BMC Genomics

Genome analysis of *Diploscapter coronatus*: insights into molecular peculiarities of a nematode with parthenogenetic reproduction

Hideaki Hiraki[1†], Hiroshi Kagoshima[1,3†], Christopher Kraus[2†], Philipp H. Schiffer[2], Yumiko Ueta[1], Michael Kroiher[2], Einhard Schierenberg[2] and Yuji Kohara[1*]

Abstract

Background: Sexual reproduction involving the fusion of egg and sperm is prevailing among eukaryotes. In contrast, the nematode *Diploscapter coronatus*, a close relative of the model *Caenorhabditis elegans*, reproduces parthenogenetically. Neither males nor sperm have been observed and some steps of meiosis are apparently skipped in this species. To uncover the genomic changes associated with the evolution of parthenogenesis in this nematode, we carried out a genome analysis.

Results: We obtained a 170 Mbp draft genome in only 511 scaffolds with a N_{50} length of 1 Mbp. Nearly 90% of these scaffolds constitute homologous pairs with a 5.7% heterozygosity on average and inversions and translocations, meaning that the 170 Mbp sequences correspond to the diploid genome. Fluorescent staining shows that the *D. coronatus* genome consists of two chromosomes (2n = 2). In our genome annotation, we found orthologs of 59% of the *C. elegans* genes. However, a number of genes were missing or very divergent. These include genes involved in sex determination (e.g. *xol-1*, *tra-2*) and meiosis (e.g. the kleisins *rec-8* and *coh-3/4*) giving a possible explanation for the absence of males and the second meiotic division. The high degree of heterozygosity allowed us to analyze the expression level of individual alleles. Most of the homologous pairs show very similar expression levels but others exhibit a 2–5-fold difference.

Conclusions: Our high-quality draft genome of *D. coronatus* reveals the peculiarities of the genome of parthenogenesis and provides some clues to the genetic basis for parthenogenetic reproduction. This draft genome should be the basis to elucidate fundamental questions related to parthenogenesis such as its origin and mechanisms through comparative analyses with other nematodes. Furthermore, being the closest outgroup to the genus *Caenorhabditis*, the draft genome will help to disclose many idiosyncrasies of the model *C. elegans* and its congeners in future studies.

Keywords: Genome assembly, Allelic expression, Nematode, Parthenogenesis, Meiosis, Cohesin

* Correspondence: ykohara@nig.ac.jp
†Equal contributors
[1]Genome Biology Laboratory, National Institute of Genetics, Mishima, Japan
Full list of author information is available at the end of the article

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 2 of 18

Background

Sexual reproduction including meiosis and subsequent mating is a characteristic feature of eukaryotes allowing them to maximize the diversity of their gene pools and preserve genetic variability. However, sexual reproduction is costly, because recombination breaks up favorable gene combinations faster than it creates new ones [1], and males require significant resources to be produced but do not generate offspring themselves. Although selected representatives of different animal phyla can reproduce parthenogenetically, in most cases this is a facultative feature, depending on environmental conditions. If these are favorable, only parthenogenetic females are found, while under stress they switch to bisexual reproduction [2]. Parthenogenesis has been considered an evolutionary dead end, since it results in the accumulation of deleterious mutations in an irreversible manner known as "Muller's ratchet" [3]. Furthermore, parthenogenetic reproduction should impair adaptability to environmental change because positive mutations present in different individuals will rarely come together within the same individual, eventually leading to extinction [4, 5]. However, certain cases in nature are in conflict with such a scenario. For example, bdelloid rotifers are thought to have followed parthenogenetic reproduction for millions of years without males and meiosis. Similar cases have been reported for certain ostracods, mites and root-knot nematodes [6]. The bdelloid rotifer genome was sequenced recently and it has been claimed that the homogenizing and diversifying roles of sex may have been compensated there by gene conversion and horizontal gene transfer [7, 8]. However, there are reports on the possibility of infrequent sexual reproduction and atypical meiosis in some bdelloid rotifers [9–11]. The plant parasitic root-knot nematode *Meloidogyne incognita* is a mitotic parthenogen, yet this species is an extremely potent and widely distributed plant parasite. Genome sequencing of *M. incognita* and several related nematodes has been accomplished [12–14]. A series of unusual features were observed in the genome, which are probably linked to plant-parasitic lifestyle, absence of meiosis and parthenogenetic reproduction [15].

Among nematodes, various modes of reproduction are found: dioecy (male and female), androdioecious hermaphrodites (self-fertilizing female) as well as parthenogenesis (development from unfertilized eggs). One such parthenogen is the Rhabditid species *Diploscapter coronatus*, a close outgroup to the model genus *Caenorhabditis*. Many parthenogenetic taxa show infrequent occurrences of males, but in *D. coronatus* no males have been reported under laboratory conditions [16], even under thermal stress increasing the occurrence of males in *C. elegans* and other nematodes [17, 18]. A distinct feature of *D. coronatus* is its truncated meiosis: In contrast to *C. elegans* (and other dioecious or hermaphroditic nematodes) with its

two consecutive meiotic divisions, polar body extrusion takes place only once [16]. Thus, *D. coronatus* appears to skip some steps in meiosis. As no parthenogenetic species has been found in the genus *Caenorhabditis* and as *Diploscapter* belongs to the closest outgroup, it is an attractive target to study the genetic basis of parthenogenesis. Moreover, *D. coronatus* is phylogenetically located between *C. elegans* and the well-studied satellite system, *Pristionchus pacificus* [19, 20]. A genomic comparison between *D. coronatus* and the two nematode model organisms will allow a deeper understanding of evolutionary changes on the cellular and molecular level.

Results

Genome assembly revealed the paired structure of the *D. coronatus* genome

Initially, we obtained more than 270,000 ESTs (expressed sequence tags, or end sequences of cDNA clones). These were classified into about 13,000 groups based on the 3′ end sequence comparison (see Methods), 48% of which showed strong homology with *C. elegans* proteins, and interestingly most of the ESTs showed clear heterozygosity within the groups. We assume that this heterozygosity is associated with parthenogenetic reproduction, where allelic differences are accumulated and maintained like in somatic cell lines [21], or originated from interspecies hybridization [5]. Subsequently, we obtained genome shotgun reads by various methods from Sanger to next generation sequencing together with fosmid end sequences (Additional file 1: Table S1). All sequence reads were assembled with the Celera assembler [22] resulting in a total genome span of 170 Mbp (511 scaffolds, N_{50}: 1.0Mbp) (Table 1). This value is consistent with the genome size calculated from the k-mer distribution (Additional file 2).

We performed a homology search among all the scaffolds. A Dot plot (Fig. 1a) shows that, in addition to a 100% match between selves (red lines), highly homologous regions (~94% similarity) are visualized as a series of purple lines that appear with an almost 45 degree slope by rearranging the order of scaffolds, suggesting a

Table 1 Statistics of the genome assembly

	Number	Length (bp)		
		Total	Max.	N_{50}
Scaffolds	511	170,470,384	3,561,896	1,007,652
Contigs	867	169,424,175	1,740,259	487,148
Paired regions[a]	6690	152,151,424	250,158	
SNVs[b]	8,685,973			
In/Dels[b]	997,343			

[a]Homologous regions identified by all-vs.-all scaffold alignment. See Methods for details
[b]Numbers detected in the 152 Mbp paired regions

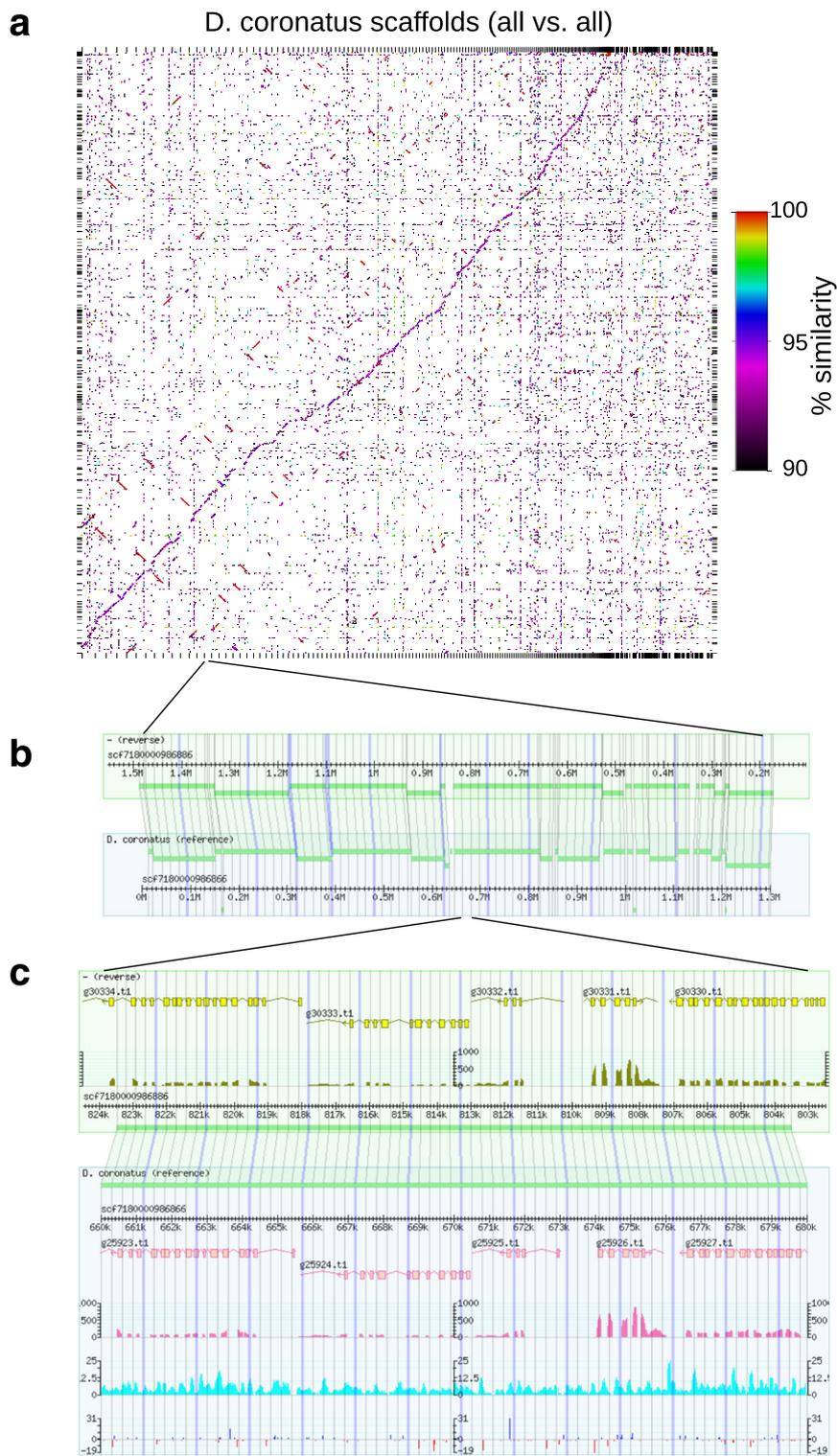Hiraki *et al. BMC Genomics* (2017) 18:478

Page 3 of 18

Fig. 1 (See legend on next page.)

