Comparative genomics among three cyst nematode species reveals distinct evolutionary histories among effector families and an irregular distribution of effector-associated promoter motifs

Joris J.M. van Steenbrugge[1*], Sven van den Elsen[1], Martijn Holterman[1,2], Jose L. Lozano-Torres[1], Vera Putker[1], Peter Thorpe[3], Aska Goverse[1], Mark G. Sterken[1], Geert Smant[1] and Johannes Helder[1]

[1] Laboratory of Nematology, Wageningen University & Research, Wageningen, The Netherlands

[2] Solynta, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands

[3] University of St. Andrews, School of Medicine, Medical & Biological Sciences, North Haugh, St Andrews, United Kingdom

***corresponding author**: joris.vansteenbrugge@wur.nl

Abstract

Potato cyst nematodes (PCNs), an umbrella term used for two species, *Globodera pallida* and *G. rostochiensis,* belong worldwide to the most harmful pathogens of potato. Pathotype-specific host plant resistances are an essential handle for PCN control. However, the poor delineation of *G. pallida* pathotypes hampers the efficient use of available host plant resistances. Long-read sequencing technology allowed us to generate a new reference genome of *G. pallida* population D383 and, as compared to the current reference, the new genome assembly is 42 times less fragmented. For comparison of diversification patterns of six effector families between *G. pallida* and *G. rostochiensis*, an additional reference genome was generated for an outgroup, the beet cyst nematode *Heterodera schachtii* (IRS population). Large evolutionary contrasts in effector family topologies were observed. While VAPs diversified before the split between the three cyst nematode species, the families GLAND5 and GLAND13 only expanded in PCN after their separation from the genus *Heterodera*. Although DNA motifs in the promoter regions thought to be involved in the orchestration of effector expression ('DOG boxes') were present in all three cyst nematode species, their presence is not a necessity for dorsal gland-produced effectors. Notably, DOG box dosage was only loosely correlated with expression level of individual effector variants. Comparison of the *G. pallida* genome with those of two other cyst nematodes underlined the fundamental differences in evolutionary history between effector families. Re-sequencing of PCN populations with deviant virulence characteristics will allow for the linking of these characteristics with the composition of the effector repertoire as well as for the mapping of PCN diversification patterns resulting from extreme anthropogenic range expansion.

Words: 262 (ME: 'about 250 words')

1. INTRODUCTION

Worldwide, affordable food and feed production depends on large-scale monocropping. For practical and economic reasons crop homogeneity in terms of yield quality and quantity is essential. At the same time, such systems are intrinsically vulnerable to damage by pests and pathogens. The highest susceptibility for biotic stressors is found in genetically homogeneous crops. Potato is the third most important staple food (Birch et al., 2012), and in most production systems clonally-propagated seed potatoes are used as starting material. Such production systems cannot do without rigorous disease management. Potato cyst nematodes (PCN), the common name for actually two species, *Globodera pallida,* and *G. rostochiensis*, are among the primary yield-limiting potato pathogens worldwide. PCN co-evolved with potato in the Andes in South America (see e.g. (Plantard et al., 2008)), and proliferation of potato as a main crop outside of its native range was unintentionally paralleled by an enormous range expansion of PCN. For decades, PCN control has mainly been dependent on the application of nematicides. Due to the non-specific nature of these nematicides they have a highly negative impact on the environment, and their use is therefore either banned or severely restricted in most parts of the world. Currently, crop rotation and the use of resistant potato varieties are the main handles in PCN control. For economic reasons, the use of plant resistances is preferred over crop rotation. However, potato resistance genes such as *H1* (Toxopeus & Huijsman, 1952), *Gro1-4* (Paal et al., 2004), *Gpa2* (Bakker et al., 2003), and *H2* (Strachan et al., 2019) are only effective against specific pathotypes of one of these PCN species. Nevertheless, there is no robust (molecular) pathotyping scheme that would allow for the matching of the genetic constitution of field populations with effective host plant resistance genes.

Effectors are proteins secreted by plant-pathogens that facilitate the manipulation of the physiology of the host plant and interfere with the host's innate immune response in favour of the invading organism (e.g. (Stergiopoulos & De Wit, 2009). Potato cyst nematode effectors have some peculiar characteristics. With at least one known exception, HYPs (Eves-van den Akker, Lilley, Jones,

& Urwin, 2014), most effectors are produced in large single-celled glands referred to as the subventral and dorsal esophageal glands. These glands empty into the pharynx lumen, and the lumen is connected to a hollow protrusible stylet with which nematodes pierce plant cell walls. Via the orifice of the stylet, effector proteins are transferred to the apoplast or the cytoplasm of infected host plant cells. It is noted that subventral gland effectors are functional during plant penetration. Subsequently, dorsal gland secretions are responsible for feeding site induction and the suppression of the host's innate immune system (Smant, Helder, & Goverse, 2018). It has been hypothesised that common transcription factors and/or common promoter motifs might facilitate coordinated expression of effectors during the infection process. Such mechanisms have been identified to regulate effector expression in plant pathogenic fungi and oomycetes (Jones et al., 2019; Roy, Kagda, & Judelson, 2013). Also, among plant-parasitic nematodes, promotor motifs have been identified upstream of effectors that could contribute to the orchestration of the infection process. In the case of the potato cyst nematode *G. rostochiensis,* a DOrsal Gland motif ('DOG box') was identified by (Eves-van den Akker et al., 2016). For the pinewood nematode *Bursaphelenchus xylophilus,* a regulatory promotor motif referred to as STATAWAARS was demonstrated to affect effector expression (Espada et al., 2018). Expression of several effectors of Clade I tropical root-knot nematodes (Tandingan De Ley et al., 2002) was suggested to be steered by a putative cis-regulatory motif 'Mel-DOG' (*Meloidogyne* DOrsal Gland, (Da Rocha et al., 2021)).

Most likely as a reflection of the co-evolution between nematodes and their host(s), effectors are typically encoded by multigene families showing family-specific levels of diversification (Masonbrink et al., 2019; Van Steenbrugge et al., 2021). Cyst nematodes harbour numerous effector families (see for instance (Pogorelko, Wang, Juvale, Mitchum, & Baum, 2020), and genome (re-)sequencing is a rigorous approach to generate comprehensive overview of PCN effector family compositions. The first genomes of *G. pallida* and *G. rostochiensis* were published by (Cotton et al., 2014) and (Eves-van den Akker et al., 2016). Although this constituted a major step forward, both genomes

are very fragmented, hampering effector family inventories. Recently, long-read technology allowed for the generation of a less fragmented and more complete reference genome for *G. rostochiensis* with - among other things - a 24 folds reduction of the number of scaffolds as compared to the initial reference genome (Van Steenbrugge et al., 2021). Here, we present a novel reference genome for the other potato cyst nematode, *G. pallida,* characterised by a 42-fold reduction of the number of scaffolds, together with a reference genome of the beet cyst nematode *Heterodera schachtii*. The *H. schachtii* genome was used to establish the polarity of effectorome contrasts between the two potato cyst nematode species. Detailed knowledge about the nematode's effector repertoire, a complete overview of variants within effector families, and insights in the evolutionary history of individual effector families are an essential ingredients for a molecular pathotyping scheme. Next to comparing effector diversification patterns, we investigated DOG box distribution and DOG box dosage (up to 16 DOG boxes were observed per putative promoter region) both within and among effectors families. Scrutinising putative effector promoter regions in three reference genomes allowed us to pinpoint the distribution of this putative regulatory motif among cyst nematode species, as well as among and within effector families. Subsequently, the impact of these new, long read technology-based reference genomes on ecological PCN diversification in general and on the development of effectorome-based pathotyping system for potato cyst nematodes in particular is discussed.

2. MATERIALS AND METHODS

2.1 DNA isolation and sequencing

Cysts from *G. pallida* line D383 were used as starting material for the collection of pre-parasitic second-stage juveniles (J2). J2's were concentrated, and sucrose centrifugation was used to purify the nematode suspension (Jenkins, 1964). After multiple rounds of washing the purified nematode suspension in 0.1 M NaCl, nematodes were resuspended in sterilised MQ water. Juveniles were lysed in a standard nematode lysis buffer with proteinase K and beta-mercaptoethanol at 60°C for 1 h as described in (Holterman et al., 2006). The lysate was mixed with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) (pH 8.0) following a standard DNA purification procedure, and finally, DNA was precipitated with isopropanol. After washing the DNA pellet with 70% ethanol for several times, it was resuspended in 10mM Tris-HCL (pH 8.0). *G. pallida* D383 DNA (10 -20 µg) was sequenced using Pacific Biosciences SMRT sequencing technology at Bioscience (Wageningen Research, Wageningen, The Netherlands). DNA (30 µg) from *H. schachtii* (IRS population) was isolated with a procedure similar to the one used for *G. pallida*, however DNA was precipitated using an ice-cold ethanol precipitation step (Jain et al., 2018). DNA fragments below 10 kb were depleted using a short read eliminator kit (Circulomics SS-100-121-01) and *H. schachtii* DNA (15 µg) was sequenced using Oxford Nanopore technology at NexOmics (Wuhan, China).

