



OPEN

Additional description and genome analyses of *Caenorhabditis auriculariae* representing the basal lineage of genus *Caenorhabditis*

Mehmet Dayi^{1,2,7}, Natsumi Kanzaki^{3,7}, Simo Sun¹, Tatsuya Ide⁴, Ryusei Tanaka¹, Hayato Masuya⁵, Kimiko Okabe⁵, Hisashi Kajimura⁶ & Taisei Kikuchi¹✉

Caenorhabditis auriculariae, which was morphologically described in 1999, was re-isolated from a *Platydemus* mushroom-associated beetle. Based on the re-isolated materials, some morphological characteristics were re-examined and ascribed to the species. In addition, to clarify phylogenetic relationships with other *Caenorhabditis* species and biological features of the nematode, the whole genome was sequenced and assembled into 109.5 Mb with 16,279 predicted protein-coding genes. Molecular phylogenetic analyses based on ribosomal RNA and 269 single-copy genes revealed the species is closely related to *C. sonorensis* and *C. monodelphis* placing them at the most basal clade of the genus. *C. auriculariae* has morphological characteristics clearly differed from those two species and harbours a number of species-specific gene families, indicating its usefulness as a new outgroup species for *Caenorhabditis* evolutionary studies. A comparison of carbohydrate-active enzyme (CAZy) repertoires in genomes, which we found useful to speculate about the lifestyle of *Caenorhabditis* nematodes, suggested that *C. auriculariae* likely has a life-cycle with tight-association with insects.

The nematode, *Caenorhabditis elegans*, has been a core model organism that is used in a wide range of biological and medical studies for the last several decades and has led to a number of key discoveries, including the molecular mechanisms of apoptosis¹ and gene silencing by small RNAs². Recently, *C. elegans* and other species of the genus *Caenorhabditis* have been developed as a useful model system for a wide range of evolutionary studies^{3,4}.

Currently, the genus *Caenorhabditis* contains 49 nominal species^{5–8}. However, because of the morphological similarity among close relatives and morphological divergence within each species, the species status is often delimited on molecular barcoding and/or hybridization analyses⁴, and several species have been described only based on molecular phylogenetic status and mating studies⁶. Meanwhile, some previously described species, including *C. anthobia*, *C. avicola*, *C. clavopapillata*, and *C. craspedocercus*, were characterized based solely on their morphological traits (Table 1, Supplementary Text). For those species, re-isolation followed by molecular profiling and/or re-characterization of typological characteristics is demanded.

Caenorhabditis auriculariae was initially described morphologically as an associate of the fruiting bodies of the basidiomycota fungus *Auricularia polytricha*⁹. Based on morphological characteristics, the species is relatively easily distinguished from other congeners⁹. For example, the species has a short stoma, bifurcated metastegostomatal teeth, a spicule structure similar to the *drosophilae* super group, and closed bursa similar to several species in the *drosophilae* and *elegans* super group species. Thus, this species seems to be rather basal in the genus, but the phylogenetic status of the species remained undetermined.

In the present study, using re-isolated materials, we examined *C. auriculariae* morphological characteristics in details, and defined the basic molecular profiles based on nuclear ribosomal sequences. Additionally, we sequenced the whole genome of the species and revealed its basic genomic features. These results revealed the

¹Department of Infectious Disease, Faculty of Medicine, University of Miyazaki, 5200 Kiyotakecho Kihara, Miyazaki-City, Miyazaki 889-1692, Japan. ²Forestry Vocational School, Duzce University, 81620 Duzce, Turkey. ³Kansai Research Center, Forestry and Forest Products Research Institute, 68 Nagaikyutaroh, Momoyama, Fushimi, Kyoto 612-0855, Japan. ⁴Department of Zoology, National Museum of Nature and Science, 4-1-1 Amakubo, Tsukuba, Ibaraki 305-0005, Japan. ⁵Forestry and Forest Products Research Institute, 1 Matsunosato, Tsukuba, Ibaraki 305-8687, Japan. ⁶Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. ⁷These authors contributed equally: Mehmet Dayi and Natsumi Kanzaki. ✉email: taisei_kikuchi@med.miyazaki-u.ac.jp

Species name with taxonomic author(s)	Type locality	Type host or habitat	Previous species code (type strain)	Reference (original description)
<i>C. afra</i> Félix, Braendle & Cutter, 2014	Begoro, Ghana	Rotting citrus fruit	sp. 7 (JU1199)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. angaria</i> Sudhaus, Kiontke & Giblin-Davis, 2011	Frog Pond, Dade County, Florida, USA	Sugarcane weevil <i>Metamasius hemipterus</i> (Olivier) (Curculionidae)	sp. 3 (PS1010)	Sudhaus et al. (2011) Nematology 13: 61–78
<i>C. anthobia</i> (W. Schneider, 1937)	South-Sumatra	Water in the inflorescences of <i>Cyrtandra glabra</i> (Gesneriaceae)		Schneider (1937) Arch Hydrobiol (Suppl.) 15: 30–108
<i>C. auriculariae</i> Tsuda & Futai, 1999	North Campus of Kyoto University, Kyoto, Japan	Fruiting bodies of the cloud ear fungus <i>Auricularia polytricha</i> (Montagne) (Auriculariaceae), generated on a trunk of the red elder <i>Sambucus racemosa</i> L. (Adoxaceae)		Tsuda & Futai (1999) Jpn J Nematol 29: 18–23
<i>C. avicola</i> Schmidt & Kuntz, 1972	Chien-shih, Hsinchu Hsien, Taiwan	Small intestine of a plumbeous water redstart <i>Rhyacornis fuliginosus affinis</i> Vigors (Muscicapidae)		Schmidt & Kuntz (1972) Proc Helminthol Soc Wash 39: 189–191
<i>C. becei</i> Stevens & Félix, 2019	Barro Colorado Island, Panama	Rotting flower of <i>Gustavia superba</i>	sp. 29 (QG704)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. bovis</i> (Kreis, 1964)	Mkwaya-Ranch of Amboni Est. Ltd., coastal area of Tanzania (East Africa)	Outer auditory canal of male zebu, <i>Bos</i> sp. (Bovidae)		Kreis (1964) Schweiz Arch Tierheilkd 106: 372–378
<i>C. brenneri</i> Sudhaus & Kiontke, 2007	Bohorok, Sumatra	Compost-like material, mostly banana leaves	sp. 4 (SB129)	Sudhaus & Kiontke (2007) Zootaxa 1456: 45–62
<i>C. briggsae</i> (Dougherty & Nigon, 1949)	Campus of Stanford University, Palo Alto, California, USA	Soil		Dougherty & Nigon (1949) J Parasitol (Suppl.) 35: 11
<i>C. castelli</i> Félix, Braendle & Cutter, 2014	“Petit Plateau” near the CNRS Biological station, Nouragues, French Guiana	Rotting <i>Micropholis cayennensis</i> fruit	sp. 12 (JU1426)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. chinkari</i> Mondal & Manna, 2015	Alipore Zoological Garden, Kolkata, West Bengal, India	Faeces of captive chinkara, <i>Gazella gazella bennettii</i> (Sykes) (Bovidae)		Mondal & Manna (2015) Proc Zool Soc (Kolkata) 68: 52–57
<i>C. clavopapillata</i> (Kreis & Faust, 1933)	New Orleans, Louisiana, USA	Faeces-soiled hair in the perianal region of a dog		Kreis & Faust (1933) Trans Amer Microscop Soc 52: 162–172
<i>C. craspedocercus</i> (Völk, 1950)	Erlangen, Germany	Earthworm, <i>Eisenia rosea</i> (Savigny) (Lumbricidae)		Völk (1950) Zool Jahrb (Systematik) 79: 1–70
<i>C. doughertyi</i> Félix, Braendle & Cutter, 2014	Angela Spice Garden a few km from Periyar, Kerala, India	Rotting cacao fruit	sp. 10 (JU1333)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. drosophilae</i> (Kiontke, 1997)	Pima County, Tucson Mountains, Arizona, USA	Decaying tissue of saguaro cactus, <i>Carnegiea gigantea</i> (Engelm.) (Cactaceae); dauer juveniles phoretic on <i>Drosophila nigrospiracula</i> Patterson and Wheeler (Drosophilidae)		Kiontke (1997) Fundam Appl Nematol 20: 305–315
<i>C. elegans</i> (Maupas, 1899)	Surroundings of Algiers, Algeria	Rich humus		Maupas (1899) Arch Zool Exp Gen 7: 563–628
<i>C. formosana</i> (Yokoo & Okabe, 1968)	Puyen Village, Changhua County, Taiwan	Amphibious snail <i>Oncomelania hupensis formosana</i> Gredler (Rissoidea)		Yokoo & Okabe (1968) Agric Bull Saga Univ 25: 69–78
<i>C. fruticicola</i> (Shinohara, 1960)	Monkara, Miyazaki, Japan	Intestine of the snail <i>Fruticicola (Acusta) sieboldiana</i> (Pfeiffer) (Helicoidea)		Shinohara (1960) J Kurume Med Assoc 23: 2777–2819
<i>C. genitalis</i> (Scheiber, 1880)	Stuhlweissenburg, Hungary	Outer genitals of a bedfast woman		Scheiber (1880) Arch Pathol Anat Physiol Klin Med 82: 161–175
<i>C. guadeloupensis</i> Félix, Braendle & Cutter, 2014	Soufrière, Forest trail, Guadeloupe, France (Overseas department)	Rotten <i>Heliconia</i> flowers	sp. 20 (NIC113)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. imperialis</i> Félix, Braendle & Cutter, 2014	Moorea, French Polynesia	Rotting horse chestnut	sp. 14 (EG5716)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. inopinata</i> Kanzaki, Tsai, Tanaka, Hunt, Liu, Tsuyama, Maeda, Namai, Kumagai, Tracey, Holroyd, Doyle, Woodruff, Murase, Kitazume, Chai, Akagi, Panda, Ke, Schroeder, Wang, Berriman, Sternberg, Sugimoto & Kikuchi, 2018	Ishigaki Isl., Okinawa, Japan	Fresh syconia of <i>Ficus septica</i>	sp. 34	Kanzaki et al. (2018) Nature Communications 9: 3216
<i>C. japonica</i> Kiontke, Hironaka & Sudhaus, 2003	Hinokuma Mountain Prefectural Park, Kanzaki Town, Saga, Japan	Dauer juveniles found on females of the shield bug <i>Parastrachia japonensis</i> Scott (Heteroptera)		Kiontke et al. (2003) Nematology 4 (2002): 933–941
<i>C. kamaaina</i> Félix, Braendle & Cutter, 2014	Kauai, Hawaii, USA	Unidentified wild rotten fruit	sp. 15 (QG122)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. latens</i> Félix, Braendle & Cutter, 2014	Juifeng Village, Wuhan City, Hubei Province, China	Soil near the lotus pond	sp. 23 (VX88)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. macrosperma</i> Félix, Braendle & Cutter, 2014	Nouragues Forest (path to Inselberg), French Guiana	Rotten <i>Hyeronima taxiflora</i> fruits (nut)	sp. 18 (JU1857)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. monodelphis</i> Slos & Sudhaus, 2017	Berlin, Germany	Galleries inside of the fungus <i>Ganoderma applanatum</i> (Polyporaceae) which grew on the stump of a tree a few centimeters above ground	sp. 1 (SB341)	Slos et al. (2017) BMC Zool 2: 4
<i>C. nigoni</i> Félix, Braendle & Cutter, 2014	Zoo/Botanical Garden of Trivandrum, Kerala, India	Rotting flowers and leaves	sp. 9 (JU1325)	Félix et al. (2014) PLoS ONE 9: e94723
Continued				