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 4 of 18

(See figure on previous page.)
Fig. 1 Paired structure of the *D. coronatus* genome. **a** Dot plot of all-vs.-all comparison of the scaffold sequences. Alignments with 90% or more identity are plotted. For most sequences, *purple line* fragments are found indicating the presence of partner sequences with about 94% identity. Both axes are ordered to emphasize paired structure of the genome by MUMmer. Along the X-axis the scaffolds that have homologous counterparts are ordered by length (longest first), and along the Y-axis the corresponding counterpart scaffolds are arranged. *Red lines* indicate trivial hits to themselves. **b** A long syntenic region (1.3Mbp) is visualized using the GBrowse syntenic browser. 65% of the scaffold scf7180000986866 (*lower box*) and its homologous region in the scaffold scf7180000986886 (*upper box*) are shown. The *thick green bars* indicate sequences that have homologous counterparts; these are linked by *light green* shading. A few unpaired short bars indicate that their counterparts are found in other parts of the genome, probably as a result of translocation. Scale unit is Mbp. **c** For a detailed view of the paired structure, a 20 kbp region is magnified. Gene models (*yellow* or *pink*, boxes show exons and *arrows* show the gene orientation) and the histograms of RNA-seq coverage are shown under the gene models in both boxes. In the *lower box*, numbers of mismatches per 100 bp window (*cyan*) and lengths of insertions and deletions (*blue* and *red*, respectively, shown at 1 bp before In/Del site) are plotted in the 3rd and bottom rows, respectively. Scale unit is kbp

linear correlation between corresponding scaffolds. In other words, most of the contigs have homologous counterparts in our assembly. In fact, 89.3% of the scaffold sequences were covered by reciprocally best alignment segments. Thus, the genome consists of pairs of allelic sequences with a high degree of heterozygosity. Figure 1b shows a blow-up of a paired region: the homologous regions between the two scaffolds are aligned. Figure 1c shows a further blow-up of a part of the homologous region: in addition to the gene models and transcriptome (see below), the frequency of the single nucleotide variations (SNV) and short insertions and deletions (In/Dels) are depicted in the bottom part. The SNV frequency is variable but shows a tendency to be relatively high in intron regions. As shown in Table 1, the overall SNV frequency is 5.7% and In/Del 0.66%. The distribution of SNV ratios in individual CDS (coding sequences), introns and intergenic regions indicates that the occurrence of SNV is relatively uniform over the genome (Additional file 3). Inversions and translocations are also found (Additional file 4).

Genome size and chromosome number
We measured the nuclear DNA content by flow cytometry to estimate the genome size using *C. elegans* (100 Mbp × 2/nucleus) and *D. melanogaster* (140 Mbp × 2/nucleus) as size references. These measurements indicate that the *D. coronatus* nuclear DNA comprises about 140 Mbp (Additional file 5). As this kind of measurement may contain considerable errors (up to ~25%) depending on the size standard used for the estimation [23, 24], the value of assembled 170 Mbp is within the error range. Microscopic measurements of fluorescently labeled nuclear DNA also indicated a similar value (Additional file 6). Thus, we conclude that our 170 Mbp assembly span closely reflects the actual genome size of *D. coronatus*.

For karyotype analysis, we fluorescently marked chromosomes at two stages: late oocytes prior to the start of cleavage, and in early blastomeres (Fig. 2). At the same stage where in the *C. elegans* oocyte 4n = 24 chromatids can be detected (data not shown), we observed 4n = 4 chromatids in the *D. coronatus* oocyte (Fig. 2c). In early embryonic cells 12 chromosomes (2n = 12) were observed in *C. elegans*, but only 2 in *D. coronatus* (Fig. 2f). These results show that in the latter the diploid set consists of 2 chromosomes. This is in accordance with the earlier report of a different isolate by Hechler [25].

Gene content of the parthenogenetic nematode *D. coronatus*
Repeat sequences and RNAs
Repeat sequences occupy 17.4% of the *D. coronatus* genome. In these repeats transposon-like sequences were identified and the number of these was comparable to *C. elegans* (Additional file 7: Table S5). We identified six 28S and five 18S rRNA genes. Two sets of these genes are found at the edges of long contigs and the others are in five short (< 9kbp) scaffolds. tRNA and other RNA families were also identified (Additional file 7: Table S3). Splice-leader sequences SL2 were found in addition to SL1 (Additional file 7: Table S4).

Protein coding genes
We obtained 140 million RNA-Seq reads from a mixed-stage population ranging from embryo to adult. Using Augustus [26, 27] with incorporation of the RNA-Seq data, we predict 34,421 protein coding genes, of which 58% are supported by our previously established EST library. Analysis with BUSCO [28] showed that the genome completeness was reasonable, taking into account the known imprecision of BUSCO for non-model organisms (Additional file 1: Table S2). Although 90 of the predicted genes showed homology only with non-metazoan sequences in the databases, we have no direct evidence for horizontally transferred genes in the *D. coronatus* genome.

Among the protein coding genes, 20,264 (59%) were shown to be orthologous to the *C. elegans* genes listed in WormPep [29] according to InParanoid [30, 31] (Table 2). Of these, 16,092 genes consisted of 8046 heterozygous pairs (doubletons), which were orthologous to a single *C. elegans* gene. They are considered as allelic
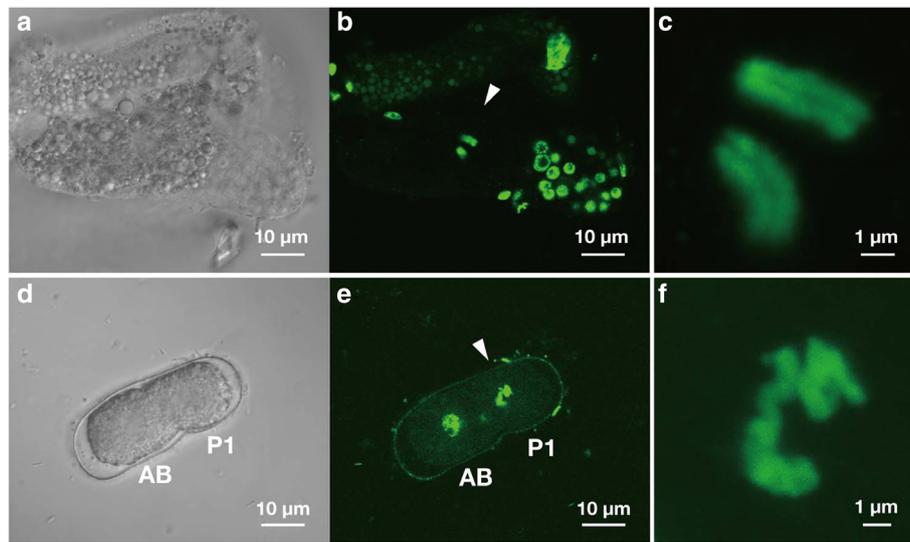
Hiraki *et al. BMC Genomics* (2017) 18:478

Page 5 of 18

Fig. 2 DAPI staining of *D. coronatus* chromosomes. **a**, **b** A gonad with a single uncleaved egg cell. **c** Magnification of the chromosomes in **b** (*arrowhead*). Note separation into chromatids. **d**, **e** A 2-cell embryo. **f** Magnification of the two condensed chromosomes in the P1 cell in **e** (*arrowhead*). **a**, **d** Nomarski images. **b**, **c**, **e**, **f** Fluorescent confocal microscopic images. Rotatable 3D images of c and f are given in Additional file 14

partners encoded in the two "haploid" genomes present in our assembly. We call this phenomenon (Dc:Ce) 2:1 relationship (Additional file 8). The other 374 pairs (748 genes) show a 2:2+ relationship, where the *D. coronatus* genes are homologous to two or more *C. elegans* genes, suggesting a gene expansion in the *C. elegans* genome. Among the remaining genes, 3214 show various relationships like 3+:1 or 3+:2+ where more than three *D. coronatus* genes that are homologous to each other are homologous to one or more *C. elegans* genes. These are classified into 559 groups. The remaining 210 genes in the *D. coronatus* genome have no homologous partners and are assigned as singletons.

Among the 14,157 (41%) protein coding genes that were not orthologous to *C. elegans* genes 5850 formed pairs. In addition, 743 genes in 191 groups formed triplets or more. The remaining 7564 genes were assigned as singletons (Table 2). The order of genes and their direction of transcription (co-linearity) are well conserved between the two allelic regions in the *D. coronatus* genome (see Fig. 1c). Thus, the counterparts of these singleton genes may have either been eliminated from the genome, or they might have been missed in our clustering analysis due to their high divergence or faulty gene predictions. Furthermore, there are still many contig gaps into which some counterparts might fall. Therefore, a closer examination of the sequences may identify more allelic counterparts.

