, deciduous forest, and dipterocarp lowland forest (Ly and Luu 2007). These ecosystems have positioned Ta Kou Nature Reserve as a critical priority landscape within the Southern Annamites due to its significant biodiversity (Baltzer et al. 2001).

Here we report a small pilot survey in Ta Kou Nature Reserve where we have isolated two independent strains of a new *Caenorhabditis* species. Additionally, based on low coverage genome sequencing we assembled and annotated its complete mitochondrial genome.

MATERIALS AND METHODS

Sampling nematodes

Field work was conducted in Dec 2022 in Ta Kou Nature Reserve, Vietnam. Sampling was focused on substrates with a greater chance of containing *Caenorhabditis* nematodes (Kiontke et al. 2011), such as decaying plant litter, fruit, leaves, and flowers, although other types (*i.e.*, dirt, animal related sample) were also collected (Table 1). To minimize cross-contamination, each sample was carefully picked up by placing a hand inside an inside-out zip-lock bag, which was then turned back to enclose the sample. For each sample, the location, collection date, GPS coordinates, and altitude were recorded. Samples were kept sealed until returned to the lab about a month later.

To isolate the nematodes from the substrate, samples were placed on an agar plate seeded with a patch of *E. coli* (strain OP50). After one to two days, any nematodes that had crawled out were transferred together to a new plate. At this point or on subsequent days, one to three gravid (*i.e.*, eggs in uterus) or mated (*i.e.*, mating plug observed) females or hermaphrodites were isolated to new plates to establish isofemale lines, or strains. After, each strain was observed under a dissecting microscope and screened for a pharynx with two round bulbs, typically indicative of free-living nematodes of the *Caenorhabditis* and closely-related genera.

Table 1. List of collected samples for *Caenorhabditis* species survey in Ta Kou Nature Reserve

Sample_ID	Species	Strain_name	Landscape	Locality	Substrate	Substrate_type	Substrate_note	Species_ID_method	Sampling_date	Lat; Lon	Altitude
takou01			rain forest	Ta Kou mountain, lying Buddha	fruit	fruit	rotten		2022/12/1	10.813618;107.894315	449
takou02			rain forest	Ta Kou mountain, lying Buddha	fruit	fruit	rotten		2022/12/1	10.813618;107.894315	449
takou03			rain forest	Ta Kou mountain, lower temple	egg fruit	fruit	rotten		2022/12/1	10.812454;107.894501	419
takou04			rain forest	Ta Kou mountain, lying Buddha	fruit	fruit	rotten		2022/12/1	10.813618;107.894315	449
takou05	non-*Caenorhabditis*		rain forest	Ta Kou mountain, lying Buddha	ficus	fruit	rotten	morphology	2022/12/1	10.813618;107.894315	449
takou06			rain forest	Ta Kou mountain, lying Buddha	fruit	fruit	rotten		2022/12/1	10.813618;107.894315	449
takou07	non-*Caenorhabditis*		rain forest	Ta Kou mountain, lying Buddha	ficus	fruit	rotten	morphology	2022/12/1	10.813618;107.894315	449
takou08			rain forest	Ta Kou mountain, lying Buddha	fruit	fruit	rotten		2022/12/1	10.813618;107.894315	449
takou09	*Caenorhabditis* sp. 71	BRC20593	rain forest	Ta Kou mountain, lying Buddha	ficus	fruit	rotten	ITS2	2022/12/3	10.813618;107.894315	449
takou10	non-*Caenorhabditis*		drought forest	Ta Kou NR ranger office	*Dipterocarpus* fruit	fruit	rotten	morphology	2022/12/3	10.832177;107.936732	52
takou11	*Caenorhabditis* sp. 71	BRC20594	rain forest	Ta Kou mountain, lying Buddha	ficus	fruit	rotten	ITS2	2022/12/3	10.813618;107.894315	449
takou12			drought forest	Bung Thi swamp	flower of *Shorea roxburghii*	flower	on sand		2022/12/2	10.772901;107.931391	43
takou13			drought forest	Bung Thi swamp	fruit	fruit	on sand		2022/12/2	10.772901;107.931391	43
takou14			drought forest	Bung Thi swamp	old ant nest	animal	on tree		2022/12/2	10.772901;107.931391	43
takou15			drought forest	Bung Thi swamp	*Shorea roxburghii* liquid	tree part	on sand		2022/12/2	10.772901;107.931391	43

ITS2 genetic barcode sequencing and genetic crosses to test reproductive compatibility

For candidate *Caenorhabditis* samples, the ITS2 barcode genetic marker was then amplified using primer pair 5.8S-1 and KK-28S-22 (Kiontke et al. 2011) and sequenced. Sequences were then used to identify the closest species using BLAST homology searches against the NCBI core nucleotide database (Altschul et al. 1990, Sayers et al. 2025).

To test if the newly isolated strains (BRC20593 and BRC20594; both *Caenorhabditis* sp. 71, see Results) were likely the same species with each other or isolates of known species, crosses for reproductive compatibility were set up. For each pair of strains (species), at least three L4 males from one strain were crossed to at least three L4 females of another strain (species) and vice versa. Pairs of strains (species) that produced no viable adult F1 progeny were scored as different species. Pairs that produced many F1s adult progeny that were also able to produce viable F2 progeny were considered the same species. In this study, the other species tested were *C. brenneri* (JU1329), *C. imperialis* (EG5716), *Caenorhabditis* sp. 61 (JU4045), *Caenorhabditis* sp. 62 (JU404), *Caenorhabditis* sp. 63 (JU4113), and *Caenorhabditis* sp. 66 (JU4121) (Table 2). All strains used in this study were provided by Marie-Anne Félix and the *Caenorhabditis* Genetics Center.

