

# Chromosome-level genome assembly reveals homologous chromosomes and recombination in asexual rotifer *Adineta vaga*

Paul Simion,<sup>1+\*</sup> Jitendra Narayan,<sup>1+</sup> Antoine Houtain<sup>1</sup> Alessandro Derzelle<sup>1</sup>  
Lyam Baudry<sup>2,3</sup> Emilien Nicolas<sup>1,4</sup> Rohan Arora<sup>1,4</sup> Marie Cariou<sup>5</sup>  
Corinne Cruaud<sup>6</sup> Florence Rodriguez Gaudray<sup>7</sup> Clément Gilbert<sup>8</sup>  
Nadège Guiglielmoni<sup>7</sup> Boris Hespeels<sup>1</sup> Djampa KL Kozlowski<sup>9</sup>  
Karine Labadie<sup>6</sup> Antoine Limasset<sup>10</sup> Marc Llirós<sup>11</sup> Martial Marbouth<sup>2</sup>  
Matthieu Terwagne<sup>1</sup> Julie Virgo<sup>1</sup> Richard Cordaux<sup>12</sup>  
Etienne GJ Danchin<sup>9</sup> Bernard Hallet<sup>13</sup> Romain Koszul<sup>2</sup>  
Thomas Lenormand<sup>14</sup> Jean-François Flot<sup>7,15\*</sup> Karine Van Doninck<sup>1,4\*</sup>

<sup>1</sup>Université de Namur, LEGE, URBE, Namur, 5000, Belgium

<sup>2</sup>Institut Pasteur, Unité Régulation Spatiale des Génomes, UMR 3525, CNRS, Paris, F-75015, France

<sup>3</sup>Sorbonne Université, Collège Doctoral, F-75005 Paris, France

<sup>4</sup>Université libre de Bruxelles (ULB), Molecular Biology and Evolution, Brussels, 1050, Belgium

<sup>5</sup>CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111,

Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007, Lyon, France.

<sup>6</sup>Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

<sup>7</sup>Université libre de Bruxelles (ULB), Evolutionary Biology and Ecology, Brussels, 1050, Belgium

<sup>8</sup>Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198, Gif-sur-Yvette, France

<sup>9</sup>INRAE, Université Côte-d'Azur, CNRS, Institut Sophia Agrobiotech, Sophia Antipolis, 06903, France

<sup>10</sup>Université de Lille, CNRS, UMR 9189 - CRISTAL, 59655 Villeneuve-d'Ascq, France

<sup>11</sup>Institut d'Investigació Biomèdica de Girona, Malalties Digestives i Microbiota, 17190 Salt, Spain

<sup>12</sup>Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des interactions, 5 rue Albert Turpaine, 86073 Poitiers, France

<sup>13</sup>Université Catholique de Louvain (UCLouvain), LIBST, Croix du Sud 4/5, Louvain-la-Neuve, 1348, Belgium

<sup>14</sup>CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

<sup>15</sup>Interuniversity Institute of Bioinformatics in Brussels - (IB)<sup>2</sup>, Brussels, 1050, Belgium

+these authors contributed equally to this work

\*To whom correspondence should be addressed; E-mail: karine.vandoninck@unamur.be, jflot@ulb.ac.be, polo.simion@gmail.com

1      **Bdelloid rotifers are notorious because they represent a speciose and ancient**  
2      **clade comprising only asexual lineages. In addition, most bdelloid species**  
3      **withstand complete desiccation and ionizing radiation, being able to repair**  
4      **their highly fragmented DNA. Producing a well-assembled reference genome**  
5      **is a critical step to unlock the understanding of the effects of long-term asexu-**  
6      **ality and DNA breakage on genome evolution. Here, we present the first high-**  
7      **quality chromosome-level genome assemblies for the bdelloid species *Adineta***  
8      ***vaga*, composed of six pairs of homologous chromosomes (i.e. diploid), with**  
9      **a footprint of paleotetraploidy. The observed large-scale losses of heterozy-**  
10     **gosity are signatures of recombination between homologous chromosomes, ei-**  
11     **ther during mitotic DNA double-strand break repair or when resolving pro-**  
12     **grammed DNA breaks during a modified meiosis. Dynamic subtelomeric re-**  
13     **gions harbor more structural diversity (e.g. chromosome rearrangements,**  
14     **transposable elements, haplotypic divergence). Our results trigger the reap-**  
15     **praisal of potential meiotic processes in bdelloid rotifers and help unravel their**  
16     **long-term asexual evolutionary success.**

## 17     **Introduction**

18     Sexual reproduction and recombination are prevalent throughout the eukaryotes, despite the  
19     substantial evolutionary costs such as the two-fold cost of males or the cost of recombination  
20     that breaks up co-adapted gene combinations (1, 2)). Several eukaryotic species appear to have  
21     evolved adaptations that reduce these costs of males, for example by producing males only fac-  
22     ultatively as in cyclical parthenogens (e.g. *Brachionus plicatilis* (3)), or by retaining a modified  
23     meiosis rescuing diploidy without fertilization by males (e.g. *Diploscapter pachys* (4)). Very  
24     few however appear to have renounced to sex and recombination completely, in which males

25 and meiosis are abolished. Theory predicts that in the absence of recombination during meiosis,  
26 the physical linkage among loci reduces the effectiveness of selection upon individual loci,  
27 resulting in a decreased rate of adaptation and the accumulation of mildly deleterious mutations (5). Obligate asexuals are therefore suitable model systems to gain general insights into  
28 the long term consequences of the lack of recombination and sexual reproduction.

30 Bdelloid rotifers are notorious ancient asexual animals. Indeed, the longevity (>60 My) of  
31 the bdelloid rotifer clade and their diversity (>400 morphospecies) challenge the expectation  
32 that obligatory asexual animal lineages, in which recombination and outcrossing are absent,  
33 are evolutionary dead-ends. Historical observations (or lack thereof) had produced a consensus  
34 that bdelloid rotifers do not produce male or hermaphrodite individuals (6), that they are  
35 strictly parthenogenetic without any meiosis (7, 8) and that the initial description of the structure  
36 of *Adineta vaga* genome, lacking colinear homologous scaffolds, was irreconcilable with  
37 meiosis (9). A draft genome assembly of the closely-related bdelloid species *Adineta ricciae*  
38 found colinearity between homologous regions but could not verify it at chromosome-scale (10),  
39 which was also the case for previous studies based on a handful of genomic regions (11–13).  
40 The presence or the absence of an ameiotic structure in bdelloids therefore remained unresolved  
41 and a chromosome-scale assembly appeared critical.

42 Besides its asexual evolution, the bdelloid rotifer *A. vaga* also became a model species for its  
43 extreme resistance to desiccation, freezing and ionizing radiation, with implications for space  
44 research (14, 15). Both prolonged desiccation, encountered in their ephemeral limno-terrestrial  
45 habitats, and ionizing radiation induce oxidative stress and massive genome breakage that *A.*  
46 *vaga* seems to handle, maintaining high survival and fecundity rates while efficiently repairing  
47 DNA damage (15–17). Maintaining such long-term survival and genome stability following  
48 DNA fragmentation likely requires the use of homologous recombination (HR) at least in the  
49 germ cells. Given the supposed absence of homologous chromosomes in *A. vaga* (9), the exact

50 nature of their double-strand break (DSB) repair mechanism remains elusive.

51 Recent studies have provided evidence for recombination in bdelloid rotifers. These in-  
52 clude a drop of linkage disequilibrium (LD) with increasing distance between genomic loci in  
53 *A. vaga* (13), signatures of gene conversion (9, 12), heterozygosity levels within the range of  
54 those reported for sexual metazoans (9, 10, 18), and reports of allele sharing between bdelloid  
55 individuals from the wild (13, 19–21). While recombination likely takes place in bdelloid ro-  
56 tifiers, its underlying mechanisms remain unknown. Recombination might theoretically occur  
57 in a mitotic or meiotic cellular context, involve short genomic regions or canonical chromo-  
58 some pairing, and take place between homologous or non-homologous loci (i.e. ectopic). The  
59 interpretation of these recombination events have yet to be reconciled with the long-standing  
60 apparent absence of males and meiosis in bdelloid rotifers (6) and to account for their ubiquity  
61 in semi-terrestrial habitats where frequent desiccation occurs, inducing DNA DSBs (14, 15, 17).

62 Here, we present a high quality chromosome-level genome assembly of *A. vaga*. This new  
63 genome is pivotal to tackle these contradictions between its putative ameiotic structure and the  
64 footprints of recombination, possibly associated to DSB repair and desiccation. We combined  
65 the use of short reads (Illumina), long reads (ONT and PacBio) and chromosome conformation  
66 capture data (Hi-C) with three assembly methods, to successfully assemble *A. vaga* genome.  
67 We provide the first telomere-to-telomere assemblies of a parthenogenetic lineage, both haploid  
68 and phased, paving the way to study genome evolution in an asexual clade. Using a newly devel-  
69 oped and publicly available tool, Alienomics, we annotated candidate horizontal gene transfers  
70 (HGTC) and confirmed that *A. vaga* possesses the highest number of HGTC across all animals.  
71 Interestingly, *A. vaga* has a diploid genome made of six pairs of homologous chromosomes,  
72 refuting the ameiotic structure previously described for this genome (9) and challenging the  
73 complete absence of meiosis in one of the most striking asexual animal clade. In addition, by  
74 observing large tracks of heterozygosity losses (LOH), we show that large-scale recombination

75 between homologous chromosomes occurs in *A. vaga*. The possibility of chromosome pairing  
76 in *A. vaga*, during a mitotic or meiotic-like process, allows for the re-interpretation of the sig-  
77 natures of LD decay and allele sharing. Until now, the lack of chromosome-scale assemblies  
78 of parthenogenetic genomes hampered the investigation of the impact of meiosis, recombi-  
79 nation, outcrossing, or their absence, on entire genomes. Moreover, characterizing homologous  
80 chromosomes as potential templates for DNA repair through HR in *A. vaga* is an important land-  
81 mark in the understanding of bdelloid extreme resistance. This high-quality genome assembly  
82 of *A. vaga* (AV20) is also timely for comparative biology within rotifers and protostomians,  
83 extending the list of chromosome-level genomes in overlooked phyla.

## 84 **Results and discussion**

85 **A diploid genome with a tetraploid past** Distinct independent genome assembly procedures,  
86 relying on different assumptions regarding ploidy levels (Bwise (22), NextDenovo (23) and  
87 Falcon (24)), were first used on a combination of short and long sequencing reads. These  
88 assemblies were then scaffolded using Hi-C data and instaGRAAL (25), revealing similar  
89 chromosome-level assemblies and genome size estimations, consistent with flow cytometry  
90 measurements (Fig. 1A and Supp. Figs. 1 and 2). All pairwise alignments of the three inde-  
91 pendent assemblies (referred to as "phased" without ploidy assumption, "haploid" and phased  
92 "diploid", see Fig. 1A) confirmed chromosome-level synteny and converged towards identical  
93 genome structure with the six longest scaffolds from the haploid assembly (hereafter named  
94 "AV20") being each colinear to exactly two long scaffolds from the phased assembly (Fig. 1B,  
95 see also Supp. Figs. 3, 4 and 5). In order to validate these assemblies, we performed fluorescent  
96 *in situ* hybridization (FISH) analyses with three pairs of fluorescent probe libraries complemen-  
97 tary to separate parts of chromosomes 2, 5 and 6 from the AV20 assembly (Figure 1B, right  
98 side). For each pair of probes (one green and one red) two individual chromosomes were la-

99 belled with little or no overlap between both signals (Figure 1C). Chromosome painting on the  
100 karyotype of 12 chromosomes of *A. vaga* (26) was consistent with our chromosome-scale as-  
101 semblies showing that the *A. vaga* genome is diploid, being composed of six pairs of colinear  
102 homologous chromosomes.

103 We compared our new AV20 assembly to the previously published draft genome assembly  
104 (hereafter named "AV13" (9)). None of the previously described colinearity breakpoints and  
105 palindromes were retrieved in the new AV20 genome, indicating that these were likely assembly  
106 artefacts resulting from erroneous scaffolding of uncollapsed haplotypes (Supp. Figs. 6 and 7).  
107 Chromosome-level colinearity, albeit weaker than between homologous chromosomes, was also  
108 observed between pairs of homoeologous (or ohnologous) chromosomes in the AV20 genome,  
109 a signature confirming the previously reported paleotetraploidy of *A. vaga* (9, 10, 12) (grey links  
110 on Figure 1B). The three chromosome pairs 1, 2 and 3 are homoeologous to the three pairs  
111 4, 5 and 6, respectively. *A. vaga* is thus a diploid, paleotetraploid species in which the level  
112 of synteny between homoeologous chromosomes is high. Notably, 30.8% of the genes have a  
113 homoeologous copy within conserved synteny blocks (see Materials & Methods section) and  
114 with an average nucleotide divergence of about 13% (Supp. Fig. 8).

115 **Recombination between homologous chromosomes causes loss of heterozygosity** The dis-  
116 covery of homologous chromosomes in the oldest known asexual animal clade represents a  
117 major shift for studies of ancient asexuals and leads us to reconsider the possibility for homol-  
118 ogous recombination in *A. vaga*. One potential genetic consequence of recombination between  
119 homologous chromosomes is large-scale loss of heterozygosity (LOH). We measured and com-  
120 pared heterozygosity along the chromosomes of three *A. vaga* samples cultured from a same  
121 ancestral laboratory strain that never underwent stresses causing recombinogenic DSBs and  
122 that were sequenced at three distinct timepoints (2009, 2015 and 2017, Fig. 2A, Supp. Ta-

ble S1). Mean single-nucleotide polymorphism (SNP) heterozygosity (i.e. divergence between homologous chromosomes) was around 1.7% (horizontal line on Fig. 2A, similar to previous reports (9, 10)). Interestingly, we observed large regions (from 100 kb to 4.5 Mb) that were fully homozygous, except for a few SNPs, in specific isolates while heterozygous in others (numbered tokens in Fig. 2A). Note that a few homozygous tracks are associated with coverage variation and could have been caused by a hemizygous deletion (when coverage drops by approximately 50%, e.g. event 5 on Fig. 2A) or by the high density of repeated sequences (e.g. event 13 on Fig. 2A). Given the genealogy of these laboratory lines, we argue that the large homozygous tracks that are associated with homogeneous median coverage are signatures of allelic recombination events causing LOH (Fig. 2B).

These LOH appeared to accumulate through time as some are shared by two samples (e.g. event 12 on Fig. 2) while others appeared in only one of these two samples (e.g. event 2 on Fig. 2). Noteworthy, no ancestral LOH was found that would be shared by all of the strains. This is likely because large LOH events increase the chance to expose recessive deleterious mutations and are thus likely selectively eliminated in nature, maintaining the relatively homogeneous heterozygosity level in the ancestral laboratory strain (Fig. 2A). Observing LOH tracks on all six chromosome pairs in the three laboratory samples over a relatively short period of time (i.e. several years, Fig. 2B) might be due to the culturing conditions allowing for possible bottlenecks and relaxed selection. Recombination occurring along the entire chromosomes, instead of being restricted to the telomeres only (27), invalidates the hypothesis that an *Oenothera*-like meiosis underlies their reproductive mode (in agreement with a recent study (13)). Overall, these LOH tracks combined with the recently reported LD decay (13) represent a clear footprint that molecular processes involving recombination between homologous chromosomes occur in the germline of *A. vaga*.

**147 Recombination could be accidental or programmed** Theoretically, recombination between  
148 homologous chromosomes resulting in inheritable LOH can occur in the germline during mi-  
149 totic repair of accidental DSBs or when handling programmed DSBs during meiosis (potentially  
150 induced by Spo11 protein (9, 28)). DSBs can be repaired by different recombination pathways  
151 but LOH of large chromosome regions without coverage reduction (e.g. events 1, 6-9, 11, 12,  
152 Fig. 2) can primarily arise from two processes, break-induced replication (BIR) and the for-  
153 mation of crossing-over (CO). BIR is a mechanism of mitotic recombination characterized by  
154 replication fork progression over hundreds of kilobases on the repair template (29). When in-  
155 volving allelic loci, it causes LOH of the segment extending from the breakpoint site until the  
156 end of the chromosome. If a double BIR (dBIR) occurs, switching templates from the homolo-  
157 gous chromosome back to the sister or the original chromatid, a LOH tract, possibly long, that  
158 does not encompass the telomere is produced (30). Such LOH could also be generated by the  
159 recombinational repair of respectively one or two DSBs leading to CO (i.e. a reciprocal genetic  
160 exchange between chromosomes). Compared to BIR, CO is however a minor pathway in mitot-  
161 ically cycling cells (31) that preferentially takes place between sister chromatids and therefore  
162 remains genetically silent (32).

**163** Alternatively, programmed DSBs during meiosis can produce large LOH tracks by favoring  
164 CO formation between homologous chromosomes (31). LOH signatures in *A. vaga* genome  
165 could therefore be acquired through meiotically-induced recombination instead of during mi-  
166 tosis. Several mechanisms of meiotic parthenogenesis, globally referred to as automixis, have  
167 been described in various species such as in *Daphnia pulex* (33), *Artemia parthenogenetica* (34)  
168 or *Apis mellifera capensis* (35). If automixis occurs in *A. vaga*, the heterozygosity patterns ob-  
169 served here (Fig. 2) in which the maternal heterozygosity is conserved along chromosomes due  
170 to the non-segregation of homologous chromosomes while large LOH tracks (likely counter-  
171 selected in nature) could result from their CO recombination, is genetically equivalent to what

is referred to as central fusion automixis (34). Nevertheless, no cytological evidence of any meiotic process has been described so far in bdelloid rotifers. Whether recombination is a key feature of the reproductive mode of bdelloids (through programmed DSBs during a modified meiosis) or whether it is mainly driven by desiccation resistance mechanisms (through accidental DSB repair in the germline), or both, remains an open question. Whichever mechanism is involved, recombination likely plays a major role in the long-term evolution of *A. vaga* genome.