Taken together, the predicted 34,421 genes are classified into 11,345 allelic pairs (22,690 genes), 7774 singletons and 750 groups of 3957 genes. It is difficult to define the gene number of the genome consisting of a pair of such

Table 2 Number of *D. coronatus* protein coding genes

Category[a]	genes		pairs/groups
Orthologs of *C. elegans* genes	20,264		
doubleton to a single *C. elegans* gene (2:1)		16,092	(8046 pairs)
doubleton to multiple *C. elegans* genes (2:2+)		748	(374 pairs)
gene family to *C. elegans* gene(s) (3+:1, 3+:2+)		3214	(559 groups)
singleton to *C. elegans* gene(s) (1:1+)		210	
Non-orthologs of *C. elegans* genes	14,157		
doubleton (2:0)		5850	(2925 pairs)
gene family (3+:0)		743	(191 groups)
singleton (1:0)		7564	
total protein coding genes	34,421		

[a]Relationships between genes of this category, e.g. (2:1) and (2:2+), are depicted in Additional file 8

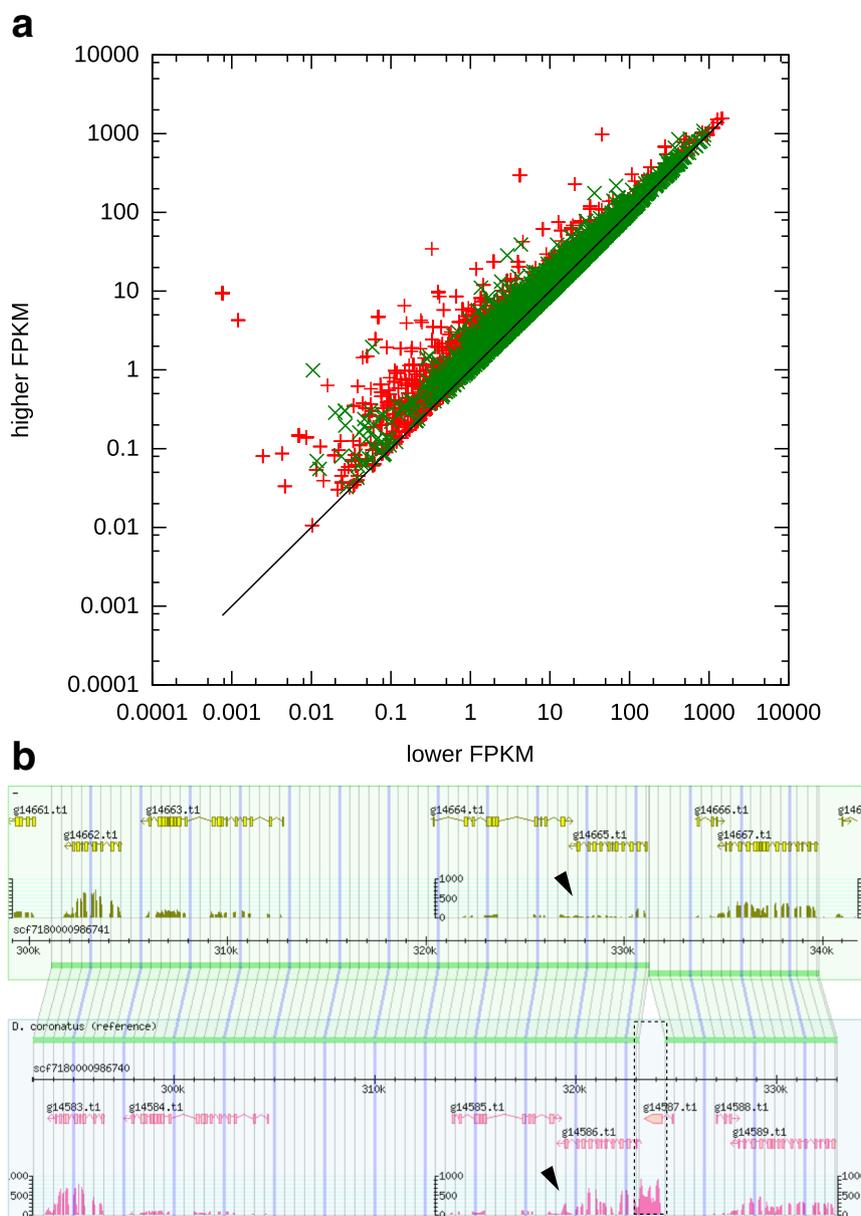Hiraki *et al. BMC Genomics* (2017) 18:478

Page 6 of 18

Fig. 3 Allelic expression analysis. **a** Comparison between allelic gene expressions. The FPKM values for allelic pairs of *D. coronatus* genes, 7306 (*green cross*) with a single *C. elegans* ortholog and 2526 (*red cross*) without *C. elegans* ortholog, are plotted in such a way that the lower value is on the X-axis and the higher value on the Y-axis. **b** Region with 10-fold difference in the allelic expression level. A 40kbp region in the scaffold scf7180000986740 (*lower box*) and its homologous region in the scaffold scf7180000986741 (*upper box*) are shown with the gene models and the histograms of RNA-seq coverage (*pink* or *yellow*). The allelic pair genes g14665.t1 and g14586.t1 show a big difference in the RNA-seq pattern (*arrowheads*), while others show almost identical patterns. The gene g14587.t1 seems to be inserted just upstream of gene g14586.t1 (*dashed box*)

diverged chromosomes. Under the simple assumption that all genes consist of allelic pairs except for singletons, the number of genes in the conventional diploid *D. coronatus* genome is 21,098 (=11,345 + 7774 + 3957/2). This number is close to what has been found in *C. elegans* [32]. It remains to be determined to what extent the paired genes may differ from each other in their function.

In parthenogenetic reproduction the allelic regions must have changed and evolved independently from each other. We calculated the base substitution ratio between the gene pairs that showed the 2:1 relationship in order to see whether there was any selection on the gene pairs. With respect to the CDS regions of the paired genes, the median identity is 97.2% in total, 93.7% at the

Table 3 Meiosis-related genes in *D. coronatus*

Gene (alias)[a]	Presence[b]	Homolog/ Ortholog[c]	Gene (alias)[a]	Presence[b]	Homolog/ Ortholog[c]
air-2	+	Aurora/IPL Kinase	mes-4	+	WHSC1; SETDB1
aspm-1	+	ASPM	met-1	+	SETD2
atm-1	−	ATM	met-2	+	SETDB1
brc-1	+	BRCA1	mix-1	+	SMC2
brc-2	−	BRCA2	mpk-1	+	ERK
capg-1	+	hCAP-G	mre-11	+	MRE11A
capg-2	+	hCAP-G2	mrg-1	+	Chromodomain protein
cep-1	−	p53	mrt-2	+	RAD1
chk-2	+	CHK2	msh-4 (him-14)	+	MSH4
cls-2	+	CLASP	msh-5	+	MSH5
cmd-1	+	CALM1	mus-81	+	MUS81
coh-3	−	Rad21/Rec8	pch-2	+	TRIP13
coh-4	−	Rad21/Rec8	plk-1	+	PLK1
com-1	−	CtIP / Sae2	plk-2	+	PLK1
cosa-1	+	CTND1	pph-4.1	+	PPP4C
cra-1	+	NAA25	prom-1	+	FBXO47
cye-1	+	CCNE1	rad-50	+	RAD50
dpy-26	+	hCAP-H	rad-51	+	Rad51; DMC1
dpy-28	+	hCAP-D2	rad-54	+	Rad54
dsb-1	−		rec-1	−	
dsb-2	−		rec-8	−	Rec8
egl-1	−		rfs-1	+	RAD51C
exo-1	+	exo1	rpa-1	+	RPA1
fcd-2	−	FANCD2	rtel-1	+	RTEL1
gsp-2	+	PP1CA	scc-2 (pqn-85)	+	SCC2
hal-2	−		scc-3	+	SCC3
hcp-6	+	hCAP-D3	slx-1	+	SLX1
helq-1 (hel-308)	+	HELQ	smc-1 (him-1)	+	SMC1
him-3	−	HORMAD1; Hop1	smc-3	+	SMC3
him-5	−		smc-4	+	SMC4
him-6	+	BLM; RecQ	smc-5	+	SMC5
him-8	−		smc-6	+	SMC6
him-17	−		spd-3	+	
him-18	+	SLX4	spo-11	+	SPO11
him-19	+		sun-1	+	SUN-domain family
hpr-9	+	RAD9	syp-1	−	
htp-1	+	HORMAD1; Hop1	syp-2	−	
htp-2	+	HORMAD1; Hop1	syp-3	−	
htp-3	−	HORMAD1; Hop1	syp-4	−	
hus-1	+	HUS1	tim-1	+	TIMELESS
kle-2	+	hCAP-H2	unc-116	+	Kinesin 1
klp-18	+		xnd-1	−	
klp-19	+	Chromokinesin	xpf-1	+	ERCC4

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 8 of 18

Table 3 Meiosis-related genes in *D. coronatus (Continued)*

lab-1	−		zhp-3	+	RNF212; Zip3
lin-5	+	NuMA	zim-1	−	
lin-41	+		zim-2	−	
mei-1	+	KATNAL1	zim-3	−	
mei-2	−		zyg-12	+	KASH-domain family

[a]Meiosis-related genes in *C. elegans* (modified from [49])
[b]Presence (+) or absence (−) of their *D. coronatus* homologs based on InParanoid analysis
[c]Names of the respective orthologous or homologous genes in other organisms

3rd letter, and 93.0% at the 4-fold degenerated site. This sequence divergence is similar to the diversity observed in *Caenorhabditis remanei* that shows one of the highest diversities in eukaryotes [33]. Of 8046 such pairs, 7306 pairs show size differences of less than 100 bp. We calculated dN/dS values (the ratio of substitution rates at non-synonymous and synonymous sites) of these pairs. Significant dN/dS values ($P < 0.05$) were obtained for 6760 pairs. The median of the values was 0.12 and none of them exceeded one, indicating that the majority of genes had diverged under negative selection and no sign of positive selection was detected.

Allelic gene expression in *D. coronatus*
Heterozygosity between the allelic genes is so high that even short sequences like 100 bp RNA-Seq reads can be assigned to either of the allelic sequences. This allowed us to analyze the expression level of individual alleles ("allelic expression analysis"). As shown in Figs. 1c and 3a, in spite of the high sequence divergence which must cause changes in regulatory sequences, most of the homologous pairs show very similar expression levels: among the 7306 gene pairs, FPKM (fragments per kilo bases of transcript per million fragments sequenced) ratio for 6736 pairs is less than 1.5-fold and the correlation coefficient of FPKM is 0.99. However, some pairs show deviant expression levels: 121 pairs show >2-fold difference, and 5 pairs show >5-fold difference (Fig. 3a). For example, between the orthologous genes of K03H6.1 (G-protein-coupled receptor), the expression level of one allele, g14586.t1, is 10-fold higher than the other allele, g14665.t1 (Fig. 3b). In this case, an insertion (or deletion at the opposite site) of the gene g14587.t1, which has the mariner transposase-like sequence, is found 370 bp upstream of the translational initiation codon of g14586.t1. This insertion might cause the increase of transcription. We searched the genome for this transposon-like sequence and found 54 loci. In one case, it is inserted in an intron and an increase of the downstream transcription is observed. These differentially expressed genes are dispersed and not clustered in the genome.

Peculiarities in the gene repertoire of a parthenogenetic nematode
We examined the presence or absence of *D. coronatus* orthologs of *C. elegans* genes for core biological processes [29, 34] by InParanoid analysis. In the following sections, we summarize the results focusing on genes related to the mode of reproduction (See Additional file 9 for others).

Genes involved in sex determination
In *C. elegans*, sex is determined by the X chromosome to autosome (X/A) ratio, which is read by chromosomal counting factors that regulate gene expression in the sex-determination cascade, i.e. *xol-1* and dosage compensation (*sdc*) genes [35–37]. The sex-determination signal is transmitted to individual cells through *her-1/tra-2* ligand-receptor genes [38–40]. At the end, the terminal transcription factor TRA-1 regulates all aspects of hermaphrodite sexual differentiation in somatic cells.