Table 2. Reproductive compatibility crossing experiments with *Caenorhabditis* sp. 71

	Species	*Caenorhabditis* sp. 71	*C.* sp 61	*C.* sp 62	*C.* sp 63	*C.* sp 66	*C. imperialis*	*C. brenneri*
	Strain	BRC20594	JU4045	JU4050	JU4113	JU4121	EG5716	JU1329
Caenorhabditis sp. 71 (BRC20593)	Male x Female (N x N)	5 × 5	5 × 5	5 × 5	5 × 5	5 × 5	3 × 3	5 × 5
	Replicate crosses	3	1	1	1	1	2	1
♂	Mating behavior[1]	Yes	Inferred from plug	Not seen	Inferred from plug	Not seen	Inferred from plug	Inferred from plug
	Mating plugs	Yes	Yes	No	Yes	No	Yes	Yes
	Eggs	> 100	< 10	0	0	0	> 100	< 10
	Larvae or adults	Yes	0	NA	NA	NA	0	0
	Male x Female (N x N)	5 × 5	5 × 5	4 × 5	5 × 5	5 × 5	3 × 3	5 × 5
	Replicate crosses	3	1	1	1	1	2	1
♀	Mating behavior[1]	Yes	Inferred from plug	Inferred from plug	Mating aggregation	Inferred from plug	Inferred from plug	Inferred from plug
	Mating plugs	Yes	Yes	Yes	No	Yes	Yes	Yes
	Eggs	> 100	0	0	0	0	> 100	0
	Larvae or adults	Yes	NA	NA	NA	NA	0	NA

Low coverage genome sequencing

Low coverage genome sequencing, or genome skimming, was used to obtain the mitochondrial genome for *Caenorhabditis* sp. 71. Since the data generated here would be included for future whole genome assembly, we decided to reduce heterozygosity by first inbreeding the reference strain BRC20593. To do this, one gravid female progeny was picked and propagated for each of five generations, to yield the 5x inbred strain BRC20616. To obtain DNA, BRC20616 worms were grown on 15 *E. coli* seeded petri dishes (Ø 9-cm), collected by washing with M9 buffer, and allowed to digest residual *E. coli* by letting worms sit in M9 at room temperature for 2 hrs. After, M9 buffer was removed as much as possible, and then Qiagen buffer G2 (Qiagen Genomic-Tip 20/G kit) was added to the worms and incubated at 56°C overnight. Finally, high molecular weight genomic DNA was extracted using the Qiagen Genomic-Tip 20/G kit following the manufacturer's instructions.

DNA was sequenced using both the Oxford Nanopore Technology (ONT) and Illumina platforms. For ONT sequencing, the DNA sequencing library was prepared using the SQK-LSK114 kit (ONT), and the library was sequenced on a single MinION flowcell (FLOW-MIN114, ONT) for two hours. Raw sequence output (fast5 format) was converted into pod5 format (pod5 python package, ONT) and then duplex basecalling was done in super high accuracy mode (Dorado ver. 0.5.3). For Illumina sequencing, a sequencing library was constructed from the same DNA and sequenced using a paired-end configuration with 150 bp read lengths (Illumina NovaSeq X Plus; Genomics, Taipei, Taiwan).

Mitochondrial genome assembly and assessment

The quality of the raw ONT sequence reads was assessed using Nanoplot (De Coster and Rademakers 2023). Given that > 70% of reads had read quality > Q7 (Fig. S1) and Illumina short reads were available to polish the ONT sequencing errors, we decided to use all the raw reads for mitochondrial genome assembly. First, all ONT reads were mapped against each other using minimap2 and then the mapping results were input into miniasm for assembly (Li 2016). The mitochondrial contig was identified based on blastn similarity searches against all *Caenorhabditis* mitochondrial genomes available at NCBI (Table S1; downloaded in April 2024).

The initial mitochondrial genome assembly was polished with Illumina short reads using Polypolish (Wick and Holt 2022) and also separately with Pilon (Walker et al. 2014). The Illumina reads were first mapped to the mitochondrial genome with bwa mem (option: "-a") (Li 2013) and then the output sam (for Polypolish) or bam (for Pilon) files were used for polishing. To determine which version of the polished mitochondrial genomes was better, after each polishing iteration, both were annotated using MITOS2 (Bernt et al. 2013) on the Galaxy webserver (Galaxy Community 2024). After two rounds of polishing, the Polypolish version had three duplicated genes (*rrnL*, *atp6*, and *nad1*) and 4 genes with translational exceptions (nad1 and atp6 genes, each with two exons with non-canonical start or stop codons), while the Pilon version had only one duplicated gene (*rrnL*) and no translational exceptions. A third round of polishing with both tools was run, but no new polishing changes were made. Thus, we continued processing completeness of the mitochondrial genome with the Pilon version.

To circularize the mitochondrial genome, we mapped all raw ONT reads onto the polished genome using MUMmer 3.0 (Kurtz et al. 2004). We visualized the mapping results using the ggplot2 package in R v4.2.3 (Wickham 2016, R Core Team 2021) and identified reads that bridged the ends of the genome, then manually checked if these reads linked both ends.

Mitochondrial genome annotation

To generate the final annotation of the mitochondrial genome, the polished genome was automatic annotated using MITOS2 (Bernt et al. 2013) and MitoZ (Meng et al. 2019). This was followed by manual curation where every automatically generated gene sequence was aligned with the published *Caenorhabditis* mitochondrial gene sequences to verify the consensus start and stop sites. In addition, some start sites were adjusted to a more 5-prime start codon using ORFfinder (NCBI) (Sayers et al. 2025), based on recommended practices (Table S2) (Ghiselli et al. 2021). The map of the mitochondrial genome was rotated so that the first gene after the putative D-loop, trnP, was placed at 12 o'clock on the α-strand (clockwise strand). This sequence was the final mitochondrial genome assembly used for all subsequent analyses.

Read Depth Analysis

To estimate the read depth of the ONT and Illumina data, raw reads were first mapped onto the twice-polished mitochondrial genome using minimap2 (options: -x map-ont and sr)(Li 2016). For Illumina reads, probable PCR duplicates were removed using samtools v1.18 (Danecek et al. 2021). Read depth was then calculated using bedtools v2.30.0 (Quinlan and Hall 2010).

Phylogenetic analysis

The phylogenetic analysis focused only on the 12 protein-coding genes (PCGs) present in nematode mitochondrial genomes. At the time of the analysis, the mitochondrial genomes of 24 *Caenorhabditis* species were available on NCBI (Table S1). These mitochondrial genomes and that of *Pristionchus pacificus*, which was used as the outgroup, were retrieved (Table S1). Among these, 18 mitochondrial genomes had already been annotated. For the remaining 7 mitochondrial genomes, annotations were produced using MitoZ and MITOS2 software. All PCGs in all annotated mitochondrial genomes were manually curated with ORFfinder (NCBI) as above for *Caenorhabditis* sp. 71.