Species name with taxonomic author(s)	Type locality	Type host or habitat	Previous species code (type strain)	Reference (original description)
<i>C. nouraguensis</i> Félix, Braendle & Cutter, 2014	Grand Plateau in Nouragues Forest, French Guiana	Rotten wide bean of <i>Barokia</i> sp.	sp. 17 (JU1825)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. oncomelaniae</i> (Yokoo & Okabe, 1968)	Kurume, Japan (isolated from laboratory-reared population of host, but the origin of the population is not specified in the description)	Snail, <i>Oncomelania hupensis nosophora</i> Gredler (Rissooidea)		Yokoo & Okabe (1968) Agric Bull Saga Univ 25: 69–78
<i>C. panamensis</i> Stevens & Félix, 2019	Barro Colorado Island, Panama	Rotting palm fruit	sp. 21 (QG702)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. parvicauda</i> Stevens & Félix, 2019	Sainte Marie Island, Madagascar	Apple-like rotten and dry fruit	sp. 21 (NIC206)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. plicata</i> (Völk, 1950)	Erlangen, Germany	<i>Hister stercorarius</i> Hoffmann (Histeridae) from carrion; phoretic on different carrion beetles (Silphidae)		Völk (1950) Zool Jahrb (Systematik) 79: 1–70
<i>C. perrieri</i> (Maupas, 1900)	Near Mitidja, Arba, Algeria	Soil at the bottom of a manure heap		Maupas (1900) Arch Zool Exp Gen 8: 463–624
<i>C. portoensis</i> Félix, Braendle & Cutter, 2014	Amares, Portugal	Rotting apple	sp. 6 (EG4788)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. quiocensis</i> Stevens & Félix, 2019	Basse-Terre Forest, Guadeloupe	Rotting fruit (species not specified)	sp. 38 (JU2745)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. remanei</i> (Sudhaus, 1974)	Freiburg, Germany	Slug <i>Arion rufus</i> L. (Arionidae)		Sudhaus (1974) Zool Jahrb (Systematik) 101: 417–465 Baird et al. (1994) Nematologica 40: 1–11
<i>C. sinica</i> Huang, Ren, Qiu & Zhao, 2014	Luofu Mountain of Huizhou City, Guangdong Province, China	Rotting fruit (species not specified)	sp. 5 ZZY0401	Huang et al. (2014) PLoS ONE 9: e110957
<i>C. sonorae</i> (Kiontke, 1997)	Pima County, Tucson Mountains, Arizona, USA	Decaying tissue of saguaro cactus, <i>Carnegiea gigantea</i> (Engelm.) (Cactaceae)		Kiontke (1997) Fundam Appl Nematol 20: 305–315
<i>C. sulstoni</i> Stevens & Félix, 2019	East Africa	Faeces of millipede <i>Archispirostreptus gigas</i> bought in spring 2013 on an insect market in Berlin	sp. 32 (SB454)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. tribulationis</i> Stevens & Félix, 2019	Danbulla, Australia	Humus below <i>Ficus destruens</i>	sp. 40 (JU2774)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. tropicalis</i> Félix, Braendle & Cutter, 2014	La Réunion, France (Overseas department)	Rotting torch ginger (rose porcelaine— <i>Etilingeria elatior</i>) flowers	sp. 11 (JU1373)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. uteleia</i> Stevens & Félix, 2019	Madre de Dios, Peru	Rotting Kumquat-like fruit	sp. 31 (JU2469)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. virilis</i> Félix, Braendle & Cutter, 2014	Orsay, France	Rotting apple	sp. 13 (JU1528)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. vivipara</i> Stevens & Félix, 2019	Chichén Itzá, Yucatán, Mexico	Rotting red berries	sp. 43 (NIC1070)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. wallacei</i> Félix, Braendle & Cutter, 2014	Cacao plantation near Sanda Center, Bali, Indonesia	Rotten cacao fruit	sp. 16 (JU1873)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. waitukubuli</i> Stevens & Félix, 2019	Morne Trois Pitons Trail, Forest, Dominica	Rotting <i>Clusia</i> fruit	sp. 39 (NIC564)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>C. yunquensis</i> Félix, Braendle & Cutter, 2014	El Yunque, Puerto Rico	Not provided (rotten fruit?)	sp. 19 (EG6142)	Félix et al. (2014) PLoS ONE 9: e94723
<i>C. zanzibari</i> Stevens & Félix, 2019	Jozani, Zanzibar	Rotting fruit of <i>Calophyllum inophyllum</i>	sp. 26 (JU2161)	Stevens et al. (2019) Evolution Letters 3: 217–236
<i>Caenorhabditis</i> sp. 2	Mediterranean islands (Mallorca, Sardinia), Madeira, Portugal and the Canary Islands, Spain	Isolated from rotting <i>Opuntia</i> cactus tissue, and probably carried by <i>Drosophila</i> sp.		Kiontke et al. (2011) BMC Evol Biol 11: 339
<i>Caenorhabditis</i> sp. 8	New Jersey, USA	Rotting tomatoes		Kiontke et al. (2011) BMC Evol Biol 11: 339

Table 1. List of nominal and characterized *Caenorhabditis* species including type locality, and type host or habitat.

C. auriculariae's basal phylogenetic position in the genus and its usefulness as an outgroup to understand the *Caenorhabditis* evolution.

Results

Morphological characteristics. *Caenorhabditis auriculariae* is a rare, neglected species which was reported only once about 20 years ago⁹. We isolated *C. auriculariae* from a mushroom beetle, *Platydemus* sp. feeding on the fruiting body of *Auricularia polytricha* in Aichi, Japan. The general morphological characteristics are as described previously⁹ and those are photo-documented and illustrated in Fig. 1 and Supplementary Figs. S1–S7. Several typological taxonomic/phylogenetic key characteristics in the male tail are explained and discussed below. Only newly found characteristics are described here to avoid redundancy.

The detailed lip and cheilostomatal structures are described for the first time as follows (Fig. 1). Lip is separated into six lip sectors, and each has an outer labial papilla. There, three pairs of neighboring (right

subventral + latelal, left subventral + lateral and right and left dorsal) lip sectors are partially fused to form three large lip sectors. Thus, these three sectors are arranged triradially in *en face* view. The short tube-like stoma is separated into three elements from anterior: cheilostom, gymnostom and stegostom. Cheilostom cuticular tube, occupies about 35–40% of total stomatal length. Anterior part of cheilostomatal wall (cheilorhabdion) extends and internally fold to form a half-circle shaped flap which covers stomatal opening like valve apparatus. Posterior end of chaeilostom overlapping gymnostom. Gymnostom is simple and short cuticular tube, occupies about 20–25% of stoma. Stegostom is separated from the other parts of stomatal element by possession of pharyngeal sleeve consisting of four subelements: pro-, meso-, meta- and telostogostom. Pro- and mesostegostom is not clearly separated and forms a simple cuticular tube. Metastegostom at the posterior end of pro-mesostegostomatal tube, forming three (two subventral and a dorsal) small bifid teeth. Telostogostom not cuticularized, connecting stoma and pharynx.

Phylogenetic status. The phylogenetic relationships of 28 *Caenorhabditis* species and one outgroup species (*Prodontorhabditis wirthi*) inferred from the full length SSU rRNA and D2-D3 regions of LSU rRNA were congruent with previously provided phylogenetic trees, except for a few terminal nodes and the placement of *C. plicata*, which was as part of the *drosophilae* supergroup with low posterior-probability support^{4,6}. *C. auriculariae* was placed close to *Caenorhabditis sonoreae*, which was isolated from the rotten cactus *Carnegiea gigantea* in the USA¹⁰, and *Caenorhabditis monodelphis*, isolated from the galleries of the fungal-feeding beetle, *Cis nitidus* inside fruiting bodies of *Ganoderma applanatum*^{3,4}. These three species formed an independent clade at the basal (outgroup) position of the other *Caenorhabditis* spp. (Fig. 2A). The GenBank accession numbers of the sequences compared are listed in Table S1.