In *D. coronatus,* a number of key components in this pathway e.g. *xol-1*, *tra-2*, were missing in our search (Additional file 10). It may not be surprising that a parthenogenetic species lost genes in the sex determination pathway. However, it is interesting that a considerable number of orthologs are retained in this pathway, i.e. SEX-1 and FOX-1 (X chromosome counting factors), and HER-1 (TRA-2 ligand). In *D. coronatus*, they might function in a pathway other than the original sex determination pathway.

Genes involved in meiosis
In *C. elegans*, sister chromatid cohesion is established during DNA replication [41, 42] by a cohesin complex that contains the meiosis-specific kleisins, REC-8 and COH-3/4. The pairing of homologous chromosomes is initiated at the pairing center by recruiting the chromosome-specific zinc-finger proteins ZIM-1/2/3 and HIM-8 [43–46]. These proteins anchor chromosomes to the nuclear envelope through binding to SUN-1 and ZYG-12 proteins present there. This process facilitates homologous chromosome pairing [47]. To form the synaptonemal complex (SC), HTP-3 localizes on the chromosomes and recruits the other axial element components HIM-3, HTP-1/2 and the transverse filaments SYP-1/2/3/4 [41, 48]. Then, meiotic
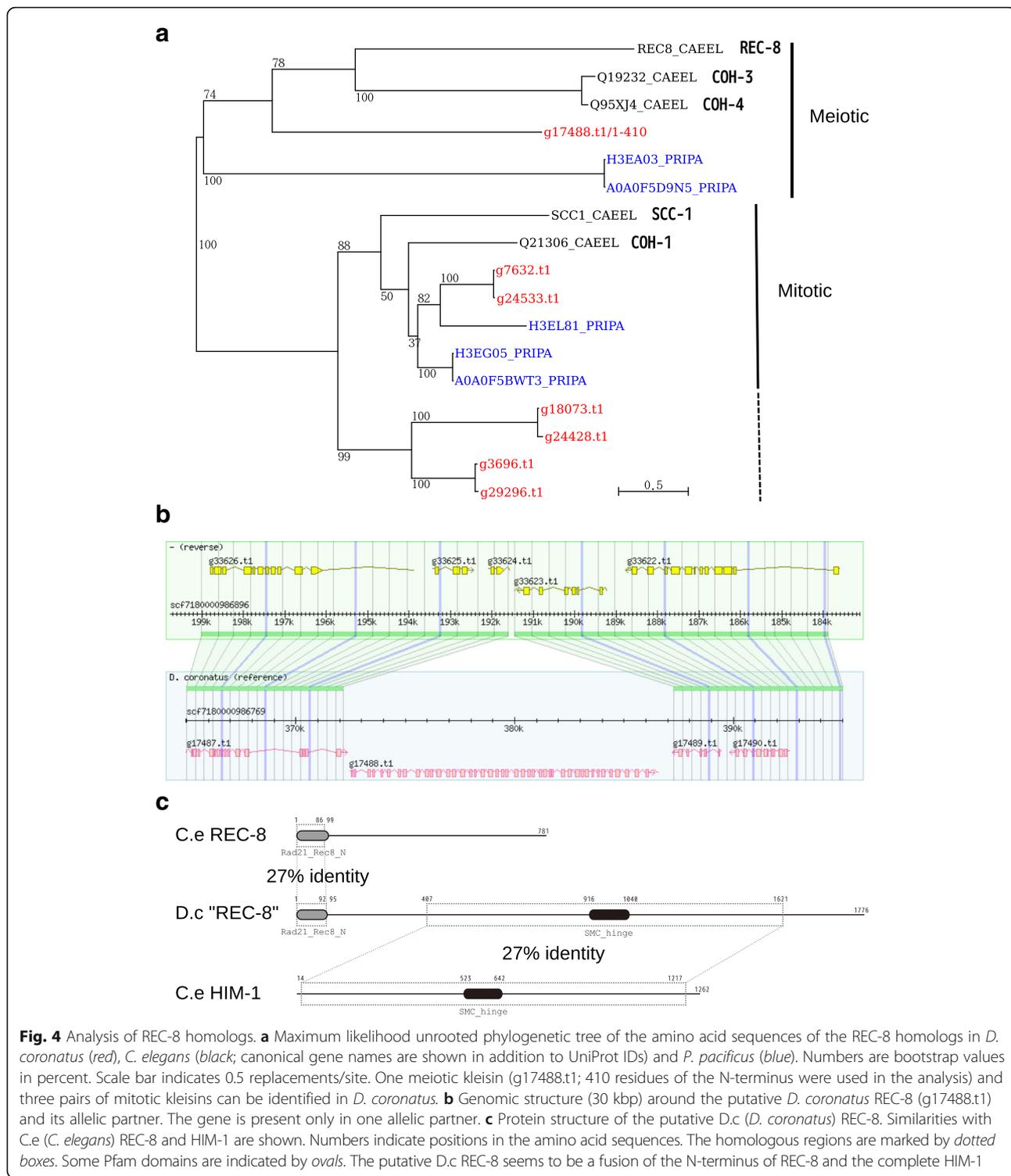
Hiraki *et al. BMC Genomics* (2017) 18:478

Page 9 of 18

Fig. 4 Analysis of REC-8 homologs. **a** Maximum likelihood unrooted phylogenetic tree of the amino acid sequences of the REC-8 homologs in *D. coronatus* (*red*), *C. elegans* (*black*; canonical gene names are shown in addition to UniProt IDs) and *P. pacificus* (*blue*). Numbers are bootstrap values in percent. Scale bar indicates 0.5 replacements/site. One meiotic kleisin (g17488.t1; 410 residues of the N-terminus were used in the analysis) and three pairs of mitotic kleisins can be identified in *D. coronatus*. **b** Genomic structure (30 kbp) around the putative *D. coronatus* REC-8 (g17488.t1) and its allelic partner. The gene is present only in one allelic partner. **c** Protein structure of the putative D.c (*D. coronatus*) REC-8. Similarities with C.e (*C. elegans*) REC-8 and HIM-1 are shown. Numbers indicate positions in the amino acid sequences. The homologous regions are marked by *dotted boxes*. Some Pfam domains are indicated by *ovals*. The putative D.c REC-8 seems to be a fusion of the N-terminus of REC-8 and the complete HIM-1

recombination takes place triggered by DNA double-strand breaks via SPO-11, followed by strand invasion mediated by RAD-51 and RAD-54 [41].

Screening our *D. coronatus* genome, we could not detect credible orthologs of many key genes in meiotic development [49], such as, *rec-8, coh-3/4, zim-1/2/3, him-8,* and *syp-1/2/3/4* by InParanoid analysis (Table 3). We thus further searched for homologs of these genes using the Pfam database [50], and found four kleisin homologs (three allelic pairs and a singleton) in the *D.*

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 10 of 18

coronatus genome. According to our phylogenetic analysis, *D. coronatus* possesses three alleles of mitotic kleisin (SCC-1/COH-1) homologs and one singleton of the meiotic kleisin (REC-8/COH-3/4) homolog (Fig. 4a). The meiotic kleisin homolog, g17488.t1 (D.c "REC-8"), however, shows atypical features: (1) It does not have an allelic counterpart in the genome while all three mitotic kleisins are present as allelic pairs (Fig. 4b), (2) It contains only the N-terminal domain of REC-8 and fuses with HIM-1(SMC-1), an interacting structural protein of kleisin in the cohesin complex (Fig. 4c), (3) It lacks the C-terminal domain of kleisin known as the Rad21/Rec8-like domain C-terminal (IPR023093). Usually, the N-terminal and C-terminal domains of kleisins interact with SMC-3 and HIM-1(SMC-1), respectively. However, in *D. coronatus* the REC-8 homolog has lost its C-terminal domain for SMC-1 binding and instead is fused directly to SMC-1. The meiotic kleisins are essential factors to hold sister chromatids together during meiosis, thus, their absence or divergence may well be related to parthenogenetic reproduction. Phylogenetic analysis of other meiosis-specific genes showed that they have orthologous counterparts in the *C elegans* genome mostly in pairs (Additional file 11).

The pairing center recognition proteins, ZIM-1/2/3 and HIM-8 are absent in *D. coronatus,* but their interacting proteins SUN-1 and ZYG-12 are present (Table 3), suggesting that unknown proteins may mediate chromosome-nuclear membrane interaction. Although SC proteins are evolutionarily variable and share only a low similarity [51], the obvious loss of the synaptonemal complex (SC) components SYP-1/2/3/4, may not be compatible with homologous chromosome pairing in *D. coronatus.*

Discussion

In the present work, we have analyzed the genome of the parthenogenetic nematode *Diploscapter coronatus*. A central question related to parthenogenesis is how such organisms are nevertheless able to preserve the necessary diversity of the gene pool. *D. coronatus* appears to be a good system to elucidate the molecular idiosyncrasies of parthenogenesis as it is a close relative of the well-studied hermaphroditic model organism *C. elegans*. A genome comparison between these two species should provide clues to the genetic basis of different reproductive modes.

Our analysis showed that the *D. coronatus* genome possesses a high degree of heterozygosity or allelic divergence. Genomes of organisms collected from the wild are often difficult to sequence due to their heterozygosity. This is because the genome assembly algorithm recognizes heterozygous regions as branch structures, leading to the termination of contig extension. Thus, some researchers resorted to inbred lines. The potato

genome project used offspring produced by parthenogenesis (containing only the haploid genome of the mother) to avoid the heterozygosity problem [52]. Alternatively, various assembly programs including PLATA-NUS have been developed to overcome the problem with wild-derived organisms [53]. In *D. coronatus,* however, we were able to successfully assemble the shotgun reads into 170 Mbp sequences using a conventional assembly software. This is probably because this genome is so heterozygous that the genome assembler recognizes the allelic regions as separate sequences. The distribution of sequence read coverage shows a normal distribution (Additional file 12), meaning that rarely two different regions are assembled together because of nearly identical sequences. However, this means, in turn, that less heterozygous or nearly identical sequences, e.g. rRNA genes, result in a branched structure and thus it becomes difficult to assemble them into long contigs/scaffolds; indeed, in our assembly all rRNA genes are located at the end of contigs. Therefore, we plan to analyze the genome structure on a larger scale, hopefully from telomere to telomere, by using a new approach such as the Irys technology [54].

As mentioned above, 89% of the 170 Mbp assembled sequences can be aligned in pairs (Fig. 1). These homologous paired sequences are on average 94.3% identical (5.7% heterozygous) at the nucleotide level, and, if focused on CDS, the identity is 97.2%. This value is comparable to that of the bdelloid rotifer *Adineta vaga* (96.2%) [7], which reproduces as a constitutive mitotic parthenogen [21]. The genome of this rotifer is degenerate tetraploid with allelic pairs sometimes found on the same chromosome, referred to as permanent translocation heterozygosity (PTH), preventing meiosis [8, 10, 55]. However, such a peculiarity is not found in the *D. coronatus* genome, raising the question of how this organism carries out parthenogenetic reproduction.