The predicted amino acid sequences of each gene were aligned using MUSCLE in MEGA 11 (Tamura et al. 2021). External gaps were marked as missing in SequenceMatrix (Vaidya et al. 2011) before phylogenic analysis. Two phylogenetic trees were constructed with IQ-TREE, employing ultra-fast bootstrap analysis with 1,000 replicates (Minh et al. 2020). In the first, individual gene trees were combined and collapsed into a single tree using ASTRAL (Zhang et al. 2018). In the alternate approach, all 12 PCGs were concatenated by species using SequenceMatrix, and an alignment of these concatenated sequences was used to build a tree (Vaidya et al. 2011). For both approaches, the best substitution model was determined with ModelFinder (Kalyaanamoorthy et al. 2017). Phylogenetic trees were visualized using FigTree v1.4.4 (Rambaut) and further refined using Adobe Illustrator.

RESULTS

Nematode collection results

In December 2022 we conducted a pilot nematode collection trip in Ta Kou Nature Reserve, which is in the south-central part of Vietnam (Fig. 1). We sampled fifteen diverse substrates from different places in the nature reserve, ranging from decomposing fruit and flowers to ant nests and tree parts (Fig. 2; Table 1). After placing these samples onto *E. coli* seeded agar plates, we recovered nematodes in five instances. Based on pharynx morphology two samples resembled *Caenorhabditis* nematodes. The other three samples had non-*Caenorhabditis* pharynxes and were not examined further. We established independent isofemale lines from the two potential *Caenorhabditis* samples under the strain IDs BRC20593 and BRC20594 (Biodiversity Research Center, Academia Sinica, Taipei, Taiwan).

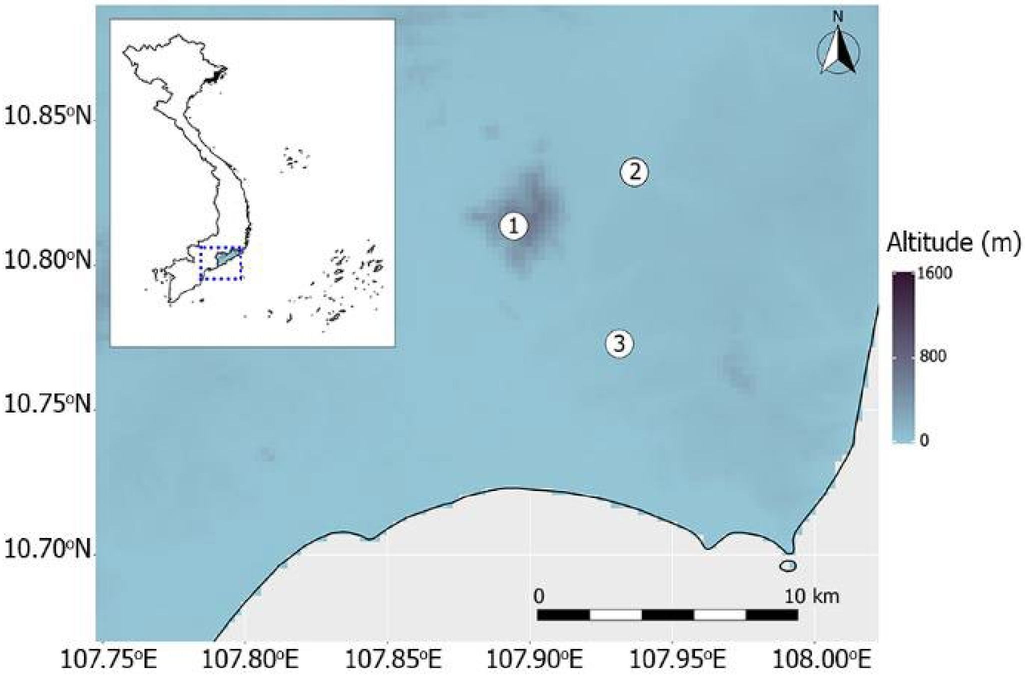

Fig. 1. Map of Ta Kou Nature Reserve region, provides key sampling and reference sites where fruit samples were collected for nematode isolation. 1. Ta Kou mountain. 2. Ta Kou NR ranger office. 3. Bung Thi swamp. The inset map shows the location of Binh Thuan province, highlighted in color.

Fig. 2. Different substrate types were collected for this Caenorhabditis species investigation. A. Rotten ficus on the ground (sample Takou09; fruit type). B. Flowers of Shorea roxburghii on sand (sample Takou12; flower type). C. An old ant nest on the tree (sample Takou14; animal type). D. Plant litter with Shorea roxburghii resin on the sand (sample Takou15; tree part type).

To determine if BRC20593 and BRC20594 were *Caenorhabditis* species, we sequenced the ITS2 region, a common barcode marker. The two sequences were nearly identical (99%, 757/766 bp identical) indicating that they were two independent isolates of the same species, which we confirmed by successful crosses (Table 2). BLASTn comparisons of the ITS2 sequence against the NCBI core-nucleotide database revealed the closest match (81.59%) was *C. brenneri* (Table S3). In *Caenorhabditis*, this level of divergence at ITS2 is indicative of separate species (Kiontke et al. 2011). Indeed, crosses with *C. brenneri* produced no progeny or only a few dead eggs (Table 2). Six other uncharacterized *Caenorhabditis* species (*Caenorhabditis* sp. 61 - 66; M.A. Félix pers. comm.) have been isolated in Vietnam and four (*Caenorhabditis* sp. 61, 62, 63, and 66) have ITS2 similarity to BRC20593 between 79 - 83% while having query coverage greater than 80% (*Caenorhabditis* sp. 64 and 65 have query coverages less than 50% and are thus very divergent; Table S4). However, given their geographical proximity we still crossed the four species with high query coverage to BRC20593. In all cases, we obtained no viable adult cross progeny and only in one case did we observe a few dead eggs (Table 2). Together these results confirm the identification of a new species. Accordingly, we have designated BRC20593 as the type strain of this novel species and given it the tentative name *Caenorhabditis* sp. 71.

Sequencing results

To obtain more genetic information we sequenced the mitochondrial genome of *Caenorhabditis* sp. 71 using a genome skimming approach (Denver et al. 2016). We used a hybrid approach combining both long and short reads from the Oxford Nanopore Technology (ONT) and Illumina platforms, respectively. First, we obtained about 114 Mb of long read sequencing data which corresponded to about 1x coverage of the presumptive nuclear genome (known *Caenorhabditis* genomes are between 63 Mb in *C. bovis* (Stevens et al. 2020) to 123 Mb in *C. inopinata* (Kanzaki et al. 2018)). ONT sequencing produced 12,732 reads with an N50 value of 29.37 kb.

