Communications Biology

https://doi.org/10.1038/s42003-026-09851-0

Article in Press

Comparative genomics of different haplotypes in *Ditylenchus destructor* provides insights into their host preferences

Received: 12 July 2025

Accepted: 2 March 2026

Cite this article as: Zhao, Z., Zhang, H., Wang, J. *et al.* Comparative genomics of different haplotypes in *Ditylenchus destructor* provides insights into their host preferences. *Commun Biol* (2026). https://doi.org/10.1038/s42003-026-09851-0

Zhengyang Zhao, Hongxia Zhang, Jiayu Wang, Ning Luo, Huixia Li, Runmao Lin & Bingyan Xie

Comparative genomics of different haplotypes in *Ditylenchus destructor* provides insights into their host preferences

Zhengyang Zhao[1, 4], Hongxia Zhang[2, 6], Jiayu Wang[5], Ning Luo[1], Huixia Li[1*], Runmao Lin[3*] & Bingyan Xie[2, 4*]

1 Biocontrol Engineering Laboratory of Crop Diseases and Pests of Gansu Province, College of Plant Protection, Gansu Agricultural University, Lanzhou 730070, China.

2 Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

3 Key Laboratory of Green Prevention and Control of Tropical Plant Diseases and Pests, Ministry of Education, School of Tropical Agriculture and Forestry, Sanya Institute of Breeding and Multiplication, Hainan University, Sanya 572025, China.

4 Center for Biosafety, Chinese Academy of Inspection and Quarantine, Sanya 572024, China.

5 School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan University, Sanya 572025, China.

6 Yongqing County Agriculture and Rural Affairs Bureau, Langfang 065600, China.

* Corresponding author: Huixia Li; E-mail: lihx@gsau.edu.cn, Runmao Lin; E-mail: linrm2010@163.com and Bingyan Xie; E-mail: xiebingyan@caas.cn

Abstract

The nematode *Ditylenchus destructor* comprises multiple haplotypes with distinct host preferences, while the genetic basis remains unclear. We generate three genomes (Haplotypes A, B, and C) using hybrid assembly and conduct comparative analysis with two published Haplotype A genomes. Integrating haplotype-resolved phylogeny, whole-genome alignments, functional annotation, orthogroup profiling, and secretome analysis shows Haplotypes B and C are more similar to each other than to Haplotype A. We identify several key genomic differences that may underlie host adaptation: Haplotype A features expanded chemosensory GPCR repertoires and GH31 glycoside hydrolases. Haplotype B possesses an abundance of cytochrome P450 domain

proteins and secretory pectate lyases. Haplotype C harbors more genes encoding NADPH reductases, oxidoreductases, ABC transporters, secreted animal haem peroxidases, C-type lectins, and Astacins. We propose that these genomic variations facilitate the nematode's adaptation to different host plants. Collectively, our findings establish a genomic framework for understanding host adaptation in *D. destructor*.

Keywords: *Ditylenchus destructor*; haplotype; comparative genomics; secretome; host preferences

Introduction

Plant-parasitic nematodes are among the main causes of plant diseases, causing significant yield losses and serious economic damage every year. Genomic studies of plant-parasitic nematodes play a key role in revealing the molecular mechanisms underlying interactions between plant-parasitic nematodes and their hosts and in the development of nematode-resistant crop varieties [1]. To date, more than 4,100 species of plant-parasitic nematodes have been identified, but genomes of fewer than 90 species have been sequenced, most of which are from root-knot nematodes (*Meloidogyne* spp.), cyst nematodes (*Heterodera* spp. and *Globodera* spp.), and pine-wood nematodes (*Bursaphelenchus* spp.) [2, 3]. Currently, ten genomes have been sequenced and made publicly available for the genus *Ditylenchus*. Five of them are from *Ditylenchus destructor*, a sexually reproducing nematode that parasitizes both plants and fungi. Three of these are newly reported in this study: BazhouSP (Haplotype A), DdB (Haplotype B), and GSAUP (Haplotype C). Three genomes are from *Ditylenchus dipsaci* and two from *Ditylenchus weischeri*, which are specialized plant parasites [4, 5].

The potato rot nematode *D. destructor*, a migratory endoparasitic plant pathogen, is of major scientific and economic importance and is subject to quarantine regulations worldwide [6]. In 1931, *D. destructor* was first reported attacking potatoes on Prince Edward Island, Canada [7]. Currently, *D. destructor* is commonly found in Europe, Asia, Africa, North America, and Oceania. It has a

wide range of hosts, infecting more than 100 plant species [8]. The common host plants include potato, sweet potato, sugar beet, carrot, American ginseng, and several other crops [9, 10, 11]. In the absence of plant hosts, it can also feed on a variety of fungi. The main fungal hosts include *Fusarium* spp., *Botrytis* spp., *Penicillium* spp., *Trichoderma* spp., and *Mucor* spp. [12]. *D. destructor* primarily damages the underground parts of host plants, including tubers, bulbs, rhizomes, and fleshy taproots [13]. Lesions caused by *D. destructor* are frequently secondarily infected by fungi, bacteria, and mites. Such secondary infections are the main cause of storage rot in underground storage organs of nematode-infected plants [14]. Therefore, this nematode causes damage during both the growing season and storage.

ITS-rDNA has been widely used for intraspecific identification of *D. destructor* populations [6, 15, 16]. Previous studies have shown extensive length variation in the ITS-rDNA of *D. destructor* [17, 18]. Secondary structure prediction of ITS1 sequences showed that these tandem repeats folded into a stem-like structure forming helix H9 [16, 17]. Haplotype A lacks helix H9, whereas Haplotypes B and C retain it [10, 16, 18].

D. destructor is a serious threat to a variety of crops, and its survival, parasitic adaptation, infection, and pathogenicity likely rely on distinct strategies and molecular mechanisms, yet few studies have addressed these aspects. Therefore, it is necessary and important to analyze the genetic makeup of *D. destructor* at the genome and transcriptome levels to understand host preference and optimize management strategies. However, due to the small size of plant-parasitic nematodes, their reference genomes are usually assembled from the genetic material of a population rather than a single individual, as in larger organisms [1]. As a result, these reference genomes contain substantial heterozygosity, which makes genome sequencing and assembly of plant-parasitic nematodes more challenging [19, 20].

Recent high-fidelity long-read platforms such as PacBio HiFi [21] can achieve base accuracies greater than 99.9%, yet their cost and data requirements were prohibitive at the time when this project was launched. Given the complexity of nematode genomes, obtaining long reads is crucial

to improving the quality of genome assembly. However, the Illumina platform provides superior base-level accuracy. Thus, combining long-read and Illumina short-read data represents the optimal strategy for nematode genome assembly. A hybrid assembly strategy has been shown to maximize contiguity while restoring single-base accuracy in nematode genomes.

Prior to this study, only two genomes were available: the scaffold-scale Dd01 [22] and the chromosome-scale A1 [23], both from Haplotype A and isolated from sweet potato. Broader surveys indicated that Haplotype A was mainly isolated from sweet potato with a minority of isolates originating from other hosts (including potato, *Angelica sinensis*, *Codonopsis pilosula*, *Dioscorea polystachya*, and other unknown hosts); Haplotype B was found on both sweet potato and potato, with a few additional isolates from *Codonopsis pilosula*; and Haplotype C was most frequently isolated from potato but also included a few isolates from sweet potato and carrot. These data suggested differences in host preference among *D. destructor* haplotypes [10, 11].

To address the lack of high-quality reference genomes for Haplotypes B and C, we generated hybrid genome assemblies and performed refined annotations for three haplotypes. BazhouSP (Haplotype A) and GSAUP (Haplotype C): we employed the well-established PacBio SMRT [24] continuous long-read (CLR) protocol to obtain 20 kb reads and polished them with Illumina HiSeq PE150 data. DdB (Haplotype B): Oxford Nanopore Technologies (ONT) was new at the time, and we used it to generate ultra-long reads, which were then assembled and polished in the same way as BazhouSP (Haplotype A) and GSAUP (Haplotype C). The purpose of introducing ONT was to evaluate its potential gain in nematode genomes and explore new assembly breakthroughs. Through a unified subsequent process (the same version of the assembler, the same round of Illumina-based polishing, and the same annotation pipeline), we obtained highly continuous and comparable genomes for the three haplotypes (BazhouSP, DdB, and GSAUP). The PacBio CLR and Illumina assembly results of BazhouSP provide a technical baseline for Haplotype A that is equivalent to those of Haplotypes B and C, thus facilitating a fair comparison of the genome architecture and gene content of the three haplotypes.

In contrast to previous studies that have primarily focused on sedentary plant-parasitic nematodes such as root-knot and cyst nematodes, this study adopts an integrative, haplotype-resolved perspective to systematically investigate host preference in the migratory endoparasitic nematode *D. destructor*. Through comparative genomics and secretome analyses across haplotypes, we find that Haplotypes B and C are closer to each other than to Haplotype A in phylogeny and genome architecture, and that different haplotypes enhance functional modules consistent with their host preferences. This multilayer comparative framework provides a genomic foundation for dissecting nematode host interactions and informs resistance breeding strategies in root crops.

Results

Phylogenetic analysis, ITS1 secondary structures, and reproductive factor of three *D. destructor* populations

In this study, we collected three distinct populations of *D. destructor*: BazhouSP was isolated from infected sweet potato (*Ipomoea batatas*) in Hebei Province, China, and GSAUP was isolated from infected potato (*Solanum tuberosum*) in Gansu Province, China. Additionally, we included the population DdB, which was previously isolated and maintained by our laboratory from potato in Gansu Province (GenBank accession MW522604.1) [18]. PCR amplification of ITS-rDNA yielded fragments of 727 bp (BazhouSP) and 915 bp (DdB and GSAUP). The nucleotide sequences obtained from BazhouSP and GSAUP were deposited in GenBank under accession numbers OL982746.1 and OL982747.1, respectively.

Phylogenetic analysis of ITS-rDNA sequences identified two well-supported clades within *D. destructor*. Haplotype A (e.g., BazhouSP) formed one distinct lineage, whereas Haplotypes B (DdB) and C (GSAUP) clustered in a separate clade (Fig. 1a). These results indicate a clear phylogenetic separation between Haplotype A and the B and C clade. Sequence alignment showed that DdB and GSAUP carry an additional ITS1 insertion composed of repetitive elements, which is absent in BazhouSP (Fig. 1c–e). This pattern was consistent with previous findings [16], which

distinguished two major *D. destructor* clades by the presence or absence of ITS1 repetitive elements. In our study, samples with the insertion (DdB and GSAUP) were isolated from potato, while this insertion was absent in BazhouSP, which was isolated from sweet potato (Fig. 1b), suggesting a potential correlation between ITS1 and host preference.

To classify ITS types precisely, ITS1 secondary structures were predicted using UNAFold [25] and refined for visualization in VARNA [26]. Adjusted minimum free energy values were –89.88 kcal mol^{-1} (BazhouSP), –160.08 kcal mol^{-1} (DdB), and –155.08 kcal mol^{-1} (GSAUP). Despite general structural conservation, key differences were identified, notably the presence or absence of helix H9. BazhouSP lacked helix H9 (Fig. 1c), consistent with ITS1 Haplotype A. In contrast, GSAUP possessed a 188 bp insertion that forms helix H9 (Fig. 1e), matching ITS1 Haplotype C. DdB (Fig. 1d), previously reported as ITS1 Haplotype B (GenBank MW522604.1) [18], also harbored an insertion that forms helix H9. These structural features allowed us to classify BazhouSP, DdB, and GSAUP as Haplotypes A, B, and C, respectively.

To assess reproductive success, we calculated reproductive factor (Rf) for the three haplotypes on sweet potato and potato. BazhouSP reproduced more successfully on sweet potato, whereas GSAUP reproduced better on potato. DdB showed no significant difference in reproduction between the two hosts (Supplementary Fig. 1).

Genome assembly and functional annotations of three haplotypes of *D. destructor*

For the three *D. destructor* isolates BazhouSP (Haplotype A), DdB (Haplotype B), and GSAUP (Haplotype C), final genome assemblies measured 139.4 Mb, 156.8 Mb, and 120.9 Mb, respectively, with scaffold N50 of 782.0 kb (BazhouSP), 356.1 kb (DdB), and 388.5 kb (GSAUP). The largest scaffold sizes were 12.0 Mb, 8.4 Mb, and 11.5 Mb, and GC contents ranged from 35.9% to 37.4%. BUSCO analysis revealed high completeness in all assemblies: 91.8% (BazhouSP), 88.8% (DdB), and 91.1% (GSAUP). We compared our assemblies with two published *D. destructor* genomes: the scaffold-level Dd01 (BUSCO 88.8%; scaffold N50 = 561.0 kb; contig N50 = 84.2

kb; 13,938 genes) and the chromosome-level A1 (BUSCO 92.7%; scaffold N50 = 32.25 Mb; contig N50 = 962.3 kb; 19,165 genes). Compared to Dd01, our assemblies demonstrated improved contiguity and higher BUSCO completeness; however, they did not reach the contiguity level or BUSCO completeness of A1.

Repeat sequences in the genomes of BazhouSP, DdB, and GSAUP accounted for 30.17%, 20.67%, and 17.31%, respectively. The proportions in BazhouSP and DdB were higher than that in Dd01 (19.50%), with only BazhouSP slightly exceeding that in A1 (28.14%). Gene prediction identified 21,283, 24,056, and 21,651 protein-coding genes in BazhouSP, DdB, and GSAUP, respectively, each exceeding the counts in Dd01 (13,938) and A1 (19,165). Table 1 presents a comparison of the genomic features of our three assemblies with those of the two published *D. destructor* genomes (Dd01 and A1).

Functional analysis based on KOG classification showed broadly conserved profiles across the seven nematode genomes examined: clusters [O] (post-translational modification, protein turnover, chaperones) and [T] (signal transduction mechanisms) each exceeded 5% within the Cellular Processes & Signaling category, and [K] (transcription) and [U] (intracellular trafficking, secretion, vesicular transport) were likewise abundant (> 5%) in *D. destructor* and *C. elegans* (Fig. 2a; Supplementary Data 1). These profiles indicate conserved demands for protein homeostasis and environmental signaling in nematodes and provide the baseline against which domain- and family-level differences are interpreted.