2.2 Genome assemblies and synteny

For *G. pallida* D383, raw PacBio reads, and for *H. schachtii* IRS, Oxford Nanopore reads were corrected to, in essence, merge haplotypes using the correction mode in Canu (Koren et al., 2017), by reducing the error rate to a maximum of 15% and the corrected coverage to a minimum of 200. Using wtdgb2 v2.3 (Ruan and Li, 2020), multiple initial genome assemblies were generated based on the corrected Nanopore reads while manually refining the parameters minimal read length, k-mer size, and minimal read depth. These parameters were optimised to generate an assembly close to

the expected genome size of *G. pallida* and *H. schachtii*. After optimisation, for *G. pallida*, a minimum read length cut-off of 5,000, minimal read depth of 6, and a k-mer size of 18 was used. To generate the assembly of *H. schachtii*, a minimum read length cut-off of 6,000, minimal read depth of 8, and a k-mer size of 21 were used. The remaining haplotigs were pruned from the assemblies using Purge Haplotigs v1.0.4 (Roach *et al.*, 2018). The contigs from the assemblies were then improved using FinisherSC v2.1 (Lam *et al.*, 2015) at default settings and scaffolded using SSPACE-Longread v1.1 (Boetzer and Pirovano, 2014). Gaps in the assemblies were then filled using GapFiller v1.0 (van Steenbrugge, 2021). For *G. pallida*, the resulting assembly was polished with Pacbio reads by three iterations of Arrow v2.3.3 (https://github.com/PacificBiosciences/GenomicConsensus), followed by five iterations of polishing with Pilon v1.23 (Walker *et al.*, 2014) using Pacbio and Illumina NovaSeq reads. Finally, the assembly of *H. schachtii* was polished with Medaka v1.4.1 model r941_prom_hac_g3210 using Nanopore sequencing reads. Repeat regions were softmasked using RepeatModeler v1.0.11 (https:// github.com/Dfam-consortium/RepeatModeler) and RepeatMasker v4.0.9 (Tarailo-Graovac and Chen, 2009). Using Braker v2.1.2 (Brůna *et al.*, 2021), gene annotations were predicted for both assemblies at default settings outputting gff3 annotations and aided by RNAseq data of different life stages (*G. pallida*: NCBI Bioproject PRJEB2896, *H. schachtii*: PRJNA767548). Full details on the generation of the genome assemblies and prediction of genes are available on https://github.com/Jorisvansteenbrugge/Gros_Gpal_Hsch. For *Globodera rostochiensis*, the Gr-Line19 genome assembly described in (Van Steenbrugge *et al.*, 2021) was used (NCBI GenBank assembly accession: GCA_018350325.1).

The synteny between the *G. rostochiensis*, *G. pallida*, and *H. schachtii* genomes was assessed by a progressive genome alignment using Mauve v2.4.0 (Darling *et al.*, 2004). The resulting alignments of regions larger than 1 kb and larger than 3kb were visualised in Circos v0.69-9 (Krzywinski *et al.*, 2009).

2.3 Identification of Effector Homologs

Effector gene families were identified in the genomes based on the predicted genes by BRAKER2. Homologs for Gr-1106/Hg-GLAND4 were identified using HMMer v3 with a custom HMM profile based on GenBank entries JQ912480 to JQ912513. SPRY homologs were identified with HMMer using a pre-calculated profile HMM in the PFAM database (PF00622). Homologs to CLE-like proteins were identified with a custom profile HMM-based on UniProt sequences (D1FNJ7, D1FNK5, D1FNJ9, D1FNK2, D1FNK8, D1FNK3, D1FNK0, D1FNK4). Homologs to Venom Allergen-like proteins (VAP) were identified with a custom profile HMM-based on Uniprot sequences (Q8MQ79, A0A0K3AST9, P90958, Q19348, A0A0K3AWG2, Q967G4, Q9BID5, A0A3Q8UEU8, Q963I7, B8LF85, A7X975, A0A7G7LJV8). Hg-GLAND5 and Hg-GLAND13 were identified with BLASTP searches with GenBank sequences KJ825716 and KJ825724, respectively, maintaining thresholds at 35% identity, 50% query coverage, and an E-value of 0.0001. Each effector homolog was tested for the presence of a signal peptide for secretion by Phobius v1.01 (Käll, Krogh, & Sonnhammer, 2007) and the presence of one or multiple DOG-box motifs in the promoter region using a custom script (available on GitHub https://github.com/Jorisvansteenbrugge/Gros_Gpal_Hsch).

For *G. rostochiensis*, effector annotations were used as described in (J. J. M. van Steenbrugge et al., 2021), except for CLE-like proteins. CLE variant Gros19_g16105.t1 was excluded since the gene model likely contains errors, and the exact location of this variant in the phylogenetic tree is therefore uncertain. Furthermore, the HMM scoring cut-off was lowered to 300 to include two more potential CLE variants.

2.4 Phylogeny

A multiple sequence alignment was generated for each effector gene family using Muscle v3.8.1551 based on gene coding sequences to infer the phylogeny of effector gene families between species. Next, phylogenetic trees were produced with RaxML v8.2.12 (Stamatakis, 2014) using the

GTRGAMMA model with 100 bootstrap replicates. The GTRGAMMA model was selected based on recommendations by ModelTest-NG v0.2.0 (Darriba et al., 2020). Finally, using Figtree V1.4.4 , the resulting trees were visualised and annotated.

2.5 DOGbox identification and Gene expression Analysis

RNAseq reads (add repository) were mapped to the *G. pallida* D383 genome, Reads xxx were mapped to the *G. rostochiensis* Gr-Line19 genome, and Reads xxx were mapped to the *H. schachtii* IRS genome using Hisat2 v2.1.0, generating alignments tailored for transcript assemblers (--dta option). Based on the alignments, abundances were estimated and normalized to transcripts per million (TPM) for the transcripts predicted by Braker2 using StringTie v2.1.7b (Kovaka et al., 2019). Expression values of SPRYSEC genes in all three cyst nematode species were extracted, and plotted against the number of DOG-box motifs in each gene. Spearman's rank-order correlation was used to determine relationship between TPM and the number of DOG-box motifs.

3. RESULTS

3.1. The use of long read sequence technologies to generate novel reference genomes

The mapping of diversification patterns of effector families requires a high-quality reference genome with preferably a low number of scaffolds and a minimal gap length. (Cotton et al., 2014) were the first to present a reference genome of the potato cyst nematode species *G. pallida*. For our specific purpose, *i.e.,* the generation of complete inventories of effector families, this reference genome was too fragmented, and the gap length was too large. PacBio long-read technology allowed us to generate a new reference genome from the *G. pallida* population D383 with a 42-fold reduction of scaffolds and a 21-fold reduction of the total gap length. As one of the results, the number of predicted genes increased from 16,403 to 18,813.

In addition, we assembled the genome sequence of the IRS population of beet cyst nematode *Heterodera schachtii*. This allowed us to compare effector family diversification among the two potato cyst nematode species, *G. pallida* and *G. rostochiensis*, and establish the polarity of these contrasts by using *H. schachtii* as an outgroup (both *Globodera* and *Heterodera* belong to the family Heteroderidae). The current genome size, 190 Mb, is slightly above the genome size estimated by flow cytometry, 160 - 170 Mb (Eves van den Akker, personal communication). It is noted that both the predicted number of genes and transcripts were about 50% higher in *H. schachtii* than in the two *Globodera species*.

Two synteny plots were generated based on the alignment of regions >1 kb and >3kb to compare the genomic organisation of the three cyst nematode species. Not unexpectedly, the two potato cyst nematode species share numerous >1 kb regions (Fig. 1A). In the *H. schachtii* genome, several homologous >1kb regions cluster together in genomic segments that span over 2 Mb (Fig. 1A, segments 1-8). It is noted that the homologous >1kb regions in these segments have equivalents in both *G. pallida* and *G. rostochiensis*. Alignment of >3kb fragments severely reduced the number of homologous regions among the three cyst nematode species (Fig. 1B). Nevertheless, the number of shared

217 >3kb regions between *G. pallida* and *G. rostochiensis* (N: 76) is considerably higher than the number
218 of shared regions between *H. schachtii* and *G. rostochiensis* (N: 23) (Fig. 1B).
219
220 3.2 Effector family selection
221 In our comparison between the three cyst nematode species, we concentrated on effectors. Cyst
222 nematodes were shown to harbour numerous effectors families (Pogorelko et al., 2020). Here we
223 concentrated on six effector families. For four of these families, one or more representatives are
224 known to be involved in the suppression of plant innate immune system (SPRYSEC (Diaz-Granados,
225 Petrescu, Goverse, & Smant, 2016; Mei et al., 2018), GLAND4 (also referred to as Gr-1106) (Barnes,
226 Wram, Mitchum, & Baum, 2018), GLAND5 (also referred to as G11A06) (Yang et al., 2019), and VAP
227 (Wilbers et al., 2018)). CLE (Wang et al., 2021) is an intriguing effector family involved in feeding site
228 induction, and the GLAND13 (Danchin, Guzeeva, Mantelin, Berepiki, & Jones, 2016) family is essential
229 in the hydrolysis of plant sugars once they are taken up by the nematode.
230
231 3.3. SPRYSECs
232 SPRYSEC is an acronym for a family of secreted effectors with an SP1a/RYanodine receptor domain.
233 This family was recently shown to be highly diverged in the potato cyst nematode *G. rostochiensis* (J.
234 J. M. van Steenbrugge et al., 2021). Fig. 2 shows a phylogenetic tree with (supposedly) all SPRYSECs
235 present in the three cyst nematode species under investigation. The number of paralogs in *G. pallida,*
236 *G. rostochiensis,* and *H. schachtii* is respectively 24, 60 and 13. Despite the poor backbone resolution
237 of the SPRYSEC tree, two moderately supported SPRYSEC clades (A and B) could be distinguished.
238 Clade A comprises SPRYSEC variants exclusively from the two potato cyst nematode species, and *G.*
239 *pallida* SPRYSEC paralogs are interspersed with *G. rostochiensis* SPRYSEC variants. SPRYSEC Clade A is
240 characterised by a dosage of 0 – 6 DOG box elements. Clade B harbours fewer SPRYSEC paralogs than
241 Clade A (27 *versus* 35 in Clade A). Notably, Gr19_g7942 is not preceded by a signal peptide for secre-

tion, whereas three DOG box elements are present in the promoter region directly upstream of this paralog. Clade B is characterised by a mix of SPRYSEC variants solely originating from *G. pallida* and *G. rostochiensis*. Compared to Clade A, Clade B is typified by an overall higher DOG box dosage (on average, 1.7 and 5.2 DOG boxes per paralog). Up to 16 DOG box elements were identified in the promoter regions of paralogs in Clade B.