Genome assembly produced eight contigs with sizes ranging from 2,341 bp to 26,915 bp. The longest, contig 2, corresponded to the partial *E. coli* genome. The small number and short sizes of the remaining contigs is consistent with the low depth of whole sequence coverage but is usually sufficient to obtain a mitochondrial genome. Indeed, BLASTn comparisons identified contig 4 as the *Caenorhabditis* sp. 71 mitochondrial sequence. The length of contig 4 was 13,612 bp, which is near the average length of *Caenorhabditis* mitochondrial genomes. Median ONT coverage for contig 4 was 65-fold.

Despite 65-fold coverage, pure ONT assemblies still contain errors. To improve accuracy, we obtained about 6 Gb of Illumina sequence data and used it to polish the mitochondrial contig with Pilon (Walker et al. 2014). Two rounds of polishing corrected a total of 133 errors (first round

120 changes; second round 13 changes with 6 reverted to the original sequence) resulting in a penultimate contig sequence. After polishing, the mitochondrial contig length increased from 13,612 bp to 13,654 bp. Finally, we confirmed the circularization of the mitochondrial genome with two independent ONT reads.

Mitochondrial genome annotation

Annotation with MITOS2 and MitoZ, followed by manual refinement (see Materials and Methods), identified 36 genes in the *Caenorhabditis* sp. 71 mitochondrial genome, including 12 protein coding genes (PCGs), 22 transfer RNA genes (tRNAs), and two ribosomal RNAs (Table 3). All annotated genes are on the α-strand (clockwise orientation) like other *Caenorhabditis* mitochondrial genomes. The gene pairs ND6 and ND4L as well as ND1 and ATP6 overlap slightly but in different reading frames, which is typical of nematode mitochondrial genomes. The base composition of the mitochondrial genome is 24% G+C, like most known *Caenorhabditis* mitochondrial genomes, which range from 22 to 26% G+C (Table S1). The map of the mitochondrial genome was rotated so that the first nucleotide corresponded to the first gene (trnP) on the α-strand after the likely control region (Fig. 3A).

Table 3. List of annotated genes in the *Caenorhabditis* sp. 71 mitochondrial genome

Gene type	Gene name	Gene label	Start	Stop	Start codon	Decoding tRNA
CDS	NADH dehydrogenase subunit 6	ND6	94	549	ATG	
CDS	NADH dehydrogenase subunit 4L	ND4L	546	785	ATA	
CDS	NADH dehydrogenase subunit 1	ND1	1765	2643	TTG	
CDS	ATP synthase F0 subunit	ATP6	2621	3238	ATG	
CDS	NADH dehydrogenase subunit 2	ND2	3421	4269	TTG	
CDS	cytochrome b	CYTB	4475	5620	ATG	
CDS	cytochrome c oxidase subunit 3	COX3	5681	6448	ATA	
CDS	NADH dehydrogenase subunit 4	ND4	6488	7738	ATT	
CDS	cytochrome c oxidase subunit 1	COX1	7801	9423	ATA	
CDS	cytochrome c oxidase subunit 2	COX2	9634	10350	ATC	
CDS	NADH dehydrogenase subunit 3	ND3	11359	11700	ATC	
CDS	NADH dehydrogenase subunit 5	ND5	11702	13285	ATT	
rRNA	ribosomal RNA small subunit	rrnS	902	1596		

rRNA	ribosomal RNA large subunit	rrnL	10411	11364			
tRNA	tRNA-Pro	trnP	1	55			UGG
tRNA	tRNA-Val	trnV	59	114			UAC
tRNA	tRNA-Trp	trnW	790	846			UCA
tRNA	tRNA-Glu	trnE	847	901			UUC
tRNA	tRNA-Ser	trnS	1597	1649			UGA
tRNA	tRNA-Asn	trnN	1650	1706			GUU
tRNA	tRNA-Tyr	trnY	1710	1767			GUA
tRNA	tRNA-Lys	trnK	3247	3309			UUU
tRNA	tRNA-Leu	trnL	3310	3364			UAA
tRNA	tRNA-Ser	trnS	3365	3420			UCU
tRNA	tRNA-Ile	trnI	4273	4334			GAU
tRNA	tRNA-Arg	trnR	4335	4390			ACG
tRNA	tRNA-Gln	trnQ	4391	4445			UUG
tRNA	tRNA-Phe	trnF	4451	4507			GAA
tRNA	tRNA-Leu	trnL	5624	5680			UAG
tRNA	tRNA-Thr	trnT	6453	6508			UGU
tRNA	tRNA-Cys	trnC	9424	9479			GCA
tRNA	tRNA-Met	trnM	9484	9543			CAU
tRNA	tRNA-Asp	trnD	9544	9598			GUC
tRNA	tRNA-Gly	trnG	9599	9654			UCC
tRNA	tRNA-His	trnH	10355	10410			GUG
tRNA	tRNA-Ala	trnA	13297	13353			UGC