To assess functional differences among *D. destructor* haplotypes, we applied Fisher's exact tests ($p < 0.05$) to Pfam- and CAZy-based count matrices across three Haplotype A genomes (A1, Dd01, BazhouSP), one Haplotype B genome (DdB), and one Haplotype C genome (GSAUP). Pairwise comparisons of Pfam domains between these haplotypes revealed 151 significant domains at $p < 0.05$ (Supplementary Data 2). Of these, 37 domains were visualized in Figure 2b, including 33 domains passing FDR correction ($q < 0.05$) and an additional 4 detoxification-related domains ($p < 0.05$) of potential biological relevance to host adaptation. Similarly, pairwise comparisons across

CAZy families revealed two significant families (Fig. 2d; Supplementary Data 3). Because Pfam identified pectate lyase and PL3 directly mediated beta-elimination cleavage of plant cell wall pectin [27], we focused on PL3.

Srd (PF10317.14), Srh (PF10318.14), Srj (PF10319.14), Sru (PF10322.14), Srv (PF10323.14), and Str (PF10326.14) are serpentine-type seven-transmembrane GPCR subfamilies classically expressed in nematode chemosensory neurons [28]. In pairwise contrasts, all six serpentine GPCR subfamilies had significantly higher gene counts in Haplotype A (A1 and BazhouSP) than in Haplotype B (DdB) and Haplotype C (GSAUP). Dd01 versus DdB and Dd01 versus GSAUP were not significant for any of the six subfamilies. Haplotype B had significantly higher gene counts than Haplotype A (A1, Dd01, BazhouSP) and Haplotype C for pectate lyase; Haplotype B also had significantly higher gene counts than Haplotype C for cytochrome P450. Haplotype C had significantly higher gene counts than Haplotype A (A1, BazhouSP) and Haplotype B for the NADPH-dependent FMN reductase domain; whereas the difference relative to Dd01 was not significant. For the pyridine nucleotide-disulphide oxidoreductase domain, Haplotype C had significantly higher gene counts than Haplotype B. Haplotype C also had significantly higher gene counts than Haplotype B and A1 for the ABC transporter (Fig. 2b; Supplementary Data 2).

When annotations from the seven genomes were pooled, CAZy annotation identified six major classes: glycosyltransferases (GT, 1,328; 59.1%); glycoside hydrolases (GH, 647; 28.8%); polysaccharide lyases (PL, 168; 7.5%); auxiliary activities (AA, 38; 1.7%); carbohydrate binding modules (CBM, 34; 1.5%); and carbohydrate esterases (CE, 32; 1.4%). GT and GH together accounted for 87.9% of all assignments. At the per genome level, a similar pattern held: GT accounted for 41.8% to 71.9% (median 58.8%), whereas GH accounted for 24.7% to 35.5% (median 30.1%). Notably, the free living nematode *C. elegans* had no PL assignments in our data (Fig. 2c). Pairwise tests showed Haplotype B had significantly higher gene counts in the PL3 family than Haplotype A and Haplotype C, whereas A versus C was not significant (Fig. 2d; Supplementary Data 3). Every plant-parasitic genome examined was significantly higher than *C.*

elegans for PL3, as expected for a free-living nematode lacking plant parasitism. The PL3 pattern of Haplotype B aligns with the Pfam result for PF03211.18, further supporting pectate lyase as an important functional component of Haplotype B.

Whole-genome alignments reveal genome-wide structural divergence among *D. destructor* haplotypes

To further investigate extensive divergence among *D. destructor* haplotypes, we conducted pairwise whole-genome BLASTN alignments among five *D. destructor* assemblies: Haplotype A (A1, Dd01, BazhouSP), Haplotype B (DdB), and Haplotype C (GSAUP) and reported directional genomic coverage for each alignment direction (Fig. 3a).

Haplotype B–C comparisons show the highest mutual coverage. When DdB was aligned against GSAUP, the coverage was 75.61%; in the reverse direction (GSAUP against DdB), the coverage was 81.71%. These high, bidirectional values indicate extensive shared sequence between Haplotype B and Haplotype C.

Haplotype A–B comparisons show directional asymmetry. When A1 was aligned against DdB, the coverage was 56.61%; in the reverse direction (DdB against A1), the coverage was 45.24%. When Dd01 was aligned against DdB, the coverage was 57.28%; in the reverse direction (DdB against Dd01), the coverage was 41.62%. When BazhouSP was aligned against DdB, the coverage was 57.11%; in the reverse direction (DdB against BazhouSP), the coverage was 44.72%. Thus, across all three Haplotype A genomes, coverage from Haplotype A to Haplotype B averaged 57.00%, while coverage from Haplotype B to Haplotype A was consistently lower, averaging 43.86%, pointing to haplotype-specific content and structural divergence.

Haplotype A–C comparisons were moderate and more balanced than Haplotype A–B. When A1 was aligned against GSAUP, the coverage was 52.91%; in the reverse direction (GSAUP against A1), the coverage was 51.35%. When Dd01 was aligned against GSAUP, the coverage was 54.77%; in the reverse direction (GSAUP against Dd01), the coverage was 47.86%. When BazhouSP was

aligned against GSAUP, the coverage was 52.54%; in the reverse direction (GSAUP against BazhouSP), the coverage was 50.43%. Taken together, coverage from Haplotype A to Haplotype C averaged 53.41%, whereas coverage from Haplotype C to Haplotype A averaged 49.88%, below the Haplotype B–C values. Whole-genome alignments showed that Haplotypes B–C had the highest bidirectional coverage; Haplotypes A–B exhibited directional asymmetry and A–C were intermediate.

The bidirectional whole-genome alignment coverage among Haplotype A isolates was higher than the inter-haplotype genome coverage (Haplotypes B and C). In the bidirectional coverage comparisons, A1–BazhouSP was highest, followed by A1–Dd01, and then BazhouSP–Dd01. When A1 was aligned against BazhouSP, the coverage was 93.35%; in the reverse direction (BazhouSP against A1), the coverage was 92.14%. When A1 was aligned against Dd01, the coverage was 87.12%; in the reverse direction (Dd01 against A1), the coverage was 91.44%. When BazhouSP was aligned against Dd01, the coverage was 85.17%; in the reverse direction (Dd01 against BazhouSP), the coverage was 90.47% (Supplementary Fig. 2a). These results indicate that there were differences among the different isolates within Haplotype A.

To visualize these genome-wide patterns, we generated Circos plots for seven pairwise comparisons: DdB–GSAUP, A1–DdB, Dd01–DdB, BazhouSP–DdB, A1–GSAUP, Dd01–GSAUP, and BazhouSP–GSAUP (Fig. 3b–h), along with three additional comparisons confined to Haplotype A: A1–BazhouSP, BazhouSP–Dd01, and A1–Dd01 (Supplementary Fig. 2b–d). Consistent with the coverage analyses, the Haplotype B–Haplotype C plots show dense, long-range syntenic ribbons with few large breaks, indicating extensive collinearity. Within Haplotype A, synteny was even denser and extends over longer ranges than in B–C. Whereas comparisons between Haplotype A genomes (A1, Dd01, BazhouSP) and Haplotypes B or C displayed fragmented and shorter synteny blocks, more frequent junctions consistent with inversions and relocations, and expanded unaligned segments, all of which point to increased structural turnover in the Haplotype A genomes relative to Haplotype B and Haplotype C.

Circos visualizations present a coherent picture: Haplotype B and Haplotype C are structurally more similar to each other than to Haplotype A, whereas A–B and A–C comparisons reveal widespread rearrangements and haplotype-specific sequences. These observations underscore extensive genome-wide structural divergence among the haplotypes and provide a structural framework for interpreting the functional annotation contrasts described above.

Gene family dynamics and evolutionary relationships of *D. destructor* haplotypes

To investigate gene family dynamics associated with the evolution of *D. destructor* and its haplotypes, we performed a comparative analysis of orthologous groups (OGs) across seven nematode taxa, including *C. elegans* (Bristol N2, outgroup), *D. dipsaci* (pooled J2), and five isolates of *D. destructor*: BazhouSP (Haplotype A), A1 (Haplotype A), Dd01 (Haplotype A), DdB (Haplotype B), and GSAUP (Haplotype C). OGs were inferred using OrthoFinder v2.4.0 [29], yielding 29,459 orthogroups comprising 165,765 genes. The proportion of core genes ranged from 44.48% (GSAUP) to 56.32% (Dd01), while accessory genes ranged from 40.93% (Dd01) to 46.74% (BazhouSP), and specific genes ranged from 2.75% (Dd01) to 13.23% (GSAUP). Within Haplotype A, the larger gene counts of A1 and BazhouSP genomes were mainly contributed by the expansion of accessory and core genes (Fig. 4a). Across haplotypes, Haplotype B possessed a greater abundance of core and accessory genes, whereas Haplotype C was enriched for specific genes.

Pfam annotation of the top 30 orthologous gene families across five *D. destructor* genomes showed domains associated with chemoreception, signal transduction, ion channels, transcriptional regulation, and transposable elements, DNA/RNA processing, protein modification, membrane transport, and diverse enzymatic activities (Supplementary Data 4).

To clarify the evolutionary relationships among the three haplotypes of *D. destructor* and related species, we constructed a maximum likelihood phylogeny from 800 single-copy core genes. The phylogenetic analysis revealed two highly supported clades: one comprising all Haplotype A

isolates, and the other comprising Haplotypes B and C as sister lineages, consistent with the whole-genome alignment results, which showed extensive sequence sharing between Haplotype B and Haplotype C. *D. dipsaci* formed a distinct lineage relative to the *D. destructor* isolates, reinforcing its separate status (Fig. 4a).

The OGs were mapped onto the phylogeny to reconstruct gene family births and deaths across key lineages (Supplementary Data 5). We identified the birth of 1,369 gene families on the branch leading to the *Ditylenchus* lineage (Fig. 4a). Analysis of gene family evolution revealed substantial gains and losses: the *D. destructor* branch gained 3,557 and lost 53 gene families, whereas the *D. dipsaci* branch experienced greater turnover, gaining 3,800 and losing 466 families. Within *D. destructor*, the Haplotype A branch gained more gene families (1,100) than the branch uniting Haplotypes B and C (524), supporting the dynamics of gene families in *D. destructor*.

We further explored OGs among the six *Ditylenchus* genomes (Fig. 4b). The analysis identified 6,387 core OGs shared by *D. destructor* haplotypes and *D. dipsaci*. Furthermore, *D. destructor* haplotypes shared 2,573 OGs, much larger than 249/151/18 OGs shared between three *D. destructor* haplotypes and *D. dipsaci*. This pattern suggested the closer evolutionary relationships between *D. destructor* haplotypes than between *D. destructor* and *D. dipsaci*. The OG analysis revealed a greater number of *D. destructor* Haplotype A specific genes (6,007) compared to Haplotypes B (1,923) and C (2,878). This difference is likely due to the higher number of Haplotype A genomes (three), versus only one genome each for Haplotypes B and C. This pattern of Haplotype A specific genes was not observed after we selected BazhouSP, one of *D. destructor* Haplotype A genomes, for OG analysis, which represented 2,689, 2,118, and 3,023 Haplotype specific genes for Haplotypes A, B, and C, respectively (Fig. 4c). Consistently, the results showed more OGs shared between Haplotypes B and C (765) than between either of them and A (520/593), supporting a closer relationship between Haplotypes B and C.

Overall, the maximum likelihood phylogeny, orthogroup distributions, and inferred gene family births and deaths outline the evolutionary relationships across *D. destructor* and provide testable

candidates for investigating host range and ecological specialization.

Secretome analysis across *D. destructor* haplotypes

Previous research has shown that secreted proteins play important roles in nematode-plant interactions [31]. Here, we explored secreted proteins in *D. destructor* haplotype genomes. Using SignalP [32] and TMHMM [33], we identified 8,191 putative secreted proteins across six *Ditylenchus* genomes, defined by the presence of a signal peptide and the absence of transmembrane helices (Fig. 5a). In the three *D. destructor* Haplotype A genomes, we identified 1,694 (A1), 881 (Dd01), and 1,832 (BazhouSP) putative secreted protein coding genes; in Haplotypes B (DdB) and C (GSAUP), 1,480 and 1,677, respectively; and in *D. dipsaci*, 627.

Then we analyzed the orthogroups of secreted proteins (OGSPs) among *Ditylenchus* haplotypes. We found that the A (A1, BazhouSP, Dd01)–C pair shared 1,152 genes, more than A (A1, BazhouSP, Dd01)–B (715) or B–C (275). Haplotype A (A1, BazhouSP, Dd01) had the largest number of haplotype-specific genes (1,402), followed by Haplotype C (333) and Haplotype B (286) (Fig. 5b). Using BazhouSP as the representative of Haplotype A, we obtained the same overall ordering of haplotype-specific counts. For shared genes, the A (BazhouSP)–C pair remained the highest at 641; however, B–C exceeded A (BazhouSP)–B, with 449 versus 384, respectively (Fig. 5c).

We further analyzed orthologous clusters of secretomes (OCSs) using OrthoVenn3 [34], yielding 1,859 orthologous clusters and 1,006 singletons (12.28%). Haplotype A (A1, BazhouSP, Dd01), B (DdB) and C (GSAUP) contained 72, 18, and 8 haplotype-specific OCSs, respectively. Pairwise intersections showed that Haplotypes B and C shared 122 OCSs, exceeding those between A and B (52) and between A and C (78), indicating greater similarity in secretome composition between B and C (Fig. 5d). From the OCSs, we identified 17 single-copy OCSs shared by all taxa and constructed a maximum likelihood phylogenetic tree from their concatenated sequences; the resulting tree showed a topology similar to the ITS and whole-genome single-copy gene

phylogenetic trees, separating Haplotype A from Haplotype B and Haplotype C (Fig. 5e).

Functional domains were annotated for the secreted proteins. Subsequently, Fisher's exact test ($p < 0.05$) was used to identify domains that were significantly enriched in pairwise comparisons between haplotypes. Haplotype B (DdB) had significantly higher counts of pectate lyase (PF03211.18) secreted proteins than Haplotype A (A1, BazhouSP, Dd01) and Haplotype C (GSAUP). Haplotype A (A1, BazhouSP) had significantly higher counts of Glycosyl hydrolases family 31 (PF01055.31, PF21365.2) secreted proteins than Haplotype B. Haplotype C possessed significantly more secreted animal haem peroxidases (PF03098.20) and C-type lectins (PF00059.26) than Haplotype B, and also more Astacins (Peptidase family M12A, PF01400.29) than Haplotype A (Dd01) (Fig. 5f; Supplementary Data 6).