The more basal part of the SPRYSEC tree (Fig. 2, part C) harbours, next to paralogs from the two potato cyst nematode species, all 13 SPRYSEC variants from *H. schachtii*. None of the promoter regions of the 35 SPRYSEC paralogs in this part of the SPRYSEC tree harbours DOG boxes. Both *H. schachtii* and *G. rostochiensis* harbour SPRYSEC paralogs with at least one transmembrane domain (gene ID in italics with lighter colour).

3.4. GLAND4

The number of GLAND4 (also referred to as Gr-1106) paralogs in Gr-Line19, Gp-D383, and Hs-IRS is respectively 10, 9, and 15. The phylogenetic analysis yields a tree with a well-supported backbone (Fig. 3) showing a clear separation between the outgroup *H. schachtii*, and both *Globodera* species. While the GLAND4 paralogs in Gr-Line19 and Gp-D383 show little divergence, they end up in separate species-specific clusters. On the other hand, Hs-IRS paralogs show more intraspecific diversification. All but two *G. pallida* paralogs (Gpal_D383_g17346.t1 and Gpal_D383_g13669) contain a signal peptide for secretion. For all but one of the GLAND4 genes in Gr-Line19, the promoter region included a DOG-box motif, while promoter regions of only four GLAND4 genes in Gp-D383 and one in Hs-IRS contained such a motif.

3.5. GLAND5

With 13 homologs, the GLAND5 effector family (also referred to as G11A06) was significantly less diversified in Gr-Line19 than in Gp-D383 and Hs-IRS with respectively 25 and 27 paralogs. In all three

267 species, the majority of the GLAND5 paralogs harbour a signal peptide for secretion. It is worth not-
268 ing that a relatively high percentage of the GLAND5 paralogs in Gr-Line19 was not preceded by a
269 signal peptide for secretion (23%). In contrast, in *H. schachtii* and *G. pallida,* respectively, 88.9% and
270 92% of the paralogs comprised a signal peptide. The phylogenetic analysis (Fig. 4) shows that
271 GLAND5 is a diversified gene family. Several well-supported branching events define a set of sub-
272 clades that either exclusively comprises *H. schachtii* or contain GLAND5 variants from both potato
273 cyst nematode species in the more distal branches. Even though the GLAND5 paralogs Gr-Line19 and
274 Gp-D383 occur together in individual subclades, no obvious sets of potential orthologs between the
275 two species could be identified. In *H. schachtii*, 82% of the paralogs contain at least one DOG-box
276 motif in the promoter region. Out of the three GLAND5 paralogs without a signal peptide for secre-
277 tion, two (Hs-IRS_g6495.t1 and Hs-IRS_g22438.t1) had at least one DOG box motif in their promoter
278 region. DOG boxes were less prominently present among the *G. rostochiensis* and *G. pallida* GLAND5
279 variants (39% and 20% of the paralogs).
280
281 3.6. VAP (Venom Allergen-like Proteins)
282 The levels of diversification in the VAP effector family were highly comparable between the three
283 cyst nematode species. In Gp-D383, Hs-IRS, and Gr-Line19, respectively, 8, 8, and 9 VAP paralogs
284 were identified. The phylogenetic analysis resulted in a tree with a well-supported backbone (Fig. 5).
285 The tree contains three clusters (Fig. 5, boxes A, B, C) with a high level of diversification between the
286 clusters. At the base of the tree, a small cluster of 4 *H. schachtii* paralogs (Fig. 6, box C) is present that
287 all lack a signal peptide for secretion. Box B harbours Gp-D383 and Gr-Line19 VAP paralogs, of which
288 all but two lack a signal peptide for secretion. Three sub-clusters are present in Box A: one with ex-
289 clusively Gr-Line19 variants, a second one with just Gp-D383 variants, and the third with an ortholo-
290 gous pair between Gr-Line19 and Gp-D383. In the largest cluster at the top of the tree (Fig. 5, box A),

291 VAP paralogs of all three species are present, including the only secreted VAP variant in Hs-IRS with a
292 DOG-box motif in the promoter region.
293
294 3.7. CLE (CLAVATA3/ESR-related peptides)
295 With 16 variants, the CLE-like effector family is considerably more diversified in *H. schachtii* than in *G.*
296 *pallida* and *G. rostochiensis* (respectively, 10 and 11 paralogs). Analysis of the CLE families on the
297 three cyst nematode species resulted in a phylogenetic tree with a reasonably well-resolved back-
298 bone (Fig. 6). It is worth noting that nearly all variants are united in species-specific clusters, and in
299 this sense, the CLE diversification patterns resemble the patterns observed for the GLAND4 (Gr-1106)
300 family (Fig. 3). Whereas Gp-D383 and Gr-Line19 are characterised by similar-sized, moderately di-
301 verged clusters of CLE paralogs, the CLE family has much more diverged in *H. schachtii.*
302 In *G. rostochiensis*, two functional classes of CLE peptides have been described, CLE-1 and
303 CLE-4 (Lu et al. 2009). The CLE-1 class (Fig. 6) comprises two Gr-Line19 paralogs that show only dis-
304 tant homology to *G. pallida* CLEs. Similarly, four *G. rostochiensis* CLE's belonging to functional CLE
305 class 4 (Fig. 6) do not have clear equivalents in *G. pallida* and *H. schachtii.* Unlike all other effector
306 families investigated so far, all CLE variants from the three cyst nematode species are preceded by a
307 signal peptide for secretion. At the same time, none of them has a DOG-box motif in the promoter
308 region.
309
310 3.8. GLAND13 (Glycosyl Hydrolase Family 32 (GH32) members)
311 GLAND13 effectors investigated so far were identified in *G. pallida* and coded for invertases belong-
312 ing to glycosyl hydrolase family 32 (GH32). While these enzymes are not secreted into the plant, they
313 are essential as they catalyse the hydrolysis of the primary type of sugar the nematode takes up from
314 its host, sucrose (Danchin et al., 2016). This gene family shows a large difference in the number of
315 paralogs present in the three species; while Gr-Line19 and Gp-D383 harbour 10 and 7 paralogs, Hs-

316 IRS holds only 3 copies. In the phylogenetic tree (Fig. 7), the paralogs in *H. schachtii* are positioned at
317 the tree's base. In Box A, paralogs of Gr-Line19 and Gp-D383 are interspersed, while in Box B, all but
318 one paralogs are from Gr-Line19. In Gr-Line19, 70% of the GLAND13 paralogs comprise a signal pep-
319 tide for secretion, slightly lower percentages (67% and 57%) were observed in Hs-IRS and Gp-D383.
320 Variants showing high similarity to each of the five GLAND13 paralogs from *G. pallida* (population
321 Lindley; indicated as GPLIN_*number*) are indicated in Fig. 7.
322 For half of the GLAND13 effector variants in Gr-Line19, a DOG-box motif in the promoter region was
323 shown. One gene, Gr19_g13610, contained the motif two times. In *H. schachtii,* this motif was pre-
324 sent in 2 out of 3 genes, while in *G. pallida*, DOG boxes were found in one variant with a signal pep-
325 tide for secretion (Gpal_D383_g17582), and in one paralog without such a signal
326 (Gpal_D383_g09388). It is noted that these DOG box motifs were found in promoter regions of *G.*
327 *pallida* effectors that are not expressed in the dorsal gland (Danchin et al., 2016).
328
329 3.9 Effects of DOG-box dosage on SPRYSEC Expression
330 Although DOG-box motifs in the promoter regions of effector variants are present in many effector
331 families, their presence is not a necessity for the functioning of an effector family. For example, none
332 of the variants of the CLE family contained DOG-box motifs, regardless of the species (Fig. 6). Dorsal
333 gland-expressed effectors can therefore be expressed and secreted without the presence of DOG-
334 box motifs. This is further illustrated in Fig. 8A, this bar chart shows DOG box distribution of over the
335 six effector families. For the three cyst nematode species under investigation it demonstrates that
336 DOG boxes can be entirely absent (CLE), present in some species only (VAP, SPRYSEC), or present in
337 all species (GLAND4, GLAND5, GLAND13). The ample presence of DOG boxes in the diversified
338 SPRYSEC family prompted us to investigate whether there is a correlation between the DOG-box
339 dosage and SPRYSEC expression levels. SPRYSEC effectors from all three cyst nematode species were
340 taking into account (Fig. 8B). A modest correlation (R^2 = 0.66) between the DOG-box dosage and the

341 expression levels based on RNA abundance is present for this family. Especially in *G. pallida* high ex-
342 pression levels of SPRYSEC variants can be reached in absence of DOG boxes in its promoter region
343 (Fig. 8B).
344

4. DISCUSSION

In our attempts to fundamentally understand the interaction between plant-parasitic nematodes and their hosts, the usefulness of high-quality reference genomes of these pathogens is beyond discussion. Keeping in mind the enormous impact of PCN in all major potato production regions of the world, it is not surprising that high priority was given to the sequencing of both the *G. pallida* (Cotton et al., 2014) and the *G. rostochiensis* (Eves-van den Akker et al., 2016) genome. This was done before long-read sequencing technologies became available. Although some research questions can very well be addressed with these reference genomes, less fragmented genomes are needed for studying effector diversification. Therefore, a new reference genome was generated from *G. pallida* population D383. As compared to the *G. pallida* Lindley genome assembly, this resulted in a 42-fold reduction in the number of scaffolds and a 24-times increase of the N50. In the comparison of the effectoromes of the two PCN species, we included a newly generated genome of the *H. schachtii* population IRS as an outgroup. It is noted that reference genomes from these obligatory sexually reproducing pathogens are actually population-based consensus genomes. Long read sequencing technologies require DNA from 10,000s genetically non-identical nematodes. While an individual of these diploid species could theoretically carry a maximum of two haplotypes per locus, a population has the potential to carry many more. It is essential to mine these haplotypes and assemble them into a single haploid assembly to generate a proper reference. This is not a trivial process and requires specialised bioinformatics software (Roach et al., 2018). As the sizes of the current genome assemblies are comparable to the genome sizes assessed by flow cytometry, and as the BUSCO duplication scores are rela-

367 tively low, we assume that the current genomes assemblies are a reasonable reflection of
368 their actual constitution.