**Dynamic subtelomeric regions** We found a low amount of transposable elements (TEs) in *A. vaga* (Fig. 3, Supp. Fig. 9). By combining two approaches to annotate both TE-like elements, including repeated sequences, as well as canonical TEs (i.e. the EDTA and REPET pipelines) we detected 6.6% of TE-like elements and 1.9% of canonical TEs, predominantly located at subtelomeric regions (Fig. 3). In addition, rotifer-specific telomeric repeats (i.e. (TGTGGG)n (36)) were detected at the extremities of every scaffold of the AV20 assembly, indicating that they indeed correspond to telomeric and subtelomeric regions and that AV20 reached a chromosome-level assembly (Supp. Fig. 10). Most consensus TE sequences were found at low copy numbers (i.e. 96% of canonical TEs consensus sequences are present in  $\leq 5x$  copies in AV20, see Supp. Fig. 9). Notably, terminal inverted repeats (TIRs) DNA transposons (i.e. Class-II) were quantitatively dominant (48% of all TEs) among the low amount of TEs in *A. vaga* genome (Supp. Figs. 9 and 11). These results are in line with previous studies of TEs in bdelloids (9, 10, 37, 38). Using sequence similarity between a TE copy and their consensus as a proxy for how recent this copy is, we found that Class-II TIRs and Class-I LINEs and LTRs had high average similarity to their consensus sequences suggesting that they have been at least recently active in *A. vaga* genome (Supp Fig. 12). Investigating putative endogenous viral elements (EVEs) in *A. vaga* revealed very few viral-like sequences (i.e. 94 loci scattered along the 6 chromosomes, Fig. 3) with potential donor candidates belonging to

196 the group of large double stranded (ds) DNA viruses. None of these EVE candidates however  
197 had definitive viral origins as their similarity was not restricted to viral sequences and there was  
198 no conservation of viral gene synteny.

199 Syntenic HGTc regions among non-homoeologous chromosomes are visible, mainly at sub-  
200 telomeric regions (violet links on Fig. 3) and may suggest chromosomal re-arrangements. Sub-  
201 telomeric regions are also the regions on which almost all divergent haplotypes (i.e. haplotigs  
202 corresponding to uncollapsed haplotypes during genome assembly process) were located (grey  
203 links on Fig. 3). Overall, these subtelomeric regions in *A. vaga* are enriched in canonical TEs,  
204 TE-like elements, HGTc, viral-like sequences but also retain a higher haplotypic divergence  
205 (i.e. uncollapsed haplotigs) and most chromosomal re-arrangements. When accounting for only  
206 coding sequences (CDS), no distinct increase or decrease of heterozygosity could however be  
207 observed at subtelomeric regions (Supp. Fig. 13). At this stage, it is therefore unclear whether  
208 homologous recombination rate covaries with telomeric proximity in *A. vaga*. Nevertheless, our  
209 results suggest that subtelomeric regions seem more prone to chromosomal re-arrangements, in-  
210 corporation of foreign DNA (TEs and HGTs) and structural variations such as putative allelic  
211 deletions (see LOH events 3 and 10 in Fig. 2), evolving faster than the rest of the genome.

212 **Horizontal gene transfers in *A. vaga* genome** The acquisition of foreign DNA has been  
213 hypothesized to play an important role in bdelloid evolution (20). HGTs could be a way to  
214 circumvent some deleterious effects of the lack of sexual outcrossing, and the occasional inte-  
215 gration of foreign DNA could trigger adaptation (10, 17, 20, 39, 40). No automated tool existed to  
216 detect HGTc, therefore we developed Alienomics, an innovative pipeline to detect both HGTc  
217 and contaminants in a genome assembly. Alienomics combines several genomic parameters  
218 such as gene taxonomy, GC content, sequencing depth, but also taking into account gene inte-  
219 gration into the genome using synteny and expression data, to detect HGTc from non-metazoan

species. In contrast with the overall low amount of TEs, many candidate HGTc (2,679, about 8.3% of all genes) were detected in the *A. vaga* AV20 genome assembly, confirming previous reports of the highest HGTc content among metazoans (9, 10, 41, 42). HGTc were enriched in subtelomeric regions as previously reported (41), although many HGTc were distributed along the chromosomes and two visible local hotspots were detected outside the subtelomeric regions (pink stars on Figure 3). Interestingly, one HGT hotspot is associated with a slight increase of interstitial telomeric repeats (Supp. Fig. 10) that could represent a signature of ancient chromosome fusion. Overall, the heterogeneity in HGTc density between subtelomeric regions and the rest of the genome could be explained either by varying rates of HGTs incorporation along the chromosomes or by varying successful integration of HGTs within the genome through selection.

Using both MCScanX and Alienomics outputs, we measured that 257 foreign genes (9.6% of all HGTc) conserved their synteny across homoeologous chromosomes, including the HGTc hotspots notably visible in homoeologous chromosomes 1 and 4 (stars on Figure 3). These horizontal transfer events therefore occurred before the ancestral tetraploidization of modern bdelloids. This amount of ancient HGTc is however likely underestimated as any loss or translocation of an ancient HGTc copy would break the ancestral synteny. When looking specifically at these HGTc that occurred prior to the tetraploidization, we observed an enrichment of genes involved DNA recombination and DNA ligation involved in DNA DSB repair, among other enriched functional categories (see Supp. Table S2). These HGTc might have set bdelloids up to resist and overcome massive DNA breakage. When analyzing all HGTc, we found that they are enriched in genes involved in oxidation-reduction and carbohydrate metabolic processes (9) as well as in the response to nitrosative stress (see Supp. Table S2). Acquisition of HGTs might therefore be central in their resistance to extreme desiccation and towards more efficient homeostasis. Overall, these results are in line with previous studies suggesting that HGTs have

245 been continuously acquired within bdelloid rotifers, even before their tetraploidization (10, 42).  
246 However, if bdelloids have the same low rate of HGT acquisition from other individuals of  
247 the same species than from non-metazoans (12.8 HGT/Myr), HGT is possibly insufficient to  
248 compensate for the plausible lack of outcrossing in bdelloid rotifers (40). Actually, a high rate  
249 of HGT acquisition from distinct species might be deleterious for *A. vaga*, which appears to  
250 rely on recombination between homologous chromosomes to maintain heterozygosity and/or  
251 genome structure.

252 **Reasoning on bdelloid rotifer reproductive mode** Bdelloid rotifer species are both suppos-  
253 edly devoid of males and prone to integrate foreign DNA (through HGTs) into their genome. In  
254 this context, several reports of allele sharing between bdelloid individuals sampled from the wild  
255 triggered a debate whether they could exchange genetic content at all and whether this might be  
256 done through HGT or through sexual reproduction (13, 19–21, 43–45). At a first glance, show-  
257 ing that homologous chromosomes exist and recombine in *A. vaga* could be viewed as a support  
258 to the hypothesis that bdelloids might undergo meiotic sexual reproduction (13). However, this  
259 hypothesis has yet to be reconciled with the absence of both males and canonical meiosis in  
260 bdelloid rotifers and here we speculate on the mechanisms of homologous recombination in  
261 *A. vaga*. The three *A. vaga* lineages analyzed here (Fig. 2) were kept in hydrated conditions,  
262 leaving few opportunities for desiccation-induced, accidental DNA DSBs. Moreover, a much  
263 lower heterozygosity than for *A. vaga* has been observed in two obligate aquatic bdelloid rotifer  
264 species (i.e. *Rotaria*, upper limit of homologous divergence ranged between 0.033 and 0.075),  
265 also described as asexual and never experiencing dessication. Both these observations are com-  
266 patible with the hypothesis that homologous recombination in bdelloids could be caused by  
267 programmed DNA DSBs during a meiotic-like process. Frequent and programmed recombi-  
268 nation would cause LOH in *A. vaga* (Fig. 2) and would have lowered heterozygosity even in the

269 obligately aquatic *Rotaria* species.

270      Whatever the underlying mechanism, the observed recombination signatures in bdelloid  
271 rotifers are compatible with the three hypotheses proposed to explain the previous reports of  
272 allele sharing patterns in bdelloid rotifers: i) allele sharing may be due to undetected contam-  
273 ination between cultures, either during colony culture itself or during sample preparation for  
274 sequencing (46–50); ii) allele sharing is the result of horizontal genetic transfers between bdel-  
275 loid individuals through unknown molecular mechanisms, possibly associated with desiccation  
276 (but not for the non-desiccating species) and potentially linked to the high propensity of bdel-  
277 loids to retain non-metazoan genes into their genomes (13, 17, 20, 40, 43, 51); iii) allele sharing  
278 is caused by cryptic sexual reproduction (52), with sex events being rare enough so that males,  
279 sperm, fertilization and meiosis were never observed, but sufficient to leave a distinctive foot-  
280 print in every population sample studied so far (13, 19, 21, 45). The mechanism behind the  
281 observed signatures of genetic exchanges between bdelloid individuals remains puzzling and  
282 therefore the significance of outcrossing in this asexual lineage remains unclear. We anticipate  
283 the chromosome-level genome assembly of *A. vaga* presented here will stimulate future popu-  
284 lation genomics studies that will help to determine the cause of these allele sharing patterns.

285 **Long-term asexual evolution** This high-quality telomere-to-telomere assembly firmly estab-  
286 lishes *A. vaga* as a model system to study long-term asexual evolution. Homologous chromo-  
287 somes are present in the bdelloid species *A. vaga* and might well occur in all bdelloid rotifers,  
288 as colinear pairs of sequenced fosmids were found in two distinct bdelloid species *A. vaga* and  
289 *Philodina roseola*, with each colinear pair in one species resembling the colinear pair in the  
290 other species (12). The observed long LOH tracks indicate the existence of long-range homol-  
291 ogous recombination (Fig. 4), whether this occurs during a meiotic-like parthenogenetic mode  
292 of reproduction or in a mitotic context during frequent repair of accidental DNA DSBs remains

293 speculative. Recombination (mitotic or meiotic) could increase the rate of gene conversion in  
294 asexual lineages, a signature previously observed in *A. vaga* (9). Gene conversion, particularly  
295 when slightly biased, could correct deleterious mutations and reduce the rate of clonal dete-  
296 rioration (53), or even speed up the fixation of beneficial mutations (54). However, besides  
297 signatures of LOH representing allelic recombination, we also observed LOH via deletions in  
298 the genome of *A. vaga*. The random accumulation of LOH events could expose deleterious  
299 recessive mutations in asexuals through loss of complementation (55). This chromosome-scale  
300 genome assembly of asexual *A. vaga* therefore is a critical tool to be able to evaluate the rela-  
301 tive benefits of these recombination events on their long-term evolution and paves the way for  
302 studies on genome dynamics in *A. vaga*.

303 In general, asexual populations suffer from the absence of gene shuffling with other individ-  
304 uals and the long-term evolutionary success of bdelloid rotifers in the absence of outcrossing  
305 therefore remains puzzling. It is important to try to discriminate between the consequences of  
306 the two aspects underlying sex: recombination and outcrossing. Theoretical work on popula-  
307 tion genetics showed that selection could be at least as efficient in automictic lineages than in  
308 sexuals under certain circumstances (e.g. effective population size, recombination rates (56)). It  
309 is therefore conceivable that the combination of potentially large populations, a relatively high  
310 level of heterozygosity (or mutation rate) and specific recombination rates might explain how *A.*  
311 *vaga* maintains a delicate balance between losing and accumulating heterozygosity, and how it  
312 adapts and persists in the long term. Unfortunately, critical knowledge about bdelloid biology is  
313 still missing (e.g. mutation and recombination rates) to determine whether they might circum-  
314 vent the lack of outcrossing through a genetic equivalent of automixis. Outcrossing through  
315 sexual reproduction might speed up adaptation by allowing the combination of independently  
316 evolved alleles within the same individual, but might not be essential for bdelloid rotifers, es-  
317 pecially if a high frequency of HGT is also taking place. Despite presenting the highest amount

318 of HGTc among animals, our results also suggest that bdelloid rotifers might have to balance  
319 the acquisition rate of HGTs, a source of functional novelties, with the maintenance of faithful  
320 homology between chromosomes for homologous recombination. Overall, our work reinforces  
321 the hypothesis that recombination is critical for lineage longevity. Ancient asexual animals  
322 without a minimal rate of recombination, programmed through meiotic processes and/or acci-  
323 dental through their life-style, might not exist at all.

## 324 Materials and Methods

325 Complete description of materials and methods can be found in Supplementary Materials.

326 **Genome sequencing and assembly** Continuous cultures of *A. vaga* AD008 lab strain were  
327 processed in order to obtain the following sequencing data: about 350x coverage of WGS 250-  
328 bp paired-end Illumina reads, 200x coverage of PacBio RSII long reads, 125x coverage of  
329 ONT long reads and 75-bp paired-end Illumina reads of Hi-C libraries. Three independant  
330 genomes were assembled using Bwise (on Illumina short reads), NextDenovo (on ONT long  
331 reads) and Falcon (on Pacbio long reads). Uncollapsed haplotypes in the ONT-based assembly  
332 were then detected and discarded using purge\_dups. The resulting assembly was polished based  
333 on Illumina short reads and Pacbio long reads using HyPo, and is here referred to as "AV20"  
334 genome assembly. All assemblies were then scaffolded using instaGRAAL (on Hi-C data).  
335 Ploidy level and genome size was confirmed using k-mers spectra using KAT, synteny analyses  
336 using MCScanX, nucmer and D-GENIES, flow cytometry measurements and FISH using three  
337 pairs of oligo datasets designed on three chromosomes.

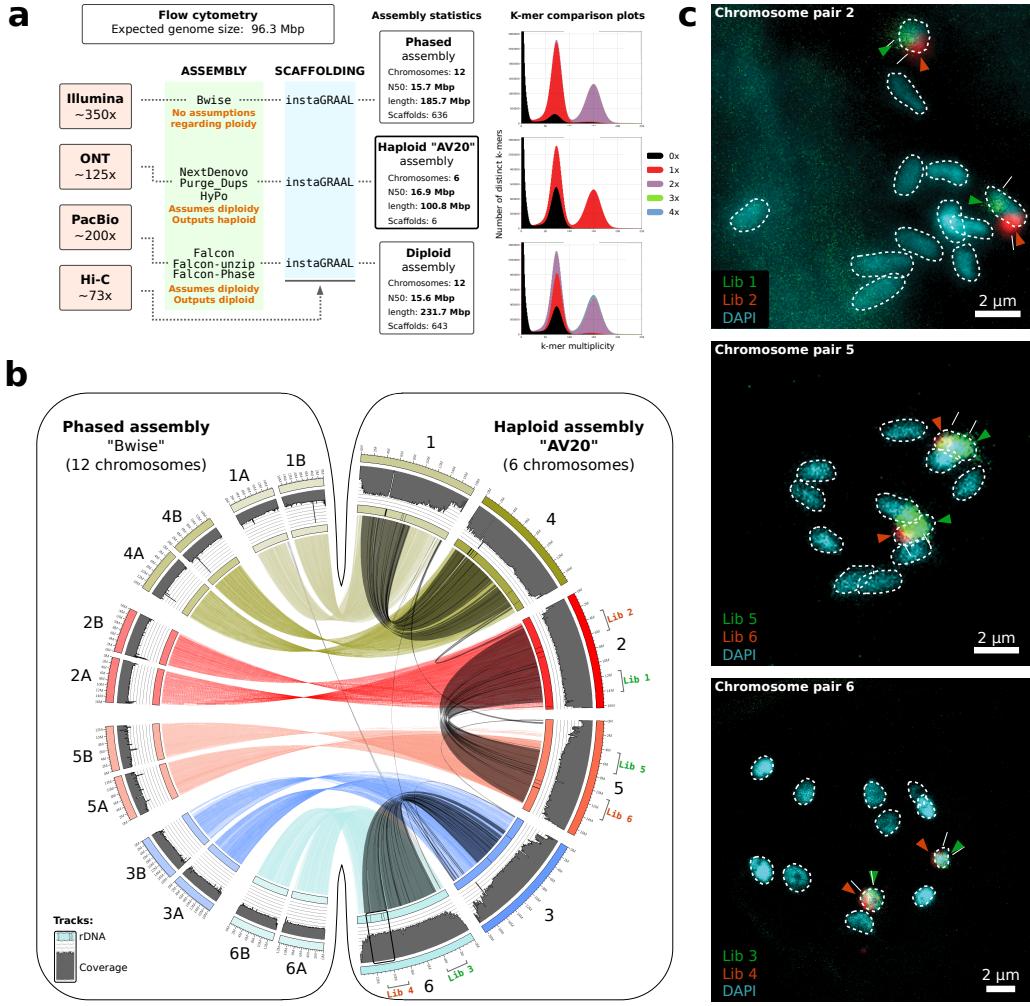
338 **Genome annotation** TE-like elements and canonical TEs consensus were build from the  
339 AV20 genome assembly using EDTA and TEdenovo pipeline, and AV20 genome was then

340 annotated using TEannot (part of the REPET pipeline). Genes were annotated using funanno-  
341 tate. For this, repeated elements previously annotated were masked using bedtools, part of the  
342 available RNA-Seq reads were mapped onto the genome while the other part of RNA-Seq reads  
343 were used to produce a de novo assembled transcriptome which was subsequently aligned onto  
344 the genome as part of the PASA pipeline. Then a combination of PASA annotations, de novo as-  
345 sembled transcriptome, metazoan BUSCO database and the proteic Uniprot database within fu-  
346 nannotate predict function. This first produced ab initio predictions using Genemark-ES, which  
347 were then used along with transcripts and proteic data to train Augustus to generate a second set  
348 of annotations. Lastly, it used Evidence Modeler as a weighted approach to combine annota-  
349 tions from PASA, Genemark and high quality predictions from Augustus into an integrated gene  
350 annotation set. We then used InterProScan5 in order to produce functional annotations to the  
351 predicted genes which were then used in combination with busco metazoan database using the  
352 funannotate annotate function with default parameters. We used Alienomics, a newly designed  
353 pipeline, in order to detect HGTC. This approach combines GC content, coverage, blasts, taxo-  
354 nomic information, expression level and synteny information in order to detect both HGTC (i.e.  
355 alien genes integrated into host scaffolds) and potential contaminants (i.e. alien genes present  
356 on alien scaffold). Note that our approach can only detect transfers from alien source outside  
357 of a given clade (i.e. here, metazoa). Viral-like genes were detected by performing a diamond  
358 blastx search on AV20 scaffolds using all viral proteins extracted from the nr database of NCBI  
359 (February 2020) to the exception of Retroviridae and Hepadnaviridae.