Similarly, Fisher's exact test ($p < 0.05$) was used to identify secreted CAZymes that were significantly enriched in pairwise comparisons between haplotypes. Haplotype B had significantly higher counts of secreted PL3 than Haplotype A (Dd01, BazhouSP) and Haplotype C, whereas Haplotype A (A1 and BazhouSP) had significantly higher counts of secreted GH31 than DdB (Fig. 5g; Supplementary Data 7).

Discussion

Previous studies reported a chromosome-scale genome for *D. destructor* A1 [23] and a scaffold-level genome for *D. destructor* Dd01 [22], both isolated from sweet potato. Based on surveys of previously reported natural populations, distinct host preferences among haplotypes have been documented: A prefers sweet potato; B shows a shared preference for sweet potato and potato; and C prefers potato [10, 11]. Consistent with these observations, differences in Rf among haplotypes on different hosts were also observed in Rf assays. Building on this foundation, we used a hybrid assembly strategy combining long- and short-read data to generate high-quality assemblies for Haplotype A (BazhouSP, isolated from sweet potato), Haplotype B (DdB, isolated from potato), and Haplotype C (GSAUP, isolated from potato), and together with the two published Haplotype A genomes (A1

and Dd01), we performed comparative genomics to provide insights into host preferences across haplotypes.

Our genome-wide analyses showed that the three Haplotype A genomes (A1, BazhouSP, and Dd01) exhibited greater pairwise sequence collinearity than that in the Haplotype B–C comparison, although sequence similarity within Haplotype A was not identical across isolates. Previous studies have reported morphometric differences among seven Haplotype A isolates [11]; here, our genome-level comparisons corroborated such differences, although the differences were insufficient to warrant redefinition of haplotypes. In addition, sequence collinearity between Haplotypes B and C was far higher than that between either of them and Haplotype A. A phylogenetic tree inferred from whole-genome single-copy genes resolved Haplotype A as a distinct clade, with Haplotypes B and C forming a separate clade, and showed high bootstrap support with values no lower than 98 (Fig. 4a). Together, these results indicate that Haplotypes B and C are more closely related to each other than to Haplotype A.

Genes within an OG share a common ancestor, and the gain and loss of these gene families are associated with evolutionary adaptation [35, 36]. Comparative gene-family analyses showed patterns useful for understanding gene annotation and evolution in *Ditylenchus* genomes. In *D. destructor*, Haplotypes B and C shared more OGs with each other than either shared with Haplotype A, supporting the phylogenetic relationships among the three haplotypes. At the genome-wide level, Haplotype A (A1 and BazhouSP) showed significantly higher gene counts of serpentine seven-transmembrane GPCRs than Haplotypes B and C. The GPCRs enable nematodes to detect volatile and soluble cues that guide chemotaxis and host finding [37]. The large number of GPCRs in Haplotype A may enhance its ability for host location. A previous study found that plant-parasitic nematodes can secrete pectate lyases to facilitate penetration and tissue migration [38]. Haplotype B had significantly higher gene counts of pectate lyase and the CAZy family PL3 than Haplotypes A and C. The large gene counts of pectate lyases in Haplotype B may facilitate its penetration and degradation of pectin-rich cell walls in sweet potato storage roots [39] and potato tubers [40].

Meanwhile, the higher counts of cytochrome P450 genes may be associated with the oxidative metabolism and detoxification of exogenous small molecules, including phytochemicals from Solanaceae [41]. The large number of NADPH-dependent FMN reductases, pyridine nucleotide-disulphide oxidoreductases and ABC transporters found in Haplotype C may help the nematode cope with early oxidative bursts in potato tubers and handle and export potato-derived small molecules [42], such as steroidal glycoalkaloids [43].

Plant-parasitic nematode effectors are secreted molecules that interact with host targets to promote parasitism [31]. The diversity of nematode effectors has been identified [44]. Analysis of the *D. destructor* secretome may provide new clues to discover significant factors contributing to host adaptation. In *D. destructor*, 500 OGSPs were shared among Haplotypes A (A1, BazhouSP, Dd01), B, and C. Meanwhile, Haplotypes A (A1, BazhouSP, Dd01), B, and C contained 858, 268, and 325 haplotype-specific OGSPs, respectively. These results suggested the diversity of secreted proteins in *D. destructor* genomes. Furthermore, we found a large number of secreted pectate lyases and CAZy family PL3 in Haplotype B, GH31 in Haplotype A, and animal haem peroxidases, C-type lectins and Astacins in Haplotype C. These secreted proteins may contribute to host adaptation in different *D. destructor* haplotypes. These patterns supported the previous discovery of the enrichment of GH31 and alpha-amylases in Dd01, the Haplotype A isolated from sweet potato [22, 45]. Alpha-amylase may act endolytically to cleave internal alpha-1,4 linkages to generate oligosaccharides [46], and GH31 may perform exolytic processing from non-reducing ends to release alpha-glucose [47]. Sweet potato storage roots commonly contain 70–80% highly branched amylopectin [48], providing abundant non-reducing ends. The enrichment of GH31 may be useful for sweet potato adaptation in Haplotype A. Potato tubers often induce reactive oxygen species (ROS) during defense [49] and their cell walls are rich in pectin with pronounced tissue compartmentalization [40]. The enrichment of animal haem peroxidases, C-type lectins and Astacins may be useful for potato adaptation in Haplotype C, as well.

Notably, A1 and BazhouSP both exhibited significantly higher GPCR domain gene counts

compared to Haplotypes B and C. For secreted GH31 family genes, A1 and BazhouSP harbored significantly more genes than Haplotype B but showed no significant difference compared to Haplotype C. In contrast, Dd01 showed no significant differences from Haplotypes B or C for either gene family. The lack of significant differences observed for Dd01 may be attributed to its lower genome assembly quality. The Dd01 genome was assembled using early short-read sequencing technology and subsequently scaffolded with long-read data for optimization [22], with a contig N50 of 84.2 kb, approximately 10% of the A1 and BazhouSP genomes' contig N50 (Table 1). Previous studies have demonstrated the limitations of short-read sequencing assemblies, which tend to produce more fragmented sequences and reduce genome completeness [50]. For the secreted PL3 family, DdB showed significant differences in gene number compared to BazhouSP, Dd01, and Haplotype C, whereas the comparison with A1 did not reach statistical significance. It is noteworthy that, despite the lack of statistical significance between DdB and A1, the number of secreted PL3 genes in DdB was approximately twice that in BazhouSP and A1, indicating that DdB still harbors higher counts of secreted PL3 genes within this family.

In particular, no secreted GH31 genes were predicted in DdB. BLAST analysis revealed that the corresponding genes in DdB carry an additional stretch of ten to twenty amino acids at the N-terminus, which likely caused the failure of signal peptide prediction. This discrepancy may result from insertion-deletion errors in Nanopore sequencing near the gene start site. These observations suggest that, even under the same assembly and annotation pipeline, different sequencing platforms, such as Nanopore and PacBio, can lead to subtle differences in gene structure prediction. Nevertheless, the closely related GSAUP and the congeneric species *D. dipsaci* each carry one predicted secreted GH31 gene, whereas Haplotype A harbors three to five such secreted genes, suggesting that secreted GH31 remains a potentially important CAZyme family in Haplotype A.

Together, these findings provide insights into the host preferences across the different haplotypes. Additionally, our work provides a foundation for expanding haplotype genomic resources for the *D. destructor* species complex.

Conclusion

The genetic basis of *D. destructor* phenotypes of Haplotype A, B, and C infesting sweet potato, sweet potato and potato, and potato, respectively, is investigated in this study. The reference assemblies for Haplotypes B and C of *D. destructor* are reported for the first time. Both ITS and genome-scale phylogenetic trees, as well as whole-genome alignments consistently show that Haplotypes B and C are closer to each other, whereas Haplotype A is relatively distinct. Comparative genomic analysis reveals gene content variations related to the patterns of host preference among three haplotypes, including chemosensory GPCRs, animal haem peroxidase, cytochrome P450s, CAZymes and proteases. These results provide a solid foundation for elucidating host use in *D. destructor* and for identifying targets for disease management.

Materials and Methods

Nematodes

In 2019, infected sweet potato storage roots were collected from Bazhou, Hebei Province, China, and infected potato tubers were collected from Dingxi, Gansu Province, China. Nematodes were extracted using a modified Baermann funnel method [51] as follows: infected plant tissues were cut into 0.5 cm cubic pieces and wrapped in double-layer filter paper, then placed at the neck of a funnel. Sterile water was slowly added dropwise to maintain a moist environment. Nematodes migrated through the filter paper pores under gravity and were collected from the tube bottom after 12 h. The two isolates were designated BazhouSP (sweet potato) and GSAUP (potato), respectively. To establish pure cultures of *D. destructor*, individual nematodes were handpicked and transferred onto *Fusarium verticillioides* grown on potato dextrose agar (PDA) plates, and maintained at 25 °C for propagation; PDA medium was prepared as follows: 200 g potatoes were cut into small pieces and boiled in 1000 mL water for 20 min, then filtered to obtain potato extract. Dextrose (20 g) and agar (15 g) were added to the potato extract, and the volume was adjusted to 1000 mL before autoclaving at 121 °C for 15 min. The isolate DdB (GenBank: MW522604.1), originally isolated

from infected potato in Gansu Province by the laboratory of Huixia Li, was included in this study for comparative analysis [18].

Molecular identification and secondary structure reconstruction of ITS-rDNA

The isolated nematodes were washed three times with sterile ddH_2O, and one was manually selected and placed into a 0.2 mL PCR tube as a sample, and five samples were obtained by repeating. Then 10 μL lysis buffer containing 10 mM Tris-HCl (pH 8.0), 50 mM KCl, 2.5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM dithiothreitol, 60 mg·mL^{-1} proteinase-K, and 0.45% Tween 20 was added into the PCR tube. Nematodes were lysed by treatment at 65 °C for 60 min and 95 °C for 10 min, and then the lysate was briefly centrifuged. Thus, the genomic DNA of a single nematode was obtained and stored at –20 °C for later use.

The region of internal transcribed spacer (ITS) rDNA sequence was amplified using primer pair AB28/TW81 (5'- ATATGCTTAAGTTCAGCGGGT-3'; 5'-GTTTCCGTAGGTGAACCTGC-3'). Using genomic DNA as template, PCR reaction was performed in 50 μL volume using Phanta Max Master Mix (Vazyme Biotech, Nanjing, China), with an initial denaturation at 95 °C for 3 min, 35 cycles of 95 °C for 15s, 55 °C for 15 s, and 72 °C for 30s and finally, elongation at 72 °C for 5 min. PCR product was purified with Easy Pure Quick Gel Extraction Kit (Vazyme Biotech, Nanjing, China), according to manufacturer's protocol. The transformants were sequenced. Amplified fragments were cloned into pUC19 vector using *pEASY*-Blunt Cloning Kit (TransGen Biotech, Beijing, China) and sequenced by TSINGKE Biological Technology (Beijing, China). The newly obtained sequences were deposited in the NCBI GenBank database.

To investigate the phylogenetic relationships among *Ditylenchus* species, we analyzed ITS-rDNA sequences from representative isolates, including BazhouSP (OL982746.1), DdB (MW522604.1) [18] and GSAUP (OL982747.1). Multiple sequence alignment was performed using MAFFT v7.525 [52] with the --auto option to ensure optimal alignment quality. Poorly aligned and ambiguous regions were filtered using Gblocks v0.91b [53] with the least stringent parameters, resulting in a

curated alignment of 870 positions (72% of the original 1,208 bp). The filtered alignment was converted to NEXUS format using Seqret from the EMBOSS v6.6.0.0 [54] suite.

Bayesian phylogenetic inference was conducted using MrBayes v3.2.7 [55]. The best-fit model of DNA substitution for BI was determined as GTR+G, selected using MrModeltest v2.3 [56] in conjunction with PAUP* under the Akaike Information Criterion (AIC). The analysis was run with four MCMC chains for 2,000,000 generations, sampling every 100 generations, and discarding the first 5,000 samples as burn-in. Posterior probabilities were used to assess node support. The secondary structures of ITS1 were predicted through the minimum free energy (MFE) approach implemented in UNAFold Web Server (http://www.unafold.org) [25]. The structures were visualized using VARNA v3-93 [26] and drawn with Inkscape 0.92.5.

Reproductive factor of *D. destructor* on sweet potato and potato hosts

Sweet potato cultivar 'Yanshu 25' and potato cultivar 'Atlantic' were selected for the reproduction factor assay of three *D. destructor* haplotypes. The substrate (soil:humus:sand = 2:2:1, v/v/v) was autoclaved at 121 °C for 20 min and transferred to 25 cm × 20 cm pots after cooling. Sweet potato slips were trimmed at the root to exclude nematodes in fibrous roots. Both sweet potato slips and potato tubers were then surface-sterilized using the same protocol: immersion in 75% ethanol for 10 s, treatment with 2.5% NaClO solution for 3 min, and thorough rinsing with sterile water. The sterilized plant materials were then transplanted into the substrate and watered regularly. Each treatment included five pots and was independently repeated three times. At 30 days after planting, nematode suspensions of BazhouSP, DdB, and GSAUP were inoculated near the root zone using a micropipette. Four holes were made evenly around each pot using a 1 mL pipette tip, and 250 nematodes were inoculated into each hole (1000 nematodes per pot in total). At 60 days post-inoculation, nematodes were extracted from storage roots (sweet potato), tubers (potato), and their respective surrounding substrate using the modified Baermann funnel method and counted. The total nematode population (from both plant tissues and substrate) was used to calculate the

reproduction factor. Data were recorded and collated using Microsoft Excel (2010). Statistical analysis and visualization were performed using Python. Comparisons between sweet potato and potato within each haplotype were performed using the Mann-Whitney U test for independent samples. A two-sided test was applied for each comparison, and $p < 0.05$ was considered statistically significant. Reproduction factor (Rf) = Pf / Pi, where Pf is the final nematode population and Pi is the initial inoculum (1000 nematodes).