369 4.1 Effector diversification

370 In our analyses we concentrated on six selected effector families, and this selection included
371 relatively widespread effector families such as the CLE, GLAND13 and the VAP family, as well
372 as families that appear to be cyst nematode lineage specific such as SPRYSEC, GLAND4 and
373 GLAND5. Although the protein architecture is distinct between lineages (see e.g. (Mitchum,
374 Wang, Wang, & Davis, 2012)), the CLE family - a category of effectors involved in feeding site
375 induction - were shown to be present as well in root-knot and reniform nematodes (Rutter
376 et al., 2014; Wubben, Gavilano, Baum, & Davis, 2015). GLAND13 effectors, members of gly-
377 cosyl hydrolase family 32, were shown to be present in a range of root knot and cyst nema-
378 todes species as well as in other plant-parasitic nematodes such as *Nacobbus aberrans* and
379 *Rotylenchus reniformis* (Danchin et al., 2016). The distribution of VAP within the phylum
380 Nematoda is even broader (Wilbers et al., 2018). Venom allergen-like proteins were discov-
381 ered in the animal parasite *Ancylostoma caninum* (Hawdon, Jones, Hoffman, & Hotez, 1996).
382 Later on it was isolated from the root-knot nematode *Meloidogyne incognita* (Ding, Shields,
383 Allen, & Hussey, 2000), and subsequently in a wide range of obligatory plant parasitic nema-
384 tode including various cyst nematode species. A number VAP variants were shown to be im-
385 plicated in the suppression of both PAMP triggered immunity and effector triggered immu-
386 nity (e.g. (Li et al., 2021) for the burrowing nematode *Radopholus similis*). Our effector fam-
387 ily selection also included families that (so far) appear to be specific to the cyst nematode
388 lineages. This includes SPRYSEC (Diaz-Granados et al., 2016), GLAND4 (also referred to as Gr-
389 1106), and GLAND5 (also referred to as G11A06). For all of these effector families it can be

said that at least a subset was shown to be involved in repression of the host plant immune system.

While comparing the overall diversification patterns of the six effector families under investigations, striking differences are observed. In case of SPRYSEC, GLAND5 (G11A06), and GLAND13 (GH32 members), virtually all *H. schachtii* paralogs appeared to be phylogenetically isolated from the *G. pallida* and *G. rostochiensis* effector family variants, while representatives from the two potato cyst nematode species were presented in mixed clusters. These results should be taken with some caution as the backbone resolution of these phylogenetic trees ranges from poor (SPRYSEC) to robust (GLAND5, GLAND13). These patterns suggest that SPRYSEC, GLAND5 and GLAND13 effectors started to diversify after the split between *Heterodera* and *Globodera.*

Effector families GLAND4 (Gr-1106) and CLE showed distinct diversification patterns; by far most paralogs are grouped in species-specific clusters. Keeping in mind that both effector families show a reasonable backbone resolution, we hypothesize that these effector families might have diverged after the split between *G. pallida* and *G. rostochiensis.*

Phylogenetic analysis of the VAP effector family in the three cyst nematode species revealed an opposite pattern as an almost complete mixtures of representative paralogs from the individual species was observed. VAPs constitute an exceptionally widespread effector family within the phylum Nematoda (Wilbers et al., 2018), and our results point at diversification of this family before the split between the cyst nematode genera *Globodera* and *Heterodera.*

4.2 Regulation of effector gene expression

412 Various stages of the parasitic life cycle of cyst nematodes such as plant invasion, feeding
413 site induction and feeding site maintenance require the carefully orchestrated expression of
414 distinct blends of effector proteins (Elashry et al., 2020; Thorpe et al., 2014). For some
415 obligatory plant-parasitic nematodes promoter elements have been identified that were
416 suggested to be involved in this orchestration (Da Rocha et al., 2021; Eves-van den Akker et
417 al., 2016). For the three cyst nematode species we showed the presence of a short DNA box
418 motif ('DOG box'; ATGCCA) in the promoter region of some members of some of the effector
419 families. The absence of DOG boxes in the CLE family, the scattered presence of DOG boxes
420 in the other 5 families and the loose correlation between DOG box dosage and expression
421 level, prompt us to conclude that DOG boxes might contribute to the orchestration of effec-
422 tor expression, but we see little evidence for a central role of this DNA motif in this process.
423 Further investigation is necessary to elucidate the function of DOG boxes in effector regula-
424 tion.
425 In case of plant pathogenic fungi, a few transcription factor have been identified that were
426 shown to steer effector expression. SIX Gene Expression 1 (Sge1), a conserved member of
427 Gti1/Pac2 protein family, was instrumental in the regulation of effector repression in a range
428 of fungal pathogens including *Verticillium dahlia* (Santhanam & Thomma, 2013), *Zymosepto-*
429 *ria tritici* (Mirzadi Gohari et al., 2014), and *F. oxysporum f. sp. cubense* (Hou et al., 2018). As
430 another example AbPf2 could be mentioned, a zinc cluster transcription factor from the ne-
431 crotrophic plant pathogen *Alternaria brassicicola.* Via a loss of function approach, this tran-
432 scription factor was shown to regulate the expression of eight putative effectors (Cho, Ohm,
433 Grigoriev, & Srivastava, 2013). Evidently, plant pathogenic fungi are only very distantly re-

434 lated to plant parasitic nematodes, and these examples should only be considered as an il-

435 lustration how effector expression is organized in other pant pathogen systems.

5. CONCLUSIONS

The potato cyst nematode *Globodera pallida* and its sibling species *G. rostochiensis* co-evolved with potato in the Andes in South America. These pathogens have been introduced unintentionally in all major potato growing regions in the world. Currently, PCN belongs to the most harmful pathogens in potato production systems, and as a result of this extreme anthropogenic range expansion potatoes worldwide can't be grown without adequate PCN management. For both *G. pallida* and *G. rostochiensis* host plant resistances belong currently to most powerful means to control these soil pathogens. However, their effectiveness depends on the proper matching between the genetic constitution of the PCN field population and the set of host pant resistances present in modern potato varieties. In 2014, (Niere, Krüssel, & Osmers, 2014) reported about *G. pallida* populations that could no longer be controlled by any of the currently used potato cultivars. This, combined with inherent imperfectness of the current *G. pallida* pathotyping system (e.g. (Phillips & Trudgill, 1983)), underlines the need for a new pathotyping system. Such a system will be based on distinctive effector variations present in any given PCN population. The availability of a high quality reference genome is a prerequisite for the development of such a system. We demonstrated that the quality of the *G. pallida* genome presented in this paper allows for the mapping of complete effector families. With this resource, re-sequencing data from pathotypically diverse *G. pallida* populations will provide insight in the ecological diversification of this extreme range expander, and enable the development of a new pathotyping system that will facilitate the targeted and durable use of precious host plant resistances against this notorious plant pathogen.

ACKNOWLEDGEMENTS

JvS, MH and SvdE were supported by a grant from the Applied and Technical Science domain (TTW) of the Netherlands Organization for Scientific Research (NWO) under grant no. 14708. PT received support from the University of St Andrews Bioinformatics Unit (AMD3BIOINF), funded by Wellcome Trust ISSF award 105621/Z/14/Z. MS benefitted from funding by a VENI grant (17282) from the NWO domain Applied and Engineering Sciences

AUTHOR CONTRIBUTIONS

SvdE performed the DNA extraction and library preparations for PCN. JL performed the DNA extraction and library preparations for *H. schachtii*. JvS, PT, and MH conceptualised the genome assembly pipeline. JvS and MH generated the genome assemblies. JvS, MS and JH conceptualised the comparative genomics analyses. JvS performed the comparative genomics / effectoromics and phylogenetic analyses. JvS and JH wrote the manuscript. JH and GS acquired the main part of the funding and supervised the project. PT, MS, AG, VP and GS substantially revised/commented on the manuscript. All author(s) read the final version of the manuscript and approved it.