Normally, during meiosis I, homologous chromosome separation takes place and the primary oocyte divides into two daughter cells (the secondary oocyte and the first polar body), each carrying two identical sister chromatids. Thus, if meiosis II were suppressed like in *D. coronatus,* even if crossing-over would take place, homozygosity would be preserved. However, we found that its genome exhibits an extraordinary high degree of heterozygosity, which corresponds to the level called "hyperdiversity" [56]. A look at the process of sister chromatid separation may help to solve the apparent contradiction.

In *C. elegans*, this critical event requires the function of the meiotic kleisins, REC-8, COH-3 and COH-4. These proteins tether sister chromatids together to assure proper separation of homologous chromosomes during meiosis I [42]. The loss of all three meiotic kleisins (REC-8, COH-3 and COH-4) results in premature

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 11 of 18

sister chromatid separation during the first meiotic division and subsequent inhibition of meiosis II [57]. Our analysis revealed that there are no meiotic kleisin orthologs in the *D. coronatus* genome other than a single atypical one. Phylogenetic analysis revealed that *D. coronatus* possesses three pairs of mitotic kleisin homologs, however, two are located in a different branch compared to the mitotic kleisin "ortholog (g7632.t1/g24533.t1)" (Fig. 4a). Although branch lengths indicate that these genes belong to the mitotic kleisins, we cannot exclude that they may take over a function in meiosis. The atypical homolog lacks the C-terminal domain of REC-8 and contains only the N-terminal domain that is directly fused to HIM-1 (SMC-1) homolog. The function of this atypical homolog is not known, but it may result in a similarly modified meiosis I as in the manipulated *C. elegans* [57]. Another possible mechanism is the so-called "inverted meiosis" where sister chromatid separation occurs during meiosis I. This phenomenon is found under natural conditions in diverse animals and plants with holocentric chromosomes [58, 59]. Such organisms face a specific kinetochore geometry problem, for which inverted meiosis is a possible solution. *C. elegans* and all studied members of neighboring nematode clades also possess holocentric chromosomes, nevertheless *C. elegans* follows the conventional meiotic order [60–63]. It remains to be tested whether *D. coronatus* makes use of this inverted meiosis.

It has been claimed that parthenogenesis commonly arises via interspecies hybridization [5]. The plant-parasitic nematode *Meloidogyne incognita*, which reproduces by obligate mitotic parthenogenesis, is thought to originate from such a hybridization event [15]. A comparative genome analysis of three *Meloidogyne* nematodes, *M. incognita*, *M. floridensis* and *M. hapla*, revealed the complex hybrid origin of the *M. floridensis* [14]. Some of the *M. floridensis* and *M. incognita* genome features are similar to those observed in *D. coronatus*. More than half (64%: 55 Mbp / 86 Mbp) of the *M. incoginita* genome consists of genomic regions in two copies [12], (*D. coronatus*: 89%: 152 Mbp / 170 Mbp). The nucleotide divergence between the pairs is 8% in *M. incognita* and 5.7% in *D. coronatus*. The *M. incognita* genome has large duplicated and rearranged regions, which may restrict recombination of chromosomes, while translocations and inversions are observed in the *D. coronatus* genome (Additional file 4). However, there are differences, too. In *M. incognita* no meiosis occurs during the production of the female gamete and the eggs are derived from unreduced oocytes by mitotic cell division, whereas *D. coronatus* executes meiosis, albeit truncated. Transposable elements and repetitive sequences, which are hypothesized to be related to the asexual mode of reproduction, comprise 36% of the *M. incognita*

genome but only 17.5% of the *D. coronatus* genome (Additional file 7: Table S5). The latter value is similar to that in nematodes showing sexual reproduction (16.5%: *C. elegans*, 22.4%: *C. briggsae*) [6, 64]. These data suggest that the mechanism of how parthenogenesis was acquired differs between these two species.

In the *D. coronatus* genome, nearly 90% of the assembled sequences that have paired structure show good co-linearity over a long range, although with many inversions and translocations (Additional file 4). It is also remarkable that expression levels are extremely similar between the allelic genes despite the high heterozygosity of the *D. coronatus* genome. If *D. coronatus* was the product of an interspecies hybridization, it must have taken place between very close relatives. Therefore, we have started to search for close relatives of this species with bisexual reproduction as potential parent species of our strain. So far, we only found representatives of the neighboring genus *Protorhabditis*. With respect to chromosomes, two of them are like *D. coronatus* (2n = 2) while another one is like *C. elegans* (2n = 12) [20]. Alternatively, whole genome duplication (WGD) followed by diversification of the gene duplicates (ohnologs) could have led to a similar situation as after interspecies hybridization. Finally, a mechanism called "Meselson effect", i.e. an independent accumulation of mutations, inversions and translocations as a consequence of parthenogenetic reproduction could be responsible for the observed diversity of the gene pool [1, 21] . Our current data do not allow us to determine the origin of heterozygosity in *D. coronatus,* however, the dN/dS ratio might give a clue. If *D. coronatus* is a result of hybridization, the considerable divergence between the two gene copies would probably indicate the original divergence between the two parent species and thus should have a strong signature of negative selection. In the WGD model, the ohnologs should acquire deleterious mutations resulting in a relatively high dN/dS ratio. The same should be true when a Meselson effect applies. The dN/dS ratio in *D. coronatus* did not exceed one and the median was 0.12, indicating negative selection. Thus, these data appear to be in favor of a hybridization origin. In any case, the genome analyzed in this work provides a solid basis to further explore the mechanism of parthenogenesis and the evolution of nematode diversity.

Conclusions

Our high-quality draft genome of *D. coronatus* reveals the genome peculiarities of a parthenogenetic nematode. We obtained a 170 Mbp draft genome in only 511 scaffolds with a N50 length of 1 Mbp. Nearly 90% of these scaffolds constitute homologous pairs with a 5.7% heterozygosity together with many inversions and translocations, and

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 12 of 18

most of the genes exist in two distinct alleles. These features mean that the 170 Mbp sequences correspond to the diploid genome. DAPI staining shows that the *D. coronatus* genome consists of two chromosomes (2n = 2). The high degree of heterozygosity allowed us to analyze the expression level of individual alleles. Most of the homologous pairs show very similar expression levels but others exhibit a 2–5-fold difference.

The draft genome provides some clues to the genetic basis for parthenogenetic reproduction. In our genome annotation, we found orthologs of 59% of the *C. elegans* genes. However, a number of genes were missing or very divergent. These include genes involved in sex determination (e.g. *xol-1, tra-2*) and meiosis (e.g. the kleisins *rec-8* and *coh-3/4*) giving a possible explanation for the absence of males and the second meiotic division.

This draft genome constitutes a solid basis for the elucidation of fundamental questions related to parthenogenesis such as its origin and underlying mechanisms in conjunction with comparative analyses of other nematodes. Furthermore, being the closest outgroup to the genus *Caenorhabditis*, our draft genome can help to disclose many idiosyncrasies of the model *C. elegans* and its congeners in future studies.

Methods
Strain and culture
Diploscapter coronatus strain PDL0010 was originally obtained from Prof. P. De Ley, Dept. of Nematology, University of California, Riverside and has been maintained in the Schierenberg laboratory [16]. The strain was cultured at 20 °C on the standard NGM agarose plates that were seeded with the OP50 strain of *Escherichia coli* as a food source [65] and covered with a thin layer of distilled water to prevent the nematodes from digging into the agar.

DNA and RNA preparation
D. coronatus were washed off the agar plates and collected on 10 μm-mesh nylon filters. The nematodes were transferred to a 1-l flask containing 100 ml of distilled water and incubated for 2 h to allow digestion of remaining food bacteria. Nematodes were collected by filtration, aliquoted ~200 mg into 2.2 ml tubes and stored at -80 °C. 200–400 mg of packed worms were ground in a mortar in liquid nitrogen and used for a single DNA/RNA preparation. Genomic DNA was purified with the Genomic-tip 500/G Kit, according to the manufacturer's instructions (Qiagen, Hilden, Germany). RNA was purified by RNAgents Total RNA Isolation System (Promega, Fitchburg, WI, USA) and polyadenylated RNA was purified with a mRNA Purification Kit (GE Healthcare Life Sciences, Buckinghamshire, UK) using an Oligo(dT)-cellulose column.

Library construction and sequencing for genomic DNA
Sanger sequencing was performed as described [66]. Briefly, for shotgun libraries, *D. coronatus* DNA was sheared randomly by Hydroshere (DIGILAB, Marlborough, MA, USA), and then the sheared DNA was end-repaired, phosphorylated and ligated into the SmaI site of pUC18 with the TaKaRa BKL Kit (Takara, Shiga, Japan). The ligated samples were purified by phenol extraction and transformed into *E. coli* DH5α by electroporation. Sequencing reactions were performed with BigDye terminator cycle sequencing kits using the M13F and M13R primers, and run on an ABI 3730xl analyzer (Applied Biosystems, Foster City, CA, USA). For fosmid sequencing, *D. coronatus* genomic DNA was randomly sheared by pipetting, and the DNA was polished and dephosphorylated by Mung Bean Nuclease, T4 DNA polymerase and alkaline phosphatase (NEB, Ipswich, MA, USA). The DNA was ligated into a pKS300 fosmid vector and packaging reactions were performed using Giga Pack III XL packaging extract (Stratagene/Agilent, Santa Clara, USA). The packaged fosmid library was transfected into *E. coli* XL1-Blue. Clones were picked randomly and sequenced in the same way as for shotgun analysis.

Next generation sequencing (NGS) was performed as described [67]. Briefly, sequencing libraries were prepared using the GS FLX Titanium Rapid Library Preparation Kit (F. Hoffmann-La Roche, Basel, Switzerland) and the TruSeq DNA Sample Prep Kit (Illumina, San Diego, USA), and these libraries were run on a GS FLX and a Miseq sequencer, respectively.

Library construction for transcriptome analysis
cDNA libraries were generated by three different full-length enriched cDNA construction methods. (1) The NDK cDNA library was prepared using the Creator SMART cDNA library construction kit with the pDNR-LIB vector (Clontech/Takara, Shiga, Japan), according to the manufacturer's protocol. (2) The NDF library was prepared by the oligo-capping method using the pME18S-FL3 vector [68]. (3) The NDV library was constructed by the vector-capping method [69] using the pGCAP10 vector (Hitachi High-Tech and Hokkaido System Science, Japan).

The RNA-Seq library was prepared with the RNA-Seq Sample Prep Kit according to the manufacturer's instructions (Illumina, USA).