Genome and transcriptome sequencing

To obtain pure *D. destructor*, nematodes were isolated from *Fusarium verticillioides* culture plates using a modified Baermann funnel method. After passing through 500 mesh screens, the nematode suspensions of mixed life stages were sterilized with 0.5% sodium hypochlorite solution for 30 seconds and then washed with distilled water three times for DNA and RNA extraction.

High molecular weight genomic DNA was prepared by the CTAB method and followed by purification with QIAGEN® Genomic Kit (Cat#13343, QIAGEN), according to the standard operating procedure provided by the manufacturer. A total amount of 5 μg DNA per sample was sent to Grandomics Biosciences (Beijing, China). DNA purity was detected using NanoDrop™ One UV-vis spectrophotometer (Thermo Fisher Scientific, USA). SMRTbell target size libraries were constructed for sequencing according to PacBio's standard protocol (Pacific Biosciences, CA, USA) using 20 kb preparation solutions. Sequencing was performed on a PacBio Sequel instrument. For the Haplotype B isolate (DdB), long read libraries were prepared and sequenced using Oxford Nanopore Technologies PromethION flow cells following the manufacturer's instructions. In addition, 150 bp paired-end genomic DNA libraries were prepared and sequenced on BGI DNBSEQ-T7 platform.

Total RNA from 100 mg mixed life stages nematodes was extracted using the Invitrogen TRIzol reagent (Life Technologies Corporation, United States) following the instructions provided by the manufacturer. The RNA samples were sent to Grandomics Biosciences. RNA integrity was

assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated.

Genome assembly

PacBio long reads were assembled de novo using wtdbg2 software with parameters "-k 15 -AS 2 -p 0 -L 5000" [57]. Assembly quality was assessed by comparing the assembled genome size to genome size estimates derived from filtered paired-end reads. Genome size estimation was performed by generating k-mer frequency histograms using Jellyfish [58] with a k-mer size of 21 ("-m 21"), and subsequent analysis of these histograms with GenomeScope [59]. The initial assemblies were polished using paired-end reads with Pilon [60] and Racon [61].

Gene prediction

Protein-coding genes were independently predicted using Augustus v3.3.3 [62] and GeneMark-ES/ET/EP v4.69_lic [63] as two separate ab initio gene prediction tools. Additionally, gene prediction was refined using the BRAKER pipeline [64] which integrated evidence from transcriptomic and protein homology data. Transcriptomic evidence was generated by aligning RNA-seq reads to genome assemblies using TopHat2 [65]. Protein homology information was inferred from invertebrate sequences in the NCBI RefSeq database using the ProtHint pipeline (https://github.com/gatech-genemark/ProtHint). Predicted protein domains were annotated by searching against the Pfam database v27.0 [66] using HMMER v3.1b1 (http://hmmer.janelia.org/). Finally, for predicted genes from Augustus [62], GeneMark-ES/ET/EP [63], and BRAKER [64] methods, we selected those supported by at least two methods and those supported by only one method but containing functional gene domains.

Genome quality control

We removed contaminated scaffolds from assemblies through multiple quality control steps (Supplementary Fig. 3), following the approach of previous studies [67]. Predicted genes were aligned against NCBI RefSeq protein sequences from invertebrates and non-invertebrates (plants, fungi, bacteria, and archaea) using BLASTP [68]. The best hits were extracted based on E-values and combined for classification into three categories: (1) genes where the E-value ratio of non-invertebrate/invertebrate BLAST hits is < 1e-50, indicating stronger similarity to non-invertebrate sequences; (2) genes aligned to non-invertebrate RefSeq but not to invertebrate RefSeq; and (3) other genes. Based on this classification, scaffolds were then categorized based on gene composition. If categories (1) and (2) combined comprised ⩾50% of genes in a scaffold, that scaffold was flagged as contaminated and removed (output to "contaminated.fa"). If categories (1) and (2) combined comprised <50% but categories (1), (2), and (3) combined comprised ⩾50% of genes in a scaffold, the scaffold was manually reviewed for potential contamination. Scaffolds passing quality control were retained and output to "genome.fa". After removing the contaminated scaffolds, repeat sequences were identified with RepeatScout v1.0.5 [69] and Repbase v19.06 [70] and masked using RepeatMasker v4.0.5 [71]. RNA-seq reads were mapped onto genome assemblies using HISAT2 v2.2.0 [72]. Genome assembly completeness and quality were evaluated using BUSCO v3.0.2 [73], employing the eukaryota_odb9 lineage dataset as the reference.

Functional annotation

Functional annotation of predicted genes was performed by aligning protein sequences against the EuKaryotic Orthologous Groups (KOG 2003-04-16) database [74] using BLASTP [68] with an E-value cutoff of 1e-5.

Protein functional domains were identified using the Pfam database v37.0 [66] with the hmmsearch program from HMMER v3.1b1 (http://hmmer.janelia.org/).

Carbohydrate-active enzymes, including glycoside hydrolase (GH), glycosyltransferase (GT), polysaccharide lyase (PL), carbohydrate esterase (CE), auxiliary activity (AA), and carbohydrate-

binding module (CBM) families, were annotated using the dbCAN3 [75] online server (https://bcb.unl.edu/dbCAN2/blast.php) with the HMMER option, applying a stringent E-value cutoff of 1e-15 and a minimum coverage threshold of 0.35.

Pairwise Fisher's exact tests were performed to compare Pfam and CAZymes annotations among five isolates representing three haplotypes, i.e., A1, BazhouSP, and Dd01 from Haplotype A; DdB from Haplotype B; GSAUP from Haplotype C. This design resulted in seven pairwise comparisons (A1 vs DdB, A1 vs GSAUP, BazhouSP vs DdB, BazhouSP vs GSAUP, Dd01 vs DdB, Dd01 vs GSAUP, DdB vs GSAUP). Annotations showing statistically significant differences ($p < 0.05$) in these comparisons were identified and compiled to characterize haplotype-specific functional signatures.

For Pfam domains, the initial analysis identified 151 domains with significant differences ($p < 0.05$). Multiple comparisons were adjusted for using the Benjamini-Hochberg FDR correction, yielding a reduced set of 33 domains ($q < 0.05$). Four detoxification-related Pfam domains (PF00067.27, cytochrome P450; PF00005.32, ABC transporter; PF03358.20, NADPH-dependent FMN reductase; PF00070.32, pyridine nucleotide-disulphide oxidoreductase) that showed statistical significance ($p < 0.05$) but did not pass the FDR threshold were additionally included in the visualization given their potential relevance to host adaptation, as potato tubers produce steroidal glycoalkaloids as defense compounds, and these domains are associated with xenobiotic metabolism and detoxification processes that may influence host preference in phytophagous organisms. Thus, Figure 2b displayed a total of 37 domains. The 151 Pfam domains with statistically significant differences ($p < 0.05$) were provided in Supplementary Data 2.

Whole-genome alignment and Circos visualization analysis of *D. destructor*

To assess genome-scale sequence conservation among the three *D. destructor* haplotypes (Haplotype A isolates: A1, BazhouSP, Dd01; Haplotype B: DdB; Haplotype C: GSAUP), pairwise whole-genome alignments were performed using BLASTN [68] with an E-value threshold of 1e-5,

comparing each Haplotype A isolate against the Haplotype B and Haplotype C genomes. Query–subject pairs were established for each comparison, and the alignment results were generated in "outfmt 6" format. These results were then used to compute genome-wide coverage for each genome in both directions. Additionally, to evaluate sequence conservation within Haplotype A, we performed pairwise whole-genome alignments among A1, BazhouSP, and Dd01 using the same BLASTN settings.

We employed a custom Python script to calculate the number of covered and uncovered bases for each genome. The script processes alignment coordinates by merging high-scoring segment pairs (HSPs) and requires ".fai" index files generated by samtools faidx to determine scaffold lengths. All BLASTN alignment blocks that passed the E-value filter were treated as HSPs and directly included in the coverage calculations, without further filtering based on alignment identity or length. Additionally, the script generates BED files marking covered and uncovered regions across scaffolds. The resulting bidirectional genome coverage percentages were used to evaluate genomic divergence among the haplotypes.

BLASTN alignments were visualized as syntenic blocks using Circos v0.699 [76]. Only scaffolds greater than 500 kb in length and syntenic blocks of at least 2 kb were included in the visualization.

Orthogroups, phylogenomics, and functional modules across *D. destructor* haplotypes

OrthoFinder v2.4.0 [29] identified orthogroups (OGs) and their constituent orthologous genes across seven genomes: BazhouSP (Haplotype A), A1 (Haplotype A) [23], Dd01 (Haplotype A) [22], DdB (Haplotype B), and GSAUP (Haplotype C), together with *D. dipsaci* (Ddi, pooled J2) [4] and *Caenorhabditis elegans* (Bristol N2, outgroup) [77]. Based on these orthogroups, we identified core, accessory, and specific genes. The core genes were presented in all seven genomes. The accessory genes were presented in at least two genomes, but in less than seven genomes. The specific genes were presented in only one genome. Pfam domain annotations used in orthogroup functional analysis were derived from the protein functional annotation pipeline described above, which

employed "hmmsearch" (HMMER v3.3.2) against the Pfam database v37.0 [66] with an E-value threshold of 1e-5. To assign dominant functional domains to each orthogroup, we processed the Pfam annotation results using a custom Perl script. For each protein, we retained the top two non-redundant Pfam domains based on the lowest E-values. Within each orthogroup, Pfam domain frequencies were calculated across all member proteins, and the top two most frequent domains were selected as the representative functions. This classification was used to summarize the functional composition of the top thirty largest orthogroups.

From these orthogroups, 800 single-copy core genes shared by all seven genomes were selected using a custom in-house Perl script. To resolve the phylogenetic placement of the *D. destructor* isolates, the protein sequences of these 800 single-copy genes were individually aligned using MUSCLE v3.8.31 [78]. Poorly aligned regions within the concatenated alignment were subsequently trimmed using Gblocks v0.91b [53] with default parameters to ensure phylogenetic accuracy. The resulting alignment was subjected to model testing with ProtTest v3.4.2 [79], and the VT+I+G+F substitution model was selected as optimal. The final phylogenetic tree was inferred using RAxML-NG v1.2.2 [30], with 500 bootstrap replicates performed to assess branch support.

Gene family birth and death events were inferred from orthogroup presence-absence across species using a custom Perl script. The Perl script generated an output file, from which gene family birth ("+" events) and death ("−" events) were manually calculated using Excel spreadsheets. This combined computational and manual approach facilitated the accurate identification of gene family changes across the phylogenetic branches of *D. destructor* and related species.

To assess the homology between different haplotypes, Haplotype A was defined as the union of OGs from BazhouSP, A1, and Dd01. Shared and unique OGs were quantified using a custom Python script. To assess the differences introduced by the union, BazhouSP was analyzed independently as a Haplotype A representative following the same pipeline.

Comparative analysis of secretomes

Secreted proteins were defined by a positive signal peptide prediction and absence of transmembrane helices, using the following pipeline:

Signal peptide prediction was performed with SignalP v6.0 [32] using "--organism eukarya" on protein FASTA inputs.

Transmembrane helices were predicted using the TMHMM v2.0 [33] online server (https://services.healthtech.dtu.dk/services/TMHMM-2.0/) with output formatted as "one line per protein."

Proteins with a positive SignalP v6.0 prediction and no TMHMM v2.0-predicted transmembrane helices were defined as secreted proteins. Results were filtered using custom Python scripts to generate the final list of secreted proteins.

We mapped the IDs of predicted secreted proteins to the genome-wide OrthoFinder v2.4.0 [29] output file "Orthogroups.txt" to extract secretome-associated orthologous groups and their member genes; we then performed counting and analysis using the same workflow as for the genome-wide orthologous groups.

Orthologous clustering of secreted proteins was performed using the OrthoVenn3 web server (https://orthovenn3.bioinfotoolkits.net/home) [34]. Within OrthoVenn3, we used the OrthoFinder algorithm for ortholog inference, running DIAMOND with an E-value cutoff of 1e-3 and a Markov Clustering (MCL) inflation value of 1.5. For phylogeny, the built-in module inferred maximum likelihood trees with FastTree [80] under the JTT+CAT model.

Annotation results for secreted proteins were extracted from Pfam [66] and dbCAN3 [75] annotations of the complete set of genes predicted from genome structural annotation. For each annotated domain or enzyme family, pairwise comparisons were performed using Fisher's exact test ($p < 0.05$). All significantly enriched annotations were compiled and summarized for downstream interpretation.

Statistics and reproducibility

For molecular identification, five individual nematodes from each of the three isolates (representing different haplotypes) were randomly selected as biological replicates and processed independently for DNA extraction and sequencing to ensure reliable species identification. For reproduction factor assays, each treatment included five biological replicates, with experiments independently repeated three times (n=15 total per treatment group). Statistical comparisons between sweet potato and potato hosts within each haplotype were performed using the Mann-Whitney U test for independent samples, with $P < 0.05$ considered statistically significant. Three assembled *D. destructor* genomes from different haplotypes and two previously reported *D. destructor* genomes were included in the comparative analysis. No data were excluded from the analyses. The investigators were not blinded to allocation during experiments and outcome assessment.

Data availability

The draft genome assemblies of *Ditylenchus destructor* isolates GSAUP (Haplotype C), BazhouSP (Haplotype A), and DdB (Haplotype B) have been deposited in GenBank (accessions JAKKQA000000000.1 and JAKKPZ000000000.1) and Genome Warehouse (accession GWHGEIJ00000000.1), respectively. The corresponding BioProjects are PRJNA800207 (GSAUP and BazhouSP) and PRJCA041684 (DdB). The Illumina sequencing data of BazhouSP, DdB, and GSAUP, and the Oxford Nanopore sequencing data of DdB have been deposited in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/browse/CRA038505). PacBio raw sequencing data for BazhouSP and GSAUP have been deposited in the Open Science Framework (https://doi.org/10.17605/OSF.IO/4T7GR). RNA-seq data for BazhouSP, DdB, and GSAUP are available in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/browse/CRA038639). Numerical source data for graphs and charts can be found in Supplementary Data 1–7 (https://doi.org/10.5281/zenodo.18540296). All other data are available from the corresponding author on reasonable request.

Code availability

The pipeline for removing contaminant sequences from nematode genomes is available at https://github.com/wangjy-ll/remove-contaminant-sequences.