476 **REFERENCES**

477 Bakker, E., Butterbach, P., Rouppe Van Der Voort, J., Van Der Vossen, E., Van Vliet, J., Bakker,
478 J., & Goverse, A. (2003). Genetic and physical mapping of homologues of the virus
479 resistance gene Rx1 and the cyst nematode resistance gene Gpa2 in potato.
480 *Theoretical and Applied Genetics, 106*(8), 1524-1531.
481 Barnes, S. N., Wram, C. L., Mitchum, M. G., & Baum, T. J. (2018). The plant-parasitic cyst
482 nematode effector GLAND4 is a DNA-binding protein. *Molecular Plant Pathology,*
483 *19*(10), 2263-2276. doi:10.1111/mpp.12697
484 Birch, P. R. J., Bryan, G., Fenton, B., Gilroy, E. M., Hein, I., Jones, J. T., . . . Toth, I. K. (2012).
485 Crops that feed the world 8: Potato: Are the trends of increased global production
486 sustainable? *Food Security, 4*(4), 477-508. doi:10.1007/s12571-012-0220-1
487 Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes
488 using long read sequence information. *BMC Bioinformatics, 15*(1), 211.
489 doi:10.1186/1471-2105-15-211
490 Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2:
491 automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS
492 supported by a protein database. *NAR Genomics and Bioinformatics, 3*(1), 108.
493 doi:10.1093/nargab/lqaa108
494 Cho, Y., Ohm, R. A., Grigoriev, I. V., & Srivastava, A. (2013). Fungal-specific transcription
495 factor AbPf2 activates pathogenicity in *Alternaria brassicicola*. *Plant Journal, 75*(3),
496 498-514. doi:10.1111/tpj.12217
497 Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., . . . Urwin, P. E.
498 (2014). The genome and life-stage specific transcriptomes of *Globodera pallida*
499 elucidate key aspects of plant parasitism by a cyst nematode. *Genome biology, 15*(3),
500 R43.
501 Da Rocha, M., Bournaud, C., Dazenière, J., Thorpe, P., Bailly-Bechet, M., Pellegrin, C., . . .
502 Danchin, E. G. J. (2021). Expression Dynamics Reveal the Parasitism Regulatory
503 Landscape ofn the Root-Knot Nematode *Meloidogyne incognita* and a Promoter
504 Motif Associated with Effector Genes. *Genes, 12*, 771.
505 doi:https://doi.org/10.3390/genes12050771
506 Danchin, E. G. J., Guzeeva, E. A., Mantelin, S., Berepiki, A., & Jones, J. T. (2016). Horizontal
507 Gene Transfer from Bacteria Has Enabled the Plant-Parasitic Nematode *Globodera*
508 *pallida* to Feed on Host-Derived Sucrose. *Molecular Biology and Evolution, 33*(6),
509 1571-1579. doi:10.1093/molbev/msw041
510 Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of
511 conserved genomic sequence with rearrangements. *Genome Research, 14*(7), 1394-
512 1403. doi:10.1101/gr.2289704
513 Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020).
514 ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein
515 Evolutionary Models. *Molecular Biology and Evolution, 37*(1), 291-294.
516 doi:10.1093/molbev/msz189
517 Diaz-Granados, A., Petrescu, A. J., Goverse, A., & Smant, G. (2016). SPRYSEC Effectors: A
518 Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity. *Front Plant Sci,*
519 *7*, 1575. doi:10.3389/fpls.2016.01575

Ding, X., Shields, J., Allen, R., & Hussey, R. S. (2000). Molecular cloning and characterisation of a venom allergen AG5-like cDNA from *Meloidogyne incognita*. *International Journal for Parasitology, 30*(1), 77-81. doi:10.1016/S0020-7519(99)00165-4

Elashry, A. M., Habash, S. S., Vijayapalani, P., Brocke-Ahmadinejad, N., Blümel, R., Seetharam, A., . . . Grundler, F. M. W. (2020). Transcriptome and Parasitome Analysis of Beet Cyst Nematode *Heterodera schachtii*. *Scientific Reports, 10*(1). doi:10.1038/s41598-020-60186-0

Espada, M., Eves-van den Akker, S., Maier, T., Paramasivan, V., Baum, T., Mota, M., & Jones, J. T. (2018). STATAWAARS: A promoter motif associated with spatial expression in the major effector-producing tissues of the plant-parasitic nematode *Bursaphelenchus xylophilus*. *BMC Genomics, 19*(1). doi:10.1186/s12864-018-4908-2

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G. J., Da Rocha, M., . . . Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology, 17*(1), 124. doi:10.1186/s13059-016-0985-1

Eves-van den Akker, S., Lilley, C. J., Jones, J. T., & Urwin, P. E. (2014). Identification and Characterisation of a Hyper-Variable Apoplastic Effector Gene Family of the Potato Cyst Nematodes. *PLoS Pathogens, 10*(9), e1004391. doi:10.1371/journal.ppat.1004391

Hawdon, J. M., Jones, B. F., Hoffman, D. R., & Hotez, P. J. (1996). Cloning and characterization of *Ancylostoma*-secreted protein: A novel protein associated with the transition to parasitism by infective hookworm larvae. *Journal of Biological Chemistry, 271*(12), 6672-6678. doi:10.1074/jbc.271.12.6672

Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., . . . Helder, J. (2006). Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular Biology and Evolution, 23*(9), 1792-1800.

Hou, X., An, B., Wang, Q., Guo, Y., Luo, H., & He, C. (2018). SGE1 is involved in conidiation and pathogenicity of *Fusarium oxysporum* f.sp. *cubense*. *Canadian Journal of Microbiology, 64*(5), 349-357. doi:10.1139/cjm-2017-0638

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., . . . Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology, 36*(4), 338-345. doi:10.1038/nbt.4060

Jenkins, W. R. (1964). A rapid centrifugal-flotation technique for separating nematodes from soil. *Plant Disease Reporter 48*(9), 48.

Jones, D. A. B., John, E., Rybak, K., Phan, H. T. T., Singh, K. B., Lin, S. Y., . . . Tan, K. C. (2019). A specific fungal transcription factor controls effector gene expression and orchestrates the establishment of the necrotrophic pathogen lifestyle on wheat. *Scientific Reports, 9*(1). doi:10.1038/s41598-019-52444-7

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research, 27*(5), 722-736. doi:10.1101/gr.215087.116

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology, 20*(1). doi:10.1186/s13059-019-1910-1

565 Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A.
566 (2009). Circos: An information aesthetic for comparative genomics. *Genome*
567 *Research, 19*(9), 1639-1645. doi:10.1101/gr.092759.109
568 Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane
569 topology and signal peptide prediction-the Phobius web server. *Nucleic Acids*
570 *Research, 35*(SUPPL.2), W429-W432. doi:10.1093/nar/gkm256
571 Lam, K. K., Labutti, K., Khalak, A., & Tse, D. (2015). FinisherSC: A repeat-aware tool for
572 upgrading de novo assembly using long reads. *Bioinformatics, 31*(19), 3207-3209.
573 doi:10.1093/bioinformatics/btv280
574 Li, J., Xu, C., Yang, S., Chen, C., Tang, S., Wang, J., & Xie, H. (2021). A Venom Allergen-Like
575 Protein, RsVAP, the first discovered effector protein of *Radopholus similis* that
576 inhibits plant defense and facilitates parasitism. *International Journal of Molecular*
577 *Sciences, 22*(9). doi:10.3390/ijms22094782
578 Lu, S. W., Chen, S., Wang, J., Yu, H., Chronis, D., Mitchum, M. G., & Wang, X. (2009).
579 Structural and functional diversity of CLAVATA3/ESR (CLE)-like genes from the potato
580 cyst nematode *Globodera rostochiensis*. *Molecular Plant-Microbe Interactions, 22*(9),
581 1128-1142. doi:10.1094/MPMI-22-9-1128
582 Masonbrink, R., Maier, T. R., Muppirala, U., Seetharam, A. S., Lord, E., Juvale, P. S., . . . Baum,
583 T. J. (2019). The genome of the soybean cyst nematode (*Heterodera glycines*) reveals
584 complex patterns of duplications involved in the evolution of parasitism genes. *BMC*
585 *Genomics, 20*(1). doi:10.1186/s12864-019-5485-8
586 Mei, Y., Wright, K. M., Haegeman, A., Bauters, L., Diaz-Granados, A., Goverse, A., . . .
587 Mantelin, S. (2018). The *Globodera pallida* SPRYSEC effector GpSPRY-414-2 that
588 suppresses plant defenses targets a regulatory component of the dynamic
589 microtubule network. *Frontiers in Plant Science, 9*. doi:10.3389/fpls.2018.01019
590 Mirzadi Gohari, A., Mehrabi, R., Robert, O., Ince, I. A., Boeren, S., Schuster, M., . . . Kema, G.
591 H. J. (2014). Molecular characterization and functional analyses of ZtWor1, a
592 transcriptional regulator of the fungal wheat pathogen *Zymoseptoria tritici*.
593 *Molecular Plant Pathology, 15*(4), 394-405. doi:10.1111/mpp.12102
594 Mitchum, M. G., Wang, X., Wang, J., & Davis, E. L. (2012) Role of nematode peptides and
595 other small molecules in plant parasitism. In*: Vol. 50. Annual Review of*
596 *Phytopathology* (pp. 175-195).
597 Niere, B., Krüssel, S., & Osmers, K. (2014). Auftreten einer außergewöhnlich virulenten
598 Population der Kartoffelzystennematoden. *Journal für Kulturpflanzen, 66*, 426-427.
599 Noon, J. B., Hewezi, T., Maier, T. R., Simmons, C., Wei, J. Z., Wu, G., . . . Baum, T. J. (2015).
600 Eighteen new candidate effectors of the phytonematode *Heterodera glycines*
601 produced specifically in the secretory esophageal gland cells during parasitism.
602 *Phytopathology, 105*(10), 1362-1372. doi:10.1094/PHYTO-02-15-0049-R
603 Paal, J., Henselewski, H., Muth, J., Meksem, K., Menéndez, C. M., Salamini, F., . . . Gebhardt,
604 C. (2004). Molecular cloning of the potato Gro1-4 gene conferring resistance to
605 pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a
606 candidate gene approach. *Plant Journal, 38*(2), 285-297. doi:10.1111/j.1365-
607 313X.2004.02047.x