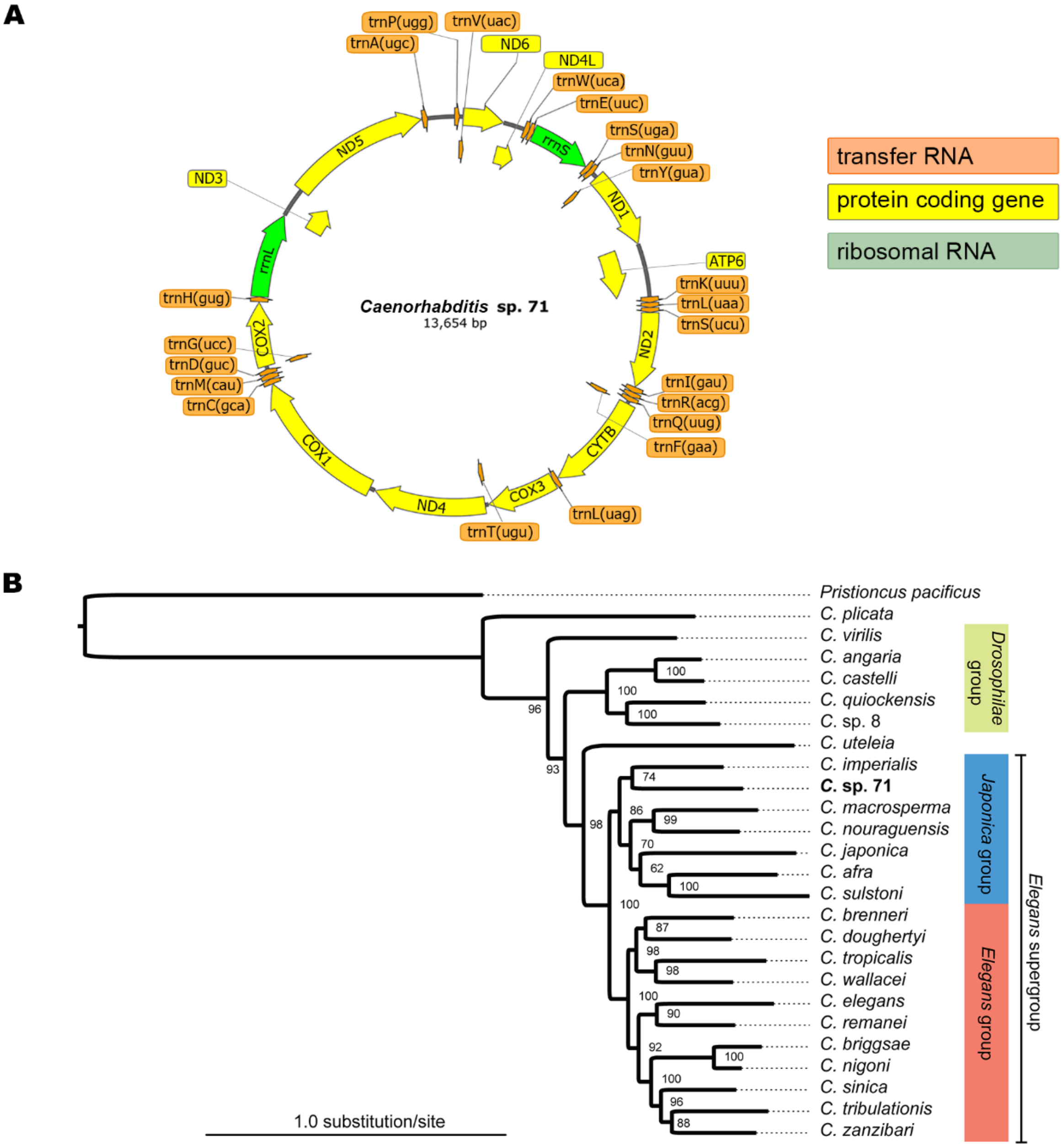

Fig. 3. Complete annotated mitochondrial genome of Caenorhabditis sp. 71 and its phylogenetic relationship with other Caenorhabditis species. A. Circular map of the mitochondrial genome of C. sp. 71, showing 12 protein-coding genes, 22 transfer RNA genes, and 2 ribosomal RNA genes. PCGs were refined using ORFfinder and extended to the longest possible open reading frames. B. Maximum likelihood phylogeny of 25 Caenorhabditis species and Pristionchus pacificus based on the concatenated amino acid sequences of 12 mitochondrial PCGs. The tree was constructed using the TVM+F+R4 model. C. sp. 71 is highlighted in bold. Bootstrap support values are indicated at the nodes. Scale bar represents substitutions per site.

Phylogenic analysis

We estimated the evolutionary relationship of *Caenorhabditis* sp. 71 within *Caenorhabditis* through phylogenetic analyses using the 12 mitochondrial PCG sequences. We included 25 mitochondrial genomes: 24 *Caenorhabditis* species and *Pristionchus pacificus* as the outgroup

(Table S1). Among these, 7 genomes lacked annotations, therefore, we annotated them (see MATERIALS AND METHODS, Table S1, Fig. S4). All annotated PCGs were re-annotated with ORFfinder to the more upstream available start codon (Fig. S5–S6).

To investigate the evolutionary relationships between *Caenorhabditis* sp. 71 and other *Caenorhabditis* species, we conducted two independent analyses. The first analysis, using a concatenated sequence of 12 mitochondrial PCG sequences under the TVM+F+R4 model, placed *Caenorhabditis* sp. 71 sister to *C. imperialis* with 74% bootstrap support (Fig. 3B). In the second, a consensus tree of the 12 individual PCG trees grouped *Caenorhabditis* sp. 71 with *C. macrosperma* and *C. nouraguensis*, with *C. imperialis* as a sister to this clade. However, the bootstrap support for the relationship between *Caenorhabditis* sp. 71 and the group of *C. macrosperma* and *C. nouraguensis* was low (47%) (Fig. S3B).

To test reproductive isolation between *Caenorhabditis* sp. 71 and its likely sister *C. imperialis*, based on the first analysis, we conducted crossing experiments to assess reproductive compatibility. The crosses produced only dead eggs, further confirming the new species status of *Caenorhabditis* sp. 71 (Table 2).

DISCUSSION

This study reports the isolation and characterization of a novel *Caenorhabditis* species, *Caenorhabditis* sp. 71, and provides the first complete mitochondrial genome for a *Caenorhabditis* species collected in Vietnam.

A nematode diversity study in French Guiana suggested that tropical environments have high level of *Caenorhabditis* diversity (Felix et al. 2013). The study conducted deep sampling of a broad range of substrates (a total of 184 samples) and revealed six different *Caenorhabditis* species, four of which were newly described and three of which have, so far, only been reported from that region. In our study, the discovery of *Caenorhabditis* sp. 71 from a small nature reserve, following a limited sampling effort (total of 15 samples) underscores the high potential for uncovering further *Caenorhabditis* diversity in the country.

Our findings suggest that Vietnam, a tropical country with complex topography and diverse microhabitats, may harbor a rich and largely undocumented assemblage of *Caenorhabditis* species, both known and undescribed. Indeed, to date, research on *Caenorhabditis* diversity in Vietnam has been limited, with only six named species reported: *C. briggsae* (Lê et al. 2023), *C. brenneri* (Lê et al. 2024), *C. sinica* (Le et al. 2023a), *C. tropicalis*, *C. zanzibari*, and *C. yunquensis* (Le et al. 2023b). Additionally, six undescribed species (*Caenorhabditis* sp. 61 to 66) have been identified by

M.A. Félix (pers. comm.). Further systematic and geographically broad surveys, incorporating both morphological and molecular tools, are likely to reveal more species and contribute significantly to global efforts to map nematode diversity.