References

1. Kikuchi T, Eves-van den Akker S, Jones JT. Genome evolution of plant-parasitic nematodes. *Annual Review of Phytopathology* **55**, 333–354 (2017).
2. Jagdale S, Rao U, Giri AP. Effectors of root-knot nematodes: an arsenal for successful parasitism. *Frontiers in Plant Science* **12**, 800030 (2021).
3. Carlton PM, Davis RE, Ahmed S. Nematode chromosomes. *Genetics* **221**, 221 (2022).
4. Mimee B, Lord E, Véronneau P-Y, Masonbrink R, Yu Q, Akker SE-vd. The draft genome of *Ditylenchus dipsaci*. *Journal of Nematology* **51**, 1–3 (2019).
5. Madani M, Tenuta M. Molecular Characterization and Phylogeny of *Ditylenchus weischeri* from *Cirsium arvense* in the Prairie Provinces of Canada. *Journal of Nematology* **50**, 163–182 (2018).
6. Jeszke A*, et al.* A comparative and phylogenetic study of the *Ditylenchus dipsaci*, *Ditylenchus destructor* and *Ditylenchus gigas populations* occurring in Poland. *Journal of Phytopathology* **162**, 61–67 (2014).
7. Baker AD. The potato-rot nematode, *Ditylenchus Destructor* thorne, 1945, attacking potatoes in Prince Edward Island. *Scientific Agriculture* **26**, 138–139 (1946).
8. Basson S, De Waele D, Meyer A. Survival of *Ditylenchus destructor* in soil, hulls and seeds of groundnut. *Fundamental and Applied Nematology* **16**, 79–85 (1993).
9. Zhao H*, et al.* Research advances of biology in *Ditylenchus destructor* Thorne,1945. *Biotechnology Bulletin* **37**, 45–55 (2021).

10. Li Y*, et al.* Molecular characterization of internal transcribed spacer (ITS) of ribosomal RNA gene, haplotypes and pathogenicity of potato rot nematode *Ditylenchus destructor* in China. *Phytopathology Research* **4**, 22 (2022).
11. Yang Y, Feng Y, Dong H, Zhao Z, Wang J, Xu Y. Morphological and genetic analysis on populations of *Ditylenchus destructor* from Shanxi Province (in chinese). *Acta Phytopathologica Sinica* **54**, 304–317 (2024).
12. Faulkner L, Darling H. Pathological histology, hosts, and culture of the potato rot nematode. *Phytopathology* **51**, 778–786 (1961).
13. Peng H*, et al.* Exploring the host parasitism of the migratory plant-parasitic nematode *Ditylenchus destuctor* by expressed sequence tags analysis. *Plos One* **8**, e69579 (2013).
14. Pridannikov M. Economic importance of the potato tuber nematode *Ditylenchus destructor* in Russia. In: *Integrated Nematode Management: State-of-the-art and visions for the future*). CABI Wallingford UK (2021).
15. Qiao Y*, et al.* Paraphyletic genus *Ditylenchus Filipjev* (Nematoda, Tylenchida), corresponding to the *D. triformis*-group and the *D. dipsaci*-group scheme. *Zookeys*, 1–12 (2016).
16. Subbotin SA, Deimi AM, Zheng J, Chizhov VN. Length variation and repetitive sequences of internal transcribed spacer of ribosomal RNA gene, diagnostics and relationships of populations of potato rot nematode, *Ditylenchus destructor* Thorne, 1945 (Tylenchida: Anguinidae). *Nematology* **13**, 773–785 (2011).
17. Marek M, Zouhar M, Douda O, Mazakova J, Rysanek P. Bioinformatics-assisted characterization of the ITS1-5·8S-ITS2 segments of nuclear rRNA gene clusters, and its exploitation in molecular diagnostics of European crop-parasitic nematodes of the genus *Ditylenchus*. *Plant Pathology* **59**, 931–943 (2010).

18. Ni C*, et al.* Diagnosis and characterization of the ribosomal DNA-ITS of potato rot nematode (*Ditylenchus destructor*) populations from Chinese medicinal herbs. *Journal of Integrative Agriculture* **22**, 1763–1781 (2023).
19. Blanc-Mathieu R*, et al.* Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLOS Genetics* **13**, e1006777 (2017).
20. van Steenbrugge JJM*, et al.* Comparative genomics of two inbred lines of the potato cyst nematode *Globodera rostochiensis* reveals disparate effector family-specific diversification patterns. *BMC Genomics* **22**, 611 (2021).
21. Wenger AM*, et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* **37**, 1155–1162 (2019).
22. Zheng J*, et al.* The *Ditylenchus destructor* genome provides new insights into the evolution of plant parasitic nematodes. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20160942 (2016).
23. Yang Y*, et al.* Chromosome-level genome assembly of the sweet potato rot nematode *Ditylenchus destructor*. *Scientific Data* **12**, 174 (2025).
24. Gao S*, et al.* Advances of sequencing and assembling technologies for complex genomes (in Chinese). *Yi Chuan* **40**, 944–963 (2018).
25. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415 (2003).
26. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
27. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal* **432**, 437–444 (2010).
28. Robertson HM, Thomas JH. The putative chemoreceptor families of *C. elegans*. *WormBook: The Online Review of C elegans Biology [Internet]*, (2006).

29. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
30. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
31. Zhao J*, et al.* A *Meloidogyne incognita* C-type lectin effector targets plant catalases to promote parasitism. *The New phytologist* **232**, 2124–2137 (2021).
32. Teufel F*, et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology* **40**, 1023–1025 (2022).
33. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567–580 (2001).
34. Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research* **51**, W397–W403 (2023).
35. Domazet-Lošo M, Široki T, Šimičević K, Domazet-Lošo T. Macroevolutionary dynamics of gene family gain and loss along multicellular eukaryotic lineages. *Nature Communications* **15**, 2663 (2024).
36. Xu Y, Guo Y. Less Is More, Natural Loss-of-Function Mutation Is a Strategy for Adaptation. *Plant Communications* **1**, 100103 (2020).
37. Liu W, Jones AL, Gosse HN, Lawrence KS, Park S-W. Validation of the Chemotaxis of Plant Parasitic Nematodes Toward Host Root Exudates. *Journal of Nematology* **51**, 1–10 (2019).
38. Chen J, Li Z, Lin B, Liao J, Zhuo K. A *Meloidogyne graminicola* Pectate Lyase Is Involved in Virulence and Activation of Host Defense Responses. *Frontiers in Plant Science* **12**, 651627 (2021).

39. Dong W*, et al.* Changes in cell wall components and polysaccharide-degrading enzymes in relation to differences in texture during sweetpotato storage root growth. *Journal of Plant Physiology* **254**, 153282 (2020).

40. Harris PJ. Chapter 3 - Cell-wall Polysaccharides of Potatoes. In: *Advances in Potato Chemistry and Technology* (eds Singh J, Kaur L). Academic Press (2009).

41. You J*, et al.* Identification of cytochrome P450 gene family and functional analysis of HgCYP33E1 from *Heterodera glycines*. *Frontiers in Plant Science* **14**, 1219702 (2023).

42. Raza A, Williams AR, Abeer MM. Importance of ABC transporters in the survival of parasitic nematodes and the prospect for the development of novel control strategies. *Pathogens* **12**, 755 (2023).

43. Razem FA, Bernards MA. Reactive oxygen species production in association with suberization: evidence for an NADPH‐dependent oxidase. *Journal of Experimental Botany* **54**, 935–941 (2003).

44. Vieira P, Gleason C. Plant-parasitic nematode effectors — insights into their diversity and new tools for their identification. *Current Opinion in Plant Biology* **50**, 37–43 (2019).

45. Chen L*, et al.* Multi-copy alpha-amylase genes are crucial for *Ditylenchus destructor* to parasitize the plant host. *PLOS ONE* **15**, e0240805 (2020).

46. van der Maarel MJEC, van der Veen B, Uitdehaag JCM, Leemhuis H, Dijkhuizen L. Properties and applications of starch-converting enzymes of the α-amylase family. *Journal of Biotechnology* **94**, 137–155 (2002).

47. Okuyama M, Saburi W, Mori H, Kimura A. α-Glucosidases and α-1,4-glucan lyases: structures, functions, and physiological actions. *Cellular and Molecular Life Sciences* **73**, 2727–2751 (2016).

48. Lyu R*, et al.* Engineering properties of sweet potato starch for industrial applications by biotechnological techniques including genome editing. **22**, 9533 (2021).

49. Lacaze A, Joly DL. Leaf and tuber treatments with PAMPs trigger organ-specific responses in potato. *Canadian Journal of Plant Pathology* **44**, 115–127 (2022).
50. Li H, Durbin R. Genome assembly in the telomere-to-telomere era. *Nature Reviews Genetics* **25**, 658-670 (2024).
51. Tintori SC, Sloat SA, Rockman MV. Rapid isolation of wild nematodes by Baermann Funnel. *Journal of Visualized Experiments*, e63287 (2022).
52. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
53. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540–552 (2000).
54. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277 (2000).
55. Ronquist F*, et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* **61**, 539–542 (2012).
56. Nylander J. MrModeltest V2. Program Distributed by the Author. *Bioinformatics* **24**, 581–583 (2004).
57. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nature Methods* **17**, 155–158 (2020).
58. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
59. Vurture GW*, et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
60. Walker BJ*, et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One* **9**, e112963 (2014).

61. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27**, 737–746 (2017).

62. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757–763 (2011).

63. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* **2**, lqaa026 (2020).

64. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. In: *Gene prediction: methods and protocols* (ed Kollmar M). Springer New York (2019).

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

66. El-Gebali S*, et al.* The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427–D432 (2018).

67. Berriman M, Coghlan A, Gordon D. Contamination screening of parasitic worm genome assemblies. *Protocol Exchange*, (2018).

68. Altschul SF*, et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

69. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).

70. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).

71. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **5**, 4.10.11–14.10.14 (2004).

72. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907–915 (2019).

73. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
74. Koonin EV*, et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology* **5**, R7 (2004).
75. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research* **51**, W115–W121 (2023).
76. Krzywinski M*, et al.* Circos: an information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645 (2009).
77. Consortium T*Ce*S. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* **282**, 2012–2018 (1998).
78. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).
79. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
80. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490 (2010).

Acknowledgements

We thank Professor Xinyue Cheng from Beijing Normal University for the help and support in this research. This work was supported by grants from the National Key R&D Program of China (2023YFD1400400), the Hainan Provincial Sanya Yazhou Bay Science and Technology Innovation Joint Project (ZDYF2025GXJS140), and the Regional Collaborative Innovation Project of Gansu Academy of Agricultural Sciences (2024GAAS04).

Author contributions

B.X. and R.L. conceived and designed the research. Z.Z., R.L., and J.W. performed the data

analysis and interpretation. H.L., H.Z., and N.L. prepared the experimental materials and carried out the sequencing. Z.Z. wrote the manuscript, R.L. and B.X. revised it. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary Figures 1–3 and Supplementary Data 1–7 (https://doi.org/10.5281/zenodo.18540296).

Fig. 1 | Phylogenetic relationships, host symptoms, and ITS1 secondary structures of *Ditylenchus destructor* haplotypes.

a The Bayesian 25% majority-rule consensus tree inferred from ITS-rDNA sequences of *D. destructor* and related species under the GTR+G model of nucleotide substitution. **b** Representative external and internal symptoms observed on host storage organs under field conditions. Potato tubers infected with DdB and GSAUP, sampled from Dingxi, Gansu Province; sweet potato storage roots infected with BazhouSP, sampled from Bazhou, Hebei Province. Both sets of images were captured and provided by Bingyan Xie. **c–e** The predicted secondary structures of ITS1 from *D. destructor* isolates, with minimum free energy (MFE; ΔG, in kcal mol^{-1}) indicated. BazhouSP, which lacks helix H9, is classified as ITS1 Haplotype A. DdB, which contains helix H9, corresponds to the previously reported ITS1 Haplotype B [18]. GSAUP, also containing helix H9, is identified as ITS1 Haplotype C. Structural classification and haplotype designations are based on the frameworks proposed by Subbotin et al. [16] and Ni et al. [18].

Fig.2 | Functional annotations of nematode genes.

a KOG functional classification of genes across seven nematode genomes. Genes are grouped into three major KOG categories: information storage and processing, metabolism, and cellular processes and signaling, illustrating conserved and diversified biological roles among species. **b** The bubble plot displays 37 Pfam protein domain families that show significant differences in gene counts in pairwise comparisons between haplotypes. Bubble size indicates the number of genes associated with each protein domain in each haplotype. These domains represent key protein functions, such as signaling, enzymatic activity, and transcriptional regulation, reflecting the functional diversity of nematode genomes. **c** Stacked bar chart displaying the overall CAZy family composition in seven nematode genomes. GT and GH dominate the CAZy repertoire, while PL, CE, CBM, and AA classes show variable abundance. **d** The bubble plot shows carbohydrate-active enzyme (CAZy) families GT31 and PL3 with significantly different gene counts between haplotypes (Fisher's exact test, $p < 0.05$). Bubble size represents the number of genes in each family for the corresponding haplotype.

Fig. 3 | Directional whole-genome coverage and macro-synteny among *Ditylenchus destructor* haplotypes.

a Genome-wide coverage analysis based on pairwise whole-genome alignments among Haplotypes A (A1, Dd01, BazhouSP), B (DdB) and C (GSAUP) isolates. Alignments were performed using BLASTN (E-value ≤ 1e-5). The percentages represent the extent of genome covered by each comparison, providing an estimate of genomic similarity and directional differences among haplotypes. **b–h** Synteny visualization using Circos plots, generated from BLASTN alignment results (E-value ≤ 1e-5). Only scaffolds ≥500 kb and syntenic blocks ≥2 kb are displayed to highlight large-scale genomic rearrangements clearly.