Voucher material. Ten male and 10 female *C. auriculariae* adults have been vouchered as permanent slides at the Forest Pathology Laboratory collection of the Forestry and Forest Products Research Institute with the material numbers *Caenorhabditis auriculariae* M01–10 and F01–10. The TAF-fixed materials and unmounted glycerol-processed materials have also been vouchered into the collection. Live cultures and a frozen stock of *C. auriculariae* has been deposited in Taisei Kikuchi's Lab. (culture code NKZ352; Miyazaki University, Miyazaki, Japan), and further genomic and transcriptomic analyses will be conducted.

Genome characteristics of *C. auriculariae*. For a deeper understanding of the phylogenetic status and biological features of *C. auriculariae*, we sequenced the genome of the species and conducted a genome comparison with other *Caenorhabditis* species. The hybrid assembly using Nanopore long reads and Illumina short reads (Table S2) resulted in a 109.5 Mb assembly composed of 491 scaffolds with high completeness values (89.8% BUSCO and 95.9/99.6% partial/complete CEGMA) (Table 2). A total of 16,279 protein-coding genes with a mean protein length of 435.34 and the largest of 8,188 amino acids were predicted on the genome assembly. This genome size is ~10% larger, but the predicted gene number is slightly smaller than those in *C. elegans*. The *C. auriculariae* genome contained of 20.8 Mb repetitive sequences that account for 19.04% of the genome, which is similar amount as *C. elegans* genome (18.45%) (Table 3). DNA repeat family was the most abundant (1.6%) followed by LINE family (0.91%) and LTR family (0.51%), though a large portion of *C. auriculariae* repeats (15.24%) were classified as “Unclassified”. Compared to *C. elegans*, more retroelements (LINEs and LTRs) were identified in *C. auriculariae*, which is consistent with the fact that the *C. auriculariae* gene models contained a higher number of transposon genes (Table 2) and RVT (reverse transcriptase) domains (see Pfam result below).

We then performed a phylogenomic analysis using 35 *Caenorhabditis* species whose draft genome sequences were available with *D. pachys* as an outgroup. A ML tree based on 97 single-copy genes showed a mostly consistent topology to the nuclear rRNA tree; species of *elegans* group and *japonica* group each formed a separated cluster, with species of *drosophila* supergroup located at more basal position of the tree. *C. auriculariae* was placed at the most basal position of *Caenorhabditis* genus with *C. monodelphis* (Fig. 2B). *C. parvicauda*, which has a morphological novelty, secondary loss of bursa, and is considered highly divergent¹¹, shows a long branch in the tree but belongs to the inner clade.

The two basal species, *C. auriculariae* and *C. monodelphis*, showed similar genome statistics to each other. For instance, *C. auriculariae*/*C. monodelphis* total assembly size are 109.5/115.1 Mb and 16,279/17,180 in the predicted gene numbers (Table 2). The gene structures of *C. auriculariae* are also similar to those of *C. monodelphis*, in which genes are generally longer, contain more exons, and a longer span of introns than *C. elegans* genes (Fig. 3), which was suggested to reflect an ancestral status of *Caenorhabditis* genome structure¹².

Comparison of protein domain (Pfam) distribution patterns in the genomes revealed that *C. auriculariae*, compared to *C. elegans*, has higher numbers of Ank, LRR, HEAT, TIL, HTH_Tnp_Tc3_2, DDE_3, RVT_1, and DEAD protein domains. The numbers of protein domains related to receptors (GPCRs, Hormone_recep, and Recep_L_domain), WD40, Collagen, Ig_3, I-set, V-set, Pkinases, EGF, Zinc finger, Shk, C2-set_2, FTH, FBA_2, Lectin_C domains are smaller in *C. auriculariae* (Fig. 4). Gene family (orthologue) analysis assigned a total of 389,541 genes (90.8%) of 18 *Caenorhabditis* species and *D. pachys* into 31,748 orthogroups. Of 31,748 orthogroups, 4971 orthogroups were shared by all species. A high number of orthogroups (9737 orthogroups) are shared by *C. auriculariae* and *C. monodelphis* with 356 unique to the clade. However, the two species still exhibit high numbers of species-specific orthogroups: 2546 and 3880 orthogroups unique to *C. auriculariae* and *C. monodelphis*, respectively (Fig. 5). *C. auriculariae* specific-orthologous include genes encoding proteins with Ank (Ankyrin), TIL (Trypsin Inhibitor like cysteine rich domain), LEA_4 (Late embryogenesis abundant), GPCR (G protein-coupled receptors), Collagen, Pkinase (Protein kinase) and Apolipoprotein domains (Table S3), suggesting that genes of those functions are highly diverged in *C. auriculariae* and possibly reflecting its unique lifestyle though it has not been revealed yet.

Carbohydrate-Active enzymes (CAZy) are involved in several biological processes, including feeding, energy metabolism, structural support, and signal transduction¹³. The repertoires in the genome generally reflects its life style. We identified a total of 312 CAZy genes (5 auxiliary activities (AA), 32 carbohydrate-binding modules (CBM), 47 carbohydrate esterases (CE), 71 glycoside hydrolases (GH) and 157 glycosyltransferases (GT)) in *C. auriculariae*, which is a comparable number with other *Caenorhabditis* species (Table S4). We found many CAZy classes are common to the 35 *Caenorhabditis* species (e.g. two AA, seven CBM, four CE, 18 GH and 30 GT) though the number of genes in each class varies, but some are species or group specific (e.g. GH131 in *C. brenneri* and GH88 in *C. guadeloupensis*). To reduce the complexity of the CAZy distribution patterns across 35 *Caenorhabditis* species, we conducted a principal component analysis (PCA). The first two principal components explained 73.4% of the overall variance (55.9% and 17.5% for principal component 1 and 2, respectively) (Fig. 6). The PCA plot (Fig. 6) clustered species largely by the taxonomic groups; species of *elegans*-group were mostly located upper right, most of *japonica*- and *drosophilae*-groups were placed lower middle, and the basal group was on the upper left. However, interestingly, this CAZy-based plot seems also highly correlated with particular lifestyles. For instance, *C. inopinata*, *C. japonica*, *C. drosophilae* and *C. bovis* were placed close to each other although they belong to *elegans*-group, *japonica*-group, the *drosophilae*-supergroup, and a separate basal clade, respectively. These four nematodes are well-known insect-associates as using insects as distributing vectors^{14–16}. *C. angaria* and *C. castelli* are phylogenetically close to each other, but they were clearly separated by PC1 and PC2. Similarly, *C. angaria* has a tendency to ride weevils¹⁷ whereas there are no reports about an insect association for *C. castelli*. We have tested if there is a relationship between the trait (insect-association) and the CAZy distribution using the phylogenetic logistic regression with the PC values and found a significant correlation between PC1 and the trait ($p < 0.01$) (Fig. 6). It is also interesting to note that the hermaphroditic species (i.e., *C. elegans*, *C. briggsae* and *C. tropicalis*) were placed together in the PCA plot although those hermaphroditism were evolved independently in the *Caenorhabditis* evolutionary history although the regression test was not statistically significant (Fig. 6).

In the PCA plot, *C. auriculariae* was placed together with *C. monodelphis* on the top left, suggesting those basal species have similar lifestyles to each other and they possibly have tight associations with insects, which is consistent with the fact that they were isolated from beetles.

Discussion

Morphological comparison with other molecularly characterized *Caenorhabditis* spp. *Caenorhabditis auriculariae* was described in 1999 before the deep-level phylogeny of the genus or the relationship between morphological characteristics and phylogenetic status had been examined⁹. Later, Kiontke et al.⁴ examined nominal and undescribed *Caenorhabditis* spp. using multiple molecular loci and coded their typological characteristics in a phylogenetic analysis.

The genus *Caenorhabditis* was separated into two supergroups (*elegans* and *drosophila* supergroups) based on molecular phylogenetic analyses and male tail characteristics. In addition, there were several species that do not fall into those supergroups^{4,6} which tentatively regarded as basal group. Basal group species including *C. auriculariae* harbour some typical characteristics from both supergroups that are hypothesized to be the stem species pattern, namely: (1) oval and anteriorly opened bursa without serratae and terminal notch on the edge of the velum, (2) nine pairs of bursal rays in which p2 reaches to the edge of the velum, p2 and p3 are clearly separate, and p1 is directed dorsally, (3) precloacal lip is rounded, (4) spicule with a slightly ventrally bent blade and complex tip, and (5) parallel mating position⁴, although several species-specific apomorphies, e.g., the secondary loss of bursa in *C. parvicauda*¹¹, has been reported. As are in other basal group species, several species-specific apomorphies (or clade, if there is a closely related cryptic species) are evident after comparing the morphological and molecular phylogenetic status of *C. auriculariae*. The typological characteristics of the *C. auriculariae* male tail are (1) wide, heart-shaped bursa with an anterior serrated-edge velum and no terminal notch, (2) nine pairs of bursal rays arranged as (p1d, p2)/P3, (p4 + p5d), p6, (p7m p8d), (ph, p9), where p2 does not reach to the edge of the velum and p2 and p3 are clearly separate, (3) precloacal lip forms a heart-shaped or bifid cap structure, (4) stout and evenly curved spicule with a complex spicule tip (possessing a dorsally oriented small projection at the distal tip), and (5) parallel mating position (not spiral).

Caenorhabditis auriculariae spicule morphology is similar to that of several *drosophilae* supergroup species (*C. drosophilae*, *C. angaria*, *C. castelli*, and *Caenorhabditis* sp. 2 and sp. 8) and *C. monodelphis*^{4,6,10,17}. The bursal velum morphology is similar to all *elegans* supergroup species and several *drosophilae* supergroup species (*C. portoensis*, *C. virilis*, and *C. latens*)^{4,6}. The arrangement of bursal rays is somewhat intermediate between the two supergroups. For example, the short p2 that does not reach the edge of the velum is similar to the *elegans* supergroup and *C. virilis*; dorsally directed p5d is shared with the *elegans* supergroup and *C. monodelphis*, clearly separate p2 and p3 are shared with all non-*elegans* group species, and dorsally directed p8d is shared with two *drosophilae* supergroup species (*C. portoensis* and *C. virilis*)^{4,6}. Additionally, the parallel mating position is similar to all known *Caenorhabditis* except three *drosophilae* supergroup species (*C. angaria*, *C. castelli*, and *Caenorhabditis* sp. 8)^{4,6}. The heart-shaped cap on the precloacal lip and the arrangement of the bursal rays (see above) are unique to *C. auriculariae*. In addition, the stomatal morphology of *C. auriculariae* is unique. Therefore, regardless of the unique characteristics, none of the nominal (and characterized) species exactly matched the typological characteristics of *C. auriculariae*. *C. auriculariae* is distinguished from all other phylogenetically characterized *Caenorhabditis* sp. based solely on typological characteristics.