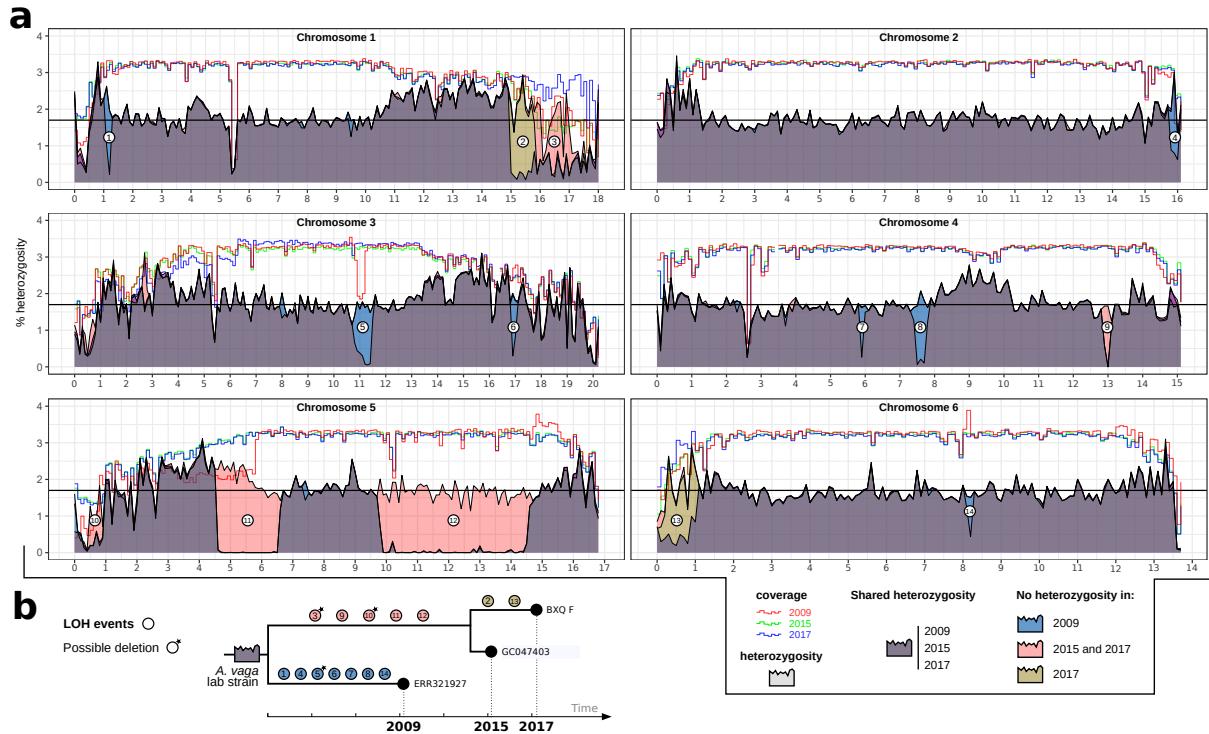
360 **Genome analyses** Coverage along AV20 scaffolds was computed using read mapping with  
361 bwa mem, and these alignments were used for genotyping three samples (i.e. GC047403,  
362 BXQF, ERR321927) using GATK (HaplotypeCaller function with -ERC GVCF option). The  
363 resulting gvcf files were combined (CombineGVCFs function) and were then jointly genotyped

364 (GenotypeGVCFs function). The variants were then filtered in order to only retain SNPs using  
365 custom bash and perl scripts. Divergence between homoeologous chromosomes was obtained  
366 by the production of a self-alignment of AV20 genome using nucmer, which was then filtered  
367 to only retain genomic alignments between homoeologous regions ranging from 500 to 10,000  
368 bp. MCSanX was used to detect synteny among HGTC, and custom scripts were used to  
369 detect strictly homoeologous HGTC synteny blocks stemming from the paelotetraploidization  
370 of bdelloids (i.e. ancient HGTC). GO terms from functional annotation were extracted for the  
371 2,422 recent HGTC and the 257 ancient HGTC were respectively compared to the entire gene set  
372 of the AV20 genome containing 32,378 proteins. Enrichment analyses were performed using  
373 topGO package with a fisher test and the "elim" algorithm.

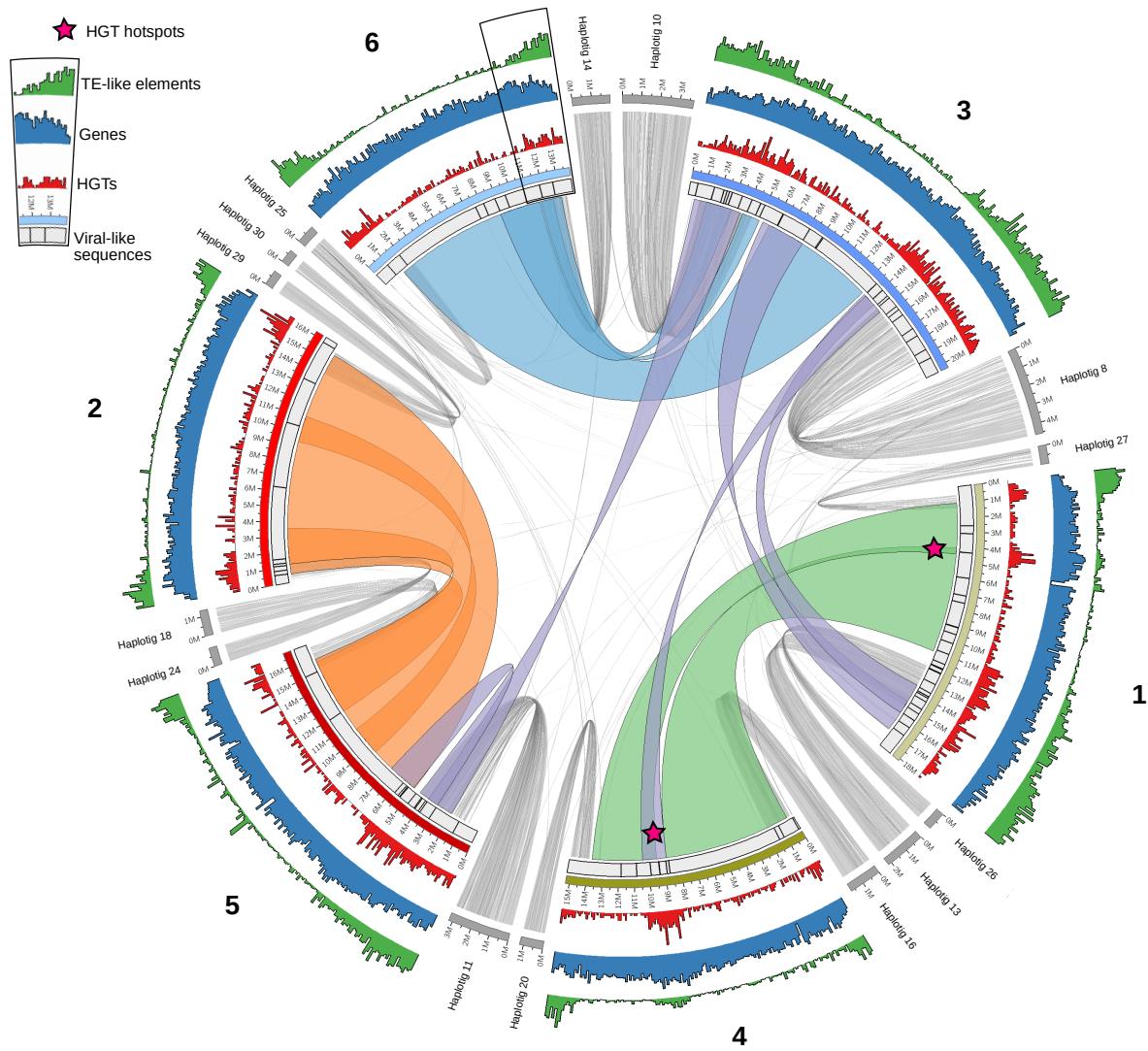
374 **Reappraisal of AV13 genome assembly** ONT reads were trimmed with porechop and were  
375 mapped onto the previous AV13 genome assembly using NGMLR. The AV20 and AV13 genome  
376 assemblies were aligned together using Sibelia. This alignment was used to detect putative  
377 breakpoint location which were then inspected using Tablet in order to evaluate whether ONT  
378 reads confirmed one of the assemblies. Synteny blocks from the Sibelia alignment between  
379 AV20 and AV13 were screened using a custom perl script (available at <https://github.com/jnarayan81/huntPalindromes>)  
380 and circos plots in order to evaluate the existence of palindromes in AV20 genome. No break-  
381 points or palindromes detected in the AV13 genome assembly could be found in the AV20  
382 genome assembly, nor be confirmed by ONT reads.



**Figure 1: Genome structure of *Adineta vaga* is diploid.** a) Outline of the three genome assembly approaches underlined by different assumptions on genome ploidy with median read coverage for all sequencing technologies indicated on the left and estimated with respect to the AV20 haploid genome assembly. The haploid genome size estimate of *A. vaga* obtained by flow cytometry is given (under the assumption that the genome is diploid) as well as the summary statistics of the genome assemblies. Number of chromosomes corresponds to the number of scaffolds longer than 10 Mbp. Ploidy levels of assemblies is indicated by the KAT plots of k-mers distribution (first and second peaks corresponds to heterozygous and homozygous k-mers, respectively; red and purple indicates haploidy and diploidy, respectively). b) Circos plot of the pairwise colinearity between the haploid AV20 and the phased Bwise genome assemblies, depicted by colored links and obtained using nucmer. Synteny blocks within AV20 genome (between homologous copies) are depicted as grey links and were obtained using MCScanX. Coverage along scaffolds of both AV20 and the phased assembly are depicted as grey histograms and were computed based on illumina reads from sample GC047403. Thin black bars on the scaffold ideograms correspond to rRNA genes. Schematic position of the FISH probes libraries on chromosome pair 2, 5 and 6 is indicated on the corresponding AV20 chromosomes. c) Karyotype of the 12 chromosomes of *A. vaga* (DAPI staining) with chromosome pairs 2, 5 and 6 highlighted by oligo painting using the FISH probe libraries depicted in panel b.



**Figure 2: Heterozygosity dynamics in *Adineta vaga*.** a) Heterozygosity and coverage distributions of three independent *A. vaga* samples from the same laboratory strain along the six chromosomes. Samples are labeled by the date of the extraction of their DNA (i.e. 2009, 2015 and 2017). Data from 2009 were used to assemble the previous version of *A. vaga* genome (9). Lines indicate short read coverage (normalized) and filled areas indicate the percentage of heterozygosity (y-axis). Chromosome lengths (x-axis) are in Mb. Mean SNP heterozygosity (1.7%) is depicted by the horizontal black line. b) Schematic reconstruction of heterozygosity evolution among 3 samples from the same initial *A. vaga* lab strain. Note that each sample had its own independent evolution and the exact sequence and timing of LOH events is unknown. LOH events noted with a small asterisk might correspond to deletions given the drop of coverage associated with the absence of heterozygosity.



**Figure 3: DNA content of haploid *Adineta vaga* genome AV20.** Synteny of HGTc is depicted by colored links between the 6 chromosome pairs. Violet links correspond to synteny block of HGTc between non-homoeologous chromosomes. Localization of alternative haplotigs, removed prior to genome scaffolding, are depicted by grey links. Distribution of repeated elements, genes, HGTc and viral-like sequences are depicted in green, blue, red and black bars, respectively. Ancient HGTc hotspots are indicated by pink stars.

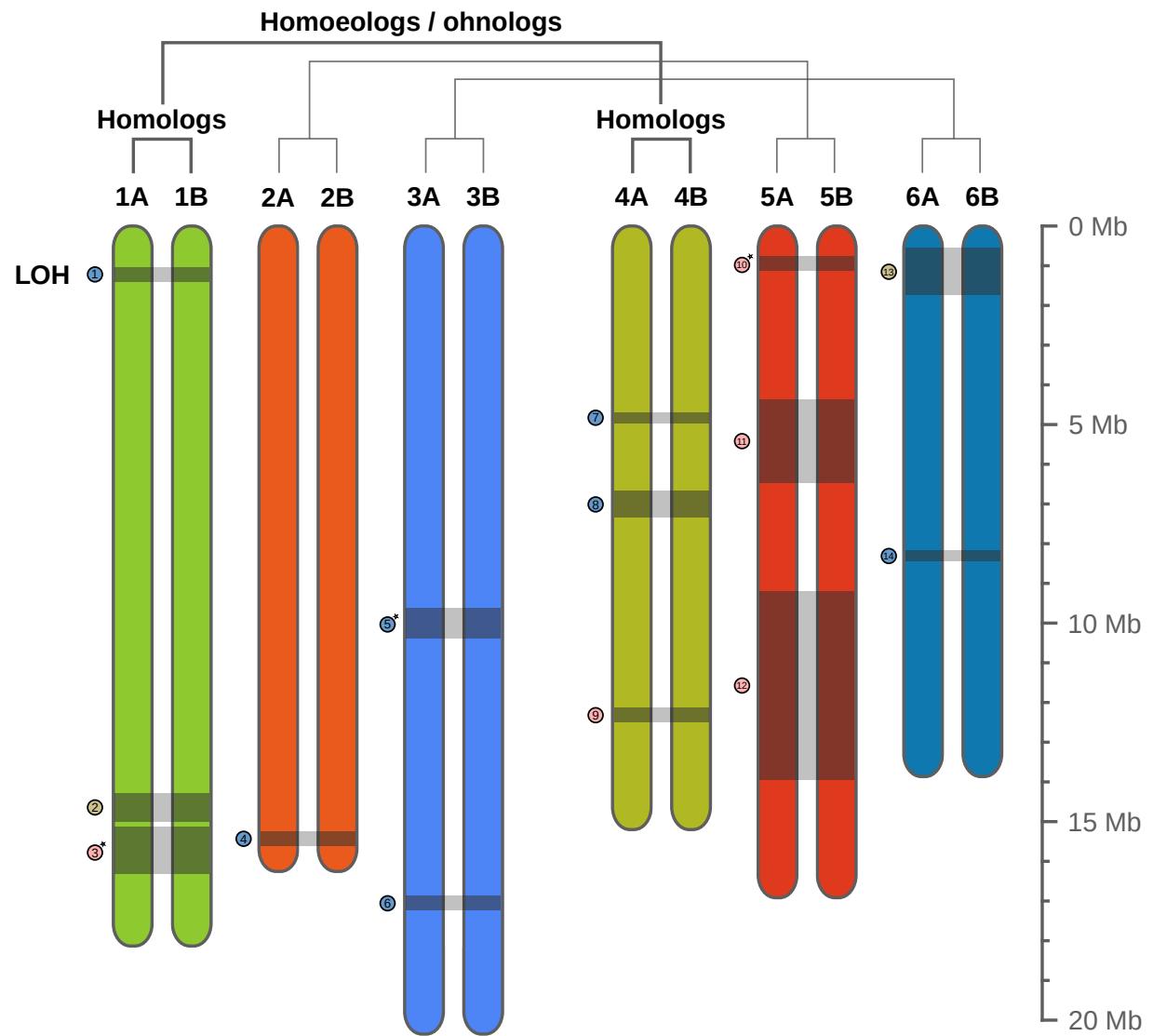


Figure 4: **Schematic representation of *A. vaga* karyotype.** The 12 chromosomes correspond to 6 pairs of homologous chromosomes (i.e. diploidy) sharing the same colour. Ancestral genome hybridization (or duplication) led to the existence of pairs of homoeologs (or ohnologs), represented in different but similar colours. Grey blocks linking homologous chromosomes and their corresponding numbered tokens depict loss of heterozygosity events produced by homologous recombination (see also Fig. 2). Length of chromosomes (in Mb) is indicated by the scale on the right.

383 **References and Notes**

- 384 1. J. Maynard Smith, *The evolution of sex* (Cambridge University Press Cambridge, 1978).
- 385 2. J. Lehtonen, M. D. Jennions, H. Kokko, *Trends Ecol Evol* **27**, 172 (2012).
- 386 3. O. Seudre, E. Vanhoenacker, S. Mauger, J. Coudret, D. Roze, *Journal of Evolutionary*  
387 *Biology* **33**, 112 (2020).
- 388 4. H. Fradin, *et al.*, *Current Biology* **27**, 2928 (2017).
- 389 5. W. G. Hill, A. Robertson, *Genetics Research* **8**, 269 (1966).
- 390 6. C. W. Birk, *Journal of Heredity* **101**, S42 (2010).
- 391 7. W. S. Hsu, *La Cellule Section* **57**, 283 (1956).
- 392 8. W. S. Hsu, *The Biological Bulletin* **111**, 364 (1956).
- 393 9. J.-F. Flot, *et al.*, *Nature* **500**, 453 (2013).
- 394 10. R. W. Nowell, *et al.*, *PLOS Biology* **16**, e2004830 (2018).
- 395 11. D. B. M. Welch, J. L. M. Welch, M. Meselson, *Proceedings of the National Academy of*  
396 *Sciences* **105**, 5145 (2008).
- 397 12. J. H. Hur, K. Van Doninck, M. L. Mandigo, M. Meselson, *Molecular Biology and Evolution* **26**, 375 (2009).
- 398 399 13. O. A. Vakhrusheva, *et al.*, *Nature Communications* **11**, 6421 (2020).
- 400 14. D. Fontaneto, N. Bunnefeld, M. Westberg, *Astrobiology* **12**, 863 (2012).
- 401 15. B. Hespeels, *et al.*, *Frontiers in Microbiology* **11**, 1792 (2020).
- 402 16. E. Gladyshev, M. Meselson, *Proceedings of the National Academy of Sciences* **105**, 5139  
403 (2008).
- 404 17. B. Hespeels, *et al.*, *Journal of Evolutionary Biology* **27**, 1334 (2014).
- 405 18. J. Romiguier, *et al.*, *Nature* **515**, 261 (2014).
- 406 19. A. Signorovitch, J. Hur, E. Gladyshev, M. Meselson, *Genetics* **200**, 581 (2015).
- 407 20. N. Debortoli, *et al.*, *Current Biology* **26** (2016).
- 408 21. A. Signorovitch, J. Hur, E. Gladyshev, M. Meselson, *Current Biology* **26**, R754 (2016).