The complete annotated mitochondrial genome of *Caenorhabditis* sp. 71 is 13,654 bp in length, consistent with the sizes observed across other *Caenorhabditis* species. Its gene content and organization—including 12 PCGs, 22 tRNA genes, and two rRNA genes—are typical of the genus, reaffirming the highly conserved nature of *Caenorhabditis* mitochondrial genomes.
However, refinement of the PCG annotations using ORFfinder across all mitochondrial genomes included in this study revealed that a notable proportion of the protein-coding genes were likely misannotated in the initial automated pipeline, particularly in terms of start codon position. In all analyzed *Caenorhabditis* mitochondrial genomes, we found that ORFfinder often identified a start codon upstream of the annotated start codons, suggesting that previous annotations may underestimate the true length of several mitochondrial proteins. These discrepancies highlight the importance of manual curation or additional validation tools when annotating mitochondrial genomes, particularly for non-model species where gene structure may deviate slightly from annotated references (Ghiselli et al. 2021). This observation also raises broader concerns about the consistency and accuracy of mitochondrial genome annotations in public databases, especially in cases where gene prediction relies solely on automated tools. For comparative genomic and phylogenetic studies, such inconsistencies can lead to errors in alignment and inference. Therefore, careful reannotation, such as that applied here, can improve the accuracy of downstream analyses and increase confidence in evolutionary interpretations.

Our phylogenetic analysis using a concatenated alignment of 12 mitochondrial PCGs is largely consistent with previous nuclear and whole-genome studies (Kiontke et al. 2011; Stevens et al. 2019; Stevens et al. 2020), notably placing *C. plicata* as the basal lineage to both the *Drosophilae* and *Elegans* supergroups (Fig. 3B). In this tree, *Caenorhabditis* sp. 71 is positioned as sister to *C. imperialis* with moderate bootstrap support (74%), a result that contrasts with ITS2 BLASTn comparisons. Conversely, individual gene trees suggest a closer relationship to the *C. macrosperma - C. nouraguensis* clade, albeit with weak support (47%), likely due to the limited phylogenetic signal from short gene sequences and potential asynchronous evolution among mitochondrial genes (Fig. S3). While mitochondrial phylogenies have previously aligned well with whole-genome analyses (Stevens et al. 2019. 2020), some branches in our phylogeny show weaker support, with bootstrap values below 90%, compared to the relationships inferred from whole-genome data. The conflicting signals in *Caenorhabditis* sp. 71 suggest that its mitochondrial and nuclear genomes may have followed different evolutionary trajectories. Thus, nuclear markers or full-genome sequencing will be essential to resolve its precise phylogenetic position (Kern et al. 2020).

CONCLUSIONS

In this study, we report a new *Caenorhabditis* species, *Caenorhabditis* sp. 71, isolated from Ta Kou Nature Reserve in Binh Thuan province, Viet Nam. The complete mitochondrial genome of the species was sequenced and assembled using ONT long-reads and polished with Illumina short-reads. The *Caenorhabditis* sp. 71 mitochondrial genome is 13,654 bp long, with 12 protein coding genes, 22 tRNAs, two rRNAs, and one control region. A phylogenic tree based on the concatenation of the 12 protein coding genes revealed that *Caenorhabditis* sp. 71 is a likely sister of *C. imperialis*.

List of abbreviations

GPS, the geographic coordinate system.
ITS2, internal transcribed spacer 2 which is the intergenic region between the 5.8S and 16S rRNA genes.
IUCN, International Union for Conservation of Nature.
ONT, the Oxford Nanopore Technology. PCG, protein coding gene.

Acknowledgments: We thank Marie-Anne Félix for some nematode strains and their ITS2 sequences, and Chen-Yi Dai for assistance with submitting raw sequence data to GenBank NCBI. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). The study is supported by VAST under the decision number 2396/QĐ-VHL for the project "Research, simulate, and recreate the biodiversity of the South-Central region of Vietnam" for V. D. D. and by NSTC, Taiwan (111-2311-B-001-036-MY3) and Biodiversity Research Center, Academia Sinica to J.W.

Authors' contributions: V.D.D. conducted sample collection, wild strain isolation, generation of the 5× inbred line, all molecular experiments, and data analysis. J.W. and J.C.H. performed the crossing experiments. V.D.D. and J.W. prepared the manuscript. All authors read and approved the final version of the manuscript.

Competing interests: All authors declare that they have no competing interests

Availability of data and materials: ITS2 sequences of the two *Caenorhabditis* sp. 71 strains **(**BRC20593, BRC20594), the mitochondrial genome sequence, and whole genome sequencing sequence data (Oxford Nanopore Technology and Illumina) have been submitted to GenBank NCBI and are available under the accession numbers PV613534, PV613535, PV773521, and PRJNA1272496.

Consent for publication:

Ethics approval consent to participate:

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol **215**:403–410. doi:10.1016/S0022-2836(05)80360-2.

Baltzer MC, Nguyen TD, Shore RG. 2001. Towards a vision for biodiversity conservation in the forests of the Lower Mekong ecoregion complex. Hanoi and Washington D.C., WWF Indochina/WWF US.

Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsch G, Putz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol **69**:313–319. doi:10.1016/j.ympev.2012.08.023.

Blair M, Nadler T, Ni O, Samun E, Streicher U, Nekaris K. 2021. *Nycticebus pygmaeus* (amended version of 2020 assessment). The IUCN Red List of Threatened Species: e. T14941A198267330.

Bresseel J, Constant J. 2018. Two new stick insect genera from Vietnam, *Nuichua* gen. nov. and *Pterohirasea* gen. nov. with two new species (Phasmida: Diapheromeridae: Necrosciinae). Belgian Journal of Entomology **70:**1–29.

Bresseel J, Constant J. 2024. Stick insects from Vietnam: The new genus *Mycovartes* gen. nov., with two new species and two new species of Neooxyartes Ho, 2018 (Phasmida: Lonchodidae: Necrosciinae). European Journal of Taxonomy **965:**1–50. doi:10.5852/ejt.2024.965.2695.