608 Phillips, M. S., & Trudgill, D. L. (1983). Variations in the ability of *Globodera pallida* to
609 produce females on potato clones bred from *Solanum vernei* or *S. tuberosum* ssp.
610 *andigena* CPC 2802. *Nematologica, 29*(2), 217-226. doi:10.1163/187529283X00465
611 Plantard, O., Picard, D., Valette, S., Scurrah, M., Grenier, E., & Mugniéry, D. (2008). Origin
612 and genetic diversity of Western European populations of the potato cyst nematode
613 (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci.
614 *Molecular Ecology, 17*(9), 2208-2218. doi:10.1111/j.1365-294X.2008.03718.x
615 Pogorelko, G., Wang, J., Juvale, P. S., Mitchum, M. G., & Baum, T. J. (2020). Screening
616 soybean cyst nematode effectors for their ability to suppress plant immunity.
617 *Molecular Plant Pathology, 21*(9), 1240-1247. doi:10.1111/mpp.12972
618 Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig
619 reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics, 19*(1),
620 460. doi:10.1186/s12859-018-2485-7
621 Roy, S., Kagda, M., & Judelson, H. S. (2013). Genome-wide Prediction and Functional
622 Validation of Promoter Motifs Regulating Gene Expression in Spore and Infection
623 Stages of *Phytophthora infestans*. *PLoS Pathogens, 9*(3).
624 doi:10.1371/journal.ppat.1003182
625 Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature*
626 *Methods, 17*(2), 155-158. doi:10.1038/s41592-019-0669-3
627 Rutter, W. B., Hewezi, T., Maier, T. R., Mitchum, M. G., Davis, E. L., Hussey, R. S., & Baum, T.
628 J. (2014). Members of the *Meloidogyne* avirulence protein family contain multiple
629 plant ligand-like motifs. *Phytopathology, 104*(8), 879-885. doi:10.1094/PHYTO-11-13-
630 0326-R
631 Santhanam, P., & Thomma, B. P. H. J. (2013). *Verticillium dahliae* sge1 differentially regulates
632 expression of candidate effector genes. *Molecular Plant-Microbe Interactions, 26*(2),
633 249-256. doi:10.1094/MPMI-08-12-0198-R
634 Smant, G., Helder, J., & Goverse, A. (2018). Parallel adaptations and common host cell
635 responses enabling feeding of obligate and facultative plant parasitic nematodes.
636 *Plant Journal, 93*(4), 686-702. doi:10.1111/tpj.13811
637 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of
638 large phylogenies. *Bioinformatics, 30*(9), 1312-1313.
639 doi:10.1093/bioinformatics/btu033
640 Stergiopoulos, I., & De Wit, P. J. G. M. (2009) Fungal effector proteins. In*: Vol. 47. Annual*
641 *Review of Phytopathology* (pp. 233-263).
642 Strachan, S. M., Armstrong, M. R., Kaur, A., Wright, K. M., Lim, T. Y., Baker, K., . . . Hein, I.
643 (2019). Mapping the *H2* resistance effective against *Globodera pallida* pathotype Pa1
644 in tetraploid potato. *Theoretical and Applied Genetics, 132*(4), 1283-1294.
645 doi:10.1007/s00122-019-03278-4
646 Tandingan De Ley, I., De Ley, P., Vierstraete, A., Karssen, G., Moens, M., & Vanfleteren, J.
647 (2002). Phylogenetic analyses of *Meloidogyne* small subunit rDNA. *Journal of*
648 *Nematology, 34*(4), 319-327.
649 Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements
650 in genomic sequences. *Current Protocols in Bioinformatics*(SUPPL. 25), 4.10.11-
651 14.10.14. doi:10.1002/0471250953.bi0410s25

652 Thorpe, P., Mantelin, S., Cock, P. J. A., Blok, V. C., Coke, M. C., Eves-van den Akker, S., . . .
653 Jones, J. T. (2014). Genomic characterisation of the effector complement of the
654 potato cyst nematode *Globodera pallida*. *BMC Genomics, 15*(1), 923.
655 doi:10.1186/1471-2164-15-923
656 Toxopeus, H. J., & Huijsman, C. A. (1952). Genotypical background of resistance to
657 *Heterodera rostochiensis* in *Solanum tuberosum*, var. *andigenum* [1]. *Nature,*
658 *170*(4337), 1016. doi:10.1038/1701016b0
659 van Steenbrugge, J. J. M. (2021). Jorisvansteenbrugge/GapFiller: GROS Assembly version.
660 *Zenodo*. doi:http://doi.org/10.5281/zenodo.4627096
661 van Steenbrugge, J. J. M., van den Elsen, S., Holterman, M., Sterken, M. G., Thorpe, P.,
662 Goverse, A., . . . Helder, J. (2021). Comparative genomics of two inbred lines of the
663 potato cyst nematode *Globodera rostochiensis* reveals disparate effector family-
664 specific diversification patterns. *BMC Genomics, 22*(1). doi:10.1186/s12864-021-
665 07914-6
666 Van Steenbrugge, J. J. M., Van den Elsen, S., Holterman, M., Sterken, M. G., Thorpe, P.,
667 Goverse, A., . . . Helder, J. (2021). Comparative Genomics of two Inbred Lines of the
668 Potato Cyst Nematode *Globodera rostochiensis* reveals disparate Effector Family-
669 specific Diversification Patterns. *bioRxiv*.
670 doi:https://doi.org/10.1101/2021.03.15.435409
671 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., . . . Earl, A. M.
672 (2014). Pilon: An integrated tool for comprehensive microbial variant detection and
673 genome assembly improvement. *PLoS ONE, 9*(11), e112963.
674 doi:10.1371/journal.pone.0112963
675 Wang, J., Dhroso, A., Liu, X., Baum, T. J., Hussey, R. S., Davis, E. L., . . . Mitchum, M. G. (2021).
676 Phytonematode peptide effectors exploit a host post-translational trafficking
677 mechanism to the ER using a novel translocation signal. *New Phytologist, 229*(1), 563-
678 574. doi:10.1111/nph.16765
679 Wilbers, R. H. P., Schneiter, R., Holterman, M. H. M., Drurey, C., Smant, G., Asojo, O. A., . . .
680 Lozano-Torres, J. L. (2018). Secreted venom allergen-like proteins of helminths:
681 Conserved modulators of host responses in animals and plants. *PLoS Pathogens,*
682 *14*(10), e1007300. doi:10.1371/journal.ppat.1007300
683 Wubben, M. J., Gavilano, L., Baum, T. J., & Davis, E. L. (2015). Sequence and spatiotemporal
684 expression analysis of CLE-motif containing genes from the reniform nematode
685 (*Rotylenchulus reniformis* Linford & Oliveira). *Journal of Nematology, 47*(2), 159-165.
686 Yang, S., Pan, L., Chen, Y., Yang, D., Liu, Q., & Jian, H. (2019). *Heterodera avenae* GLAND5
687 effector interacts with pyruvate dehydrogenase subunit of plant to promote
688 nematode parasitism. *Frontiers in Microbiology, 10*(JUN).
689 doi:10.3389/fmicb.2019.01241
690

691 **Table 1**. Comparative genome statistics of four cyst nematode genome assemblies. In bold, data from the current paper, data on *Globodera pallida* Lindley and *G. rostochiensis* Line 19 ge-

692 nomes were published by respectively (Cotton et al., 2014) and (Van Steenbrugge et al., 2021)

Nematode species population	**Size (Mb)**	**Number of Scaffolds**	**N50 (Mb)**	**N90 (Mb)**	**Number of Gaps**	**Gap length**	**Number of Genes**	**Number of Transcripts**
Globodera pallida Lindley	124.7	6,873	0.122	0.011	6,873	19,990,795	16,403	16,403
G. pallida Pa2 - D383	**113**	**163**	**2.9**	**0.515**	**22,788**	**945,137**	**18,813**	**27,787**
G. rostochiensis Line 19	92	173	1.70	0.582	2,733	130,000	17,928	21,037
Heterodera schachtii IRS	**190**	**705**	**0.5**	**0.132**	**705**	**4,285,731**	**29,851**	**31,564**
	BUSCO							

G. pallida, Lindley	**C:48.2%[S:43.9%,D:4.3%],F:22.4%,M:29.4%,n:255**
G. pallida, D383	**C:79.6% [S:78.9%, D:0.7%], F:10.6%, M9.8%, n:303**
G. rostochiensis Line 19	C:83.9 [S:82.2%, D: 1.7%], F: 7.9%, M8.2%, n: 303
H. schachtii, IRS	**C:86.3% [S:80.8%, D:5.5%], F:7.1%, M:6.6%, n:255**