Fig. 4 | Phylogenetic relationships, orthologous group distributions, and gene family changes among *Ditylenchus destructor* haplotypes and related species.

a Maximum likelihood phylogeny of *D. destructor* and related nematodes based on a concatenated alignment of 800 single-copy core protein-coding genes. Bootstrap support values are calculated from 500 replicates using RAxML-NG v1.2.2 [30]. Host sources from which *D. destructor* isolates were obtained are indicated in parentheses. At the branch of each lineage, the '+' number indicates family birth events, and the '−' number indicates family death events, which represent gene family changes for the lineages. For example, the birth of 3,557 gene families (+3,557) and the death of 53 gene families (–53) are found in *D. destructor* compared with protein families in *D. dipsaci*. While the birth of 3,800 gene families (+3,800) and the death of 466 gene families (–466) are found in *D. dipsaci* compared with *D. destructor*. Among *D. destructor* isolates, Haplotype A isolates show the birth of 1,100 gene families (+1,100) and the death of 84 families (–84), whereas Haplotypes B and C display the birth of 524 gene families (+524) and the death of 112 families (–112). These family dynamics may highlight molecular innovations or reductions associated with ecological adaptation and functional diversification. **b–c** Venn diagrams showing the distribution of orthologous groups among four *Ditylenchus* spp. In panel b, Haplotype A is represented by the union of predicted proteomes of BazhouSP, A1, and Dd01; in panel c, by BazhouSP alone. Haplotype B (DdB), Haplotype C (GSAUP), and *D. dipsaci* (Ddi) are shown in both Venn diagrams. Orthologous groups and associated gene counts are identified using OrthoFinder v2.4.0 [29].

Fig. 5 | Comparative secretome analysis of *Ditylenchus destructor* haplotypes.

a Stacked bar chart showing protein categories per isolate: (i) predicted secreted proteins (SignalP +, TMHMM −), (ii) secreted membrane proteins (SignalP +, TMHMM +), and (iii) other proteins (SignalP −). **b−c** Figure b and c show Venn diagrams of secretome orthologous groups across three *Ditylenchus* haplotypes. In panel b, Haplotype A is represented by the union of the predicted secretomes of BazhouSP, A1, and Dd01; in panel c, by BazhouSP alone. Haplotype B (DdB) and Haplotype C (GSAUP) are included in both panels. **d** UpSet plot showing shared and unique

secretome orthologous cluster intersections across six nematode isolates, generated using OrthoVenn3 [34]. Vertical bars represent the number of orthologous clusters in each intersection set, with the connected dots below indicating which isolates are included in each intersection. Horizontal bars (left) show the total number of clusters in each isolate. **e** A maximum likelihood tree built from the concatenated alignment of 17 strictly single-copy secretome orthologous clusters (JTT+CAT model), confirming the divergence between haplotypes isolated from sweet potato (A) and those isolated from potato (B and C). **f** Pfam domains of predicted secreted proteins are extracted from the full Pfam annotation dataset. Bubble plot showing protein domains with significantly different gene counts between haplotypes ($p < 0.05$, Fisher's exact test). Bubble size indicates the number of genes associated with each protein domain in each haplotype. **g** Pairwise Fisher's exact tests ($p < 0.05$) are used to compare gene counts between haplotypes. The bubble plot shows the CAZy families PL3 and GH31 with significantly different gene counts between haplotypes. Bubble size represents the number of genes in each family in the corresponding haplotype.

Table 1. Statistics of the whole-genome assemblies of *Ditylenchus destructor*

Haplotype	**A**			**B**	**C**
Feature	**BazhouSP**	**Dd01[a]**	**A1[b]**	**DdB**	**GSAUP**
Scaffold sizes (Mb)	139.4	111.1	133.8	156.8	120.9
Scaffold number	1,236	1,761	79	1,724	1,697
Scaffold N50 (kb)	782.0	561.0	32,249.5	356.1	388.5
Longest scaffold (Mb)	12.0	3.6	44.2	8.4	11.5
Contig sizes (Mb)	139.4	108.8	133.7	149	120.9
Contig number	1,236	3,886	441	1,724	1,697
Contig N50 (kb)	782.0	84.2	962.3	356.1	388.5
GC content (%)	37.4	36.8	37.14	35.9	36.7

Complete BUSCOs (%)	91.8	88.8	92.7	88.8	91.1
Repeat sequences (%)	30.17	19.50	28.14	20.67	17.31
Number of predicted genes	21,283	13,938	19,165	24,056	21,651

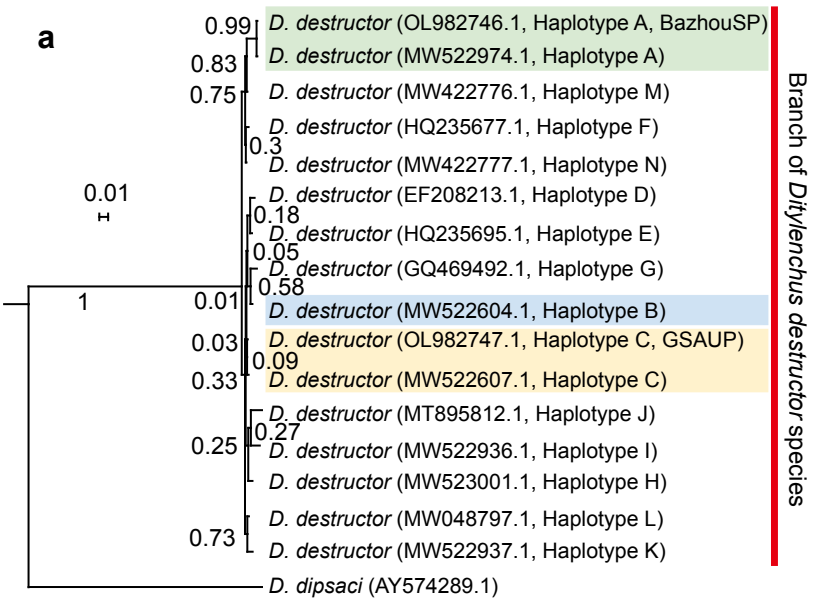
a Dd01, Zheng et al., 2016; b A1, Yang et al., 2025.

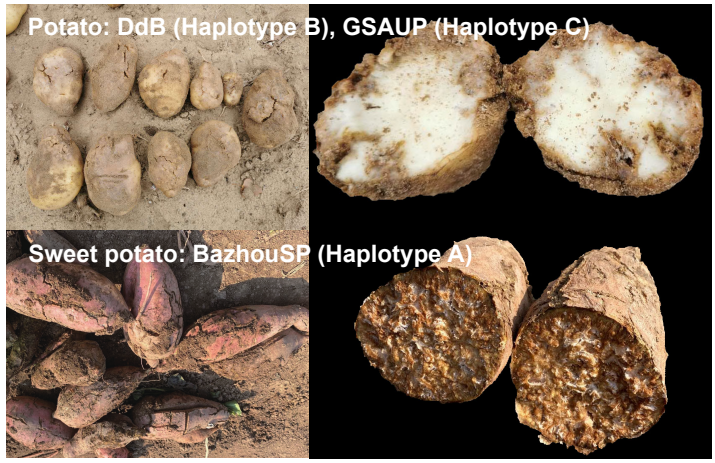
Editor's Summary

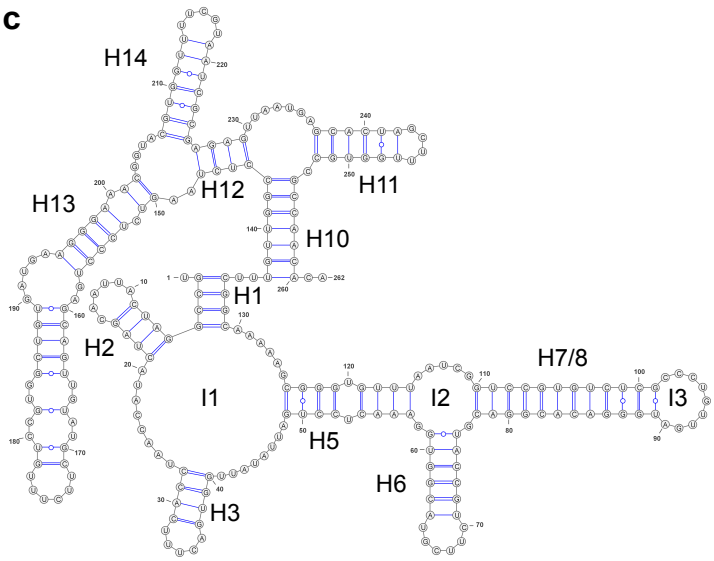
The genetic basis of *Ditylenchus destructor* haplotypes with host preferences remains unclear. This study investigates the genomes of three distinct haplotypes and identifies key genomic variations that may underlie host adaptation.

Peer review information

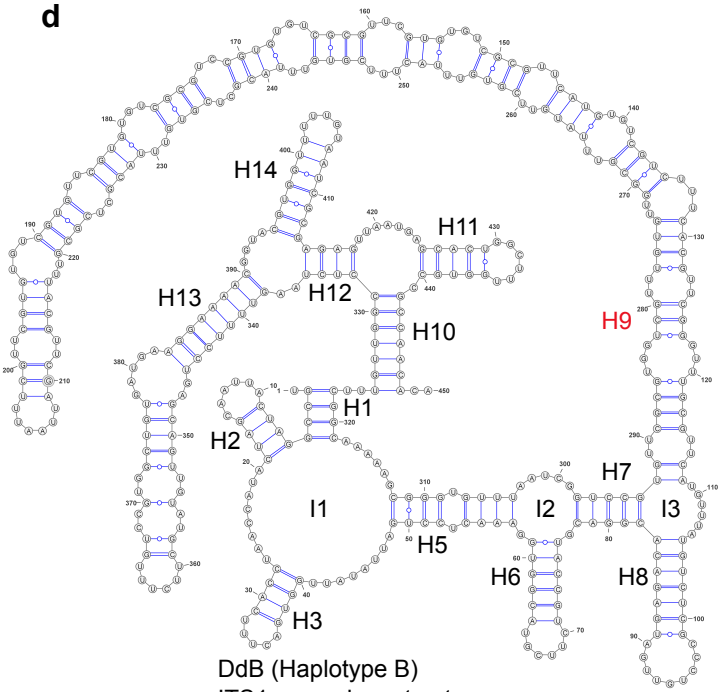
Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Rosie Bunton-Stasyshyn & David Favero.



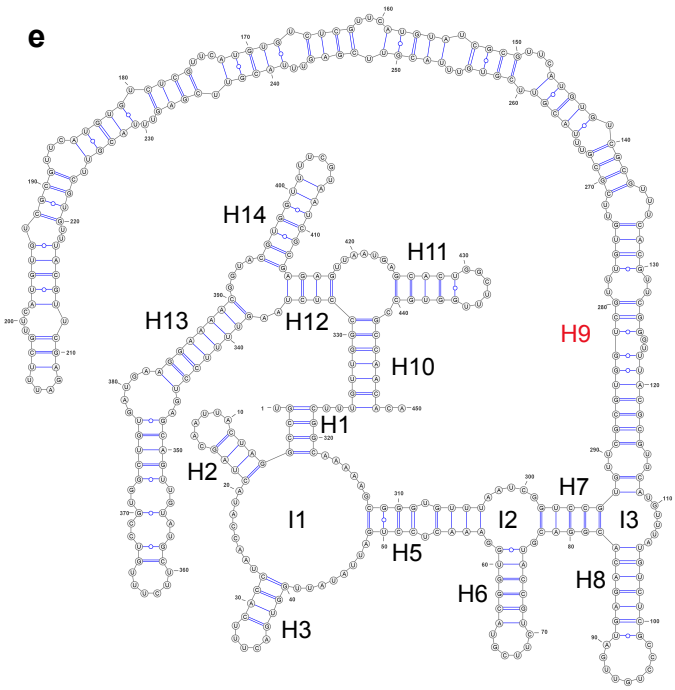




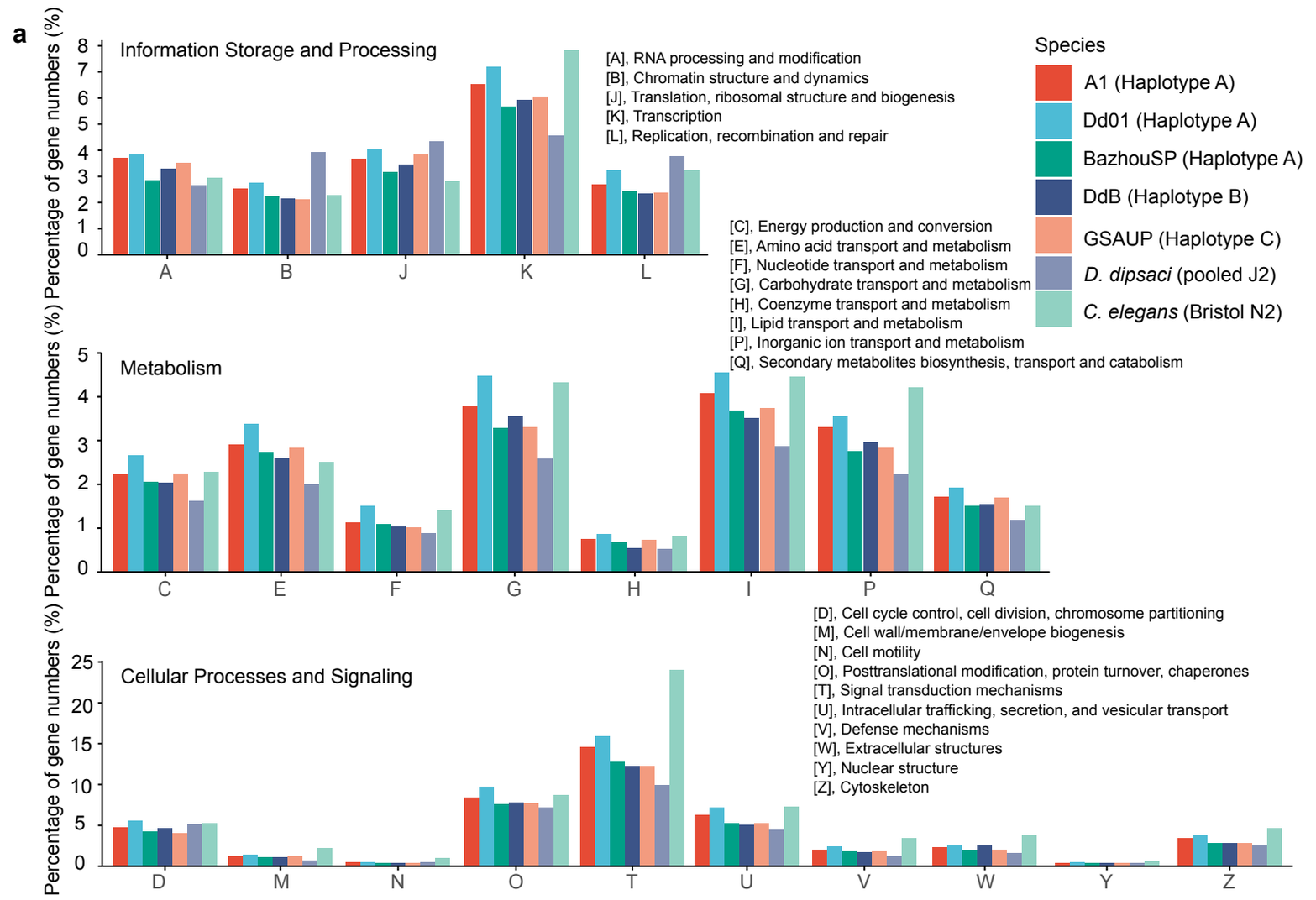
BazhouSP (Haplotype A)
ITS1 secondary structure
Minimum Free Energy (MFE): –94.50 kcal mol^{-1}
Displayed structure energy: –89.88 kcal mol^{-1}