Chen IS, Nguyen AT, Huynh QT. 2024. A new freshwater goby of *Rhinogobius* Gill, 1859 (Teleostei, Gobiidae) from South Central Vietnam. Zootaxa **5550:**354–361. doi:10.11646/zootaxa.5550.1.35.

Constant J, Pham HT. 2024. Sixteen issid planthopper species in one day in Dong Son-Ky Thuong Nature Reserve in North Vietnam: Eight new species, one new genus and additional new records (Hemiptera: Fulgoromorpha: Issidae). European Journal of Taxonomy **919:**81–87. doi:10.5852/ejt.2024.919.2407.

Crombie TA, Battlay P, Tanny RE, Evans KS, Buchanan CM et al. 2022. Local adaptation and spatiotemporal patterns of genetic diversity revealed by repeated sampling of *Caenorhabditis elegans* across the Hawaiian Islands. Mol Ecol **31:**2327–2347. doi:10.1111/mec.16400.

Crombie TA, Zdraljevic S, Cook DE, Tanny RE, Brady SC et al. 2019. Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. Elife **8:**e50465. doi:10.7554/eLife.50465.

Cumming RT, Bank S, Bresseel J, Constant J, Le Tirant S, Dong Z, Sonet G, Bradler S. 2021. *Cryptophyllium*, the hidden leaf insects - descriptions of a new leaf insect genus and thirteen species from the former celebicum species group (Phasmatodea, Phylliidae). Zookeys **1018:**1–179. doi:10.3897/ZOOKEYS.1018.61033.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V et al. 2021. Twelve years of SAMtools and BCFtools. Gigascience **10:**1–4. doi:10.1093/gigascience/giab008.

De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics **39:**btad311. doi:10.1093/bioinformatics/btad311.

Denver DR, Brown AM, Howe DK, Peetz AB, Zasada IA. 2016. Genome skimming: a rapid approach to gaining diverse biological insights into multicellular pathogens. PLoS Pathog **12:**e1005713. doi:10.1371/journal.ppat.1005713.

Félix MA, Braendle C. 2010. The natural history of *Caenorhabditis elegans*. Curr Biol **20:**R965–R969. doi:10.1016/j.cub.2010.09.050.

Felix MA, Braendle C, Cutter AD. 2014. A streamlined system for species diagnosis in *Caenorhabditis* (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. PLoS ONE **9:**e94723. doi:10.1371/journal.pone.0094723.

Felix MA, Duveau F. 2012. Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and *C. briggsae*. BMC Biol **10:**59. doi:10.1186/1741-7007-10-59.

Felix MA, Jovelin R, Ferrari C, Han S, Cho YR et al. 2013. Species richness, distribution and genetic diversity of *Caenorhabditis* nematodes in a remote tropical rainforest. BMC Evol Biol **13:**10. doi:10.1186/1471-2148-13-10.

Ferrari C, Salle R, Callemeyn-Torre N, Jovelin R, Cutter AD et al. 2017. Ephemeral-habitat colonization and neotropical species richness of *Caenorhabditis* nematodes. BMC Ecol **17:**43. doi:10.1186/s12898-017-0150-z.

Galaxy Community. 2024. The Galaxy platform for accessible, reproducible, collaborative data analyses: 2024 update. Nucleic Acids Res **52:**W83–W94. doi:10.1093/nar/gkae410.

Ghiselli F, Gomes-dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J et al. 2021. Molluscan mitochondrial genomes break the rules. Philos T Roy Soc B **376:**20200159. doi:10.1098/rstb.2020.0159.

Hansen MF, Ang A, Trinh TTH, Sy E, Paramasivam S et al. 2022. *Macaca fascicularis* (amended version of 2022 assessment). Page e.T12551A221666136. doi:10.2305/IUCN.UK.2022-2.RLTS.T12551A221666136.en.

Hoang MD, Tran VN, Covert HHH, Luu HT, Tran QT. 2010. Conservation status of primates in Ta Kou Nature Reserve. Conservation of primates in Indochina, pp. 91–98.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods **14**:587–589. doi:10.1038/nmeth.4285.

Kanzaki N, Tsai IJ, Tanaka T, Hunt VL, Liu D et al. 2018. Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. Nat Commun **9:**3216. doi:10.1038/s41467-018-05712-5.

Kern EM, Kim T, Park J-K. 2020. The mitochondrial genome in nematode phylogenetics. Front Ecol Evol **8:**250. doi:10.3389/fevo.2020.00250.

Kiontke KC, Felix MA, Ailion M, Rockman MV, Braendle C et al. 2011. A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. BMC Evol Biol **11:**339. doi:10.1186/1471-2148-11-339.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M et al. 2004. Versatile and open software for comparing large genomes. Genome Biol **5:**R12. doi:10.1186/gb-2004-5-2-r12.

Lê TS, Bùi TMH, Hà BH, Nguyễn TT. 2024. Sự phân bố không đồng nhất của tuyến trùng *Caenorhabditis brenneri* ở vườn quốc gia Cát Tiên và Cúc Phương. Vietnam Journal of Forest Science.

Lê TS, Hà BH, Bùi TMH, Nguyễn TT. 2023. Đa dạng di truyền của tuyến trùng *Caenorhabditis briggsae* ở tỉnh Ninh Bình, Đồng Nai và Lâm Đồng, Việt Nam. Vietnam Journal of Forest Science.

Le TS, Nguyen THG, Nguyen TT, Do THL. 2023a. Wild-type *Caenorhabditis sinica*, a model nematode for speciation and evolution, massively found in Vietnam. Vietnam Journal of Biotechnology **21:**543–548. doi:10.15625/1811-4989/19494.

Le TS, Nguyen TTH, Nguyen TT. 2023b. Three rare dioecious *Caenorhabditis* nematode species (*C. tribulationis*, *C. yunquensis*, *C. zanzibari*) do not live in the same habitats. Vietnam J Biotechnol **21:**759–764. doi:10.15625/1811-4989/19552.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997v2.

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics **32:**21032110. doi:10.1093/bioinformatics/btw152.

Luu HT, Hsieh C-L, Chuang C-R, Chen C-W, Tran NT et al. 2023. *Langbiangia*, a new genus of Gesneriaceae endemic to Langbiang Plateau, southern Vietnam and a taxonomic endeavor to achieve key targets of the post-2020 global biodiversity framework. PLoS ONE **18:**e0284650. doi:10.1371/journal.pone.0284650.