The rRNA and genome-based phylogeny suggested the closeness of *C. auriculariae* with *C. sonorae*, and *C. monodelphis*, as these three species formed a well-supported independent clade at the basal position of the genus. They, however, can be clearly separated by their typological characteristics, as the bursal velum, ray characteristics, and precloacal lip structure differ from each other^{4,10}. In addition, *C. auriculariae* has a quite unique stomatal

morphology, with a long flap-like cuticular extension on the cheilostom and three bifid metastegostomatal teeth. Although the stomatal characteristics of *C. monodelphis* have not been described in detail, both species have a long and narrow stoma, and *C. sonorae* has a three triangular teeth, which is common in the genus, but *C. monodelphis* does not have glottoid apparatus^{5,10}. The unique stomatal structure of *C. auriculariae* could be a species (or clade) specific apomorphy.

Biological features and genome. *Caenorhabditis* nematodes have been isolated from many different environments and animals, such as rotting fruit^{3,4,10}, rich soil and manure^{18–21}, mushrooms^{3,9}, insects^{14,17,22}, soil and freshwater invertebrates^{22,23}, and vertebrates potentially including humans^{24–26}. Some vertebrate associations could be due to insect carriers associated with the “host” vertebrates. *C. monodelphis* and *C. auriculariae* were originally isolated from *G. applanatum* in Berlin, Germany and from *A. polytricha* in Kyoto, Japan, respectively, and are associated with fungal-feeding beetles^{3,9}. In the present study, *C. auriculariae* was isolated from a fungal beetle, *Platydemia* sp. Although the detailed carrier association, e.g., the beetle species is primary carrier of *C. auriculariae*, beetle body organ harbouring the nematode, and number and association rate of nematode in individual beetles, was not clarified in this study, at least the ability of insect association (phoresy) was confirmed for *C. auriculariae*. Because of the limited data, we cannot conclude that the fungal (mushroom) associations of these species are related to clade-specific habitat preferences or carrier insects. However, the present results will be useful to further isolate strains of those species. Diplogastrids nematodes, *Pristionchus* spp. were considered as the soil-inhabiting free-living nematodes for long time, and its close insect association has been confirmed recently^{27,28}. The close rotten fruits-association of *Caenorhabditis* spp. has not been recognized for a long time⁴. After findings of these associations, the number of new species isolation increased dramatically^{3,4,29}. Similarly, this study and recent reports on insect-associated *Caenorhabditis* spp. should enhance new species identification of the genus by surveys of nematodes around insects.

The genome comparison revealed the presence of highly diverged or unique genes encoding GPCRs in *C. auriculariae*. GPCRs work as primary receptors to detect a wide variety of environmental signals and are therefore highly diverged in organisms or even among individuals^{30,31}. The unique repertoires of *C. auriculariae* GPCRs probably reflect the need to detect environmental signals specific for its lifestyle, such as mushroom and insect associations. The genome comparison also found diverged LEA proteins in *C. auriculariae*. LEA proteins were initially discovered accumulating late in embryogenesis of cotton seeds and later shown to have a role to protect proteins against aggregation due to desiccation or osmotic stresses in some plants, bacteria and invertebrates^{32,33}. Further functional investigation is necessary, but this may reflect its life-cycle in which the nematode encounters relatively dry condition compared to *C. elegans*.

CAZy distribution-based PCs separated insect associated species from non- or less- associates regardless of their phylogenetic relationships. Furthermore, this method roughly separated hermaphroditic species from gonochoristic even when two closely related sister species have contrastive reproduction modes. Therefore, this method can be of particular usefulness to speculate on the lifestyle of newly isolated species with non-detailed ecological information. For example, based on the fact that *C. pamanensis* was placed in the insect-associate ellipse (Fig. 6), we could speculate that the worm has a lifestyle with a tight insect-association, although no such records were reported (Table 1). Indeed, there are several rare *Caenorhabditis* species with unclear ecological status, such as *C. yunguensis* (Table 1).

This study provided a high-quality genome reference for *C. auriculariae*. A genome of *C. monodelphis* was recently published as an outgroup reference for *Caenorhabditis*¹². *C. auriculariae* is also phylogenetically placed at the basal position of the genus and shared several genome features with *C. monodelphis*. However, the distance of the two species is substantially long, and each genome contained a number of species-specific genes. Therefore, *C. auriculariae*, together with *C. monodelphis*, provides a powerful resource to perform deep evolutionary studies in the genus *Caenorhabditis*.

Methods

Nematode materials. Potential carrier insects of nematodes were collected in the field in Nagoya, Aichi, Japan on 17 June 2015. The samples were collected under an official permit from the Nagoya City local governmental office. Several species of coleopteran insects (beetles) were collected, brought back to the laboratory, morphologically identified, and dissected to examine their association with nematodes. The dissected insect bodies were placed in 2.0% water agar to allow propagation of phoretic microbe-feeding species and examined occasionally. No endangered or protected species were collected in the present study.

A *Caenorhabditis* sp. was isolated from the dissected body of *Platydemia* sp. (Coleoptera: Tenebrionidae); the nematode was not confirmed during the dissection but propagated on the dissected body of its carrier beetle. The nematode was observed under a light microscope (Eclipse 80i: Nikon, Tokyo, Japan) to determine its feeding habits. It was then transferred to nematode growth medium (NGM) and kept as a laboratory strain with culture code NKZ352.

Morphological observations and micrographs. Live and TAF-fixed *C. auriculariae* material from 2-week-old cultures was observed under a light microscope using the methodologies defined by Kanzaki³⁴. The nematode were identified to species based on typological characteristics when compared with the original description⁹. Thereafter, the TAF-fixed material was processed into glycerin according to a modified Seinhorst's method³⁵ and deposited as morphological vouchers.

Several morphological characteristics that were not provided in the original description, e.g., detailed stomatal structure, were drawn using a drawing tube, and other general characteristics were photo-documented using a digital camera system (DS-Ri1, Nikon) connected to a microscope.

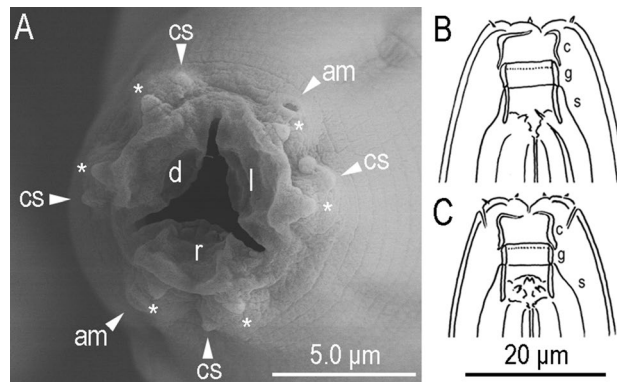


Figure 1. Stomatal structure of *Caenorhabditis auriculariae*. (A) En face view where laval censilla (asterisk), cephalic sensilla (c), amphids (a) and cheilostomatal flap bearing on dorsal (d), right subventral (r) and left subventral (l) sectors are suggested; (B) left lateral view; (C) ventral view. Cheilostom (c), gymnostom (g) and stegostom (s) are suggested in (B,C).

Scanning electron microscope (SEM) observation. For SEM observation, adult nematodes were treated with the pre-fixation solution (2% paraformaldehyde, 2.5% glutaraldehyde, 0.1 M Cacodylate, pH 7.4) for 2 h at 4 °C followed by incubation in the fixation solution (1% OsO₄, 0.1 M Cacodylate, pH 7.4) for 1 h at 4 °C. Samples were then dehydrated in ethanol (50% to 100%, gradually). They were substituted by isoamyl acetate and were dried by using a freeze-drying device (Eiko ID-2). Dried nematodes were coated with Platinum by using ION SPUTTER (HITACHI E-1045) and were observed by using SEM (Hitachi S-4800) operating at 20 kV.

Molecular profiles and preliminary phylogenetic analyses. Prior to genome wide phylogenetic analysis of selected species, the phylogenetic status of *C. auriculariae* within the genus was analysed based on the ribosomal RNA genes. Nematode lysate material was prepared for use as a polymerase chain reaction (PCR) template according to the protocol developed by Kikuchi et al.³⁶ and Tanaka et al.³⁷. The molecular sequences of small subunit ribosomal RNA (SSU rRNA) and D2-D3 regions of large subunit ribosomal RNA (LSU rRNA) were sequenced with the PCR direct sequencing methods developed by Ye et al.³⁸ and Kanzaki and Futai³⁹.

A Bayesian molecular phylogenetic analysis was conducted based on SSU and D2-D3 LSU as previously described⁴⁰. The sequences were aligned using MAFFT⁴¹ and the base substitution model was determined using Modeltest ver. 3.7⁴² under the Akaike information criterion model selection criterion. Then, a Bayesian analysis was performed to infer the tree topology of each gene using MrBayes 3.2⁴³, four chains were run for 4×10^6 generations. Markov chains were sampled at intervals of 100 generations⁴⁴. Two independent runs were performed, and the remaining topologies were used to generate a 50% majority-rule consensus tree after confirming convergence of runs and discarding the first 2×10^6 generations as burn-in.