- 409 22. A. Limasset, Novel approaches for the exploitation of high throughput sequencing data,  
410 PhD thesis, Université Rennes 1 (2017).
- 411 23. Nextdenovo (2020). <Https://github.com/Nextomics/NextDenovo>.
- 412 24. C.-S. Chin, *et al.*, *Nature Methods* **13**, 1050 (2016).
- 413 25. L. Baudry, *et al.*, *bioRxiv* p. 2019.12.22.882084 (2019).
- 414 26. J. L. Mark Welch, M. Meselson, *Hydrobiologia* **387**, 403 (1998).
- 415 27. H. Golczyk, A. Massouh, S. Greiner, *The Plant Cell* **26**, 1280 (2014).
- 416 28. B. J. Hecox-Lea, D. B. Mark Welch, *BMC Evolutionary Biology* **18**, 177 (2018).
- 417 29. C. J. Sakofsky, A. Malkova, *Critical Reviews in Biochemistry and Molecular Biology* **52**,  
418 395 (2017).
- 419 30. E. Yim, K. E. O'Connell, J. St. Charles, T. D. Petes, *Genetics* **198**, 181 (2014).
- 420 31. M. Bzymek, N. H. Thayer, S. D. Oh, N. Kleckner, N. Hunter, *Nature* **464**, 937 (2010).
- 421 32. L. C. Kadyk, L. H. Hartwell, *Genetics* **132**, 387 (1992).
- 422 33. C. Hiruta, C. Nishida, S. Tochinai, *Chromosome Research* **18**, 833 (2010).
- 423 34. O. Nougué, *et al.*, *Journal of Evolutionary Biology* **28**, 2337 (2015).
- 424 35. F. Goudie, B. P. Oldroyd, *Apidologie* **45**, 306 (2014).
- 425 36. J. M. Mason, T. A. Randall, R. Capkova Frydrychova, *Chromosoma* **125**, 65 (2016).
- 426 37. I. Arkhipova, M. Meselson, *BioEssays* pp. 76–85 (2005).
- 427 38. R. W. Nowell, *et al.*, *eLife* **10**, e63194 (2021).
- 428 39. C. Boschetti, *et al.*, *PLOS Genetics* **8**, e1003035 (2012).
- 429 40. I. Eyres, *et al.*, *BMC biology* **13**, 90 (2015).
- 430 41. E. A. Gladyshev, M. Meselson, I. R. Arkhipova, *Science* **320**, 1210 (2008).
- 431 42. B. Hespeels, J.-F. Flot, A. Derzelle, K. Van Doninck, *Evolutionary Biology: Genome Evo-*  
432 *lution, Speciation, coevolution and Origin of Life*, P. Pontarotti, ed. (Springer International  
433 Publishing, 2014), pp. 207–225.
- 434 43. J.-F. Flot, N. Debortoli, B. Hallet, K. V. Doninck, *Current Biology* **26**, R755 (2016).

- 435 44. T. Schwander, *Current Biology* **26**, R233 (2016).
- 436 45. V. N. Laine, T. Sackton, M. Meselson, *bioRxiv* p. 2020.08.06.239590 (2020).
- 437 46. M. Ballenghien, N. Faivre, N. Galtier, *BMC Biology* **15**, 25 (2017).
- 438 47. P. Simion, *et al.*, *BMC Biology* **16**, 28 (2018).
- 439 48. C. G. Wilson, R. W. Nowell, T. G. Barraclough, *Current Biology* **28**, 2436 (2018).
- 440 49. M. Prous, K. M. Lee, M. Mutanen, *Molecular Phylogenetics and Evolution* **143**, 106670  
441 (2020).
- 442 50. P. Simion, F. Delsuc, H. Philippe, *Phylogenetics in the Genomic Era*, C. Scornavacca,  
443 F. Delsuc, N. Galtier, eds. (<https://hal.inria.fr/PGE/>, 2020), pp. 2.1:1–2.1:34.
- 444 51. J.-F. Flot, N. Debortoli, B. Hallet, J. Narayan, K. Van Doninck, *BioRxiv* p. 368209 (2018).
- 445 52. L. Boyer, R. Jabbour-Zahab, M. Mosna, C. R. Haag, T. Lenormand, *Evolution Letters* **5**,  
446 164 (2021).
- 447 53. O. Khakhlova, R. Bock, *The Plant Journal* **46**, 85 (2006).
- 448 54. M. A. Mandegar, S. P. Otto, *Proceedings. Biological Sciences* **274**, 1301 (2007).
- 449 55. M. Archetti, *Journal of Evolutionary Biology* **17**, 1084 (2004).
- 450 56. J. Engelstädter, *Genetics* **206**, 993 (2017).
- 451 57. L. Lazar-Stefanita, *et al.*, *The EMBO Journal* **36**, 2684 (2017).
- 452 58. E. Lieberman-Aiden, *et al.*, *Science* **326**, 289 (2009).
- 453 59. D. B. Mark Welch, M. Meselson, *Biological Journal of the Linnean Society* **79**, 85 (2003).
- 454 60. M. D. Bennett, I. J. Leitch, H. J. Price, J. S. Johnston, *Annals of Botany* **91**, 547 (2003).
- 455 61. J. L. Mark Welch, D. B. Mark Welch, M. Meselson, *Proceedings of the National Academy  
456 of Sciences* **101**, 1618 (2004).
- 457 62. B. J. Beliveau, *et al.*, *Proceedings of the National Academy of Sciences* **109**, 21301 (2012).
- 458 63. B. J. Beliveau, *et al.*, *Proceedings of the National Academy of Sciences* **115**, E2183 (2018).
- 459 64. B. J. Beliveau, *et al.*, *Nature Communications* **6**, 7147 (2015).
- 460 65. B. D. Fields, S. C. Nguyen, G. Nir, S. Kennedy, *eLife* **8**, e42823 (2019).

- 461 66. R. Chikhi, A. Limasset, P. Medvedev, *Bioinformatics* **32**, i201 (2016).
- 462 67. A. Limasset, J.-F. Flot, P. Peterlongo, *Bioinformatics* **36**, 1374 (2020).
- 463 68. A. Limasset, B. Cazaux, E. Rivals, P. Peterlongo, *BMC Bioinformatics* **17**, 237 (2016).
- 464 69. A. V. Zimin, *et al.*, *Bioinformatics* **29**, 2669 (2013).
- 465 70. D. Guan, *et al.*, *Bioinformatics* **36**, 2896 (2020).
- 466 71. H. Li, *Bioinformatics* **34**, 3094 (2018).
- 467 72. F. Cabanettes, C. Klopp, *PeerJ* **6**, e4958 (2018).
- 468 73. R. Kundu, J. Casey, W.-K. Sung, *bioRxiv* p. 2019.12.19.882506 (2019).
- 469 74. H. Li, *arXiv:1303.3997 [q-bio]* (2013).
- 470 75. H. Li, *et al.*, *Bioinformatics* **25**, 2078 (2009).
- 471 76. A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, P. Prins, *Bioinformatics* **31**, 2032 (2015).
- 472 77. S. Koren, *et al.*, *Genome Research* **27**, 722 (2017).
- 473 78. C. Matthey-Doret, *et al.*, hicstuff (2020). [Https://github.com/koszullab/hicstuff](https://github.com/koszullab/hicstuff).
- 474 79. B. Langmead, S. L. Salzberg, *Nature Methods* **9**, 357 (2012).
- 475 80. K.-K. Lam, K. LaButti, A. Khalak, D. Tse, *Bioinformatics* **31**, 3207 (2015).
- 476 81. A. L. Delcher, S. L. Salzberg, A. M. Phillippy, *Current Protocols in Bioinformatics* **00**, 10.3.1 (2003).
- 477 82. S. Ou, *et al.*, *Genome Biology* **20**, 275 (2019).
- 478 83. T. Flutre, E. Duprat, C. Feuillet, H. Quesneville, *PLOS ONE* **6**, e16526 (2011).
- 479 84. H. Quesneville, *et al.*, *PLOS Computational Biology* **1**, e22 (2005).
- 480 85. J. Palmer, J. Stajich, *nextgenusfs/funannotate: funannotate v1.5.3* (Zenodo, 2019).
- 481 86. A. R. Quinlan, I. M. Hall, *Bioinformatics* **26**, 841 (2010).
- 482 87. A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **30**, 2114 (2014).
- 483 88. M. G. Grabherr, *et al.*, *Nature Biotechnology* **29**, 644 (2011).
- 484 89. B. J. Haas, *et al.*, *Nucleic Acids Research* **31**, 5654 (2003).

- 486 90. A. Lomsadze, V. Ter-Hovhannisyan, Y. O. Chernoff, M. Borodovsky, *Nucleic Acids Re-*  
487 *search* **33**, 6494 (2005).
- 488 91. P. Jones, *et al.*, *Bioinformatics* **30**, 1236 (2014).
- 489 92. T. Seemann, Barrnap 0.9 (2018). [Https://github.com/tseemann/barrnap](https://github.com/tseemann/barrnap).
- 490 93. B. Buchfink, C. Xie, D. H. Huson, *Nature Methods* **12**, 59 (2015).
- 491 94. D. Mapleson, G. Garcia Accinelli, G. Kettleborough, J. Wright, B. J. Clavijo, *Bioinfor-*  
492 *matics* **33**, 574 (2017).
- 493 95. S. Kurtz, *et al.*, *Genome Biology* **5**, R12 (2004).
- 494 96. A. McKenna, *et al.*, *Genome Research* **20**, 1297 (2010).
- 495 97. R. Poplin, *et al.*, *bioRxiv* p. 201178 (2018).
- 496 98. Y. Wang, *et al.*, *Nucleic Acids Research* **40**, e49 (2012).
- 497 99. A. Alexa, J. Rahnenführer, T. Lengauer, *Bioinformatics* **22**, 1600 (2006).
- 498 100. R. R. Wick, L. M. Judd, C. L. Gorrie, K. E. Holt, *Microbial Genomics*, **3**, e000132 (2017).
- 499 101. F. J. Sedlazeck, *et al.*, *Nature Methods* **15**, 461 (2018).
- 500 102. I. Minkin, A. Patel, M. Kolmogorov, N. Vyahhi, S. Pham, *Algorithms in Bioinformatics*,  
501 A. Darling, J. Stoye, eds. (Springer, 2013), pp. 215–229.
- 502 103. I. Milne, *et al.*, *Bioinformatics* **26**, 401 (2010).

503 **Data availability**

504 All data and genome assembly are available under the project accession number PRJNA680543.

505 **Acknowledgments**

506 We thank Alexandre Mayer for his help with scores computation and transformation within  
507 Alienomics, Mathilde Colinet for her help with FISH experiments as well as Nathalie Ver-  
508 bruggen for kindly providing *Arabidopsis* plantlets for genome size estimation. The Morphim  
509 imaging platform of UNamur is acknowledged for his technical help with the FISH image anal-  
510 yses.

511 **Funding**

512 This project received funding from the Horizon 2020 research and innovation program un-  
513 der the European Research Council (ERC) grant agreement 725998 (RHEA) and from BEL-  
514 SPO PRODEX for the ESA selected ILSRA-2014-0106 Project to KVD; from the Fédération  
515 Wallonie-Bruxelles via an 'Action de Recherche Concertée' (ARC) grant to KVD and BH;  
516 under the Marie Skłodowska-Curie grant agreement 764840 to JFF (ITN IGNITE, [www.itn-ignite.eu](http://www.itn-ignite.eu)); from the Fédération Wallonie-Bruxelles via an 'Action de Recherche Concertée'  
517 (ARC) grant to JFF; from the European Research Council under the Horizon 2020 Program  
518 (ERC grant agreement 260822) and JPI-EC-AMR STARCS ANR-16-JPEC-0003-05 grant to  
519 RK; through funding by the French Government "Investissement d'Avenir" program FRANCE  
520 GENOMIQUE (ANR-10-INBS-09); EN obtained a FSR UNamur fund; AH, AD and RA are  
521 Research Fellows of the Fonds de la Recherche Scientifique – FNRS.  
522

523 **Author contributions**

524 Conceptualization : PS, JN, TL, JFF, EN, KVD, AH, AD, MC, RK, AL, MT.  
525 Data curation: PS, JN, AH, AD, RA.  
526 Formal Analysis: PS, JN, TL, JFF, EN, AH, CG, FR, AD, MC, LB, DK.  
527 Funding acquisition: JFF, KVD.  
528 Investigation: PS, JN, TL, JFF, KVD, AH, EN, MC, ML, AD, LB, RK, ED, DK, RA, AL.  
529 Methodology: PS, JN, TL, JFF, EN, AH, CG, RC, ML, AD, MC, LB, RK, DK, MM, NG, AL.  
530 Project administration: JFF, KVD.  
531 Resources: LB, CC, MM, BH, KL, JV, JFF, KVD.  
532 Software: JN, PS, JFF, AH, AD, MC, DK, NG, AL.  
533 Supervision: TL, JFF, KVD, RK, ED, RC, BH.  
534 Validation: JN, PS, JFF, KVD, AH.  
535 Visualization: PS, JN, JFF, EN, KVD, AH, DK.  
536 Writing – original draft: PS, JN, JFF, KVD, AH, MT, ED, TL.  
537 Writing – review & editing: all authors.

538 **Competing Interests**

539 All authors declare that they have no competing interests.

540 **Supplementary Materials**

541 **Data generation**

542 **Strain culture, library preparation and DNA sequencing** We continuously cultivated *A.*  
543 *vaga* individuals from AD008 strain (i.e. same strain as in (9), COI sequence accession number  
544 is KM043184) since 2007 in Petri dishes using Spa water, feeding them with sterile extract of  
545 lettuce juice and stocking well-grown cultures at -80°C. *A. vaga* individuals were thawed before  
546 proceeding to DNA extraction using QIAGEN Gentra Puregene Tissue Kit. Genomics Core  
547 (UZLeuven) produced PCR-free 250-bp paired-end Illumina reads that were sequenced with a  
548 depth of approximately 350x on a HiSeq 2500 sequencing platform. The same procedure was  
549 followed in order to obtain high molecular weight DNA using Macherey-Nagel NucleoBond  
550 HMW procedure that was subsequently sent to the Genomics Core (UZLeuven) to generate a  
551 depth of 200x of PacBio RSII sequencing data. Around 30 µg of high molecular weight DNA  
552 was also extracted from living *A. vaga* individuals using the QIAGEN Gentra Puregene Tissue  
553 Kit and then sent to the Genoscope sequencing center (François Jacob Institute of Biology)  
554 which produced 5 ONT libraries, each starting from 2 to 5 µg of DNA, using the 1D ligation  
555 sequencing kit (SQ-LSK108) and R9.4 (or R9.4.1) flowcells. This resulted in a sequencing  
556 depth of 125x long-reads using Oxford Nanopore Technology (ONT). All samples ID and SRA  
557 accession numbers are detailed in Supp. Table S1.

558 **Chromosome conformation capture: Hi-C** The Hi-C library construction protocol was  
559 adapted from (57, 58). Briefly, individuals from the *A. vaga* AD008 strain were chemically  
560 cross-linked for 20 min at room temperature and 30 min at 4°C (with gentle stirring) using  
561 formaldehyde (final concentration: 5% in milliQ water; final volume: 50 ml). After fixation the  
562 sample was centrifuged for 10 min at 4000 rpm at 4°C. The formaldehyde was then quenched  
563 for 5 min at RT and 15 min at 4°C (with gentle stirring) by adding 50 ml of 250mM glycine.  
564 The cells were recovered by centrifugation for 10 min at 4000rpm at 4°C, supernatant was re-  
565 moved and pellet stored at -80°C until use. The Hi-C library was then prepared as follows. Cells  
566 were resuspended in 1.2 mL of 1X DpnII buffer (NEB), transferred to a VK05 tubes (Precellys)  
567 and disrupted using the Precellys apparatus and the following program ([20 sec – 6000 rpm,  
568 30 sec – pause] 9x cycles). The lysate was recovered (around 1.2 mL) and transferred to two  
569 1.5 mL tubes. SDS was added to a final concentration of 0.3% and the 2 reactions were incu-  
570 bated at 65°C for 20 minutes followed by an incubation of 30 minutes at 37°C. A volume of 50  
571 µL of 20% triton-X100 was added to each tube and incubation was continued for 30 minutes.  
572 DpnII restriction enzyme (150 units) was added to each tube and the reactions were incubated  
573 overnight at 37°C. Next morning, reactions were centrifuged at 16,000 x g for 20 minutes. The  
574 supernatants were discarded and the pellets were resuspended in 200 µL of NE2 1X buffer and  
575 pooled (final volume = 400 µL). DNA extremities were labelled with biotin using the following  
576 mix (50 µL NE2 10X buffer, 37.5 µL 0.4 mM dCTP-14-biotin, 4.5 µL 10mM dATP-dGTP-dTTP  
577 mix, 10 µL Klenow 5 U/µL) and an incubation of 45 minutes at 37°C. The labelling reaction

578 was then split in two for the ligation reaction (ligation buffer – 1.6 mL, ATP 100 mM – 160  
579  $\mu$ L, BSA 10 mg/mL – 160  $\mu$ L, ligase 5 U/ $\mu$ L – 50  $\mu$ L, H<sub>2</sub>O – 13.8 mL). The ligation reactions  
580 were incubated for 4 hours at 16°C. After addition of 200  $\mu$ L of 10% SDS, 200  $\mu$ L of 500 mM  
581 EDTA and 200  $\mu$ L of proteinase K 20 mg/mL, the tubes were incubated overnight at 65°C. DNA  
582 was then extracted, purified and processed for sequencing as previously described (57). Hi-C  
583 libraries were sequenced on a NextSeq 550 sequencer (2 $\times$ 75 bp, paired-end Illumina NextSeq  
584 with the first ten bases acting as barcodes).

585 **Genome size estimation** The genome assemblies produced by all three methods (Bwise, Flye  
586 and Falcon) were markedly smaller than expected based on the generally admitted genome size  
587 of 0.25 pg per (non-reduced) oocyte ([http://www.genomesize.com/result\\_species.php?id=5369](http://www.genomesize.com/result_species.php?id=5369)),  
588 equivalent to 244 Mbp for a diploid assembly or 122 Mbp for a haploid assembly. As there is  
589 considerable confusion in the literature considering the genome size of *Adineta vaga* (e.g. re-  
590 port of a nuclear DNA content of about 0.7 pg (59), nearly 3 times higher than in the Animal  
591 Genome Size database although the entry there refers to this article), we decided to perform  
592 an independent assessment of the genome size of *Adineta vaga* using flow cytometry, with  
593 *Arabidopsis thaliana* ecotype Colombia (for which a haploid genome size of 157 Mbp was pre-  
594 viously measured (60) as a genome-size standard for comparison. Nuclei from both species  
595 were isolated according to the protocol from the Cystain Pi absolute T (SYSMEX #05- 5023)  
596 kit. Briefly, we chopped them together in the same extraction buffer (500  $\mu$ L), after which the  
597 material was filtered through a 30  $\mu$ m nylon membrane. After RNase treatment (80  $\mu$ g/ml),  
598 the DNA was labeled for 1h in the dark with 2 ml of staining buffer containing 120  $\mu$ L of  
599 propidium iodide. The labeled nuclei were then analyzed on the CyFlow Space flow cytome-  
600 ter (Sysmex) of the research unit "Evolutionary Biology & Ecology" of the Université libre de  
601 Bruxelles (ULB). We used a blue laser with an excitation wavelength of 488 nm. The whole  
602 procedure was performed three times on different days, using different batches of rotifers and  
603 leaves from different *A. thaliana* plants every time, and the .FCS files were analyzed using the  
604 FlowJo v10.6.2 software. The estimated haploid genome size is presented in Supp. Fig. 1.