Quantification of nuclear DNA by flow cytometry
D. coronatus and *C. elegans* (genome size: 100 Mbp) were washed out and collected with a 10 μm nylon filter. Nematodes were transferred to a 300 ml flask containing 50 ml of distilled water and incubated for 60 min to reduce ingested food bacteria. Five head parts of *Drosophila melanogaster* were also prepared as the standard for genome size (140 Mbp). Worms and fly heads were

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 13 of 18

homogenized in sodium citrate buffer (pH 7) using a Dounce homogenizer by hand for 10 strokes. The homogenate was centrifuged at 400×g for 3 min to remove debris. The supernatants were treated with trypsin in a spermine tetrahydrochloride detergent buffer and stained with 125μg/ml propidium iodide (PI) (for details, see Cycle TEST PLUS DNA Reagent Kit manual (BD Biosciences, Franklin Lakes, NJ, USA)). The standards and *D. coronatus* samples were analyzed individually, and their mixture was analyzed by flow cytometry. Flow cytometry was performed with a Desktop cell sorter JSAN (Bay bioscience, Tokyo, Japan).

Chromosome staining
Adult *D. coronatus* were transferred to a drop of M9 buffer [65] containing 25 μM levamisole and 0.1 μg/ml 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI), and the gonad was dissected by nicking with a scalpel blade behind the pharynx. The slides were frozen on dry ice, and thawed at room temperature before microscopic observation. Images were recorded and analyzed with FV1200 confocal microscope using 100× UPlanSApo objective (Olympus, Tokyo, Japan). *C. elegans* worms were examined as a control with the same protocol. We performed a closer inspection of 11 oocytes (4n was observed in two oocytes, 2n was in six) and 6 embryos (2n was in three).

Data analysis
Data processing was done in the NIG supercomputer facility [70] using BioPerl [71] (version 1.6.1), EMBOSS [72] (version 6.4), BEDtools [73] (version 2.16.2), SAMtools [74] (version 0.1.18) and the other programs described below, which are installed in the super computer system as standard software. The sequence data were assigned BioProject accession PRJDB3143.

EST clustering
This was carried out by an in-house UNIX shell script with a short program written in C (Additional file 13). Briefly, first we take one clone and compare its 3′ end sequence with the 2nd clone using the FASTA program. If there is a match above a threshold (usually 90% considering the errors in EST sequencing), they are grouped, and if not, they are assigned a different group. The 3rd clone is compared with the previous ones, and, if there is match, it is included in the existing group, and if not, it is assigned a new group. Repeating this process, we classify EST clones based on the 3′ end sequences.

Genome assembly
The genome sequence was assembled from all the four libraries together by the Celera assembler [22] (options: "gkpFixInsertSizes = 0 bogBadMateDepth = 1000

cgwDemoteRBP = 0" and "doTrim_initialMerBased = 0 doTrim_initialQualityBased = 1" for Illumina Miseq reads, version: 7.0). The obtained sequences were 177,655,898 bp in 971 scaffolds consisting of 1817 contigs and 12,242,269 bp in 38,996 degenerate (meaning unused repeats) contigs. The statistics of the reads used (trimmed in the assembly process) are found in Additional file 1: Table S1. Miseq reads were re-mapped to the scaffolds and degenerates by BWA [75] (version 0.6.1-r104). Based on the results, 520 scaffolds, which were mapped at more than 0.01 reads/bp and longer than 2 kbp, were selected. A long scaffold apparently derived from the food bacterium *E. coli* OP50 was thus discarded at this stage. Furthermore, nine scaffolds turned out to represent the mitochondrial genome as described below. As a result, the remaining 511 scaffolds were considered to represent the *D. coronatus* genome. All analyses were performed with these.

The mitochondrial genome sequence was assembled manually using Consed [76] (version 29, with phrap version 1.090518) with the reads collected from the shotgun Sanger sequencing library based on the homology to the mitochondrial genome sequence of *C. elegans* [77]. A circular genome sequence of 13,378 bp was obtained. Covariance models for 22 tRNA genes were built from the alignments of Nematoda tRNA sequences in the mitotRNAdb database [78] and were searched for by Infernal [79] (version 1.1rc2). rRNA and protein coding genes were identified, such that the ranges on the whole genome alignment were similar to the *C. elegans* genes, assuming TTT as a start codon and T (with polyadenylation after transcription) as a stop codon [80, 81].

Paired structure of the genome
At first, the MUMmer package [82] (version 3.23) was used for the whole genome alignment. The scaffold sequences were aligned by nucmer (options –maxmatch –nosimplify). Trivial hits (alignments to themselves) were removed, delta-filter (option –1) was applied, and the scaffolds were reordered by mummerplot (option –fat) such that the resultant 1-to-1 alignments were emphasized by placing them diagonally on a Dot plot. In parallel, the alignments of minimal sequence identity 90% were filtered by delta-filter (option -i 90) from the nucmer result. Figure 1a is the plot of the >90% identity alignments with the order emphasizing the 1-to-1 alignments.

Next, longer alignments were obtained by the LAST package [83] (version 460). The scaffold sequences without masking were aligned by lastal (option -e1000). After trivial hits had been removed, the reciprocally best alignment segments were obtained by applying last-split (option -m 1) twice with maf-swap.py in-between. The alignments covered 152,151,424 bp (89.3% of the genome) and 8,685,973 bp (5.7% of them) were mismatches. To compare the partners

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 14 of 18

of the alignments visually, the Syntenic Browser of GBrowse [84, 85] (version 2.55) was set up [86]. The length of each gap on the alignments and the number of mismatches on each 100 bp window were counted and loaded onto the browser in addition to the annotations (Fig. 1c).

Repeat contents

Repeat sequences were identified de novo in the 971 scaffolds (before cleaned up) by RepeatModeler [87] (version 1.0.7 with RepeatScout version 1.0.5, RECON version 1.07). The obtained 754 repeat sequences were used by RepeatMasker [88] (options: -s -gccalc, version: 4.0.1 with RMBlast version 2.2.27, HMMER version 3.1-snap20121016.1 and TRF version 4.0.4) and 17.4% of the 511 scaffolds (after cleaned up) were masked by the de novo modeled repeats or simple repeats. The 754 repeat sequences were analyzed by REPCLASS [89] (version 1.0.1 with RepBase version 22.03, blast version 2.3.0, options for tblastx: -evalue 0.0001 -num_descriptions 10,000,000 -num_alignments 10,000,000 -seg yes and options for blastn: -task blastn -gapopen 2 -gapextend 1 -reward 1 -penalty −3 -dust no). 423 out of the 754 sequences, which was longer than 100 bp and whose copy number in the genome was greater than nine, were subjected to the classification procedure.

RNA genes

tRNA and rRNA genes were predicted by tRNAscan-SE [90] (version 1.23) and RNAmmer [91] (version 1.2 with HMMER version 2.3.2) respectively. RNA families in the Rfam database [92] (release 11.0) were searched by Infernal.

Protein coding genes

From the paired-end reads of the RNA-seq library, the adapter sequences were removed by SeqPrep [93] (version 1.1). Because almost all (92.8%) of the paired end reads could be merged to single sequences by this process, we used only the merged sequences. The merged reads were mapped to the genome sequence by TopHat [94] (options: –min-intron-length 5 –min-segment-intron 5, version: 2.0.5 with bowtie version 2.0.0-beta7). 91% of the reads could be mapped and 88% of the mapped reads were uniquely mapped.

Protein coding genes were predicted by Augustus [26, 27] (options: –species = caenorhabditis –allow_hinted_splice-sites = atac –alternatives-from-evidence = false –min_intron_len = 8, version: 2.7) using the hints from the mapping result of the RNA-seq (bam2hints with options –intronsonly –maxgaplen = 0 –minintronlen = 8 –maxintronlen = 10,000, bam2wig and wig2hints.pl with options –width = 10 –margin = 10 –minthresh = 2 –minscore = 4 –prune = 0.1 –radius = 4.5 were used to prepare the hints

and the configuration file "extrinsic.M.RM.E.W.cfg" in the Augustus package was applied). 33,459 genes were obtained.

Additional genes were predicted from the mapping result of RNA-seq by Cufflinks [95, 96] (options: –min-intron-length 5 –max-intron-length 25,000 –overlap-radius 5, version 2.0.0). If the prediction was placed intergenic of the Augustus predicted genes (class_code "u" was assigned by cuffcompare) and its longest open reading frame was longer than 89 bp, the model was adopted. 962 genes were obtained in this way. Together with the Augustus predicted genes, 34,421 protein coding genes were predicted.

Gene expression levels

The expression levels of the total 34,421 gene models were estimated again by Cufflinks without predicting new isoforms (options -G -b -u).

EST sequences were cleaned up by SeqClean [97] (option -v, version x86_64) using spliced leader (SL) sequences of Nematoda [98]. The sequences removed of poly-A and SL were mapped to the genome by exonerate [99] (options -m est2genome -bestn 1, version 2.2.0). The coding sequences of 20,003 genes were overlapped with (or supported by) the EST mapping.

Homology analyses

Protein categories were predicted by InterProScan [100] (option -goterms, version 5.3-46.0 with PANTHER version 8.1 data and Phobius version 1.01). 5859 InterPro entries were assigned to 20,264 genes.

The homologous (orthologous, in the conventional sense) gene groups between *D. coronatus* and *C. elegans* were obtained by InParanoid [30, 31] (options: seq_overlap_cutoff = 0 segment_overlap_cutoff = 0, version: 4.1 with BLAST version 2.2.26). The longest isoforms of 20,520 *C. elegans* genes (version wormpep230 [29]) were used in the analysis. As a result, 9189 homologous groups consisting of 20,264 genes of *D. coronatus* and 11,003 genes of *C. elegans* were obtained. To estimate total gene number, the remaining protein sequences of *D. coronatus* were clustered by cd-hit [101, 102] (option -g 1, version 4.6.1) with a threshold identity of 90%.

The coding sequences of the gene pairs of Dc:Ce = 2:1 were aligned by prank [103](option -codon, version 140110) and dN/dS were calculated by KaKs Calculator [104](option -m MLWL, version 1.2).

From the *D. coronatus* genes which belong to the orthologous groups of Dc:Ce = 2:1, 7306 pairs of genes, whose predicted CDS lengths differ by less than 100 bp, were selected as "allelic pairs". The expression levels of the paired genes, indicated by their FPKM values, were compared and Pearson's correlation coefficient of the higher and lower FPKM values was calculated ($P < 2.2e\text{-}16$).