Ly N, Luu H. 2007. Ecological characteristics of forest types on Ta Kou Mountain, Binh Thuan Province, pp. 556–562. *In*: Proceedings of the 2nd National Scientific Conference on Ecology and Biological Resources. Hanoi.

Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res **47:**e63. doi:10.1093/nar/gkz173.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD et al. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol **37:**1530–1534. doi:10.1093/molbev/msaa015.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:**841–842. doi:10.1093/bioinformatics/btq033.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Available at: https://www.R-project.org/.

Rambaut A. 2016. FigTree. Molecular Evolution, Phylogenetics and Epidemiology. Last modified November 28, 2018. Available at: https://github.com/rambaut/figtree.

Sayers EW, Beck J, Bolton EE, Brister JR, Chan J et al. 2025. Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Res **53:**D20–D29. doi:10.1093/nar/gkae979.

Sloat SA, Noble LM, Paaby AB, Bernstein M, Chang A et al. 2022. *Caenorhabditis* nematodes colonize ephemeral resource patches in neotropical forests. Ecol Evol **12:**e9124. doi:10.1002/ece3.9124.

Stevens L, Felix MA, Beltran T, Braendle C, Caurcel C et al. 2019. Comparative genomics of 10 new *Caenorhabditis* species. Evol Lett **3:**217–236. doi:10.1002/evl3.110.

Stevens L, Rooke S, Falzon LC, Machuka EM, Momanyi K et al. 2020. The Genome of *Caenorhabditis bovis*. Curr Biol **30:**1023–1031.e1024. doi:10.1016/J.CUB.2020.01.074.

Sudhaus W. 2011. Phylogenetic systematisation and catalogue of paraphyletic "Rhabditidae" (Secernentea, Nematoda). Journal of nematode morphology and systematics **14:**113–178.

Sun S, Kanzaki N, Dayi M, Maeda Y, Yoshida A et al. 2022. The compact genome of *Caenorhabditis niphades* n. sp., isolated from a wood-boring weevil, *Niphades variegatus*. BMC Genomics **23:**765. doi:10.1186/s12864-022-09011-8.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol **38:**3022–3027. doi:10.1093/molbev/msab120.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics **27:**171–180. doi:doi.org/10.1111/j.1096-0031.2010.00329.x.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE **9:**e112963. doi:10.1371/journal.pone.0112963.

Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol **18:**e1009802. doi:10.1371/journal.pcbi.1009802.

Wickham H. 2016. Programming with ggplot2. *In*: ggplot2: elegant graphics for data analysis, pp. 241–253. Cham: Springer International Publishing.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics **19:**153. doi:10.1186/s12859-018-2129-y.

Supplementary materials

Table S1. Mitochondrial genomes of *Pristionchus pacificus* and *Caenorhabditis* species available in NCBI that were used in the study.

Table S2. Positions of protein-coding genes (PCGs) in the mitochondrial genome of *Caenorhabditis* sp. 71. Genes were initially annotated using MITOS2 and subsequently refined with ORFfinder to extend the coding regions to the earliest available upstream start codon.

Table S3. BLASTn results of the ITS2 sequence of *Caenorhabditis* sp. 71 against the NCBI core nucleotide database. The results support the distinctiveness of *Caenorhabditis* sp. 71, with the highest similarity (81.59%) observed with *C. brenneri*, indicating substantial divergence consistent with species-level differentiation.

Table S4. ITS2 blast results against uncharacterized *Caenorhabditis* spp. isolated in Vietnam

Fig. S1. Distributions of read lengths and read qualities of the ONT run on whole *Caenorhabditis* sp. 71 genomic DNA.

Fig. S2. Read coverage of Illumina and ONT data on the assembled *Caenorhabditis* sp. 71 mitochondrial genome. A. Illumina data. B. ONT data. Left, read coverage at each position. Right, violin plots of read coverage over the whole mitochondrial genome.

Fig. S3. Phylogenetic analysis of *Caenorhabditis* sp. 71 and related species. A, Maximum likelihood phylogeny constructed using a concatenated alignment of 12 mitochondrial protein-coding genes. *Caenorhabditis* sp. *71* is placed as the sister species to *C. imperialis* with moderate bootstrap support, indicating a potential close evolutionary relationship. B, Consensus phylogeny inferred by collapsing individual trees generated from each of the 12 mitochondrial protein-coding genes. In this analysis, *Caenorhabditis* sp. 71 is recovered as sister to the *C. nouraguensis–C. macrosperma* clade, suggesting alternative evolutionary placement. Both analyses place *Caenorhabditis* sp. 71 within the Japonica group. Numbers at nodes represent bootstrap support values, indicating confidence in each split.

Fig. S4. Published *Caenorhabditis* mitochondrial genomes that were annotated in this study. A. *Caenorhabditis imperialis* OY751545. B. *Caenorhabditis* OY754348 C. *Caenorhabditis sinica*

OZ009944 Annotated mitochondrial genome. D. *Caenorhabditis* sp. 8 OY751552 E. *Caenorhabditis sulstoni* OZ016481. F. *Caenorhabditis uteleia* OY752139. G. *Caenorhabditis zanzibari* OZ016474.

Fig. S5. Published annotated *Caenorhabditis* mitochondrial genomes that were refined with ORFfinder in this study. A. *Caenorhabditis afra* NC_035252. B. *Caenorhabditis angaria* NC_035246 C. *Caenorhabditis brenneri* NC_035244. D. *Caenorhabditis briggsae* NC_009885 E. *Caenorhabditis castelli* NC_035253. F. *Caenorhabditis doughertyi* NC_035247. G. *Caenorhabditis elegans* NC_001328. H. *Caenorhabditis macrosperma* NC_035251. I. *Caenorhabditis nigoni* KP259621. J. *Caenorhabditis nouraguensis* NC_035250. K. *Caenorhabditis plicata* NC_035245. L. *Caenorhabditis remanei* NC_035243. M. *Caenorhabditis quiockensis* KY552911. N. *Caenorhabditis tribulationis* OL362111. O. *Caenorhabditis tropicalis* NC_025756. P. *Caenorhabditis virilis* NC_035248. Q. *Caenorhabditis wallacei* NC_035249.

Fig. S6. Published annotated *Pristionchus pacificus* mitochondrial genome (NC_015245) that was refined with ORFfinder in this study.