605 **Chromosome painting (FISH)** To assess the colinearity between two chromosomal mark-  
606 ers, FISH experiments were performed on samples containing well resolved condensed chro-  
607 mosomes. As bdelloids are eutelic, such condensed chromosomes are only found in embryos  
608 undergoing nuclear divisions. Particularly, young embryos containing only few nuclei usually  
609 exhibit the nicest karyotypes (61). To collect young, ideally one-cell, embryos, about 200 ro-  
610 tifiers bearing a single egg were first isolated in a petri dish containing a 1% agarose pad and  
611 ice-cold Spa® spring water. The agarose pad avoids the embryos to stick at the bottom of the  
612 plate and ease their isolation. The rotifers were starved for 24 hours at 4°C and, the next day,  
613 about half of the water was removed and replaced by the same volume of fresh water at RT  
614 containing lettuce filtrate. Rotifers were incubated at 25°C and, about 3 hours later, all individ-  
615 uals were laying eggs almost synchronously. Immediately after laying, the eggs were collected  
616 and fixed in methanol (Merck Millipore®, 1070182511): acetic acid glacial (VWR™, 20104-

617 243) (3:1) solution on ice. After isolation of all eggs, they were collected by centrifugation  
618 (14,000 rpm, 2 min, RT), fixed again with methanol: acetic acid glacial (3:1) and stored at 4°C  
619 until slide preparation. About 100 embryos bearing one or few nuclei can be collected by this  
620 method. For the FISH probe synthesis, we used the Oligopaint strategy that consists in the use  
621 of libraries of short single-stranded oligonucleotides (oligos) that are fluorescently labeled to  
622 visualize megabases (Mbs) of genomic regions (62). The design of the probes was performed  
623 using the OligoMiner pipeline (63) that selects for oligos having similar parameters such as  
624 melting temperature (Tm) or the absence of secondary structures. The selected oligos have  
625 a 30-42 nt region of genomic homology with a Tm of 42°C flanked by constant nongenomic  
626 sequences at the 5' end (5'-ccc-gcg-tta-acc-ata-cac-cg-3') and at 3' end (5'-ggt-agc-cac-acg-ctt-  
627 cga-tg-3'). These sequences are necessary for the labeling and the amplification of the libraries  
628 by PCR (see below). We ordered 6 libraries from GenScript®: (i) library 1 (9.2k oligos) targets  
629 the chromosomes 2a/b from 13 to 16 Mbs; (ii) library 2 (7.7k oligos) targets the chromosomes  
630 2a/b from 2 to 6 Mbs; (iii) library 3 (7.7k oligos) targets the chromosomes 6a/b from 2 to 6  
631 Mbs; (iv) library 4 (8.0k oligos) targets the chromosomes 6a/b from 8 to 12 Mbs; (v) library  
632 5 (7.9k oligos) targets the chromosomes 5a/b from 3 to 7 Mbs; and (vi) library 6 (7.8k oligos)  
633 targets the chromosomes 5a/b from 9 to 13 Mbs. The probes were labeled and amplified ac-  
634 cording to the 'One-day' probe synthesis protocol using lambda exonuclease described in (64)  
635 (<https://oligopaints.hms.harvard.edu/protocols>). The oligo libraries were first amplified and la-  
636 beled by PCR. Twenty-four PCR reactions (24 x 50 µl) were performed with 1 U of Q5 high-  
637 fidelity polymerase (New England Biolabs®, M0491), 200 µM dNTPs, 0.5 µM of fluorescently  
638 labeled forward primer (5'-Fluo/ccc-gcg-tta-acc-ata-cac-cg-3'), 0.5 µM of phosphorylated re-  
639 verse primer (5'-Phos/cat-cga-agc-gtg-tgg-cta-cc-3'), and 1.25 ng of Oligopaint library. The  
640 primers were ordered from IDT®. To perform the two-color FISH experiments, libraries 1,  
641 3 and 5 were labeled with Atto488N (green) and the libraries 2, 4, and 6 were labeled with  
642 Atto565N (red). The PCR reactions were incubated at 98°C for 5 min, followed by 40 cycles of  
643 30 sec at 98°C, 30 sec at 56°C, and 15 sec at 72°C, and a final extension at 72°C for 5 min. The  
644 PCR reactions were then collected and concentrated using the Zymo DNA clean concentrator  
645 kit (Zymo research®, D4032). The concentration was performed according to the manufacturer  
646 protocol and the libraries were eluted in 2,800 µl of RNase/DNase free water. Lambda ex-  
647 onuclease (New England Biolabs®, M0262) was then used to hydrolyze the 5'-phosphorylated  
648 strand of the double-stranded amplicons. DNA eluant (2,200 µl) was processed by 250 U of  
649 lambda exonuclease at 37°C for 30 min, and then stopped by incubation at 75°C for 10 min.  
650 The single-stranded labeled probes were finally cleaned up using the Monarch PCR & DNA  
651 cleanup kit (New England Biolabs®, T1030) following the oligonucleotide cleanup protocol.  
652 Probes were eluted in 20 µl of RNase/DNase free water and stored protected from light at  
653 -20°C until use. The hybridization of the probes on embryos was adapted from previous proto-  
654 cols (61,65). At least 100 embryos stored in methanol: acetic acid glacial were dropped onto an  
655 uncoated and clean microscope slide (VWR™, 631-1550) and let dry on a wet paper for 30 min.  
656 Then, a cover slip (VWR™, 631-1572) was placed over the embryos and they were squashed  
657 by gentle pressure on the slide. All following treatments of embryos on slides were conducted

in Coplin jars. Embryos were permeabilized in 0.1% saponin (Sigma-Aldrich®, 47036)/0.1% triton X-100 (Sigma-Aldrich®, T8787) in PBS (Lonza®, 17516Q) for 10 min, followed by 2 washes of 5 min in PBS. Samples were incubated for 20 min in PBS containing 20% of glyc- erol (Carl Roth®, 7530.1) and washed again 2 times in PBS. Slides were incubated for 5 min in 2x SSC (SSC 20X, Invitrogen 15557-036) supplemented with 0.1% of Tween-20 (Sigma- Aldrich®, P1379) (i.e., 2x SSCT), and then for 5 min in 2x SSCT supplemented with 50% of formamide (Sigma-Aldrich®, 47671). The slides were then put on top of a thermoblock at 92°C for 2.5 min and transferred in a Coplin jar containing 2x SSCT-50% formamide at 60°C for 20 min. The jar was then removed from 60°C and placed at RT for 1 hour. The hybridization mixture (50  $\mu$ l) composed of 2x SSC, 50% formamide, 1  $\mu$ l of RNase A (Sigma-Aldrich®, R4642), 10% dextran sulfate (Sigma-Aldrich®, S4030), and 10  $\mu$ l of each labeled oligo libraries, was placed on a clean cover slip and the slide containing the embryos was inverted onto this cocktail of hybridization. For the two-colors FISH, the oligo library 1 (green) was mixed with the oligo library 2 (red), the oligo library 3 (green) was mixed with the oligo library 4 red), and the oligo library 5 (green) was mixed with the oligo library 6 (red). The cover slip was sealed with rubber cement and let dry for 5 min at RT. The mounted slide was denatured at 92°C for 2.5 min on a thermoblock, transferred to a dark humidified chamber, and incubated O/N at 37°C. The next day, the cover slip was removed carefully from the slides. The slides were then washed in 2x SSCT at 60°C for 15 min, and in 2x SSCT at RT for 10 min. Chromosomes were counterstained for 20 minutes with 1  $\mu$ g/ml DAPI (4',6-diamidino-2-phenylindole; ThermoFisher Scientific, D3571) in 2x SSC. Slides were washed twice in 2x SSC for 10 min, and mounted under a 24  $\times$  32 mm cover slip in Mowiol 40-88 (Sigma-Aldrich®, 324590). Chromosomes and FISH signals were observed under a Leica TCS SP5 fluorescence confocal microscope using the 488 nm laser to capture the green signal, the 561 nm laser for the red signal and the 405 nm laser line for the DAPI signal. Images were captured in Z-stacks with the LAS AF software and they were finally processed and analyzed with Fiji (ImageJ, version 2.0.0).

## Chromosome-level genome assemblies

**Phased assembly: Bwise** The Bwise assembler v0.1 (<https://github.com/Malfoy/BWISE>) was used on high-coverage Illumina data (sample ID GC047403, see Supp. Table 1) to produce a draft phased genome assembly. We selected a Kmer size parameter of 63 (-k 63) as this produced the most contiguous assembly over the range of tested Kmer sizes: 63, 73, 101, 201. Other parameters were left as default. Bwise rests on a different paradigm than most assemblers: it starts by generating a de Bruijn graph from the reads to assemble (66), then cleans the graph by removing tips caused by sequencing errors (67), remaps the initial reads on this corrected de Bruijn graph (68), transforming them in super-reads (69). Finally, the resulting super-reads are assembled in a greedy fashion whenever they overlap unambiguously by one or several unitigs. This approach was devised in order to produce an assembly that reflects faithfully the unknown ploidy level of the organism sequenced. Therefore, Bwise will produce haploid assemblies whenever the organism sequenced is haploid, diploid assemblies whenever

697 the organism sequenced is diploid, triploid... etc.

698 **AV20 Haploid assembly: NextDeNovo** Reads quality control was performed with FastQC  
699 v0.11.8, seqkit v0.11.0 stats and MultiQC v1.7. PacBio reads (GC032883  
700 ~ 235x) were scanned for SmartBells adapters that remained in the .fastq files with the mod-  
701 ule removesmartbells.sh from bbmap v38.73. Then modules rmdup and rename  
702 from seqkit v0.11.0 were used to rename duplicated read names. PacBio reads were sep-  
703 arated in two sets: **Assembly** (AS) and **Polishing** (PS). AS reads were filtered with filtlong  
704 v0.2.0 on length and cropped on quality. Length threshold was set at 42Kbps, quality  
705 thresholds ( $\min\_mean\_q \leq 9$  and  $\min\_window\_q \leq 8$ ) were chosen based on the FastQC  
706 v0.11.8 stats. PS reads were filtered and cropped using the same parameters except for the  
707 length threshold that was set at 5 Kbps. In both AS and PS, the worst 10% reads were dis-  
708 carded. After filtering: 0.3x remained in AS and 208.5x in PS. Oxford Nanopore minion reads  
709 (BXQ\_F ~ 175x) were used only for assembly. Reads were filtered with filtlong v0.2.0  
710 on length and cropped on quality. Length threshold was set at 42Kbps, quality thresholds  
711 ( $\min\_mean\_q \leq 16$  and  $\min\_window\_q \leq 15$ ) were chosen based on the FastQC v0.11.8  
712 stats. Since ~ 114x Illumina HiSeq 2x240bp was also available for this biosample, the 16-mer  
713 spectrum from these reads was used to crop and split long reads using filtlong parameters  
714 split 250 and trim. The worst 10% reads were discarded. After filtering: 38.3x remained.  
715 Illumina HiSeq2500 reads (GC047403 ~ 357x) were not filtered nor cropped. The sequencing  
716 facility (GenomicsCore KUL) provided pre-processed sequences without adapters. Assem-  
717 bly was performed with <https://github.com/Nextomics/NextDenovo>, using AS reads (PacBio  
718 and ONT). Many parameters were adjusted empirically, see the provided configuration file for  
719 details. After this step, all assembly files between each step were sorted and renamed us-  
720 ing the funannotate v1.5.0-12dd8c7 sort module. The raw assembly contained  
721 uncollapsed diploid sequences called haplotigs. Haplots were removed with purge\_dups  
722 v1.0.0 (70) using PacBio PS (GC032883 ~ 208x). The configuration file was modified to  
723 skip the facultative BUSCO and KCM steps. The obtained purged assembly was self aligned  
724 with minimap2 v2.17-r941 -DP (71) and then visualized with dgenies (72) to find  
725 haplotype alignments. Haplots were not visible compared to the unpurged assembly. The  
726 purged assembly was then polished using HyPo v1.0.3 (73) with Illumina (GC047403 ~  
727 357x) and PacBio reads from the polishing set (GC032883 ~ 208x). PacBio long reads align-  
728 ments were performed using minimap2 v2.17-r941. Illumina reads were mapped using  
729 bwa mem v0.7.17-r1188 (74) and bam indexing and sorting was done with samtools  
730 v1.10 (75). Illumina coverage estimates for HyPo was done using the average coverage of the  
731 mapping file computed with sambamba v0.6.8 depth base (76) (106x) and expected  
732 haploid genome size (96m) was set based on the last available flow cytometry genome size  
733 estimations results.

734 **Diploid assembly: FALCON** The *de novo* assembly of *Adineta vaga* genome was carried out  
735 with diploid-aware long-read assembler FALCON version 0.7.0, FALCON-Unzip and partial

736 FALCON-Phase (only FALCON-Phase Workflow steps 1, 2 and 3) (24). Prior to the assem-  
737 bly, Canu error correction module (77) was used for read error correction based on raw PacBio  
738 reads. The FALCON software is highly optimised for eukaryote genomes, and uses hierarchical  
739 genome assembly process (HGAP). More specifically, reads longer than 15 kb were selected by  
740 Falcon as "seed" reads to generate consensus sequences with high accuracy. The pre-assembly  
741 steps in FALCON uses DALigner to do all-by-all alignments of the corrected PacBio reads.  
742 Long reads were then trimmed at regions of low coverage with FALCON sense parameters (-  
743 minidt 0.70 -mincov 4 -maxnread 200) and sensitive DALigner parameters were selected (-h60  
744 -e.96 -l500 -s1000) for pre-assembly process. The FALCON pre-assembly resulted in 331 pri-  
745 mary contigs of total length 125 Mb, contig N50 of 6 Mb and an additional 36 Mb of "associate  
746 contigs" that represent divergent haplotypes in the genome. FALCON-unzip was then used to  
747 phase the pre-assembly, producing contiguous leading contigs (named "primary") and associ-  
748 ated contigs (i.e. phased, alternate haplotypes). The genome assembly was polished as part of  
749 the FALCON-Unzip pipeline using haplotype-phased reads. The haplotigs contain one of the  
750 two allelic copies of the heterozygous regions; in this respect, the haplotigs serve as phasing in-  
751 formation for the haploid representation. The FALCON-Unzip assembly had 241 primary con-  
752 tigs and 999 haplotigs. FALCON-Phase (<https://github.com/phasegenomics/FALCON-Phase>)  
753 was developed to resolve haplotype switching in diploid genome assemblies. The FALCON-  
754 Phase haplotig placement defines phased blocks in the FALCON-Unzip assembly. The Falcon-  
755 Phase Workflow steps 1 and 2 were used to place the haplotigs along primary contigs. Once the  
756 haplotig placement file and phase block pairings are done, the primary contigs are cut up into  
757 very small pieces at phase block boundaries with Falcon-phase workflow step 3.

758 **Assemblies scaffolding: instaGRAAL** Hi-C contact maps were generated from paired-end  
759 reads using the hicstuff pipeline (78) for processing generic 3C data, available at  
760 <https://github.com/koszullab/hicstuff>. The backend uses bowtie2 (79) in paired-end mode (with  
761 the following options: -{}-maxins 5 -{}-very-sensitive-local). We discarded alignments with  
762 mapping quality lower than 30. The remainder was converted to a sparse matrix representing  
763 contacts between each pair of DpnII restriction fragments. The instaGRAAL program (25)  
764 was used in conjunction with the contact maps to scaffold the genomes. Prior to running it,  
765 restriction fragments are filtered based on their size and total coverage. Fragments shorter than  
766 fifty base pairs are discarded. Then, fragments with coverage lesser than one standard deviation  
767 below the mean of the global coverage distribution are also removed from the initial contact  
768 map. These fragments were reintegrated later after the scaffolding step. The instaGRAAL  
769 scaffolder uses a Markov Chain Monte Carlo (MCMC) method: briefly, the contact data is  
770 fitted on a simple three-parameter polymer model. The 3D contacts are exploited and used by  
771 the program to infer the relative 1D positions of the sequences and thus the genome structure.  
772 To do so, the program attempts to perform a number of operations between each sequence and  
773 one of its neighbours (e.g. flipping, swapping, merging or splitting contigs) and the operation  
774 is either accepted or rejected with a certain probability depending on the likelihood shift. The  
775 model parameters are then also updated and a new iteration begins. A set of computations

776 whereby every sequence of the genome has been iterated over this way is called a *cycle*. The  
777 scaffolder was run for 100 cycles on the phased and the diploid genome and was run for 50  
778 cycles on the AV20 haploid genome, after which convergence in both genome structure and  
779 model parameters was evidently apparent. The scaffolded assemblies were then refined using  
780 instaGRAAL’s instaPolish module, with the aim of correcting the small artefactual inversions  
781 sometimes produced by instaGRAAL. The resulting contact map can be seen in Supp. Fig. 2.

782 **Post-treatment of scaffolded assemblies** Post-treatment of the diploid assembly (Falcon):  
783 we used the repeat-aware finisherSC tool (80) to upgrade the *de novo* phased genome assem-  
784 bly of *Adineta vaga*. Final round of polishing were performed with the Pilon corrector using  
785 Illumina data (sample ID GC047403, see Supp. Table 1). Post-treatment of the phased as-  
786 semby (Bwise): : to resolve a remaining fragmentation of one single chromosome (i.e. chro-  
787 mosome 5B) after scaffolding with instaGRAAL based on Hi-C data, we established a novel  
788 comparative approach that incorporates computational methods to transform fragmented con-  
789 tigs into near-chromosome fragments. First, Bwise contigs were aligned against themselves  
790 using NUCmer v4.0 (81). Ploidy pairing was evaluated using the online visualization tool,  
791 DOT (<https://dnanexus.github.io/dot/>) and we were able to anchor fragmented contigs into a  
792 single chromosome using its homologous template (i.e. chromosome 5A).

## 793 AV20 genome annotation

794 **Transposable elements annotation** TE-like elements, including transposable elements (TEs),  
795 were predicted using a combination of two complementary tools: EDTA v1.7.8 (82) and TEde-  
796 novo (part of the REPET pipeline) (83, 84). The former relies on structure-based programs  
797 allowing for the detection of even single-copy elements, while the latter relies on sequence re-  
798 peatedness. The TE-like elements consensus sequences they both produced were then merged  
799 and subsequently filtered by performing a basic annotation of the genome with TEannot from  
800 the REPET pipeline, and retrieving only consensus sequences with at least one full length copy  
801 annotated and discarding sequences corresponding to potential host genes. The 521 retained  
802 consensus sequences (293 from EDTA, 124 from TEedenovo) were then used as input for the  
803 subsequent genome annotation with TEannot. This resulted in a draft annotation of 8,590 TE-  
804 like elements covering 6.57% of the genome. A series of filters were then applied to these  
805 annotations using in-house script: i) conserving only retro-transposons and DNA-transposons;  
806 ii) with minimal copy length of 250 bp; iii) with minimum identity with consensus of 85%;  
807 iv) with a minimal proportion of the consensus overlapped of 33%; v) resolving overlapping  
808 annotation. These filtering steps resulted in a final annotation of 841 putative canonical TEs  
809 covering 1.98% of the genome. Proportions of TE-like sequences and TEs are shown in Supp.  
810 Fig. 9.