Hiraki et al. BMC Genomics (2017) 18:478

Page 15 of 18

Search for REC-8 homolog

The domain model of the N-terminus of the Rad21/Rec8 like protein, Rad21_Rec8_N, was retrieved from the Pfam database [50] (release 27) and the proteins of *D. coronatus* were searched by hmmsearch [105] (option –max, version 3.1b1). The proteins of *C. elegans* and *P. pacificus* were retrieved from the UniProt database [106] by the query expression 'database:(type:pfam Rad21_Rec8_N) AND (organism:6239 OR organism:54,126)'. The sequences were aligned by MAFFT [107] (option –linsi, version 6.864b), the alignment was trimmed by trimAl [108] (option -automated1, version 1.2rev59) and the maximum likelihood unrooted tree with bootstrap values was constructed by RAxML [109] (options -f a -m PROTGAMMAAUTO -N autoMRE, version 8.1.17). The amino acid substitution model LG [110] with empirical amino acid frequencies and 200 replicates for bootstrapping were assigned. The phylogenetic tree was drawn by SeaView [111] (version 4.4.2).

Low complexity regions of the *D. coronatus* proteins were masked by segmasker [112] (version blast 2.2.25). The *D. coronatus* proteins were searched for the *C. elegans* REC-8 protein by ssearch [113] (options -s BP62 -S, version 36.3.5c). The N-terminus of *C. elegans* REC-8 and the *D. coronatus* homolog could be aligned but the score was quite insignificant (E-value 4.8). The *C. elegans* HIM-1 protein could be aligned to the *D. coronatus* REC-8 homolog with high significance, though the rank in the search was third (the top and second hits formed a 2:1 group with *C. elegans* HIM-1 in the InParanoid analysis).

Additional files

Additional file 1: Table S1. Sequence reads for analysis. The numbers of genome shotgun reads, RNA-seq reads and ESTs are shown. **Table S2.** Genome assembly statistics. Information on genome assembly, paired regions between scaffolds, SNVs in the paired regions, In/Dels in the paired regions and BUSCO assessment results are shown. (DOCX 36 kb)

Additional file 2: k-mer distribution analysis. k-mer distribution in the Miseq library was analyzed with kmerspectrumanalyzer (version b584039 with jellyfish version 2.0.0, numpy version 1.8.1 and scipy version 0.12.0) [114]. The frequency of each 21-mer in the library was measured and the number of distinct 21-mers for each frequency is plotted (red cross). This frequency spectrum was fitted as a mixture of over-dispersed Poisson (negative binomial) distributions (green line). The two peaks at k-mer frequency 35.9 and 71.8 correspond to single copy (heterozygous) regions and two copy (homozygous) regions, respectively. This spectrum indicates that 63% of the genome are present as single copy and 32% as two copies and that the genome size is 164 Mbp. (PDF 19 kb)

Additional file 3: Distribution of SNV ratio in the *D. coronatus* genome. The numbers of SNV in the paired regions were plotted against the lengths of the regions. Individual circles in a) and b) are the data from CDS and introns of individual genes. The circles in c) are the data from individual intergenic regions. The lines show mean densities of SNV: 3.7% for CDS, 7.1% for intron and 5.6% for the other intergenic regions. (PDF 1749 kb)

Additional file 4: Example of structural variations. A part of paired scaffolds (GBrowse screen capture) is shown. The horizontal green solid bars show the regions that have paired counterparts in the genome. Most of the green bars show a good synteny between the two paired scaffolds, but there are some variations with respect to translocation and inversion. The red circle on the right indicates an inversion, and the red circle on the left shows that the paired sequence corresponding to this green bar is located at another scaffold (translocation). (PDF 24 kb)

Additional file 5: Estimation of nuclear DNA amount by flowcytometry. The histogram of relative DNA content was obtained after flow cytometric analysis of propidium iodide-stained nuclei of a) *C. elegans* and b) *D. coronatus*, c) *D. melanogaster* and d) *D. coronatus* + *D. melanogaster*. *C. elegans* (100 Mbp ×2/nuleus) and *D. melanogaster* (140 Mbp ×2/nuleus) served as reference standard. X-axis: relative nuclear DNA content and Y-axis: number of events. (PDF 111 kb)

Additional file 6: Microscopic measurements of fluorescently labeled nuclear DNA. The fluorescent intensity of ventral nerve cord (VNC) nuclei of *D. coronatus* and sperm and VNC of *C. elegans* stained by Hoechst 33342 were measured as shown in Additional file 9. The y-axis indicates the fluorescent intensity (arbitrary unit: AU) of VNC nuclei in *D. coronatus* (grey circle), sperm and VNC of *C. elegans* (white and black circle, respectively). The amount of *D. coronatus* nuclear DNA was estimated to be 146 Mbp by interpolation of the average fluorescent intensities using *C. elegans* sperm (100 Mbp) and VNC (200 Mbp) as internal standards. (PDF 27 kb)

Additional file 7: Table S3. RNA gene annotations. Genes for rRNA, tRNA and RNA families are listed. **Table S4.** Splice leader sequences found in EST libraries. **Table S5.** List of transposon-like sequences. RepeatModeler identified 754 repeat sequences in the *D. coronatus* genome. The repeat sequences were filtered and classified by REPCLASS. As a result, 423 sequences were retained based on the criteria described in [89] and 104 sequences were classified into 4 categories of transposon-like sequences. (DOCX 35 kb)

Additional file 8: Schematic representation of orthologous gene relationships between *D. coronatus* (Dc) and *C. elegans* (Ce). (2:1): Paired genes (doubleton) in Dc are orthologous to a single gene in Ce. (2:2): Paired genes in Dc is orthologous to a gene family in Ce probably duplicated in the *C. elegans* lineage. (3+2): Two doubletons in Dc are orthologous to a gene family in Ce. (1:1): A gene without homologous partner (singleton) in Dc is orthologous to a single gene in Ce. (2:0): Paired genes in Dc do not have an ortholog in Ce. (PDF 128 kb)

Additional file 9: Supplementary text and methods. Supplementary description for signal transduction pathways and RNAi pathways, and Supplementary methods for microscopic measurements of fluorescently labeled nuclear DNA are given. (DOCX 32 kb)

Additional file 10: Protein coding gene annotations. **Table S6.** The predicted protein coding genes of *D. coronatus* are listed with their annotations. **Table S7.** The orthologs for *C. elegans* genes are shown. (XLSX 19699 kb)

Additional file 11: Phylogenetic analysis for meiosis-specific genes. Phylogenetic analysis of (1) MSH-2, (2) MSH-4 (HIM-14), (3) MSH-5, (4) MSH-6, (5) RAD-51 and (6) SPO-11 are shown, indicating that orthologs of these *C. elegans* meiotic-specific genes are present in the *D. cornatus* genome mostly in pairs. (PDF 151 kb)

Additional file 12: Histogram of sequence read coverage. The sequence reads from the Miseq library were remapped to the scaffolds by bwa (option: mem, version: 0.7.13-r1126). The mean depth of sequence reads for every 1kbp segments were counted and are shown as a histogram. The distribution of sequence reads is unimodal with the peak at 41.9X coverage. There is no significant peak at twofold higher coverage (84X), indicating that rarely two different regions are assembled together. (PDF 12 kb)

Additional file 13: 3′EST clustering program (cluster3.sh). This shell script originally written for *C.elegans* EST analysis was used. (TXT 5 kb)

Additional file 14: 3D images of DAPI stained chromosomes in oocytes and embryos of *D. coronatus*. 3D images were reconstructed by 3DView software in the FLUOLOVIEW system (Olympus, Tokyo, Japan) from the confocal images obtained for Fig. 2c and f. **Movie S1** and **Movie S2** correspond to Fig. 2c (oocyte) and Fig. 2f (embryo), respectively. (PPTX 437 kb)

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 16 of 18

Authors' contributions
HH performed all the computer analysis including genome assembling, gene model prediction and allelic gene expression, HK prepared materials for genome and transcriptome analysis and performed genome annotation, CK and PHS performed computer analysis for genome annotation, YU performed karyotype analysis and RNAi, and MK assisted in genome annotation, ES provided materials and information on *D. coronatus*, YK supervised the entire project and wrote the paper together with HH, HK, CK, PHS and ES. HH, HK and CK contributed equally to this work. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

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

Author details
[1]Genome Biology Laboratory, National Institute of Genetics, Mishima, Japan. [2]Zoologisches Institut, Universität zu Köln, Cologne, NRW, Germany. [3]Transdisciplinary Research Integration Center, Research Organization of Information and Systems, Tokyo, Japan.

References
1. Butlin R. Evolution of sex: the costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet. 2002;3:311–7.
2. Simon J-C, Rispe C, Sunnucks P. Ecology and evolution of sex in aphids. Trends Ecol Evol. 2002;17:34–9.
3. Müller HJ. Some genetic aspects of sex. Am Nat. 1932;66:118–38.
4. Schurko AM, Logsdon JM Jr. Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. BioEssays. 2008;30:579–89.
5. Simon J-C, Delmotte F, Rispe C, Crease T. Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biol J Linn Soc. 2003;79:151–63.
6. Danchin EGJ, Flot J-F, Perfus-Barbeoch L, Van Doninck K. Genomic perspectives on the long-term absence of sexual reproduction in animals. In: Pontarotti P, editor. Evolutionary biology – concepts, biodiversity, macroevolution and genome evolution. Berlin: Springer-Verlag; 2011. p. 223–42.
7. Flot JF, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EG, et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature. 2013;500:453–7.
8. Hur JH, Van Doninck K, Mandigo ML, Meselson M. Degenerate tetraploidy was established before bdelloid rotifer families diverged. Mol Biol Evol. 2009;26:375–83.
9. Schwander T. Evolution: the end of an ancient asexual scandal. Curr Biol. 2016;26:R233–5.
10. Signorovitch A, Hur J, Gladyshev E, Meselson M. Allele sharing and evidence for sexuality in a mitochondrial Clade of Bdelloid rotifers. Genetics. 2015;200:581–90.
11. Debortoli N, Li X, Eyres I, Fontaneto D, Hespeels B, Tang CQ, et al. Genetic exchange among Bdelloid rotifers is more likely due to horizontal Gene transfer than to meiotic sex. Curr Biol. 2016;26:723–32.
12. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, Deleury E, et al. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne Incognita. Nat Biotechnol. 2008;26:909–15.
13. Blanc-Mathieu R, Perfus-Babeoch L, Aury J-M, Da Rocha M, Gouzy J, Sallet E, Martin-Jimenez C, Castagnone-Sereno P, Flot J-F, Kozlowski D, Cazareth J, Couloux A, Da Silva C, Guy J, Rancurel C, Schiex T, Abad P, Wincker P, Danchin E. Peculiar hybrid genomes of devastating plant pests promote plasticity in the absence of sex and meiosis. bioRxiv 2016; doi: https://doi.org/10.1101/046805
14. Lunt DH, Kumar S, Koutsovoulos G, Blaxter ML. The complex hybrid origins of the root knot nematodes revealed through comparative genomics. PeerJ. 2014;2:e356.
15. Castagnone-Sereno P, Danchin EG. Parasitic success without sex - the nematode experience. J Evol Biol. 2014;27:1323–33.
16. Lahl V, Sadler B, Schierenberg E. Egg development in parthenogenetic nematodes: variations in meiosis and axis formation. Int J Dev Biol. 2006;50:393–8.
17. Riddle DL. The genetics of development and behavior in *Caenorhabditis elegans*. J Nematol. 1978;10:1–16.
18. Ali R, Amin B, Adachi T, Ishibashi N. Host and temperature preference, male occurrence and Morphometrics of Fungivorous nematode, *Aphelenchus avenae* isolates from Japan. Nematol Res. 1999;29:7–17.
19. Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, et al. The Pristionchus Pacificus genome provides a unique perspective on nematode lifestyle and parasitism. Nat Genet. 2008;40:1193–8.
20. Kiontke K, Fitch DH. The phylogenetic relationships of *Caenorhabditis* and other rhabditids. WormBook. 2005; p. 1-11. doi:10.1895/wormbook.1.11.1.
21. Mark Welch D, Meselson M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science. 2000;288:1211–5.
22. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, et al. A whole-genome assembly of Drosophila. Science. 2000;287:2196–204.
23. Bennett MD, Leitch IJ, Price HJ, Johnston JS. Comparisons with *Caenorhabditis* (approximately 100 Mb) and *Drosophila* (approximately 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be approximately 157 Mb and thus approximately 25% larger than the *Arabidopsis* genome initiative estimate of approximately 125 Mb. Ann Bot. 2003;91:547–57.
24. Doležel J, Bartoš J. Plant DNA flow cytometry and estimation of nuclear genome size. Ann Bot. 2005;95:99–110.
25. Hechler HC. Postembryonic development and reproduction in *Diploscapter coronata* (Nematoda, Rhabditidae). Helminth Soc Washington. 1968;35:24–30.
26. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637–44.
27. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19(Suppl 2):ii215–25.