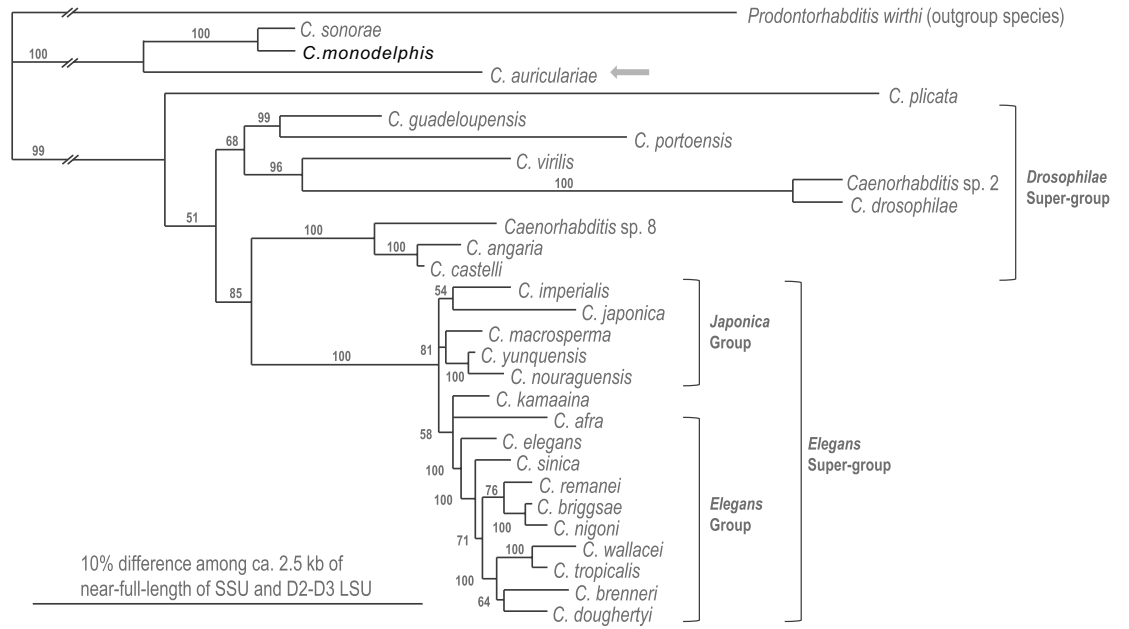
DNA/RNA isolation and sequencing. For whole genome analyses, nematodes were propagated on NGM plates implemented with *E. coli* Op50 strain. After 2 weeks of incubation at 20 °C, nematodes were collected from the plate, washed five times with M9 buffer and the genomic DNA was extracted using Genomic-tip (Qia-gen) following the manufacturer's protocol. Paired-end and Mate-pair sequencing libraries were prepared using the Nextera DNA Sample Prep kit (Illumina) and TruSeq DNA Library Preparation kit, respectively, according to the manufacturer's instructions and sequenced using Illumina MiSeq sequencer with the v3 kit (301 cycles \times 2 or 76 cycles \times 2) (Illumina) (Supplementary Table S2).

Two μ g of genomic DNA was used to prepare Nanopore sequencing library using the Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies) according to the manufacturer's protocol. The library was sequenced with a single 24 h run with FLO-MIN106 R9 MinION flowcell (Oxford Nanopore Technologies). Base calling for R9 runs was performed with Guppy v.3.1.5 using the 'dna_r9.4.1_450bps_fast' model and obtained 771,594 reads (~3 Gb) (Supplementary Table S2).

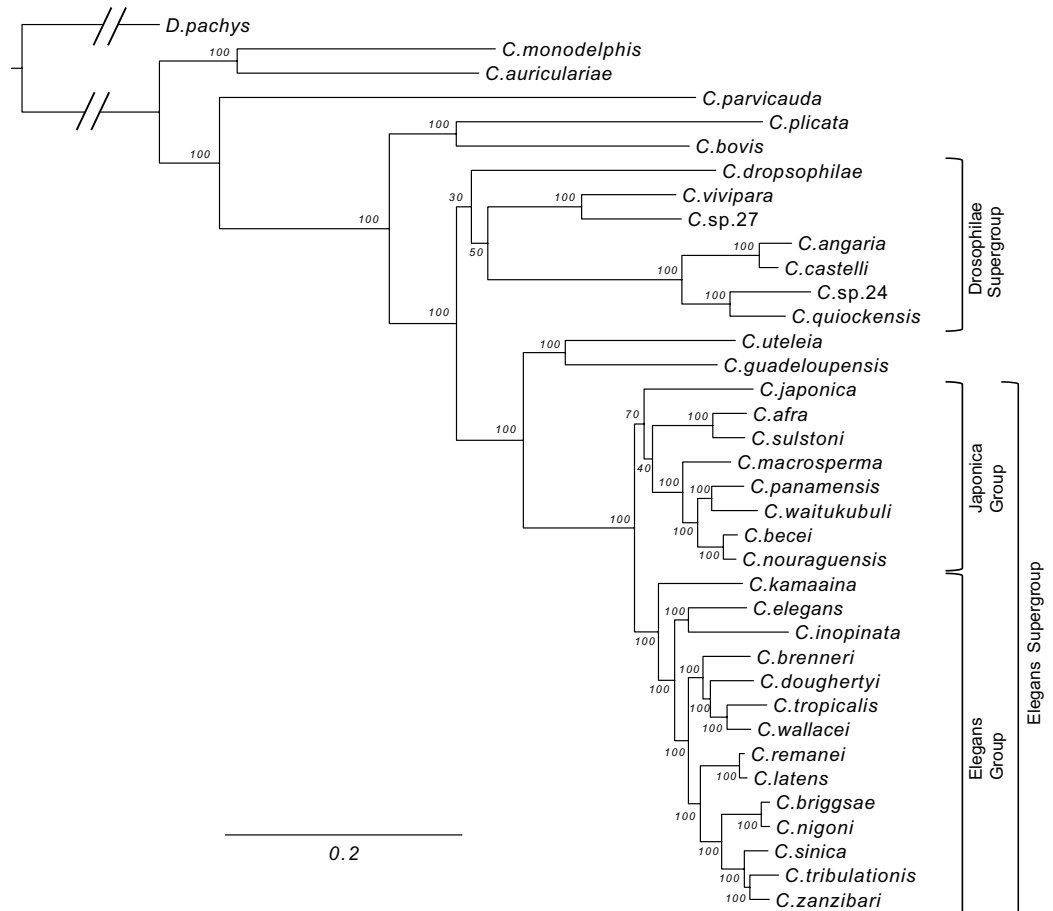
For mRNA-seq analysis, RNA was extracted from fresh mixed-stage nematodes using TRI reagent according to the manufacturer's instructions. Total RNA samples were qualified using Bioanalyzer 2100 (Agilent Technology, Inc.) and only samples with an RNA integrity value (RIN) greater than 8.0 were used for library constructions. One hundred ng of total RNA was used to produce an Illumina sequencing library using the TruSeq RNA-seq Sample Prep kit according to the manufacturer's recommended protocols (Illumina). The RNA libraries were sequenced using Illumina MiSeq sequencer with the v3 kit (301 cycles \times 2) (Illumina) (Supplementary Table S2).

Genome assembly. Three de novo assemblers were used to generate initial assemblies. The Nanopore reads (~771 K reads, N50=4.7 kb) were assembled with Flye (v.2.7.1)⁴⁵ in raw nanopore mode using -g 100 m or Canu⁴⁶ using genomeSize=100 m, both followed by base correction by Illumina DNA reads using Pilon (v.1.22)⁴⁷. Spades (v.3.7.1)⁴⁸ was separately used to generate a hybrid assembly of Nanopore, Illumina pair-end and mate-pair reads (Supplementary Table S1) with the default options after trimming of Illumina reads for

A)



B)



◀ **Figure 2.** Phylogenetic status of *Caenorhabditis auriculariae*. (A) The combined Bayesian tree inferred from near-full-length SSU and D2-D3 LSU. The substitution model and parameters for SSU and D2-D3 are GTR + I + G (AIC = 16,046.5234; lnL = 8013.261; freq A = 0.2452; freq C = 0.2064; freq G = 0.2661; freq T = 0.2823; R(a) = 1.4800; R(b) = 3.4097; R(c) = 1.9899; R(d) = 0.5864; R(e) = 5.6910; R(f) = 1; Pinva = 0.4188; Shape = 0.4851) and GTR + I + G (AIC = 9579.5820; lnL = 4779.7910; freq A = 0.2163; freq C = 0.2034; freq G = 0.3165; freq T = 0.2638; R(a) = 0.7283; R(b) = 2.7572; R(c) = 1.3283; R(d) = 0.4052; R(e) = 5.6136; R(f) = 1; Pinva = 0.2046; Shape = 0.4580), respectively. Posterior probability values exceeding 50% are given for the appropriate clades. (B) Maximum Likelihood Phylogeny inferred using a total of 299 one-to-one single copy orthologous by IQTREE v2 with 1000 bootstrap values under LG + F + R5 substitution model.

Species	<i>C. auriculariae</i>	<i>C. monodelphis</i>	<i>C. angaria</i>	<i>C. japonica</i>	<i>C. elegans</i>	<i>C. inopinata</i>	<i>C. briggsae</i>
Taxonomic group	Basal group	Basal group	<i>Drosophilae</i> super group	<i>Japonica</i> group	<i>Elegans</i> group	<i>Elegans</i> group	<i>Elegans</i> group
Mating type	Gonochoresitic	Gonochoresitic	Gonochoresitic	Gonochoresitic	Hermaphroditic	Gonochoresitic	Hermaphroditic
Version	v1.1	v1	WBPS15	WBPS15	WS276	WBPS15	WS271
Strain	NKZ352	JU1667	NA	DF5081	N2	NKZ35	AF16
Span (Mb)	109.5	115.1	105.9	156.5	100.2	122.9	108.3
Scaffolds (n)	433	6864	34,621	662	6	6	367
N50 (Mb)	0.95	0.05	0.07	0.86	17.49	20.59	17.49
Genes (n)	16,279	17,180	27,970	29,935	19,999	21,443	20,829
Num. transposon proteins	486	176	157	3578	114	3512	148
GC (%)	38.2	43.9	35.7	39.2	35.4	38.5	37.4
BUSCO %	89.8	89.8	77.9	92.4	98	96.6	97
CEGMA complete/partial %	95.97/99.60	93.95/ 97.58	79.44/95.16	80.65/ 96.37	98.39/100	98.79/100	99.19/100

Table 2. Genome and gene model statistics for *C. auriculariae* and comparisons of other *Caenorhabditis* species.