811 **Gene annotation** Gene prediction and annotation of AV20 genome were done according to  
812 current integrative approaches based on several independent lines of evidence. We first dis-

813 carded scaffolds shorter than 1000 bp using funannotate clean function (85). Repeats in the  
814 genome were then soft-masked using bedtools (86) maskfasta function using the draft annota-  
815 tion of repeated elements as described above. RNA-Seq data from several cultured clones (see  
816 Supp. Table 1) were used to produce *de novo* a transcriptomic assembly with trimmomatic (87)  
817 and trinity (88) both under default parameters. This transcriptomic assembly as well as addi-  
818 tional RNA-Seq data directly mapping on the genome (see Supp. Table 2) were used as input  
819 for the funannotate train function that wrap the PASA pipeline (89) which relies on RNA-Seq  
820 to produce high quality annotations. Then, we used a combination of PASA annotations, *de*  
821 *novo* assembled transcriptome, metazoan BUSCO database and the proteic Uniprot database  
822 within funannotate predict function. This first produced *ab initio* predictions using Genemark-  
823 ES (90), which were then used along with transcripts and proteic data to train Augustus to  
824 generate a second set of annotations. Lastly, it used Evidence Modeler as a weighted approach  
825 to combine annotations from PASA, Genemark and high quality predictions from Augustus into  
826 an integrated gene annotation set. We then used InterProScan5 in order to produce functional  
827 annotations to the predicted genes (91) which were then used in combination with busco meta-  
828 zoan database using the funannotate annotate function with default parameters. In addition,  
829 Ribosomal RNA genes were predicted from the AV20 genome assembly using barrnap (92)  
830 (parameters: –kingdom euk). Note that the number of genes annotated differs greatly from  
831 the number of genes annotated previously (9). This is mainly due to the structure of the two  
832 genome assemblies: the AV13 genome was phased (many pairs of annotated genes correspond  
833 to alleles) while the genome assembly we present here is haploid.

834 **Detecting HGTc with a new tool: Alienomics** We used a newly developed tool, named  
835 Alienomics, in order to detect Horizontal gene transfers candidate (HGTc). This tool is be-  
836 ing submitted and described in detail elsewhere. Briefly, its approach first integrates several  
837 lines of quantitative evidence into a score for every predicted gene. This gene score is based  
838 on several blast results (i.e. against Uniref50 database, a user-defined set of closely-related  
839 reference genomes, bacterial rRNA database, BUSCO database) as well as on read coverage  
840 and GC content. It represents how "alien" or "self" a given gene is. Note that when con-  
841 sidering blast results the taxonomy of multiple best-hits are parsed and evaluated in order to  
842 determine whether the query origin belong to "self" or to "alien". We then superimpose this  
843 qualitative synteny information to the quantitative gene score in order to discriminate if alien  
844 genes stemmed from contaminant or from HGT. For this, scaffolds are being given a score  
845 based on the integration of all the gene scores, slightly modified using expression level based  
846 on RNA-Seq (in order not to penalize scaffolds including many true HGTs). This scaffold score  
847 represents whether it originated from a contaminant or from the genome under study. Synten-  
848 y is then taken into account by comparing gene scores to their respective scaffold scores to  
849 validate a HGTc. For example, an "alien" gene on a "self" scaffold corresponds to a HGTc  
850 while an "alien" gene on an "alien" scaffold is a contaminant. Alienomics is available here:  
851 <https://github.com/jnarayan81/Alienomics>. Within Alienomics, results for each criteria (e.g.  
852 blast bitscores, GC content, coverage) are transformed into criteria scores ranging from -1 to

853 +1. Criteria scores from blast results are turned into negative values if the taxon id from the best  
854 representative match among all hits do not belong to a user-defined clade (such as "metazoa").  
855 Gene scores result from the combination of criteria scores and correspond to the hyperbolic tan-  
856 gent of the sum of criteria scores multiplied by a ratio that depends on the number of informative  
857 criteria (e.g. number of criteria for which the value is different from "0"). Scaffold scores result  
858 from the combination of gene scores (with the addition of expression score based on RNA-Seq  
859 data) and correspond to the hyperbolic tangent of the sum of gene scores multiplied by the  
860 square root of the number of genes and normalized by gene lengths. Coverage information  
861 was computed from raw ONT reads using minimap2 (parameters as follows: -ax map-ont -c  
862 -Y). Alienomics was run here under the following parameters: level\_up to = metazoa; gc\_filter  
863 = 26:38 ; evalue = 1e-01; qcovper = 0; bitscoreCutoff = 150; coverage = 100; ignoretaxid =  
864 104782—10195—96448—249248—1813166—104781—4513—112509—9606—7574—42192—29159—283  
865 HGTC were categorized as such under the following default thresholds: genescore = 0.5; scaf-  
866 foldscore = 0.5.

867 **Endogenous viral elements detection** Sequences showing similarity to viral genes were  
868 searched in the AV20 genome assembly by performing a diamond (93) blastx search (options: –  
869 max-target-seqs 1 –range-culling –min-score 40 –more-sensitive -F 15) using AV20 scaffolds as  
870 queries all viral proteins extracted from the nr database of NCBI (February 2020). Proteins from  
871 two viral families were excluded from this database (i.e. Retroviridae and Hepadnaviridae) to  
872 avoid false-positive blast hits corresponding to the reverse-transcriptases of *A. vaga* retrotrans-  
873 posons. All *A. vaga* sequences showing similarity to a viral sequence were then used as queries  
874 to perform a reciprocal diamond blastx search against the entire NCBI nr protein database. All  
875 sequences aligning with a higher score to a viral sequence than to a non-viral sequence were  
876 annotated as viral-like sequence.

## 877 **AV20 genome analyses**

878 **Ploidy, synteny and colinearity among the three *A. vaga* genome assemblies** Genome as-  
879 sembly tools rely on various assumptions including the ploidy level of the organism under  
880 study. In order to circumvent potential impact of such ploidy assumptions on genome struc-  
881 ture, we compared our three new genome assemblies. First we evaluated the classical genome  
882 assembly statistics using in-house script (see Figure 1A). We then used the illumina reads (i.e.  
883 GC047403, see Supplementary Table 1, as input for the *comp* function of the KAT software (94)  
884 which uses k-mers distribution in order to explore ploidy levels of *A. vaga* genomes (see 1A).  
885 Genomes were aligned pairwise using nucmer 3.1 (using –maxmatch option) (95), the results  
886 of which were converted into paf format using minimap2 paftools script (71). We then used the  
887 online tool D-GENIES (72) to visualise the three pairwise alignment as dotplots (see Supp. Fig.  
888 3, 4 and 5).

889 **Read depth and heterozygosity** Average coverage on AV20 genome assembly was computed  
890 independently for the Illumina PE reads, ONT reads and PacBio reads, on 100 Kbs windows.  
891 The mapping was performed with bwa mem 0.7.17 (74) on default settings for the short-reads  
892 reads. Heterozygosity analysis was performed using GATK 4.1.0.0 (96) on Illumina PE reads  
893 for genotyping all sites (HaplotypeCaller function with -ERC GVCF option). This was done  
894 for all samples analyzed (i.e. GC047403, BXQF, ERR321927). The resulting gvcf files were  
895 combined (CombineGVCFs function) and were then jointly genotyped (GenotypeGVCFs func-  
896 tion) (97). Distribution of heterozygous sites are shown on Figure 2A.

897 **Homoeologous divergence** The self alignment of AV20 genomic sequences obtained using  
898 nucmer 3.1 (-maxmatch option, see methods on synteny and colinearity above) (95) was re-  
899 used in order to evaluate the genomic divergence between homoeologous chromosomes. The  
900 paf alignment file was filtered using custom script in order to only retain genomic alignments  
901 between homoeologous regions ranging from 500 to 10,000 bp (see Supp. Fig 8). Note that  
902 the divergence between homoeologues (and alleles) measured here are much lower than the  
903 measures previously reported (9). This is because we aligned genomic regions at the nucleotide  
904 level using nucmer. On the contrary, previous study aligned CDS at the proteic level using MC-  
905 ScanX which then guided corresponding alignment at the nucleotide level, producing additional  
906 indels due to the existence of frameshifts.

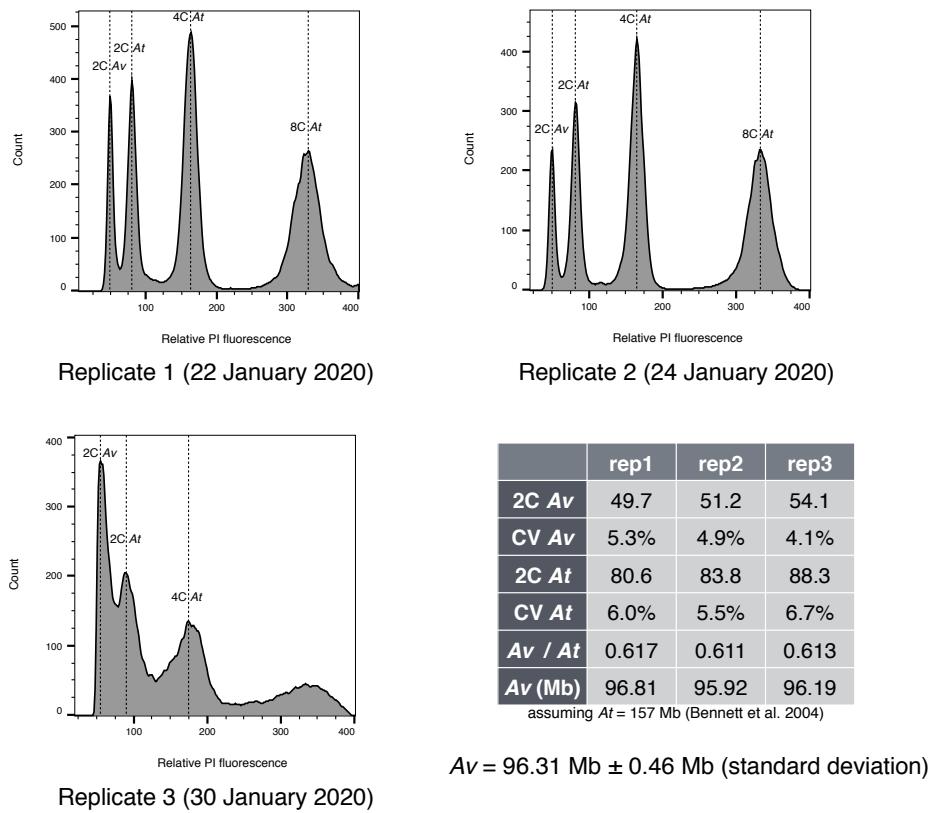
907 **Detecting gene synteny** Protein sequences from annotated genes were used as input for MC-  
908 ScanX (98) in order to detect blocks of gene synteny (parameters: -s 5 -b 1). Home-made  
909 script was used in order to only retain genes that composed syntenic blocks between homoeo-  
910 ologous chromosomes. Among the 31,582 annotated proteins in *A. vaga*, 9,726 of them had  
911 a proteic counterpart in a synteny block on their respective homoeologous chromosome (i.e.  
912 30.79% of proteins). Note that the number of gene existing prior to the tetraploidization of  
913 the genome is very likely larger than this estimate, as any gene loss, translocation or structural  
914 re-arrangements would break gene synteny. All synteny blocks are depicted as grey links on  
915 Figure 1B. The same procedure was followed to detect colinear blocks of synteny using only  
916 the 2,679 HGTC (corresponding synteny blocks are depicted as colored links on Figure 3).

917 **Gene enrichment analyses** GO terms from functional annotation of the haploid genome  
918 were extracted from gene annotation (see gene annotation section above). The 2,422 recent  
919 HGTC and the 257 ancient HGTC were respectively compared to the entire gene set of the AV20  
920 genome containing 32,378 proteins. Enrichment analyses were performed using topGO pack-  
921 age with a fisher test and the "elim" algorithm (99). Results are presented in Supplementary  
922 Table 2.

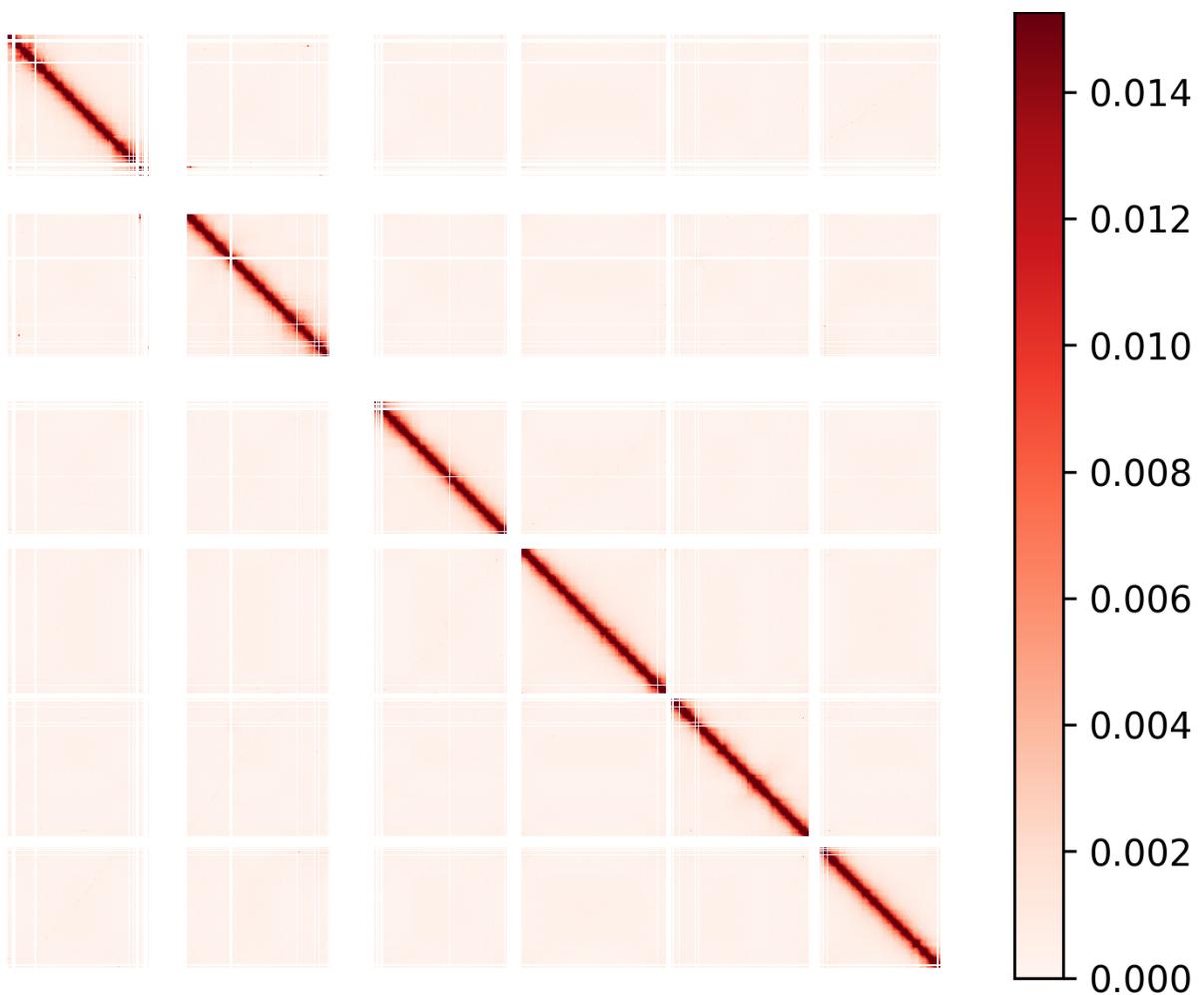
923 **Re-appraisal of the AV13 Genome**

924 **Investigating AV13 Breakpoints** The previously identified synteny breakpoints in the genome  
925 of *A. vaga* 2013 (AV13 (9)) were verified by mapping the ONT reads (median size: 4,149 kb;  
926 max size: 353,147 kb) produced in this study onto the AV13 genome according to the following  
927 procedure: i) ONT reads were filtered with Porechop (100) to discard long reads containing  
928 adapters. This discarded 1,202 out of the 1,634,477 reads; ii) Reads were mapped onto the  
929 AV13 genome using NGMLR (101) with default parameters. This tool was selected for its  
930 accuracy when aligning long reads in a context of structural variation; iii) The scaffold of inter-  
931 est (i.e. scaffold1 from AV13) was aligned against the rest of the AV13 genome using Sibelia  
932 v3.0.7 (102) with the following parameters: '-s loose -m 10000 --gff'. iv) The new  
933 AV20 haploid genome assembly was aligned against the AV13 genome using the same proce-  
934 dure as in the previous step; v) Synteny block from Sibelia were used to determine the genomic  
935 windows containing the putative breakpoints described previously (9). These regions were man-  
936 ually screened using Tablet (103) to visualize the alignment of ONT reads. We notably checked  
937 for the presence of clipped regions. Every window contained at least one clipped region (i.e. a  
938 position that is not supported by a single long read) which we reported as screenshots in Supp.  
939 Fig. 6.