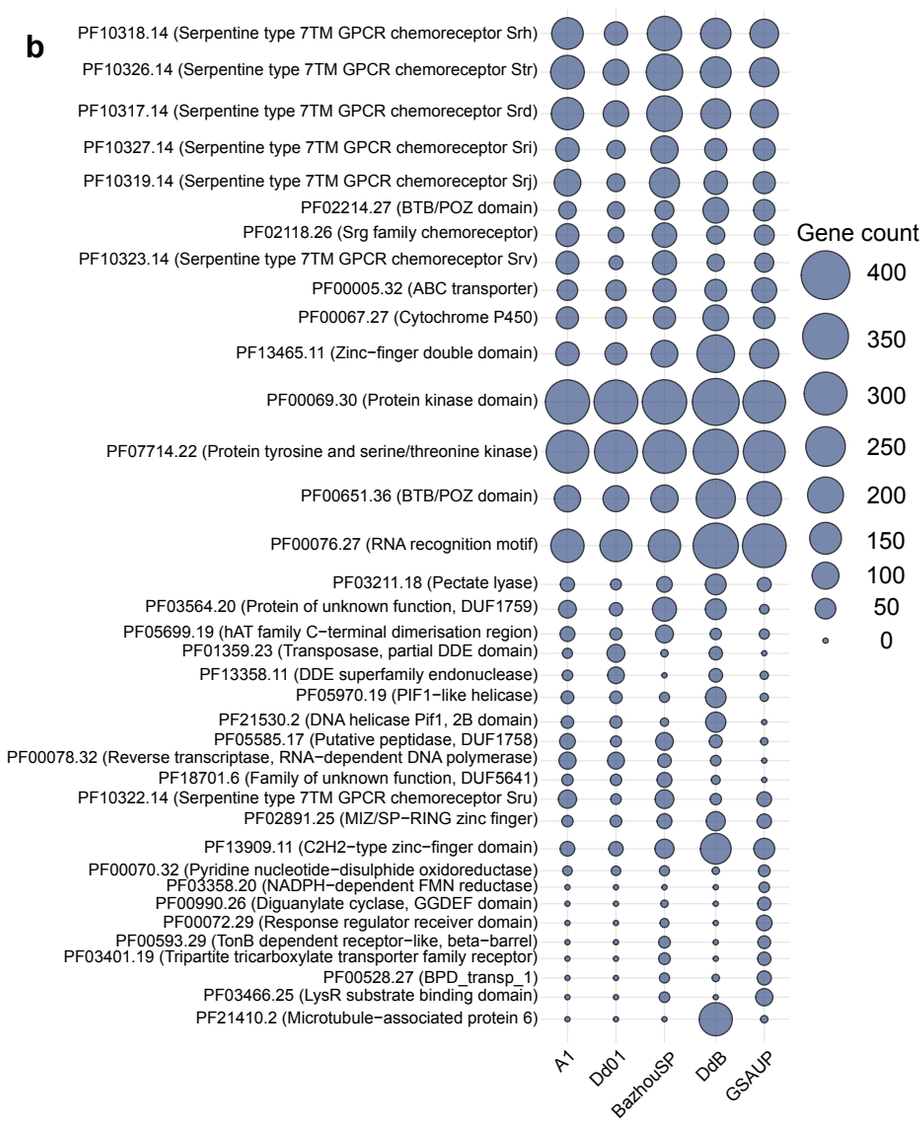


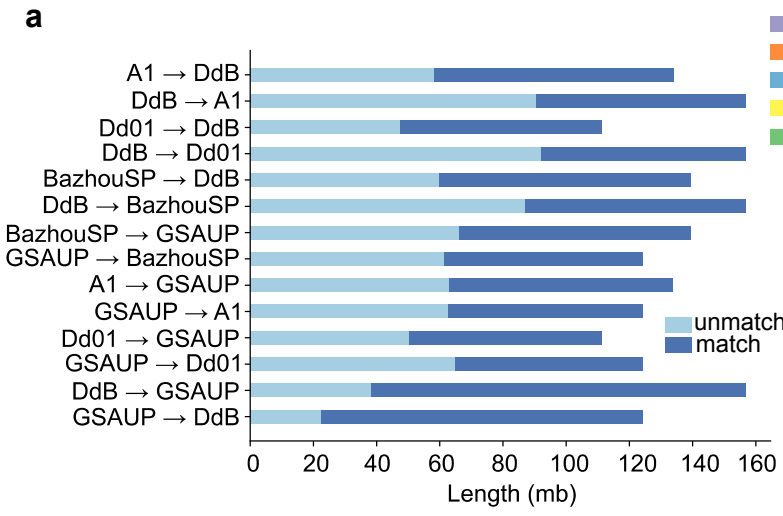
DdB (Haplotype B)
ITS1 secondary structure
Minimum Free Energy (MFE): –165.60 kcal mol^{-1}
Displayed structure energy: –160.08 kcal mol^{-1}

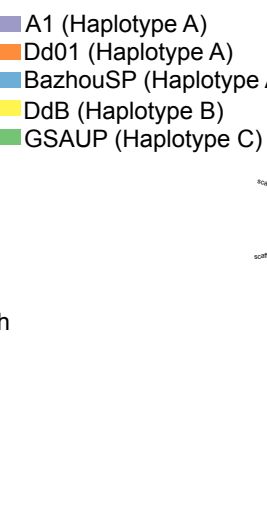


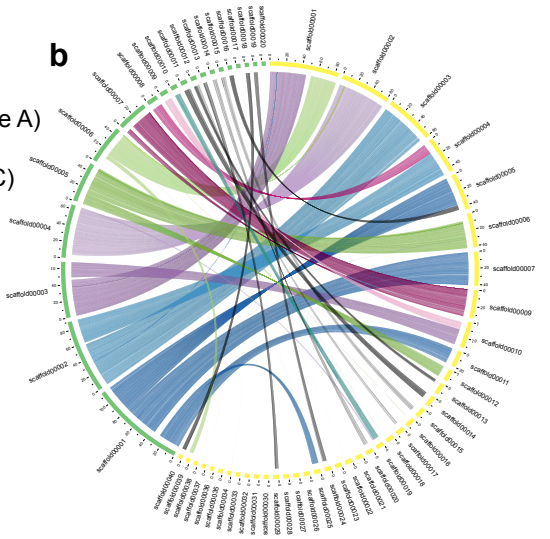
GSAUP (Haplotype C)
ITS1 secondary structure
Minimum Free Energy (MFE): –160.60 kcal mol^{-1}
Displayed structure energy: –155.08 kcal mol^{-1}


a
Information Storage and Processing
[A], RNA processing and modification
[B], Chromatin structure and dynamics
[J], Translation, ribosomal structure and biogenesis
[K], Transcription
[L], Replication, recombination and repair
Species
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)
C. elegans (Bristol N2)
[C], Energy production and conversion
[E], Amino acid transport and metabolism
[F], Nucleotide transport and metabolism
[G], Carbohydrate transport and metabolism
[H], Coenzyme transport and metabolism
[I], Lipid transport and metabolism
[P], Inorganic ion transport and metabolism
[Q], Secondary metabolites biosynthesis, transport and catabolism
Metabolism
Cellular Processes and Signaling
[D], Cell cycle control, cell division, chromosome partitioning
[M], Cell wall/membrane/envelope biogenesis
[N], Cell motility
[O], Posttranslational modification, protein turnover, chaperones
[T], Signal transduction mechanisms
[U], Intracellular trafficking, secretion, and vesicular transport
[V], Defense mechanisms
[W], Extracellular structures
[Y], Nuclear structure
[Z], Cytoskeleton
Percentage of gene numbers (%)

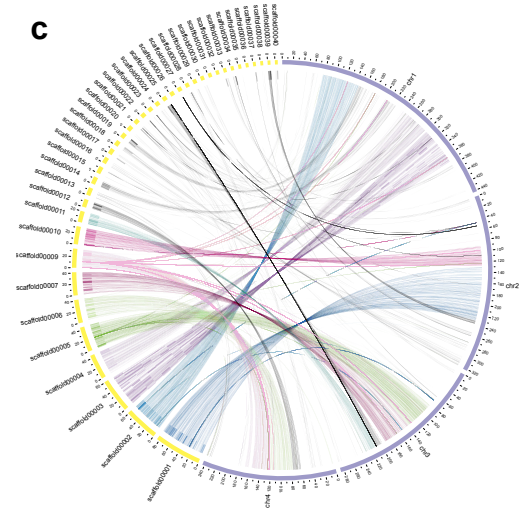

b
PF10318.14 (Serpentine type 7TM GPCR chemoreceptor Srh)
PF10326.14 (Serpentine type 7TM GPCR chemoreceptor Str)
PF10317.14 (Serpentine type 7TM GPCR chemoreceptor Srd)
PF10327.14 (Serpentine type 7TM GPCR chemoreceptor Sri)
PF10319.14 (Serpentine type 7TM GPCR chemoreceptor Srj)
PF02214.27 (BTB/POZ domain)
PF02118.26 (Srg family chemoreceptor)
PF10323.14 (Serpentine type 7TM GPCR chemoreceptor Srv)
PF00005.32 (ABC transporter)
PF00067.27 (Cytochrome P450)
PF13465.11 (Zinc−finger double domain)
PF00069.30 (Protein kinase domain)
PF07714.22 (Protein tyrosine and serine/threonine kinase)
PF00651.36 (BTB/POZ domain)
PF00076.27 (RNA recognition motif)
PF03211.18 (Pectate lyase)
PF03564.20 (Protein of unknown function, DUF1759)
PF05699.19 (hAT family C−terminal dimerisation region)
PF01359.23 (Transposase, partial DDE domain)
PF13358.11 (DDE superfamily endonuclease)
PF05970.19 (PIF1−like helicase)
PF21530.2 (DNA helicase Pif1, 2B domain)
PF05585.17 (Putative peptidase, DUF1758)
PF00078.32 (Reverse transcriptase, RNA−dependent DNA polymerase)
PF18701.6 (Family of unknown function, DUF5641)
PF10322.14 (Serpentine type 7TM GPCR chemoreceptor Sru)
PF02891.25 (MIZ/SP−RING zinc finger)
PF13909.11 (C2H2−type zinc−finger domain)
PF00070.32 (Pyridine nucleotide−disulphide oxidoreductase)
PF03358.20 (NADPH−dependent FMN reductase)
PF00990.26 (Diguanylate cyclase, GGDEF domain)
PF00072.29 (Response regulator receiver domain)
PF00593.29 (TonB dependent receptor−like, beta−barrel)
PF03401.19 (Tripartite tricarboxylate transporter family receptor)
PF00528.27 (BPD_transp_1)
PF03466.25 (LysR substrate binding domain)
PF21410.2 (Microtubule−associated protein 6)
A1
Dd01
BazhouSP
DdB
GSAUP
Gene count
c
CAZy Count
GT
GH
PL
AA
CBM
CE
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)
C. elegans (Bristol N2)
d
Gene count
GT31
PL3



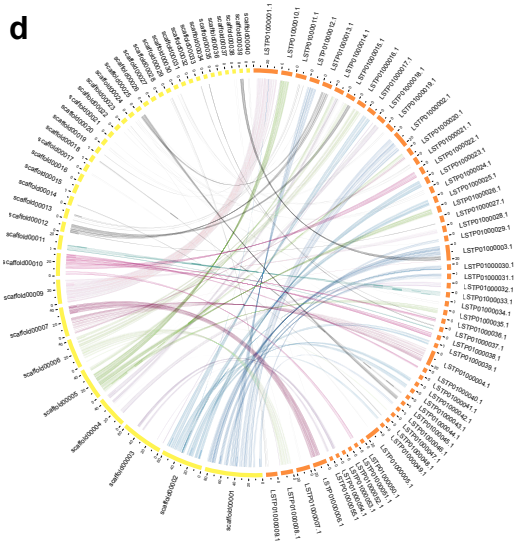




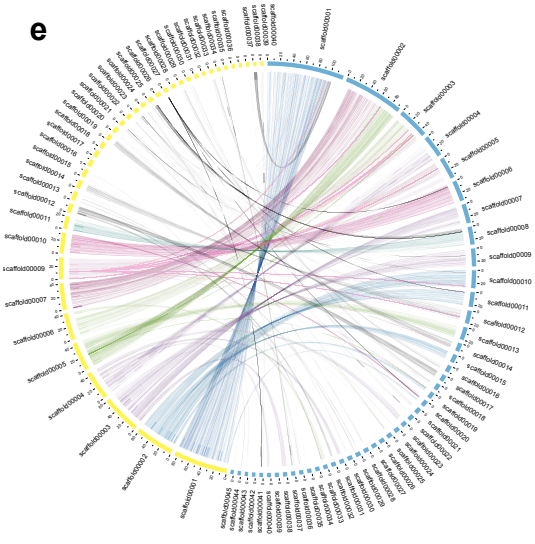
DdB (Haplotype B) vs GSAUP (Haplotype C)