	<i>Caenorhabditis auriculariae</i>		<i>Caenorhabditis elegans</i>	
	num element	% in bp	num element	% in bp
SINEs	531	0.08	724	0.12
LINEs	2504	0.91	1041	0.49
LTR element	1770	0.51	1412	0.58
DNA element	5774	1.55	55,152	9.28
Rolling-circles	582	0.13	7289	1.88
Small RNA	174	0.02	551	0.05
Satellites	247	0.02	5326	1.44
Simple repeat	10,945	0.38	20,390	0.96
Low complexity	4000	0.16	5916	0.28
Unclassified	96,519	15.24	20,703	3.37
Total	19.04%		18.45%	

Table 3. Repeat contents in *C. auriculariae* and *C. elegans*.

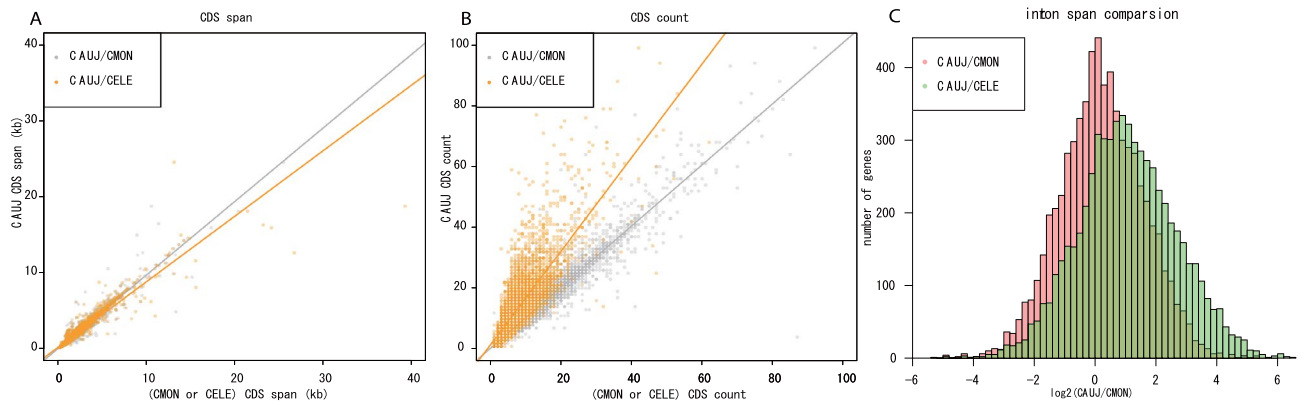


Figure 3. Comparison of gene structure in single-copy orthologues between *C. auriculariae*, *C. elegans* and *C. monodelphis*. Whereas total CDS length per gene is similar in the three species (A), *C. auriculariae* and *C. monodelphis* have larger CDS counts (B) and longer intron span per gene (C) than *C. elegans*. CAUJ; *C. auriculariae*, CMON; *C. monodelphis*, CELE; *C. elegans*.

after trimming for low quality and adaptor contamination using Trimmomatic (v.0.32)^{49,50}. The three assemblies were merged using MetaAssembler (v.1.5)⁵¹ with the Flye assembly as a reference. Haplomerger2 (20151106)⁵² was run on the merged assembly to remove remaining haplotypic sequences. Further base corrections were performed by ICORN2⁵³ using ~5G base of the Illumina pair-end reads. Contigs derived from bacteria or other organisms contaminations were identified and removed from the assembly using Blobtools⁵⁴ and BlastN search against NCBI bacterial nt database. CEGMA v2⁵⁵ were used to assess the completeness of the assemblies.

Gene prediction. RNA-seq read pairs were aligned to the *C. auriculariae* assembly using Hisat2 v2.1.0⁵⁶ with default parameters and used to generate intron hints using bam2hints script in Augustus v3.3.2⁵⁷. Protein-coding genes on the assembly were predicted using BRAKER2⁵⁸ with the intron hints and protein homology hints from ~78,000 proteins of 9 nematode species (*Brugia malayi*, *Bursaphelenchus xylophilus*, *Caenorhabditis elegans*, *C. briggsae*, *Necator americanus*, *Pristionchus pacificus*, *Strongyloides ratti*, *Trichinella spiralis*, and *Trichuris muris*). Protein domain annotations were performed on the gene models using Pfam search (ver. 28.0)⁵⁹ with HMMER v3.1b2⁶⁰ with e-value cutoff (1e−5).

Carbohydrate-active enzyme analysis. Carbohydrate-active enzyme (CAZy) were detected using CAZy database⁶¹ and HMMER v3.1b2 under e-value cutoff (1e−5). Possible contaminations of bacteria or fungi were removed from the detected CAZy genes using BlastP search results against NCBI nr database. CAZy genes of each species were then counted for auxiliary activities, carbohydrate-binding modules, carbohydrate esterases, glycoside hydrolases, polysaccharide lyases and glycosyltransferases, separately.

Principal component analysis was performed for CAZy distribution of 35 *Caenorhabditis* species using the prcomp function and the results were visualised by Factoextra package⁶² both implemented in R (<https://www.r-project.org/>). Phylogenetic logistic regressions for ecological traits (reproduction modes or insect-associations) were performed with the Phylolm R package⁶³ using the principal component values (PC1 to PC4) as explanatory variables and the tree shown in Fig. 2B as phylogenetic information under the logistic_IG10 method and the best models were selected by Akaike's entropy-based Information Criterion (AIC).

Orthologous relationship of *C. auriculariae* with other *Caenorhabditis* species and constructing phylogenetic tree. Orthologous analysis of *C. auriculariae* with 17 selected *Caenorhabditis* species and *Diploscapter pachys* as an outgroup was performed using OrthoFinder v2.3.11⁶⁴ with default parameters using the longest isoform set of each species. Orthologous distribution among species was visualised using the UpSetR package⁶⁵.

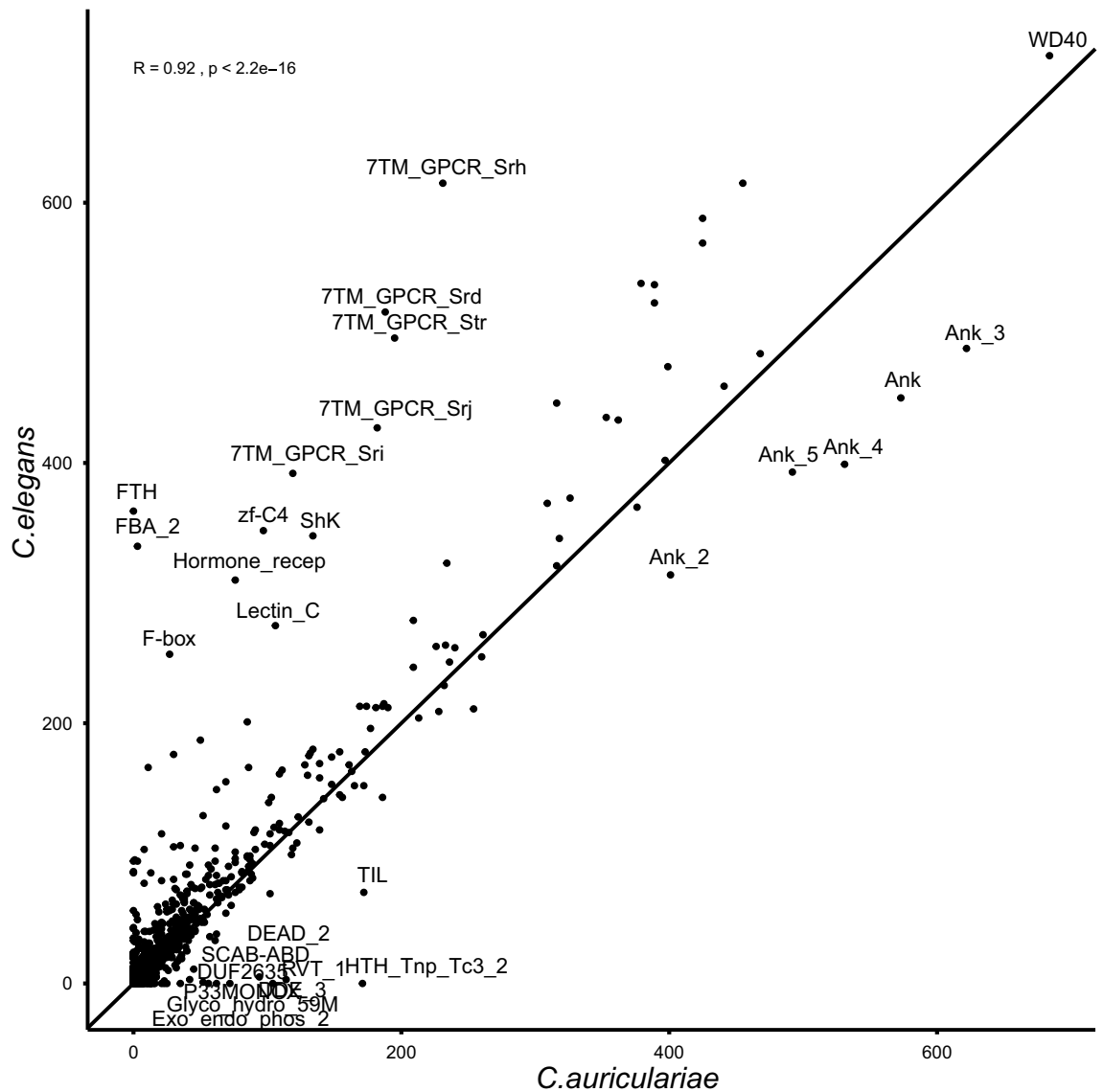


Figure 4. Pfam domain abundance in *C. auriculariae* and *C. elegans*. The x-axis represents abundance of Pfam domains in *C. elegans* and the y-axis represents abundance of the same domains in *C. auriculariae*. Linear regression is plotted alongside their respective equations and correlation coefficients. PFAM domains enriched were labelled with the domain names.