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 17 of 18

28. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
29. Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, et al. WormBase 2014: new views of curated biology. Nucleic Acids Res. 2014;42:D789–93.
30. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, et al. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 2010;38:D196–203.
31. Remm M, Storm CE, Sonnhammer EL. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol. 2001;314:1041–52.
32. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 1998;282:2012–8.
33. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 2012;10:e1001388.
34. The *C. elegans* Research Community ed. WormBook. The online review of *C elegans* biology. [http://www.wormbook.org].
35. Meyer BJ. X-Chromosome dosage compensation. WormBook. 2005; p. 1-14. doi:10.1895/wormbook.1.8.1
36. Powell JR, Jow MM, Meyer BJ. The T-box transcription factor SEA-1 is an autosomal element of the X:a signal that determines *C. elegans* sex. Dev Cell. 2005;9:339–49.
37. Zarkower D. Somatic sex determination. WormBook. 2006; p. 1-12. doi:10.1895/wormbook.1.84.1.
38. Baldi C, Cho S, Ellis RE. Mutations in two independent pathways are sufficient to create hermaphroditic nematodes. Science. 2009;326:1002–5.
39. Goodwin EB, Ellis RE. Turning clustering loops: sex determination in *Caenorhabditis elegans*. Curr Biol. 2002;12:R111–20.
40. Kuwabara PE, Okkema PG, Kimble J. *tra-2* encodes a membrane protein and may mediate cell communication in the *Caenorhabditis elegans* sex determination pathway. Mol Biol Cell. 1992;3:461–73.
41. Lui DY, Colaiácovo MP. Meiotic development in *Caenorhabditis elegans*. In: Schedl T, editor. Germ Cell Development in C elegans, vol. 757. New York: Springer; 2013. p. 133–70.
42. Schvarzstein M, Wignall SM, Villeneuve AM. Coordinating cohesion, co-orientation, and congression during meiosis: lessons from holocentric chromosomes. Genes Dev. 2010;24:219–28.
43. Phillips CM, Dernburg AF. A family of zinc-finger proteins is required for chromosome-specific pairing and synapsis during meiosis in *C. elegans*. Dev Cell. 2006;11:817–29.
44. Phillips CM, Meng X, Zhang L, Chretien JH, Urnov FD, Dernburg AF. Identification of chromosome sequence motifs that mediate meiotic pairing and synapsis in *C. elegans*. Nat Cell Biol. 2009;11:934–42.
45. Phillips CM, Wong C, Bhalla N, Carlton PM, Weiser P, Meneely PM, et al. HIM-8 binds to the X chromosome pairing center and mediates chromosome-specific meiotic synapsis. Cell. 2005;123:1051–63.
46. Sanford C, Perry MD. Asymmetrically distributed oligonucleotide repeats in the *Caenorhabditis elegans* genome sequence that map to regions important for meiotic chromosome segregation. Nucleic Acids Res. 2001;29:2920–6.
47. Rog O, Dernburg AF. Chromosome pairing and synapsis during *Caenorhabditis elegans* meiosis. Curr Opin Cell Biol. 2013;25:349–56.
48. Burger J, Merlet J, Tavernier N, Richaudeau B, Arnold A, Ciosk R, et al. CRL2(LRR-1) E3-ligase regulates proliferation and progression through meiosis in the *Caenorhabditis elegans* germline. PLoS Genet. 2013;9:e1003375.
49. Hillers KJ, Jantsch V, Martinez-Perez E, Yanowitz JL. Meiosis. WormBook. 2015; p. 1-54. doi:10.1895/wormbook.1.178.1.
50. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222–30.
51. Grishaeva TM, Bogdanov YF. Conservation and variability of synaptonemal complex proteins in phylogenesis of eukaryotes. Int J Evol Biol. 2014;2014:856230.
52. Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, et al. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475:189–95.
53. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014;24:1384–95.
54. Xiao M, Phong A, Ha C, Chan TF, Cai D, Leung L, et al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Res. 2007;35:e16.
55. Golczyk H, Massouh A, Greiner S. Translocations of chromosome end-segments and facultative heterochromatin promote meiotic ring formation in evening primroses. Plant Cell. 2014;26:1280–93.
56. Cutter AD, Jovelin R, Dey A. Molecular hyperdiversity and evolution in very large populations. Mol Ecol. 2013;8:2074–95.
57. Severson AF, Ling L, van Zuylen V, Meyer BJ. The axial element protein HTP-3 promotes cohesin loading and meiotic axis assembly in *C. elegans* to implement the meiotic program of chromosome segregation. Genes Dev. 2009;23:1763–78.
58. Melters DP, Paliulis LV, Korf IF, Chan SW. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosom Res. 2012;20:579–93.
59. Viera A, Page J, Rufas JS. Inverted meiosis: the true bugs as a model to study. Genome Dyn. 2009;5:137–56.
60. Albertson DG, Thomson JN. The kinetochores of *Caenorhabditis elegans*. Chromosoma. 1982;86:409–28.
61. Albertson DG, Thomson JN. Segregation of holocentric chromosomes at meiosis in the nematode, *Caenorhabditis elegans*. Chromosom Res. 1993;1:15–26.
62. Dernburg AF. Here, there, and everywhere: kinetochore function on holocentric chromosomes. J Cell Biol. 2001;153:F33–8.
63. Goday C, Ciofi-Luzzatto A, Pimpinelli S. Centromere ultrastructure in germ-line chromosomes of Parascaris. Chromosoma. 1985;91:121–5.
64. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'Eustachio P, Fitch DH, et al. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 2003;2:E45.
65. Brenner S. The genetics of *Caenorhabditis elegans*. Genetics. 1974;77:71–94.
66. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, et al. The medaka draft genome and insights into vertebrate genome evolution. Nature. 2007;447:714–9.
67. Nikaido M, Noguchi H, Nishihara H, Toyoda A, Suzuki Y, Kajitani R, et al. Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Res. 2013;23:1740–8.
68. Suzuki Y, Sugano S. Construction of full-length-enriched cDNA libraries. The oligo-capping method. Methods Mol Biol. 2001;175:143–53.
69. Kato S, Ohtoko K, Ohtake H, Kimura T. Vector-capping: a simple method for preparing a high-quality full-length cDNA library. DNA Res. 2005;12:53–62.
70. Ogasawara O, Mashima J, Kodama Y, Kaminuma E, Nakamura Y, Okubo K, et al. DDBJ new system and service refactoring. Nucleic Acids Res. 2013;41:D25–9.
71. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 2002;12:1611–8.
72. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16:276–7.
73. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
74. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
75. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.
76. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Res. 1998;8:195–202.
77. Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics. 1992;130:471–98.
78. Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, Putz J. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res. 2009;37:D159–62.
79. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5.
80. Hu M, Chilton NB, Gasser RB. The mitochondrial genome of *Strongyloides stercoralis* (Nematoda) - idiosyncratic gene order and evolutionary implications. Int J Parasitol. 2003;33:1393–408.
81. Lemire B. Mitochondrial genetics. WormBook. 2005; p. 1-10. doi:10.1895/wormbook.1.25.1.
82. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
83. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Res. 2011;21:487–93.
84. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, et al. The generic genome browser: a building block for a model organism system database. Genome Res. 2002;12:1599–610.
85. McKay SJ, Vergara IA, Stajich JE. Using the generic Synteny browser (GBrowse_syn). Curr Protoc Bioinformatics. 2010;Chapter 9:Unit 9 12.
86. *D. coronatus* genome project. [http://nematode.lab.nig.ac.jp/d_coronatus/].
87. RepeatModeler Open-1.0. [http://www.repeatmasker.org].
88. RepeatMasker Open-4.0. [http://www.repeatmasker.org].

Hiraki *et al. BMC Genomics* (2017) 18:478

Page 18 of 18

89. Feschotte C, Keswani U, Ranganathan N, Guibotsy ML, Levine D. Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. Genome Biol Evol. 2009;1:205–20.

90. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.

91. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.

92. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 2013;41:D226–32.

93. SeqPrep. [https://github.com/jstjohn/SeqPrep].

94. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

95. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011;12:R22.

96. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.

97. Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, et al. The TIGR Gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res. 2005;33:D71–4.

98. Guiliano DB, Blaxter ML. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLoS Genet. 2006;2:e198.

99. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31.

100. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

101. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT. Accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

102. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.

103. Löytynoja A, Goldman N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science. 2008;320:1632–5.

104. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs calculator: calculating ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics. 2006;4:259–63.

105. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.

106. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43:D204–12.

107. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 2008;9:286–98.

108. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.

109. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

110. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25:1307–20.

111. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27:221–4.

112. Wootton JC, Federhen S. Statistics of local complexity in amino acid sequences and sequence databases. Comput Chem. 1993;17:149–63.

113. Pearson WR. Effective protein sequence comparison. Methods Enzymol. 1996;266:227–58.

114. Williams D, Trimble WL, Shilts M, Meyer F, Ochman H. Rapid quantification of sequence repeats to resolve the size, structure and contents of bacterial genomes. BMC Genomics. 2013;14:537.