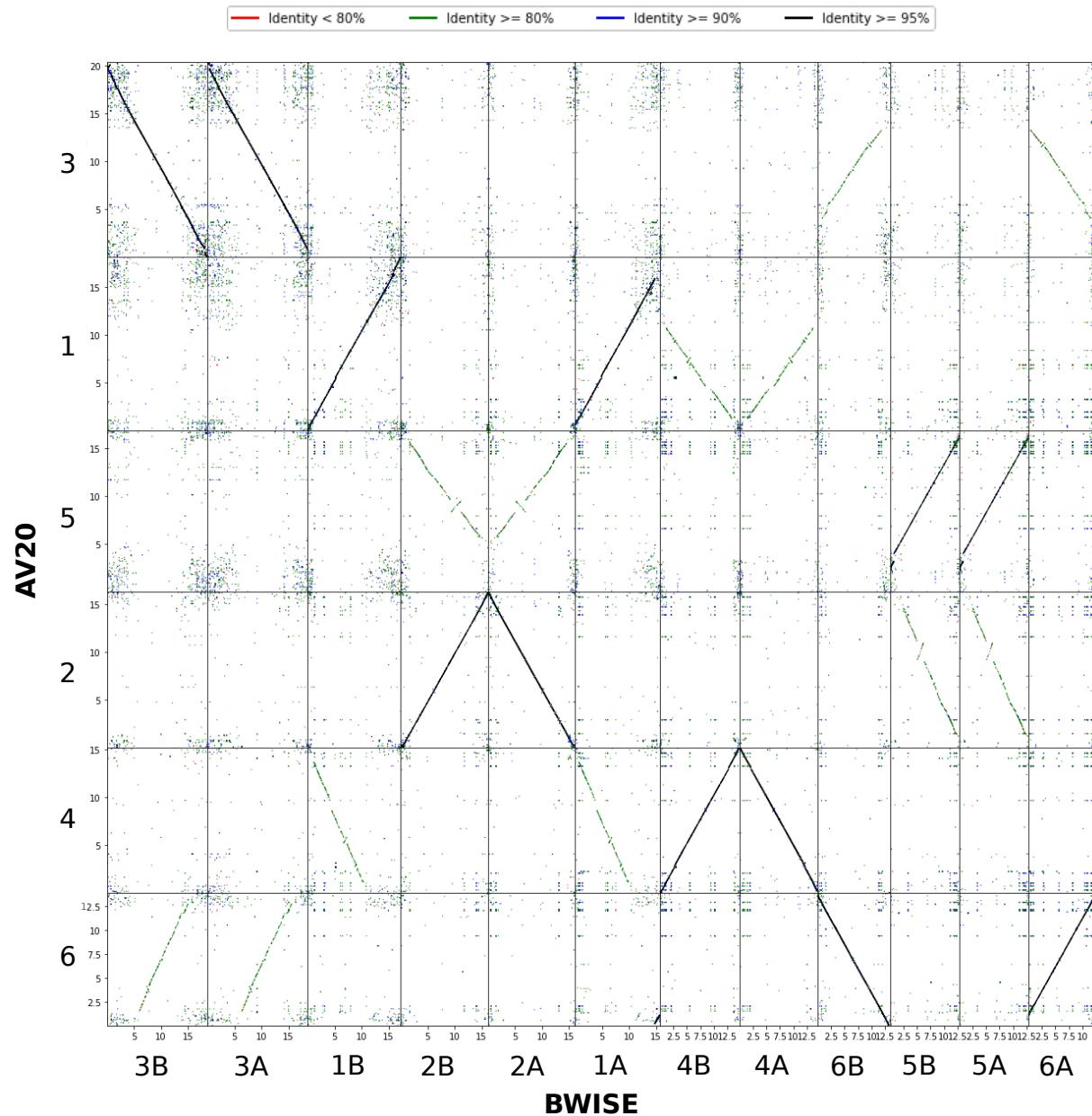
940 **Investigating AV13 Palindromes** Palindromes previously reported in the AV13 genome were  
941 investigated in the light of our new AV20 assembly. We first *de novo* determined the location  
942 of palindromes in AV13 by filtering ONT long reads, mapping them onto AV13 genome using  
943 NGMLR (101) with default parameters and subsequently detecting the palindromic breakpoints  
944 (PBR) using a in-house tool, huntPalindrome (available at  
945 <https://github.com/jnarayan81/huntPalindrome>). Each PBR location was extended by 2.5 kbp  
946 on both sides to produce PBR windows within which we checked for clipped long reads us-  
947 ing in-house script. Additionally, we used the alignment between AV13 and AV20 genomes (as  
948 described in the previous paragraph) to show how these 20 palindromes from AV13 were assem-  
949 bled in AV20 (see Supp. Fig. 7). All these palindromes were collapsed into non-palindromic  
950 regions in the new AV20 genome assembly.



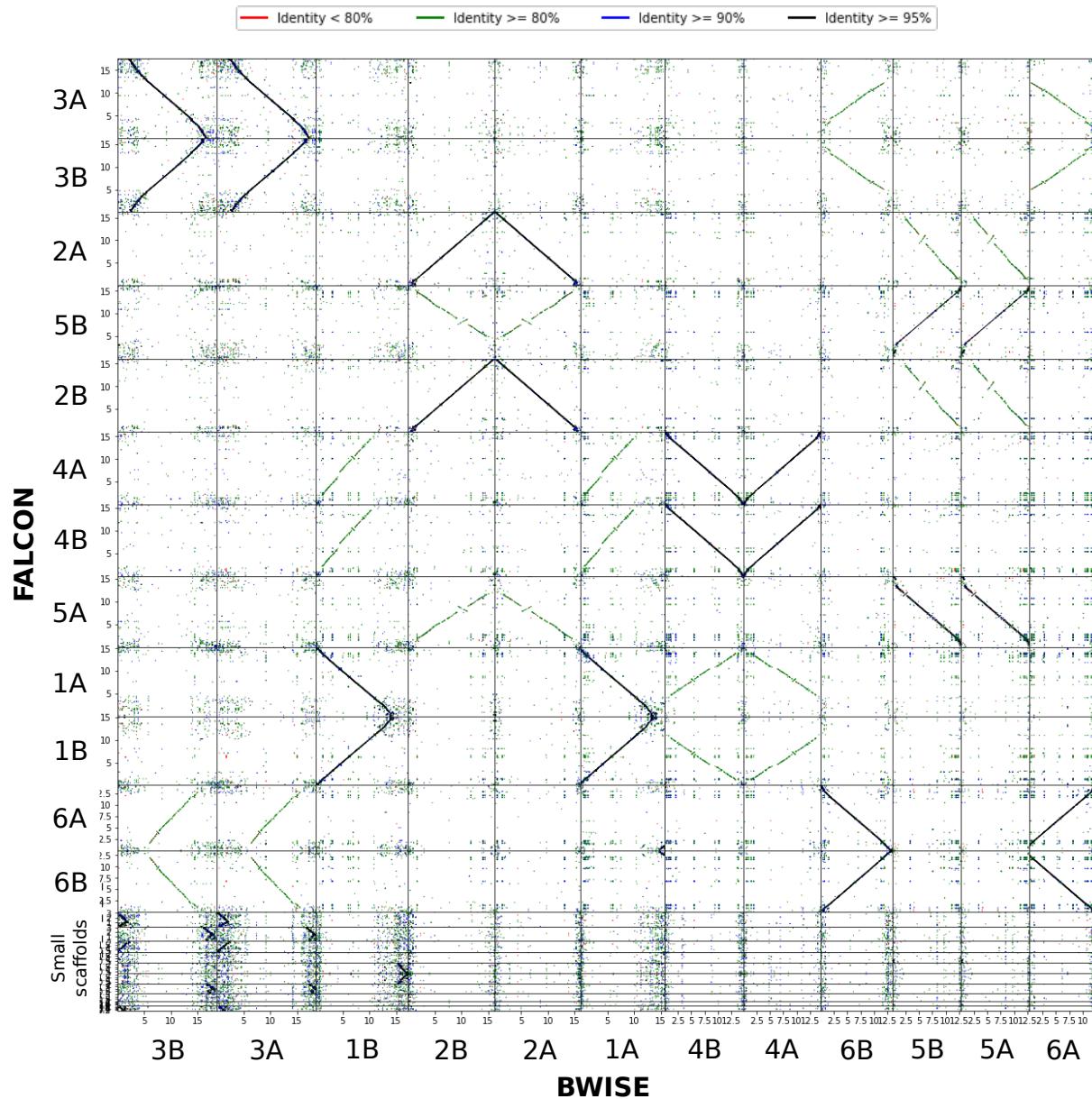
Supplementary Figure S1: *Genome size estimation.* Flow cytometry measurement of the genome size of *Adineta vaga* (*Av*) by comparison to *Arabidopsis thaliana* cultivar Colombia (*At*). Genome size length of *A. thaliana* 1C is 157 Mbp. Assuming *A. vaga* is diploid, the ratio between the two species (i.e. *Av/At*) is about 0.61, leading to the estimation that *A. vaga* 1C genome size is 96.3 Mbp



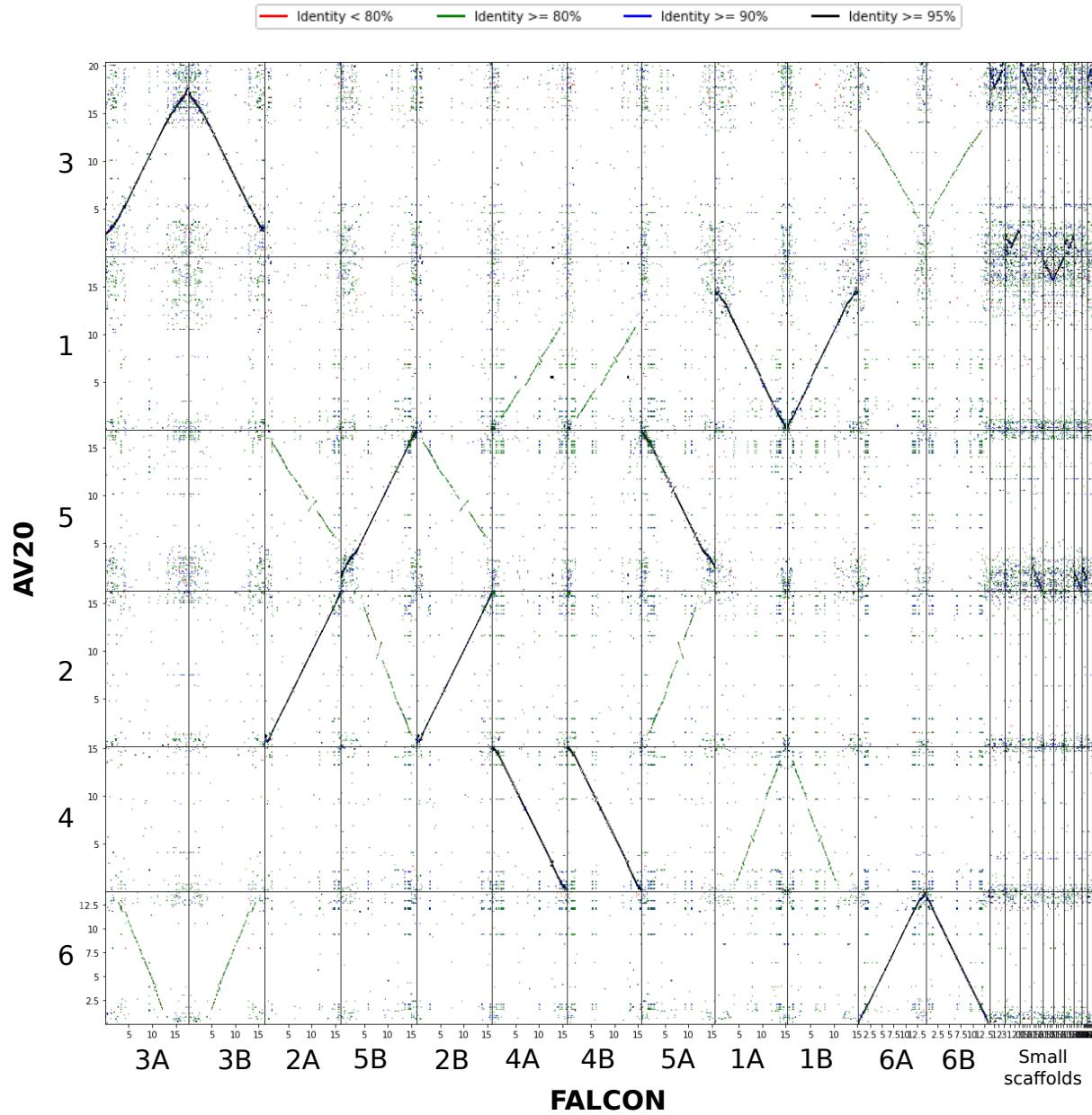
Supplementary Figure S2: *AV20 contact map*. Proximity ligation sequencing data (Hi-C) contact map on AV20 assembly after scaffolding using instaGRAAL and instapolish.



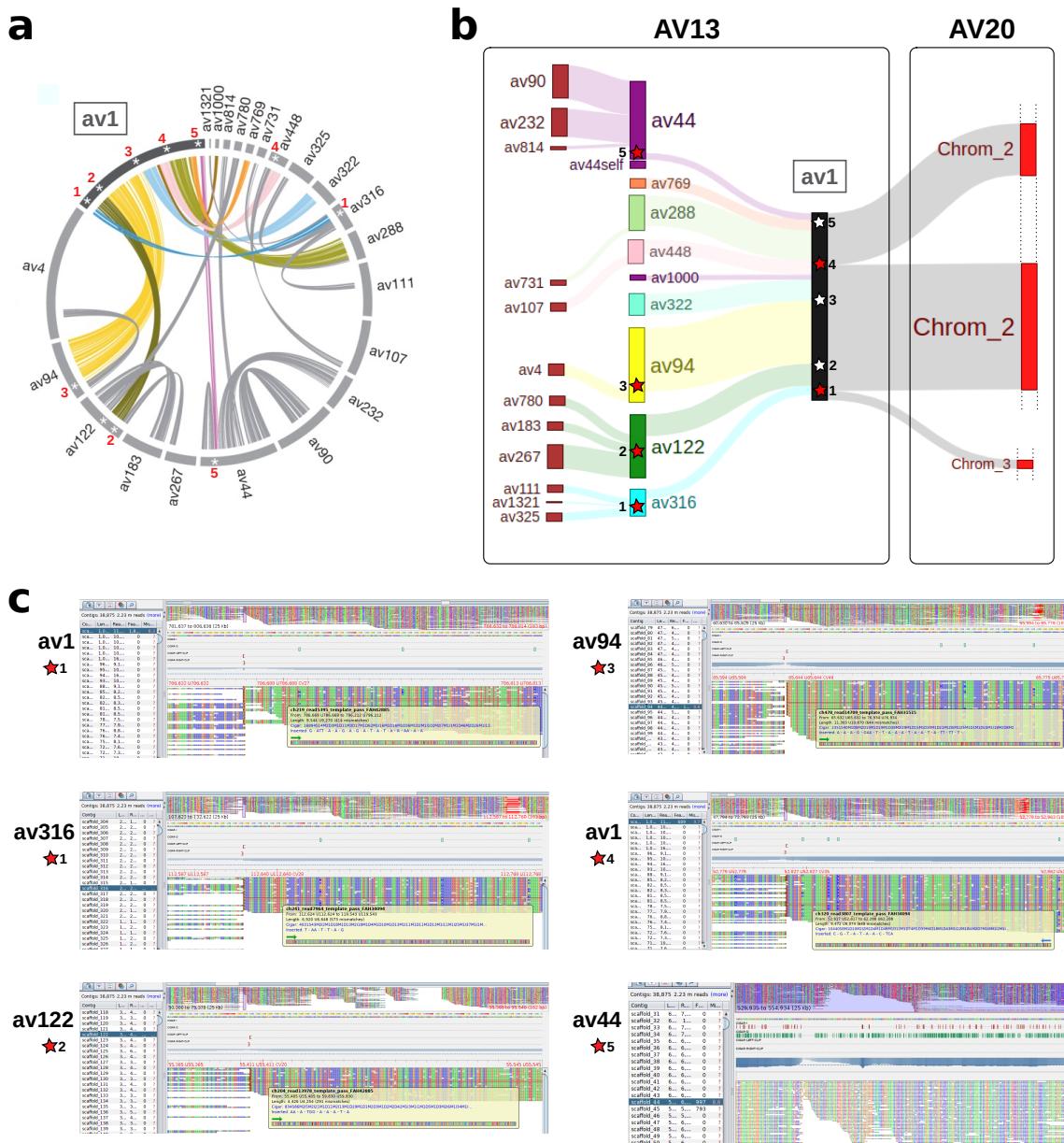
Supplementary Figure S3: *Haplid versus phased genome dotplot*. Pairwise alignment of the haploid assembly (AV20) against the phased assembly (B-WISE), visualised using D-GENIES.



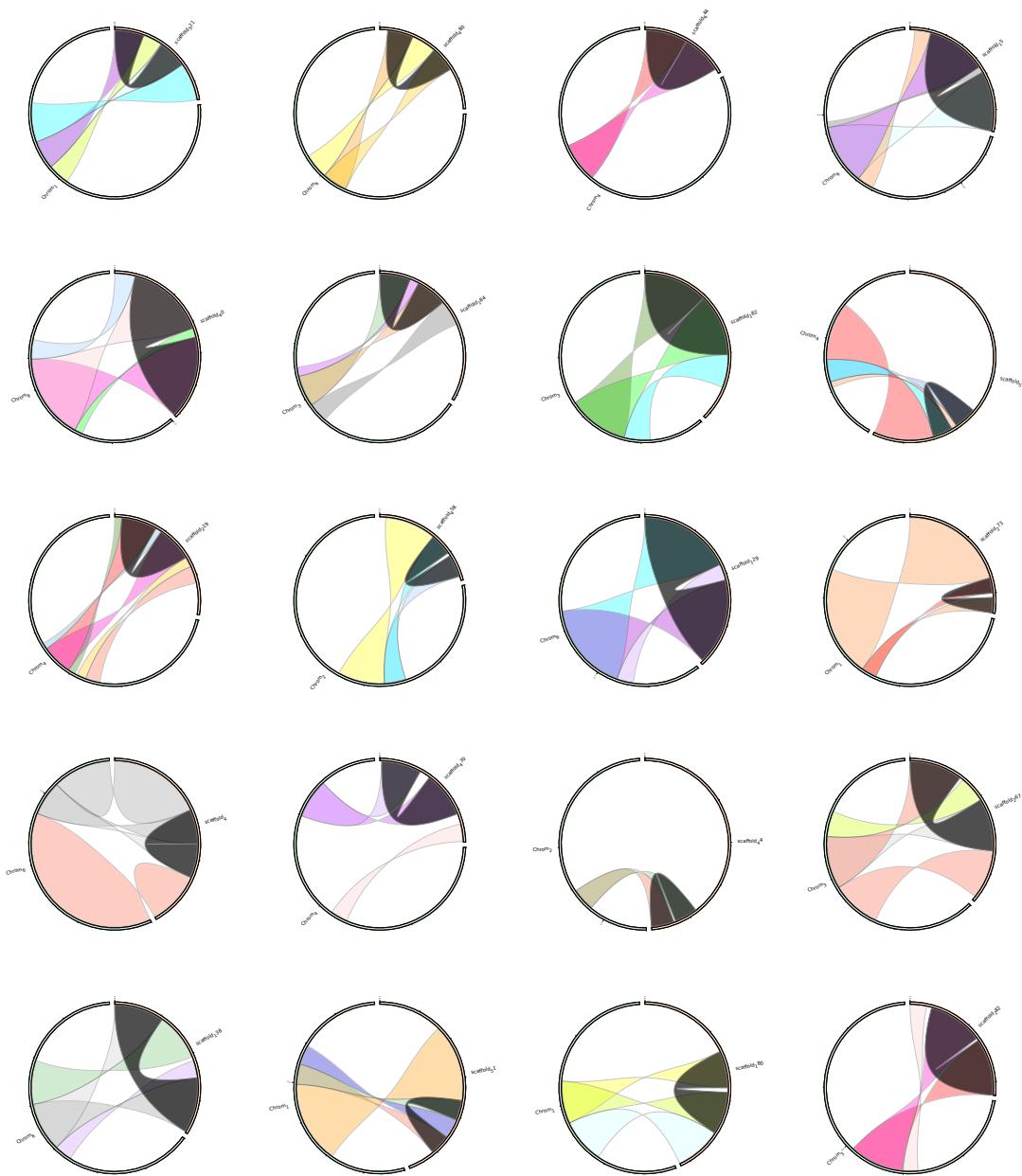
Supplementary Figure S4: *diploid versus phased genome dotplot*. Pairwise alignment of the diploid assembly (Falcon) against the phased assembly (BWise), visualized using D-GENIES.