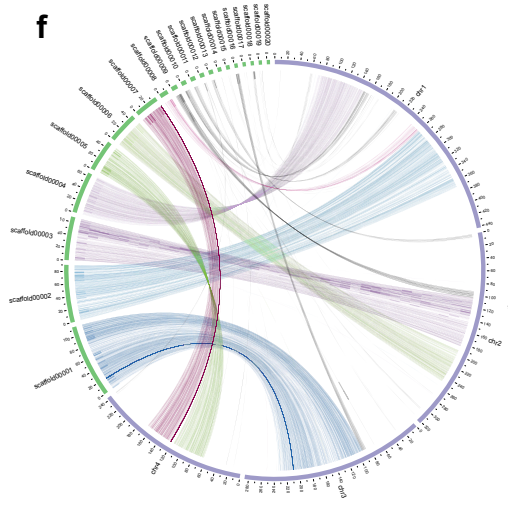
A1 (Haplotype A) vs DdB (Haplotype B)



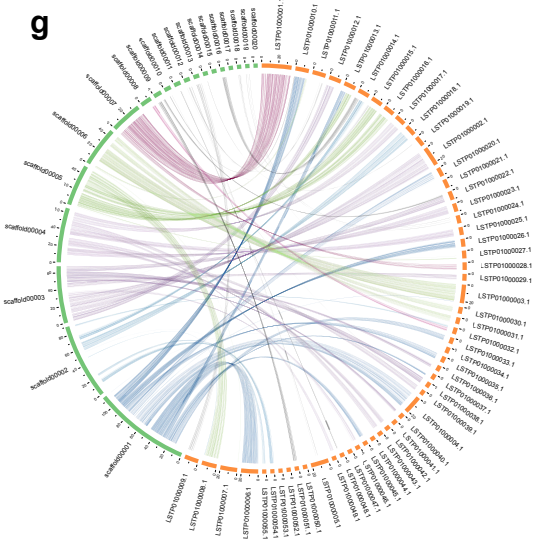
Dd01 (Haplotype A) vs DdB (Haplotype B)



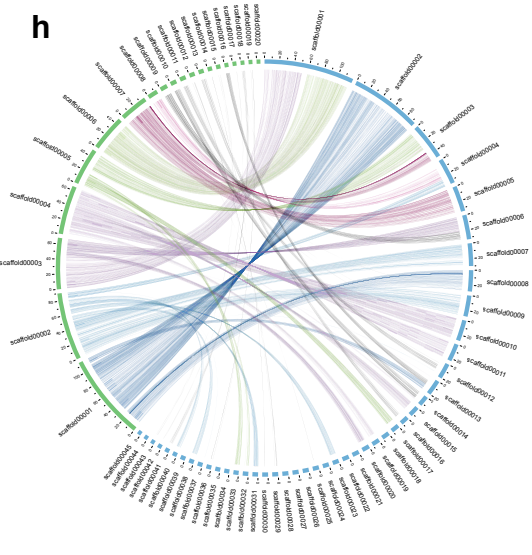
BazhouSP (Haplotype A) vs DdB (Haplotype B)



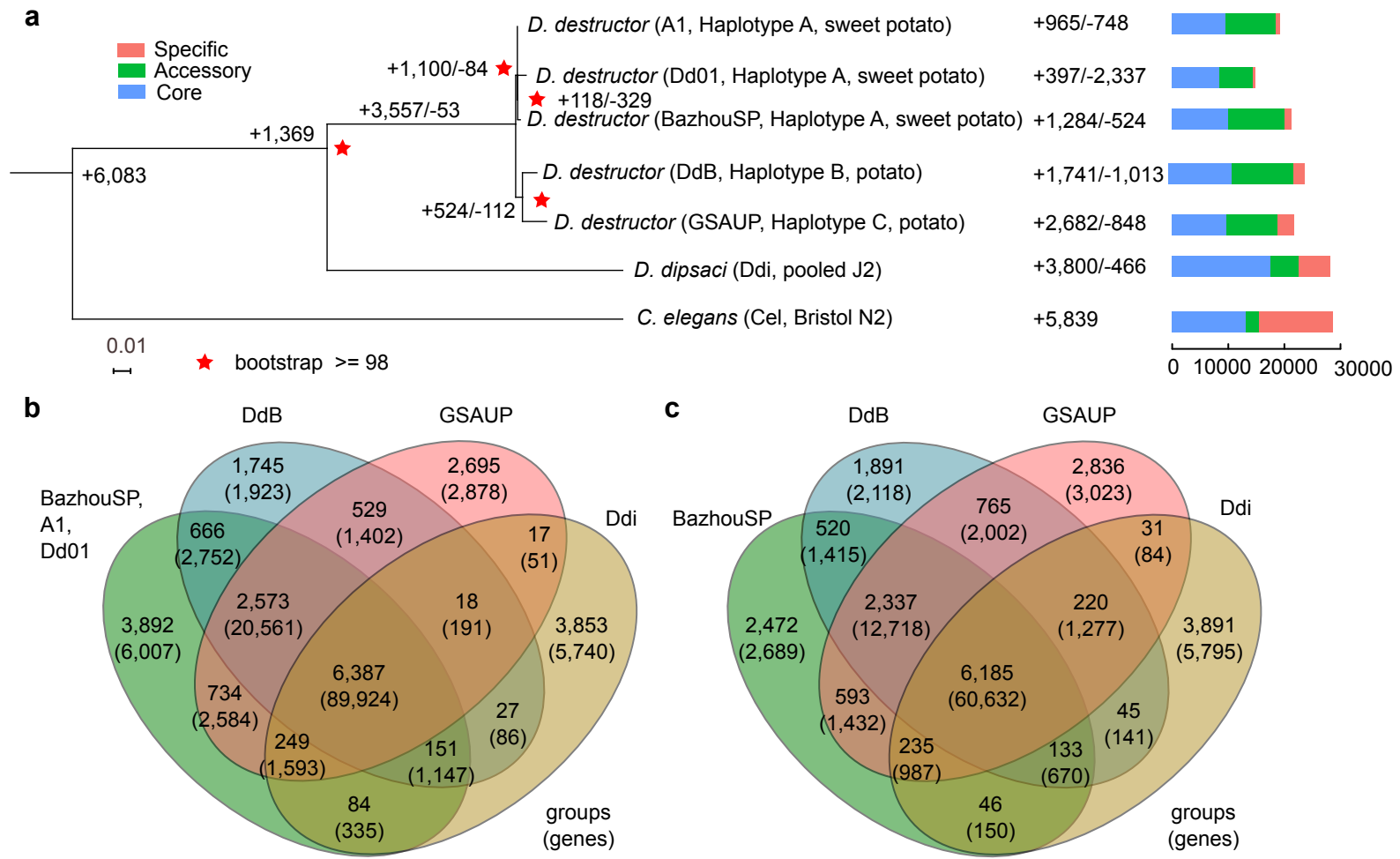
A1 (Haplotype A) vs GSAUP (Haplotype C)

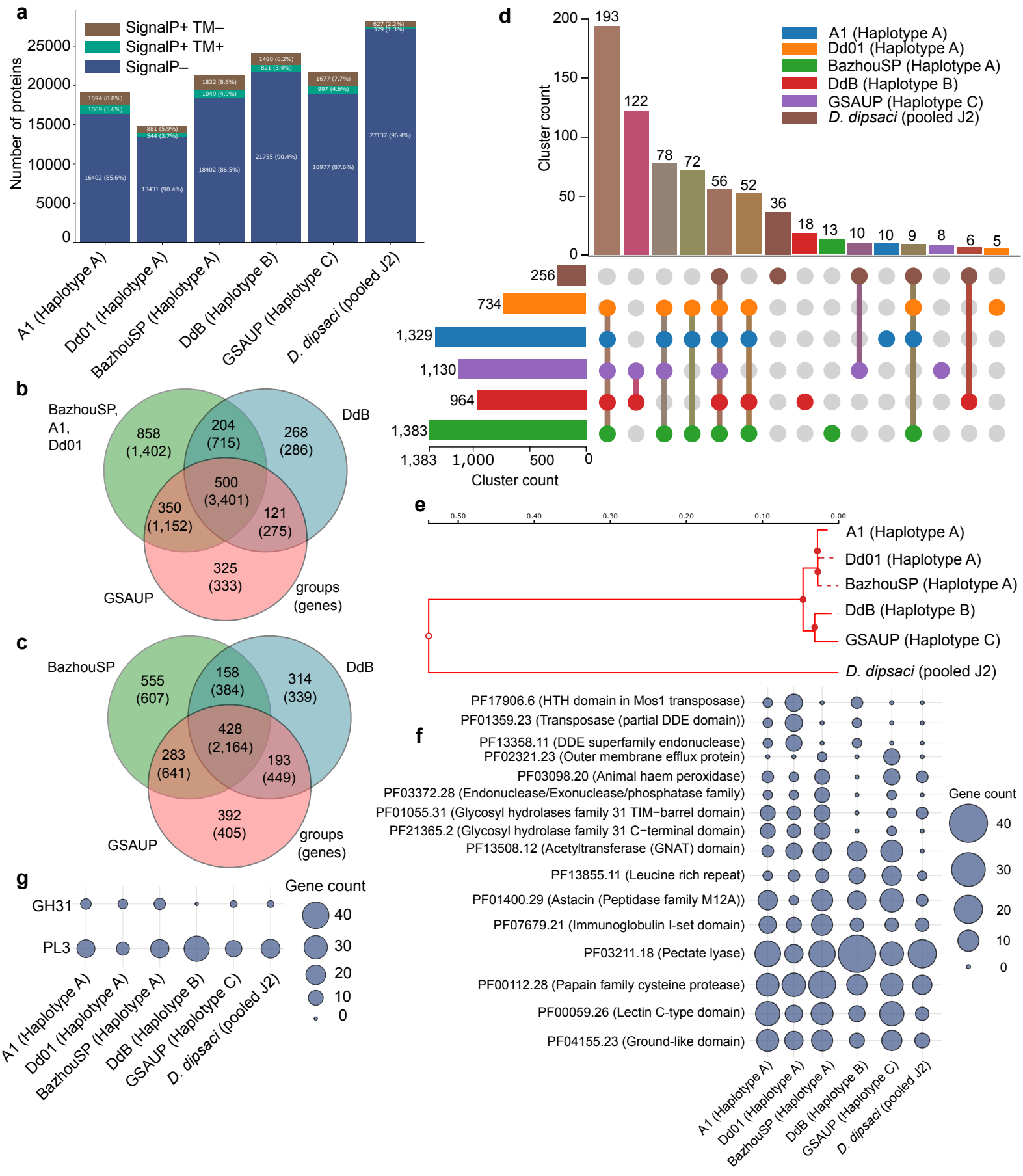


Dd01 (Haplotype A) vs GSAUP (Haplotype C)



BazhouSP (Haplotype A) vs GSAUP (Haplotype C)


a
Specific
Accessory
Core
+6,083
+1,369
+3,557/-53
+1,100/-84
+118/-329
+524/-112
D. destructor (A1, Haplotype A, sweet potato)
D. destructor (Dd01, Haplotype A, sweet potato)
D. destructor (BazhouSP, Haplotype A, sweet potato)
D. destructor (DdB, Haplotype B, potato)
D. destructor (GSAUP, Haplotype C, potato)
D. dipsaci (Ddi, pooled J2)
C. elegans (Cel, Bristol N2)
+965/-748
+397/-2,337
+1,284/-524
+1,741/-1,013
+2,682/-848
+3,800/-466
+5,839
0.01
bootstrap >= 98
0
10000
20000
30000
b
DdB
GSAUP
Ddi
groups (genes)
BazhouSP, A1, Dd01
1,745 (1,923)
2,695 (2,878)
529 (1,402)
666 (2,752)
17 (51)
2,573 (20,561)
18 (191)
3,892 (6,007)
3,853 (5,740)
6,387 (89,924)
734 (2,584)
27 (86)
249 (1,593)
151 (1,147)
84 (335)
c
DdB
GSAUP
Ddi
groups (genes)
BazhouSP
1,891 (2,118)
2,836 (3,023)
765 (2,002)
520 (1,415)
31 (84)
2,337 (12,718)
220 (1,277)
2,472 (2,689)
3,891 (5,795)
6,185 (60,632)
593 (1,432)
45 (141)
235 (987)
133 (670)
46 (150)


a
SignalP+ TM–
SignalP+ TM+
SignalP–
Number of proteins
25000
20000
15000
10000
5000
0
1694 (8.8%)
1069 (5.6%)
16402 (85.6%)
881 (5.9%)
544 (3.7%)
13431 (90.4%)
1832 (8.6%)
1049 (4.9%)
18402 (86.5%)
1480 (6.2%)
821 (3.4%)
21755 (90.4%)
1677 (7.7%)
997 (4.6%)
18977 (87.6%)
627 (2.2%)
379 (1.3%)
27137 (96.4%)
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)
b
BazhouSP, A1, Dd01
DdB
GSAUP
858 (1,402)
204 (715)
268 (286)
500 (3,401)
350 (1,152)
121 (275)
325 (333)
groups (genes)
c
BazhouSP
DdB
GSAUP
555 (607)
158 (384)
314 (339)
428 (2,164)
283 (641)
193 (449)
392 (405)
groups (genes)
d
Cluster count
200
150
100
50
0
193
122
78
72
56
52
36
18
13
10
10
9
8
6
5
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)
256
734
1,329
1,130
964
1,383
1,383
1,000
500
0
Cluster count
e
0.50
0.40
0.30
0.20
0.10
0.00
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)
f
PF17906.6 (HTH domain in Mos1 transposase)
PF01359.23 (Transposase (partial DDE domain))
PF13358.11 (DDE superfamily endonuclease)
PF02321.23 (Outer membrane efflux protein)
PF03098.20 (Animal haem peroxidase)
PF03372.28 (Endonuclease/Exonuclease/phosphatase family)
PF01055.31 (Glycosyl hydrolases family 31 TIM−barrel domain)
PF21365.2 (Glycosyl hydrolase family 31 C−terminal domain)
PF13508.12 (Acetyltransferase (GNAT) domain)
PF13855.11 (Leucine rich repeat)
PF01400.29 (Astacin (Peptidase family M12A))
PF07679.21 (Immunoglobulin I-set domain)
PF03211.18 (Pectate lyase)
PF00112.28 (Papain family cysteine protease)
PF00059.26 (Lectin C-type domain)
PF04155.23 (Ground-like domain)
Gene count
40
30
20
10
0
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)
g
GH31
PL3
Gene count
40
30
20
10
0
A1 (Haplotype A)
Dd01 (Haplotype A)
BazhouSP (Haplotype A)
DdB (Haplotype B)
GSAUP (Haplotype C)
D. dipsaci (pooled J2)