For a genome-wide phylogenetic analysis, amino acid sequences of 96 single-copy orthologous in 37 species were aligned using MAFFT v7.221⁴¹ with auto options. Poorly aligned regions were removed using Gblocks v0.91b⁶⁶ with the parameters (-t=p, -b4=10, -b5=n, -b6=y, -s=y, -p=y, -e=-gb). The alignments were concatenated and used to generate a maximum-likelihood tree using RAxML v8.0.26⁶⁷. For the RAxML analysis, alignments were partitioned by gene with the PROTGAMMAAUTO model (the best-fitting model for each gene) used for all partitions. The topological robustness was assessed with 100 replicates of fast bootstrapping. Resulting phylogenetic tree was visualized in FigTree v1.4.4⁶⁸.

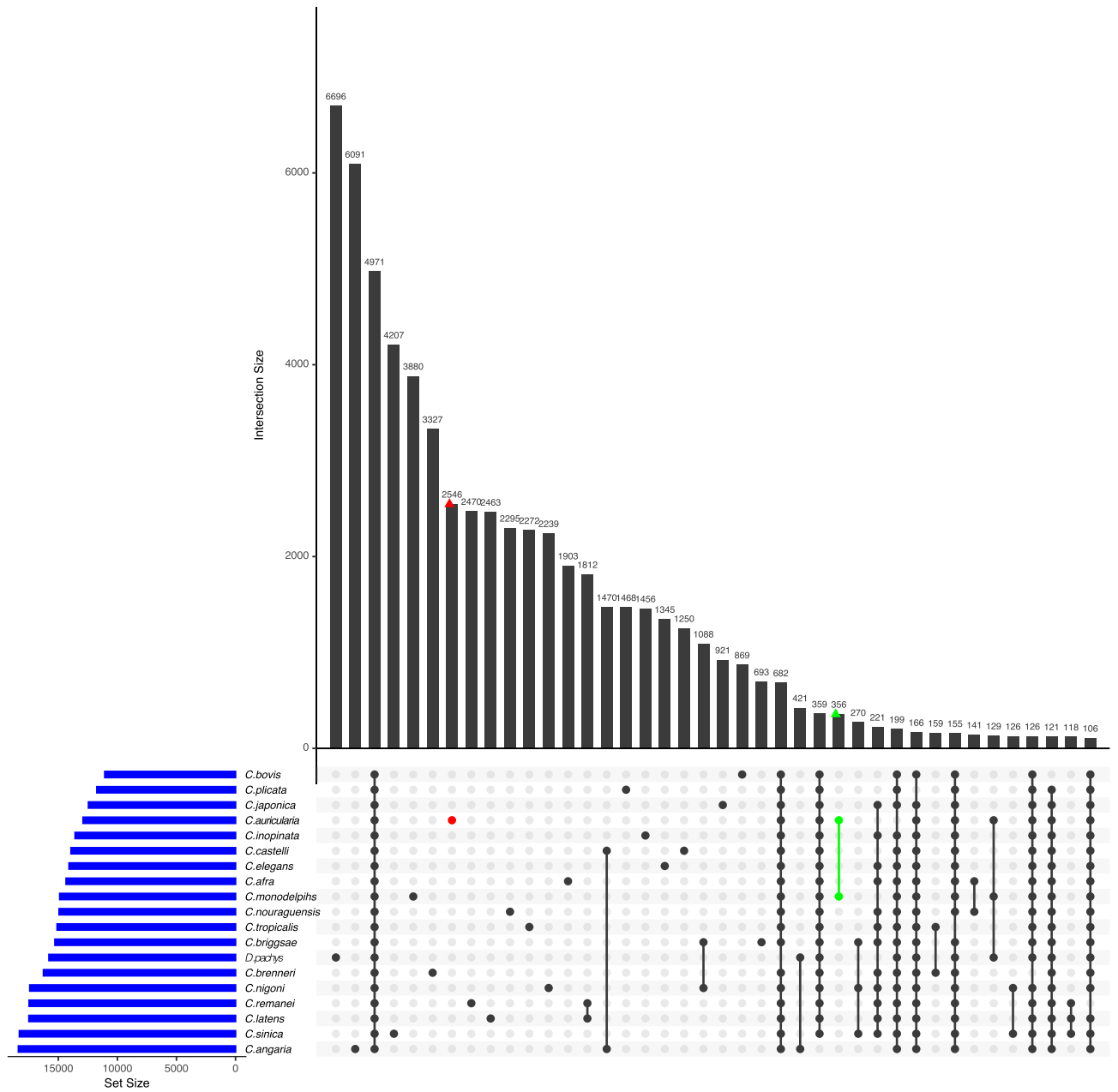


Figure 5. Orthologue comparison across the 18 *Caenorhabditis* and *D. pachys* genomes. UpSetR plot showing unique and overlapping protein ortholog clusters. The intersection matrix is sorted in descending order. Blue bars represent the orthogroup size for each genome and connected dots represent intersections of overlapping orthogroups while vertical bars show the size of each intersection. Orthogroups unique to *C. auriculariae* and *C. auriculariae-C. monodelphis* are shown in red and green, respectively.

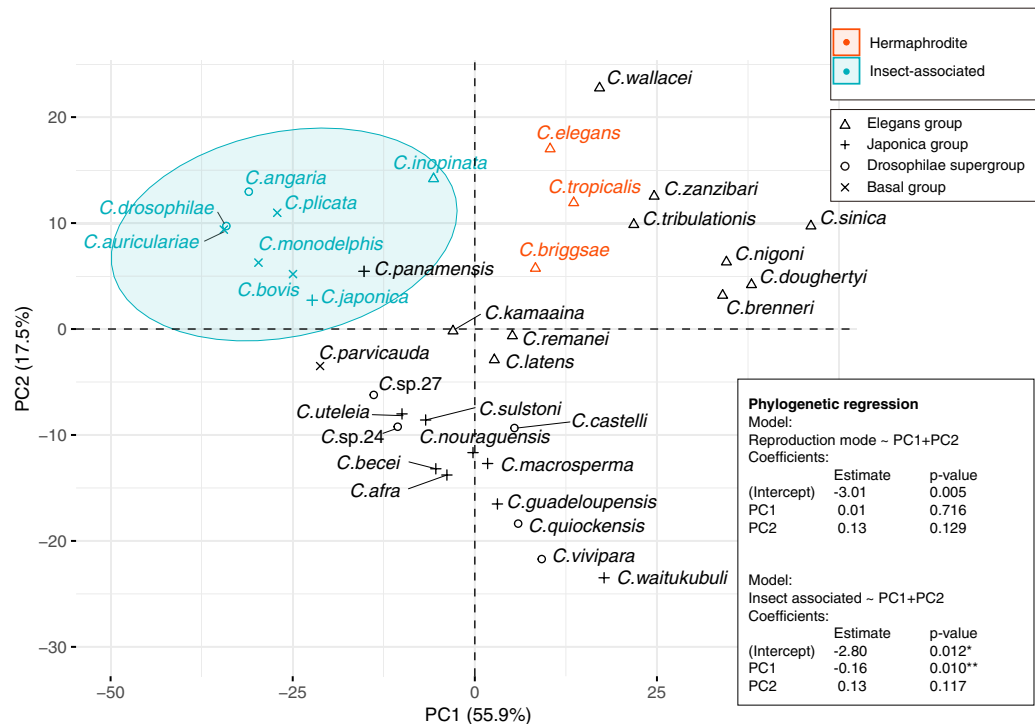


Figure 6. Principle component analysis (PCA) score plot of Carbohydrate active enzyme (CAZy) distribution. The first two axes explain 55.9% and 17.5% of total variance, respectively. Taxonomic groups and ecological traits (hermaphroditism and insect association) are indicated by point shapes and colours, respectively. The ellipse in the plot illustrates the 95% prediction region. The best models of phylogenetic logistic regression for the ecological traits with PCs were shown in box.

Data availability

The raw sequencing data have been deposited to the DNA Data Bank of Japan Sequence Read Archive under the BioProject PRJDB10634. The *C. auriculariae* assembly was deposited in the DDBJ/EMBL/GenBank under Project PRJEB40642 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB40642>).