693

694 **Figure legends.**
695
696 **Fig. 1.** Synteny between *G. pallida* (population D383), *G. rostochiensis* (Gr-Line19) and *H. schachtii*
697 (population IRS) based on a progressive genome alignment in Mauve. In Fig. 1 A syntenic regions
698 larger than 1 kb are shown, in panel B syntenic regions larger than 3 kb are shown. In panel A, *H.*
699 *schachtii* genome regions are indicated where multiple syntenic regions cluster together into seg-
700 ments spanning over 2 Mb (Fig. 1A, segments 1-8). It is noted that these segments have equivalents
701 in both *G. pallida* and *G. rostochiensis*.
702
703 **Fig. 2**. Phylogeny of SPRYSEC effector genes (see e.g. (Diaz-Granados et al., 2016)) of *G. pallida* (popu-
704 lation D383) (ochre), *G. rostochiensis* (Gr-Line19) (green) and *H. schachtii* (population IRS) (purple). A
705 multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree
706 was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap
707 values < 50 % are indicated by a "-". Gen IDs in italics in lighter shades of ochre, green or purple are
708 used to indicate effector variants with at least one predicted transmembrane domain. Boxed clusters
709 (A and B) highlight two moderated supported subclades with on average moderate (A) and high DOG
710 box dosages. C refers to the basal part of the SPRYSEC tree.
711
712 **Fig. 3.** Phylogeny of GLAND 4 (equivalent to 1106, see, e.g., (Noon et al., 2015)) effector genes of *G.*
713 *pallida* (population D383) (ochre), *G. rostochiensis* (Gr-Line19) (green) and *H. schachtii* (population
714 IRS) (purple). A multiple sequence alignment was made using MUSCLE on the coding sequence. A
715 phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap
716 replicates. Bootstrap values < 50 % are indicated by a "-". Bootstrap values < 50 % are indicated by a
717 "-". Gen IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a
718 signal peptide for secretion.
719
720 **Fig. 4.** Phylogeny of GLAND 5 (equivalent to G11A06, see, e.g., (Noon et al., 2015)) effector genes of
721 *G. pallida* (population D383) (ochre), *G. rostochiensis* (Gr-Line19) (green) and *H. schachtii* (population
722 IRS) (purple). A multiple sequence alignment was made using MUSCLE on the coding sequence. A
723 phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap
724 replicates. Bootstrap values < 50 % are indicated by a "-". Bootstrap values < 50 % are indicated by a
725 "-". Gen IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a
726 signal peptide for secretion.
727
728 **Fig. 5.** Phylogeny of VAP (Venom Allergen-like Protein, see, e.g., (Wilbers et al., 2018)) effector genes
729 of *G. pallida* (population D383) (ochre), *G. rostochiensis* (Gr-Line19) (green) and *H. schachtii* (popula-
730 tion IRS) (purple). A multiple sequence alignment was made using MUSCLE on the coding sequence.
731 A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap
732 replicates. Bootstrap values < 50 % are indicated by a "-". Bootstrap values < 50 % are indicated by a
733 "-". Gen IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a
734 signal peptide for secretion.
735
736 **Fig. 6.** Phylogeny of CLE (CLAVATA3/ESR-related peptides, see, e.g., (Lu et al., 2009)) effector genes
737 of *G. pallida* (population D383) (ochre), *G. rostochiensis* (Gr-Line19) (green) and *H. schachtii* (popula-
738 tion IRS) (purple). A multiple sequence alignment was made using MUSCLE on the coding sequence.
739 A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap
740 replicates. Bootstrap values < 50 % are indicated by a "-". Bootstrap values < 50 % are indicated by a
741 "-". Gen IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a
742 signal peptide for secretion.
743

Fig. 7. Phylogeny of GLAND13 (invertases, see, e.g., (Danchin et al., 2016)) effector genes of *G. pallida* (population D383) (ochre), *G. rostochiensis* (Gr-Line19) (green) and *H. schachtii* (population IRS) (purple). A multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values < 50 % are indicated by a "-". Bootstrap values < 50 % are indicated by a "-". Gen IDs in lighter shades of ochre, green or purple are used to indicate effector variants lacking a signal peptide for secretion.

Fig. 8. This distribution of a DOrsal Gland promoter element motif ('DOG box, (Eves-van den Akker et al., 2016)) among a selection of cyst nematode species. Fig. 8A shows the percentages of variants per effector family with one or more a DOG boxes in their promoter region. In Fig. 8B the relationship between DOG box dosage and expression level (expressed as log2 TPM (Transcript Count Per Million) is presented.