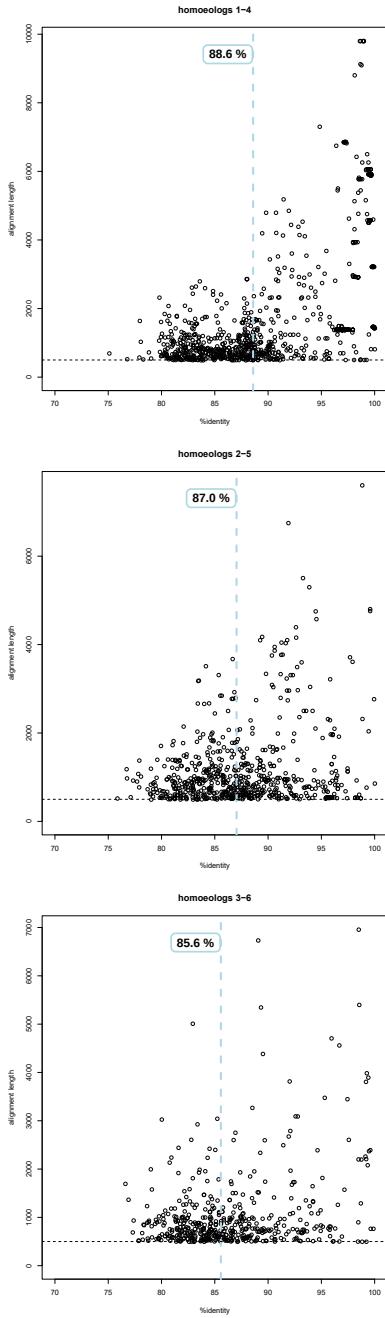
Supplementary Figure S5: *Haplid versus diploid genome dotplot.* Pairwise alignment of the haploid assembly (AV20) against the diploid assembly (FALCON), visualized using D-GENIES.



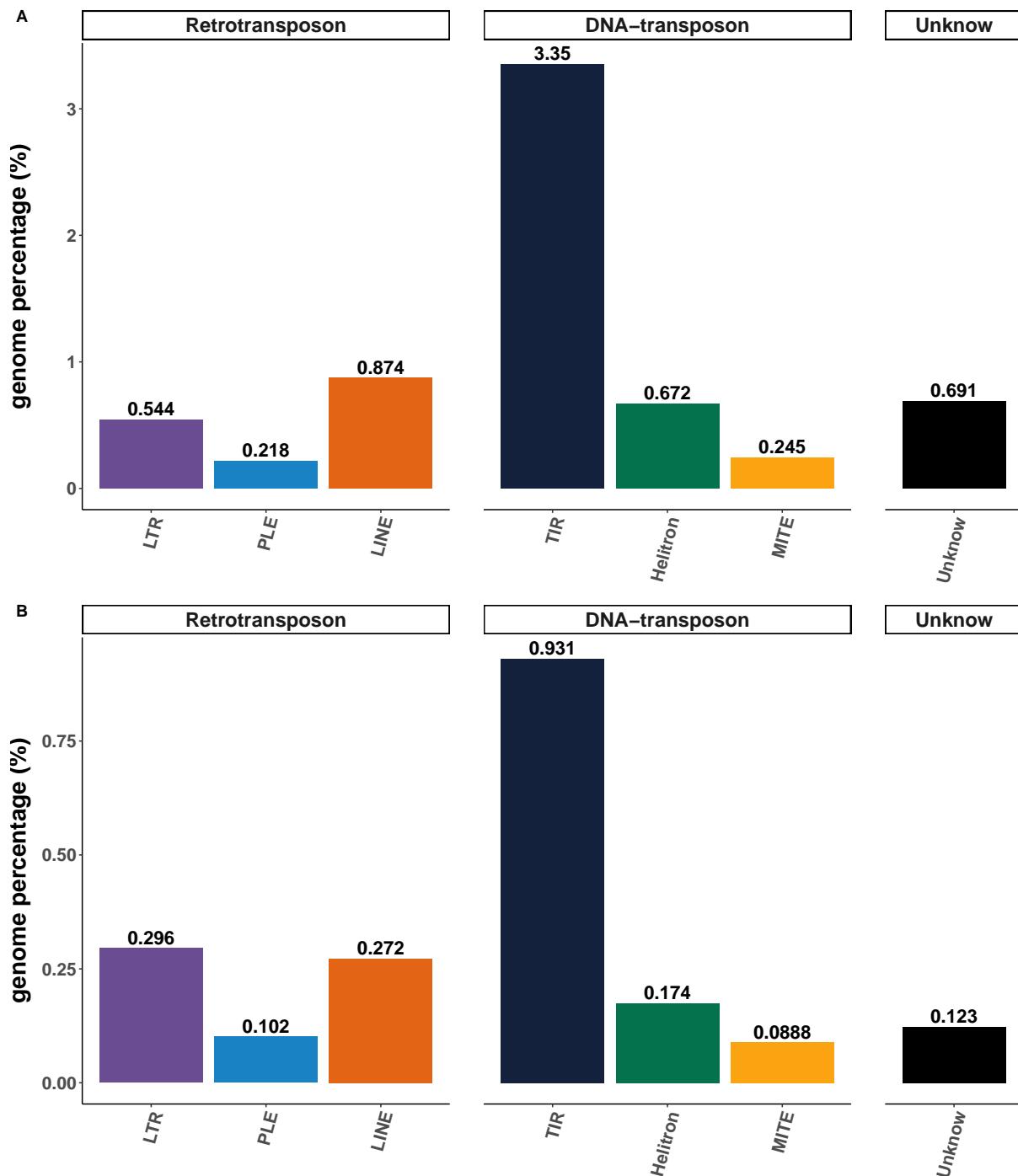
Supplementary Figure S6: *Invalidating AV13 breakpoints*. a) Schematic view of AV13 genome synteny depicting five putative colinear breakpoints on the scaffold av1 and its homologous counterparts (adapted from (9)). b) Schematic view of synteny alignment between the scaffold av1 from AV13 and the new AV20 genome. The 5 putative colinear breakpoints corresponding to panel a are also depicted. Red stars indicate genomic region in AV13 assembly that are not supported by long reads, while white stars indicate regions supported by long reads. Note that regions supported by long-reads in scaffold av1 (white stars 2, 3 and 5) systematically corresponded to a red star in their homologous counterparts in AV13 (red stars on other AV13 scaffolds), indicating that the colinear breakpoint was, in fact, not supported. c) Screenshots of the alignment of long-reads on AV13 assembly (using Tablet) depicting clipped regions at the location of putative colinearity breakpoints. These cases correspond to the red stars depicted on panel b.



**Supplementary Figure S7: Invalidating AV13 palindromes.** Alignment of the AV13 genome assembly (9) against the new AV20 genome assembly shows the total absence of previously reported palindromes. Orange bars represent scaffolds from 2013 assembly and green bars represents chromosomes assembled in the present study. Palindromic regions in 2013 assembly are shown in dark grey.



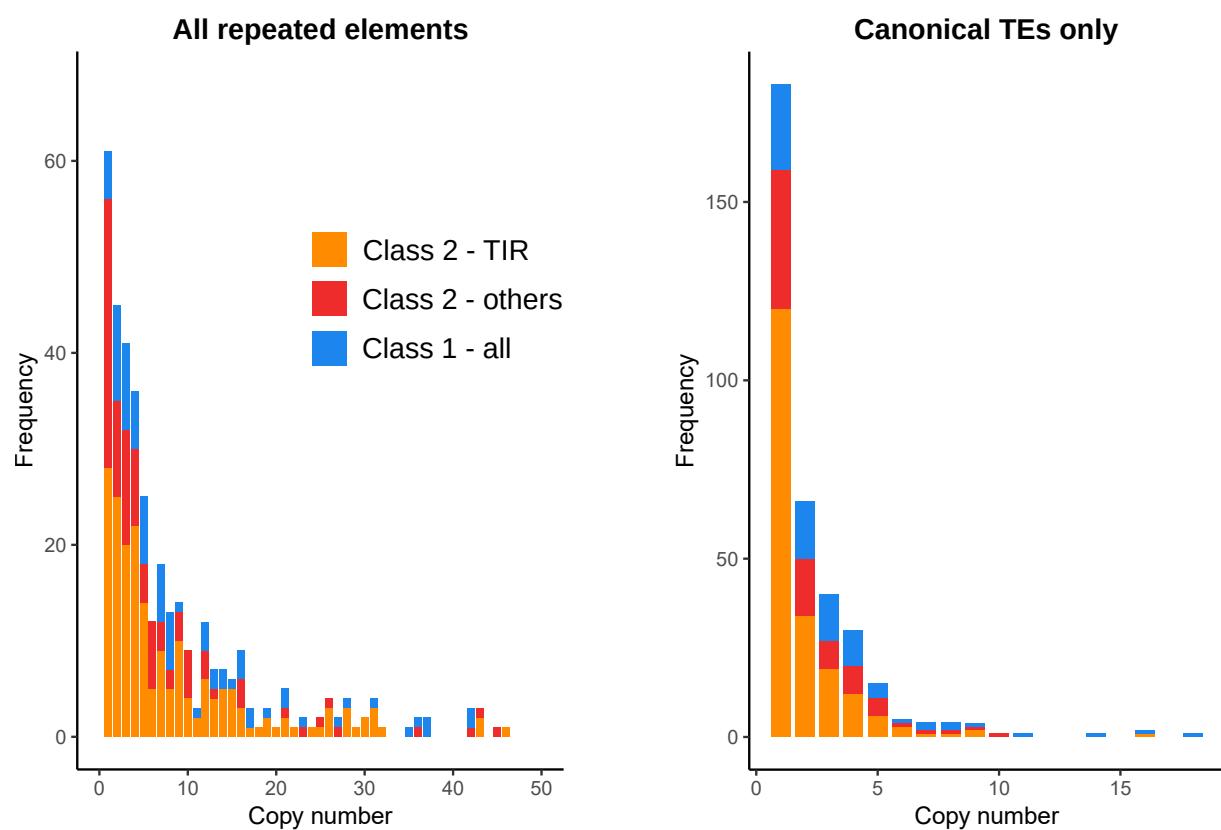
**Supplementary Figure S8: Genomic divergence between homoeologous chromosomes.** Chromosome pairwise alignments length and identity percentage for every pair of homoeologous chromosomes (i.e. chromosomes 1 and 4, 2 and 5, 3 and 6). Median identity percentage between homoeologous chromosome is indicated on each plot (i.e. vertical dotted blue line). Alignments shorter than 500 bp (i.e. horizontal dotted black line) and longer than 10,000 bp were discarded.



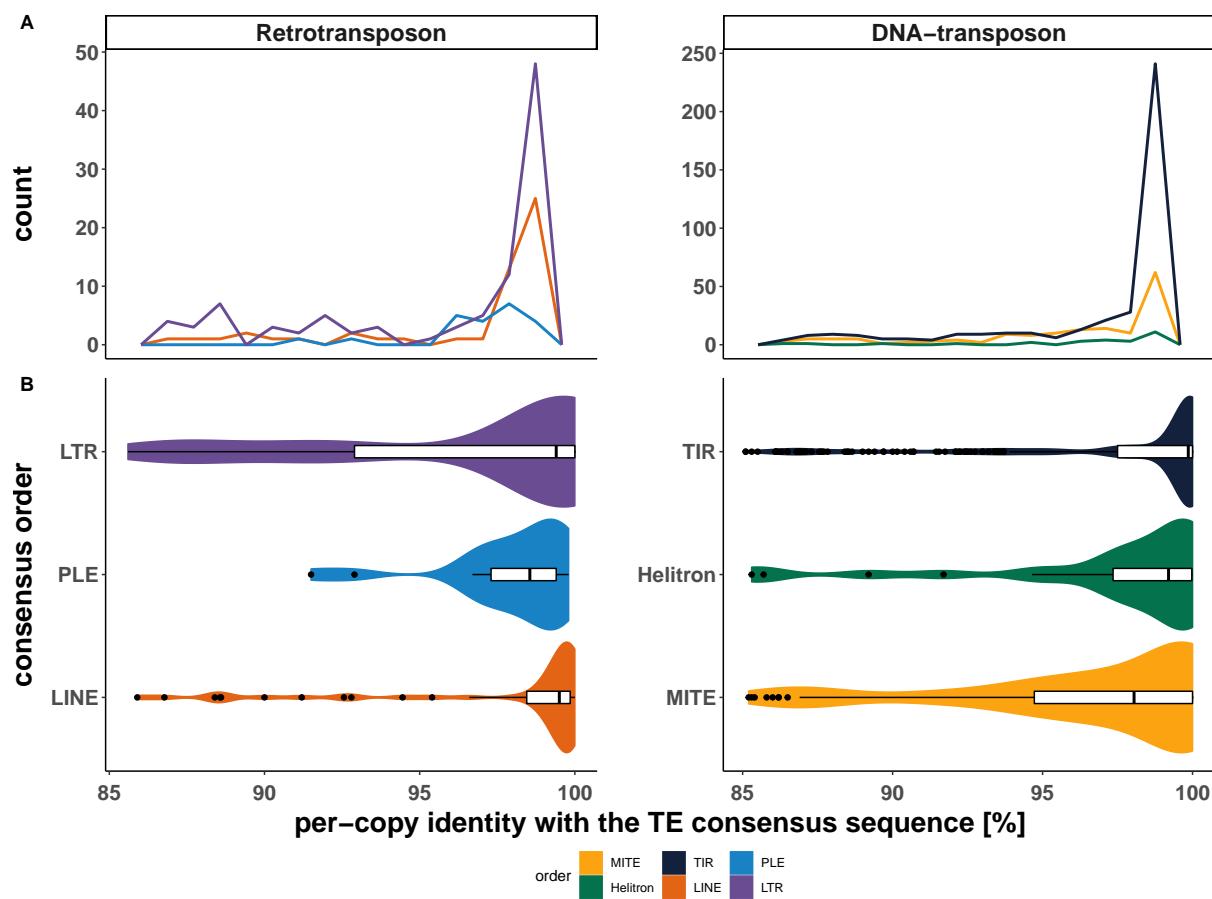
Supplementary Figure S9: *Repeated and transposable elements*. Proportion of the genome covered by each TE order for: a) draft annotation including all repeated elements likely to be related to transposable elements; b) filtered annotation, including only putative canonical TEs.



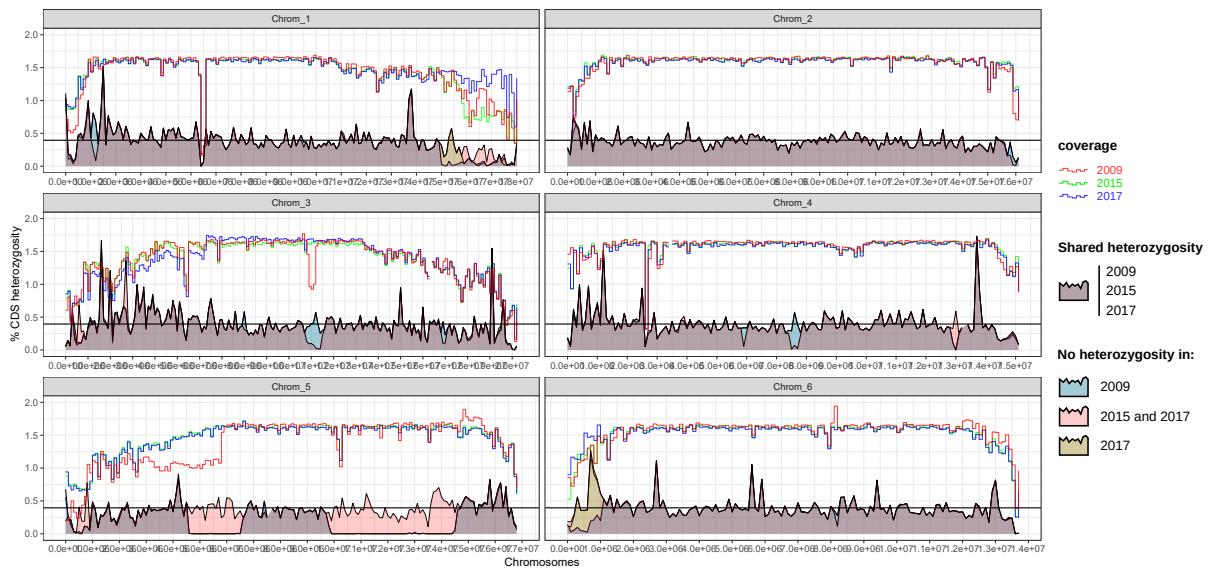
Supplementary Figure S10: *Telomeric repeats and HGTc distribution*. Telomeric repeats (blue) are mostly found in telomeres and subtelomeric regions. Note the small increase of telomeric repeats colocalizing with local hotspot of HGTc (red) on chromosome 4.



Supplementary Figure S11: *Transposable elements per-consensus copy number distribution.*



Supplementary Figure S12: *Transposable elements per-copy identity with consensus.*



Supplementary Figure S13: *CDS Heterozygosity distribution*. Heterozygosity level was normalized by the density of CDS per windows. Sample names and colors as in Figure 2.