Received: 21 October 2020; Accepted: 8 March 2021

Published online: 24 March 2021

References

- Conradt, B., Wu, Y. C. & Xue, D. Programmed cell death during *Caenorhabditis elegans* development. *Genetics* **203**, 1533–1562 (2016).
- Grishok, A. *Advances in Genetics* Vol. 83, 1–69 (Elsevier, 2013).
- Kiontke, K. & Sudhaus, W. Ecology of *Caenorhabditis* species. *WormBook* **9**, 1–14 (2006).
- Kiontke, K. C. *et al.* A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol. Biol.* **11**, 339 (2011).
- Sudhaus, W. Phylogenetic systematisation and catalogue of paraphyletic “Rhabditidae” (Secernentea, Nematoda). *J. Nematode Morphol. Syst.* **14**, 113–178 (2011).
- Felix, M. A., Braendle, C. & Cutter, A. D. A streamlined system for species diagnosis in *Caenorhabditis* (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. *PLoS ONE* **9**, e94723 (2014).
- Huang, R.-E., Ren, X., Qiu, Y. & Zhao, Z. Description of *Caenorhabditis sinica* sp. n. (Nematoda: Rhabditidae), a nematode species used in comparative biology for *C. elegans*. *PLoS ONE* **9**, e110957 (2014).
- Mondal, S. & Manna, B. *Caenorhabditis chinkari* sp. n. (Nematoda: Rhabditida) from Chinkara of Alipore Zoological Garden, Kolkata, West Bengal, India. In *Proceedings of the Zoological Society*. Vol. 68. No. 1. (Springer, India, 2015).
- Tsuda, K. & Futai, K. Description of *Caenorhabditis auriculariae* n. sp. (Nematoda: Rhabditida) from fruiting bodies of *Auricularia polytricha*. *Nematol. Res. (Jpn. J. Nematol.)* **29**, 18–23 (1999).
- Kiontke, K. Description of *Rhabditis (Caenorhabditis) drosophilae* n. sp. and *R. (C.) sonoreae* n. sp. (Nematoda: Rhabditida) from saguaro cactus rot in Arizona. *Fund. Appl. Nematol.* **20**, 305–315 (1997).
- Stevens, L. *et al.* Comparative genomics of 10 new *Caenorhabditis* species. *Evol. Lett.* **3**, 217–236 (2019).
- Slos, D., Sudhaus, W., Stevens, L., Bert, W. & Blaxter, M. *Caenorhabditis monodelphis* sp. n.: Defining the stem morphology and genomics of the genus *Caenorhabditis*. *BMC Zool.* **2**, 1–15 (2017).
- Benini, S. Carbohydrate-Active Enzymes: structure, activity, and reaction products. *Int. J. Mol. Sci.* **21**(8), 2727 (2020).
- Kiontke, K., Hironaka, M. & Sudhaus, W. Description of *Caenorhabditis japonica* n. sp. (Nematoda: Rhabditida) associated with the burrower bug *Parastrachia japonensis* (Heteroptera: Cydnidae) in Japan. *Nematology* **4**, 933–941 (2002).
- Kanzaki, N. *et al.* Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nat. Commun.* **9**, 3216 (2018).
- Stevens, L. *et al.* The genome of *Caenorhabditis bovis*. *Curr. Biol.* **30**, 1023–1031 (2020).

17. Sudhaus, W., Giblin-Davis, R. & Kiontke, K. Description of *Caenorhabditis angaria* n. sp. (Nematoda: Rhabditidae), an associate of sugarcane and palm weevils (Coleoptera: Curculionidae). *Nematology* **13**, 61–78 (2011).
18. Maupas, E. La mue et lenkystement chez les nématodes. *Arch. Zool. Exp. Gén.* **7**, 563–628 (1899).
19. Maupas, E. Modes et formes de reproduction des nématodes. *Arch. Zool. Expt. e. Gen.* **8**, 578–582 (1900).
20. Dougherty, E. & Nigon, V. A new species of the free-living nematode genus *Rhabditis* of interest in comparative physiology and genetics. *J. Parasitol.* **35**, 11 (1949).
21. Sudhaus, W. Zur Systematik, Verbreitung, Ökologie und Biologie neuer und wenig bekannter Rhabditiden (Nematoda). 2. Teil. *Zool. Jahrbücher* **101**, 417–465 (1974).
22. Volk, J. *Die Nematoden der Regenwürmer und aasbesuchenden Kafer* (G. Fischer, 1951).
23. Yokoo, T. & Okabe, K. Two new species of genus *Rhabditis* (Nematoda: Rhabditidae) found in the intermediate host of *Schistosoma japonica*, *Oncomelania hupensis* nosophora and *Oncomelania hupensis* formosana. *Agric. Bull. Saga* **43**, 69–78 (1968).
24. Scheiber, S. Ein Fall von mikroskopisch kleinen Rundwürmern—*Rhabditis genitalis*—im Urin einer Kranken. *Arch. für Pathol. Anat. Physiol. für klinische Med.* **82**, 161–175 (1880).
25. Kreis, H. A. Beiträge zur Kenntnis parasitischer Nematoden. *Z. Parasitenkd.* **16**, 36–50 (1953).
26. Schmidt, G. & Kuntz, R. *Caenorhabditis avicola* sp. n. (Rhabditidae) found in a bird from Taiwan. *Proc. Helminthol. Soc. Washington* **39**, 189–191 (1972).
27. Herrmann, M., Mayer, W. E. & Sommer, R. J. Sex, bugs and Haldane's rule: The nematode genus *Pristionchus* in the United States. *Front. Zool.* **3**, 1–15 (2006).
28. Herrmann, M., Mayer, W. E. & Sommer, R. J. Nematodes of the genus *Pristionchus* are closely associated with scarab beetles and the Colorado potato beetle in Western Europe. *Zoology* **109**, 96–108 (2006).
29. Ragsdale, E. J., Kanzaki, N. & Herrmann, M. *Pristionchus pacificus* 77–120 (Brill, 2015).
30. Vilaradaga, J.-P. *Signal Transduction Protocols* 133–148 (Springer, 2011).
31. Kroeze, W. K., Sheffler, D. J. & Roth, B. L. G-protein-coupled receptors at a glance. *J. Cell Sci.* **116**, 4867–4869 (2003).
32. Hong-Bo, S., Zong-Suo, L. & Ming-An, S. LEA proteins in higher plants: Structure, function, gene expression and regulation. *Colloids Surf. B* **45**, 131–135 (2005).
33. Hundertmark, M. & Hinch, D. K. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118 (2008).
34. Kanzaki, N. Simple methods for morphological observation of nematodes. *Nematol. Res. (Jpn. J. Nematol.)* **43**, 15–17 (2013).
35. Minagawa, N. & Mizukubo, T. A simplified procedure of transferring nematodes to glycerol for permanent mounts. *Nematol. Res. (Jpn. J. Nematol.)* **24**, 75–75 (1994).
36. Kikuchi, T., Aikawa, T., Oeda, Y., Karim, N. & Kanzaki, N. A rapid and precise diagnostic method for detecting the pinewood nematode *Bursaphelenchus xylophilus* by loop-mediated isothermal amplification. *Phytopathology* **99**, 1365–1369 (2009).
37. Tanaka, R., Kikuchi, T., Aikawa, T. & Kanzaki, N. Simple and quick methods for nematode DNA preparation. *Appl. Entomol. Zool.* **47**, 291–294 (2012).
38. Ye, W., Giblin-Davis, R. M., Braasch, H., Morris, K. & Thomas, W. K. Phylogenetic relationships among *Bursaphelenchus* species (Nematoda: Parasitaphelenchidae) inferred from nuclear ribosomal and mitochondrial DNA sequence data. *Mol. Phylogenet. Evol.* **43**, 1185–1197 (2007).
39. Kanzaki, N. & Futai, K. A PCR primer set for determination of phylogenetic relationships of *Bursaphelenchus* species within the *xylophilus* group. *Nematology* **4**, 35–41 (2002).
40. Kanzaki, N., Giblin-Davis, R. M., Gonzalez, R., Duncan, R. & Carrillo, D. Description of *Ruehmaphelenchus juliae* n. sp. (Tylenchina: Aphelenchoididae) isolated from an ambrosia beetle, *Xylosandrus crassiusculus* (Motschulsky), from South Florida. *Nematology* **17**, 639–653 (2015).
41. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
42. Posada, D. & Crandall, K. A. Modeltest: Testing the model of DNA substitution. *Bioinformatics (Oxford, England)* **14**, 817–818 (1998).
43. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
44. Larget, B. & Simon, D. L. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16**, 750–759 (1999).
45. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
46. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
47. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
48. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Kajitani, R. *et al.* Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
51. Wences, A. H. & Schatz, M. C. Metassembler: Merging and optimizing de novo genome assemblies. *Genome Biol.* **16**, 207 (2015).
52. Huang, S., Kang, M. & Xu, A. HaploMerger2: Rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics (Oxford, England)* **33**, 2577–2579 (2017).
53. Otto, T. D., Sanders, M., Berriman, M. & Newbold, C. Iterative correction of reference nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* **26**, 1704–1707 (2010).
54. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287 (2017).
55. Parra, G., Bradnam, K. & Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)* **23**, 1061–1067 (2007).
56. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
57. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).
58. Bruna, T., Hoff, K., Lomsadze, A., Stanke, M. & Borodovsky, M. In *Plant and Animal Genome XXVII Conference (January 12–16, 2019)*. (PAG).
59. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36**, D281–D288 (2007).
60. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
61. Cantarel, B. L. *et al.* The carbohydrate-active enzymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
62. Kassambara, A. & Mundt, F. Package 'factoextra'. In *Extract and Visualize the Results of Multivariate Data Analyses*, Vol. 76 (2017).

63. Ho, L. S. T. *et al.* Package ‘phylolm’ (2016) (accessed February 2018); <http://cran.r-project.org/web/packages/phylolm/index.html>.
64. Emms, D. M. & Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
65. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
66. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
67. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
68. Rambaut, A. *FigTree v1. 4.0. A Graphical Viewer of Phylogenetic Trees* (2012) (accessed 11 March 2021). <http://tree.bio.ed.ac.uk/software/figtree>.

Acknowledgements

We thank Noriko Shimoda, Atsuko Matsumoto and Akemi Yoshida for their technical assistance in culturing the nematodes, preparing the permanent mounts and sequencing. This study was supported, in part, by Grants-in-Aid for Scientific Research (B), nos. 26292083, 26292178 and 19H03212 from the Japan Society for the Promotion of Science, Environment Research and Technology Development Fund (4-1401) from the Ministry of the Environment, Japan and JST CREST Grant Number JPMJCR18S7.

Author contributions

N.K., T.I., H.M., K.O., H.K. and T.K. conceived and designed the study and collected the materials. N.K. performed the taxonomic study on *Caenorhabditis auriculariae*. R.T. prepared SEM materials. M.D., S.S. and T.K. analysed genome data. N.K. and T.K. wrote the manuscript with inputs from all the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-85967-z>.

Correspondence and requests for materials should be addressed to T.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021