Fig. 1

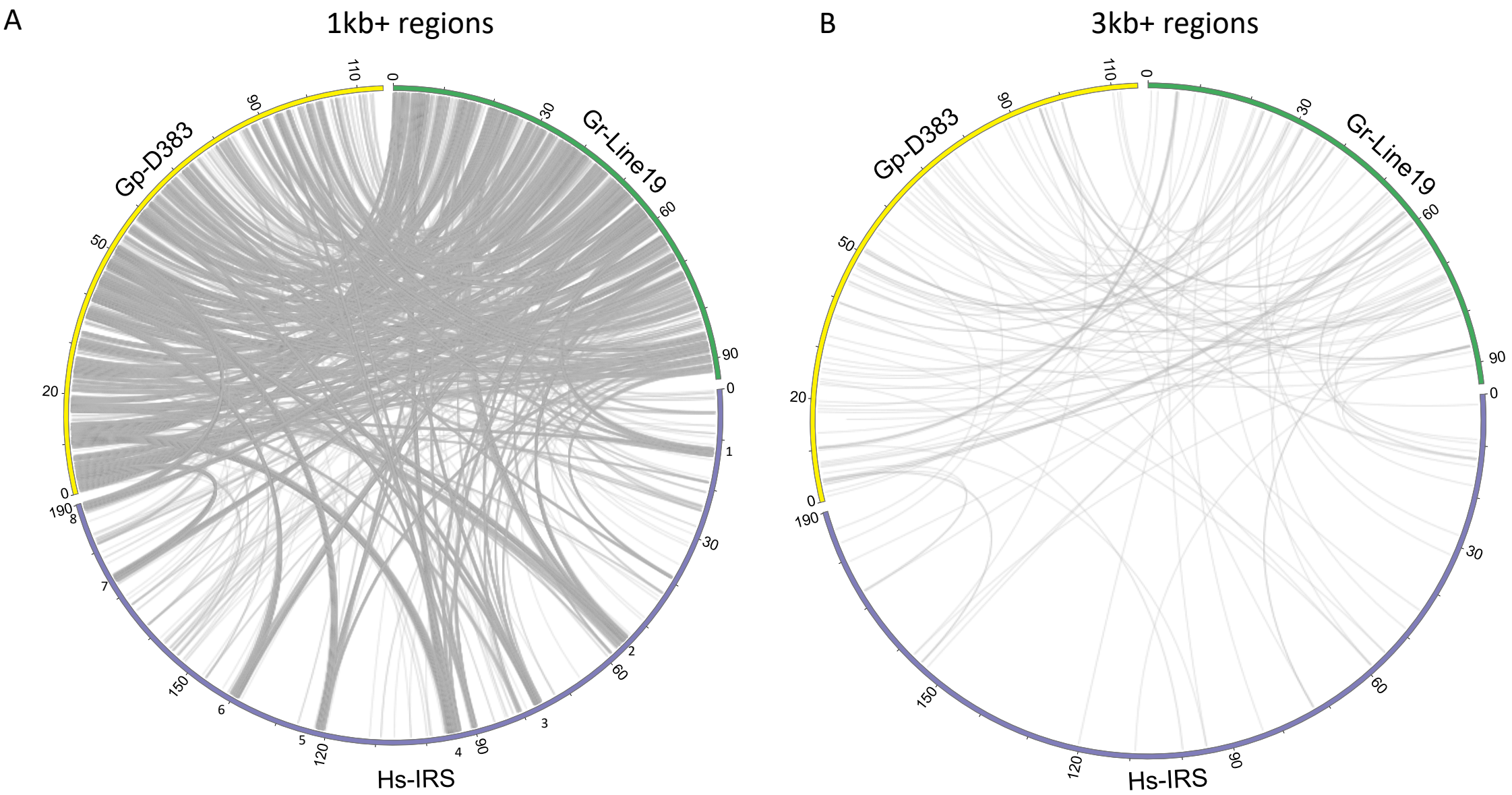

A
1kb+ regions
Gp-D383
Gr-Line19
Hs-IRS
B
3kb+ regions
Gp-D383
Gr-Line19
Hs-IRS

Fig. 2

SPRYSEC

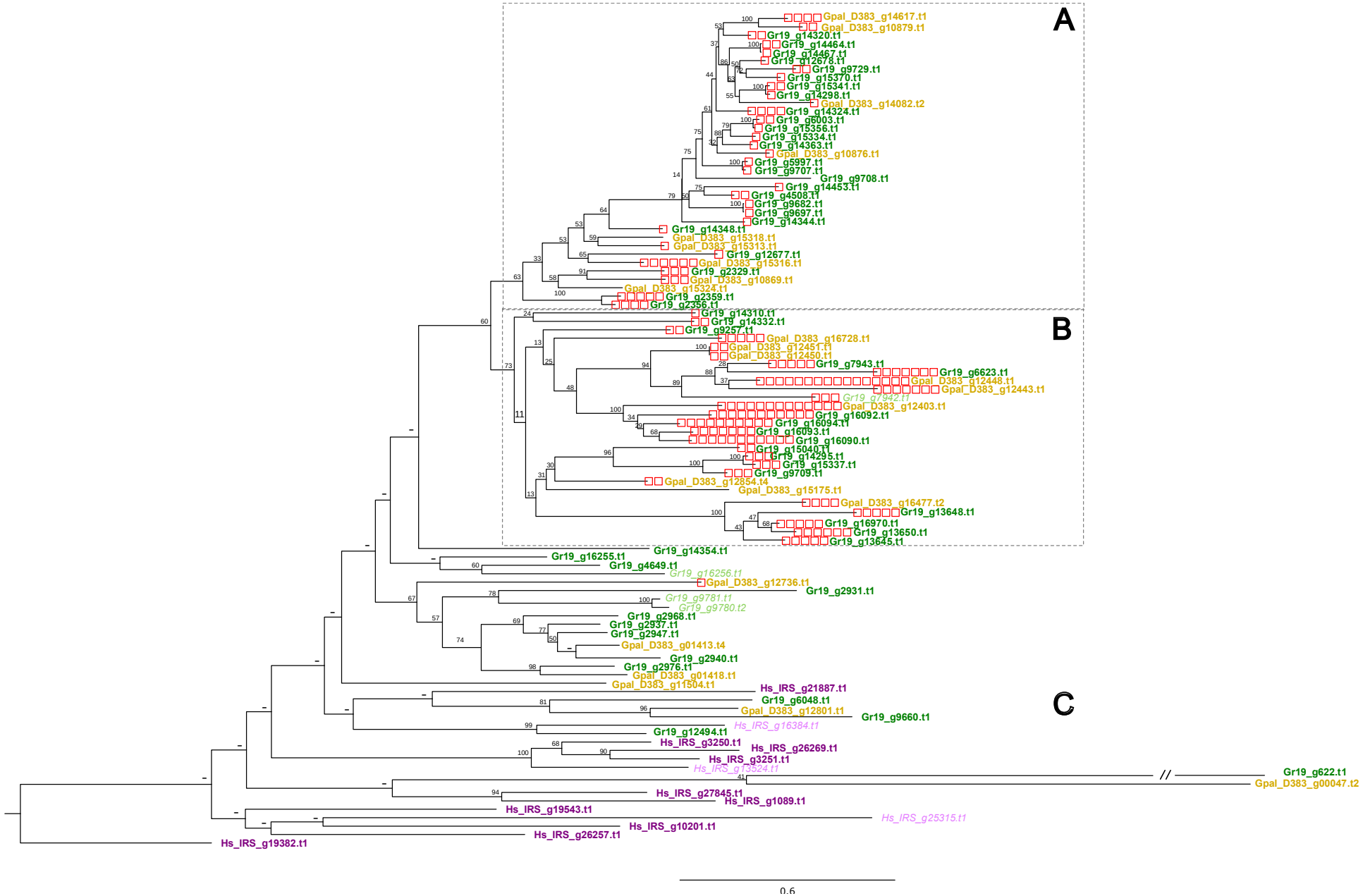


Fig. 3


GLAND4 / 1106
Gpal_D383_g17346.t1
Gpal_D383_g17350.t1
Gpal_D383_g13670.t1
Gpal_D383_g13671.t1
Gpal_D383_g13668.t1
Gpal_D383_g13669.t1
Gpal_D383_g13667.t1
Gpal_D383_g13714.t1
Gpal_D383_g13703.t1
Gr19_g2068.t1
Gr19_g2108.t1
Gr19_g2106.t1
Gr19_g2102.t1
Gr19_g2104.t2
Gr19_g2107.t1
Gr19_g15298.t1
Gr19_g2075.t1
Gr19_g15312.t1
Gr19_g732.t1
Hs_IRS_g21474.t1
Hs_IRS_g20983.t1
Hs_IRS_g13217.t1
Hs_IRS_g3574.t1
Hs_IRS_g24286.t1
Hs_IRS_g20982.t1
Hs_IRS_g24437.t1
Hs_IRS_g21472.t1
Hs_IRS_g3134.t1
Hs_IRS_g21252.t1
Hs_IRS_g10197.t1
Hs_IRS_g10186.t1
Hs_IRS_g10183.t1
Hs_IRS_g10181.t1
Hs_IRS_g24669.t1
0.08

Fig. 4

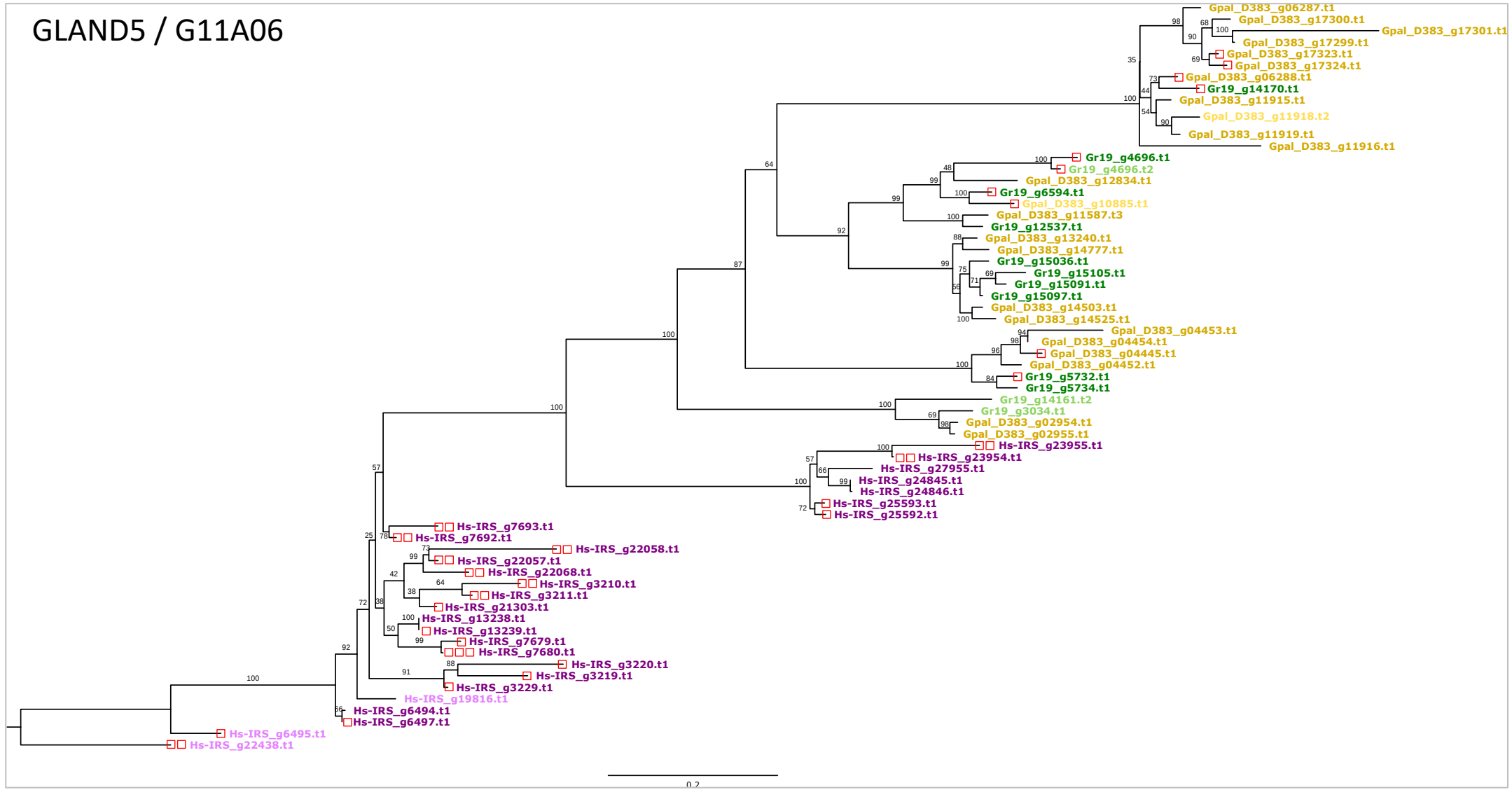

GLAND5 / G11A06
Gpal_D383_g06287.t1
Gpal_D383_g17300.t1
Gpal_D383_g17301.t1
Gpal_D383_g17299.t1
Gpal_D383_g17323.t1
Gpal_D383_g17324.t1
Gpal_D383_g06288.t1
Gr19_g14170.t1
Gpal_D383_g11915.t1
Gpal_D383_g11918.t2
Gpal_D383_g11919.t1
Gpal_D383_g11916.t1
Gr19_g4696.t1
Gr19_g4696.t2
Gpal_D383_g12834.t1
Gr19_g6594.t1
Gpal_D383_g10885.t1
Gpal_D383_g11587.t3
Gr19_g12537.t1
Gpal_D383_g13240.t1
Gpal_D383_g14777.t1
Gr19_g15036.t1
Gr19_g15105.t1
Gr19_g15091.t1
Gr19_g15097.t1
Gpal_D383_g14503.t1
Gpal_D383_g14525.t1
Gpal_D383_g04453.t1
Gpal_D383_g04454.t1
Gpal_D383_g04445.t1
Gpal_D383_g04452.t1
Gr19_g5732.t1
Gr19_g5734.t1
Gr19_g14161.t2
Gr19_g3034.t1
Gpal_D383_g02954.t1
Gpal_D383_g02955.t1
Hs-IRS_g23955.t1
Hs-IRS_g23954.t1
Hs-IRS_g27955.t1
Hs-IRS_g24845.t1
Hs-IRS_g24846.t1
Hs-IRS_g25593.t1
Hs-IRS_g25592.t1
Hs-IRS_g7693.t1
Hs-IRS_g7692.t1
Hs-IRS_g22058.t1
Hs-IRS_g22057.t1
Hs-IRS_g22068.t1
Hs-IRS_g3210.t1
Hs-IRS_g3211.t1
Hs-IRS_g21303.t1
Hs-IRS_g13238.t1
Hs-IRS_g13239.t1
Hs-IRS_g7679.t1
Hs-IRS_g7680.t1
Hs-IRS_g3220.t1
Hs-IRS_g3219.t1
Hs-IRS_g3229.t1
Hs-IRS_g19816.t1
Hs-IRS_g6494.t1
Hs-IRS_g6497.t1
Hs-IRS_g6495.t1
Hs-IRS_g22438.t1
0.2

Fig. 5

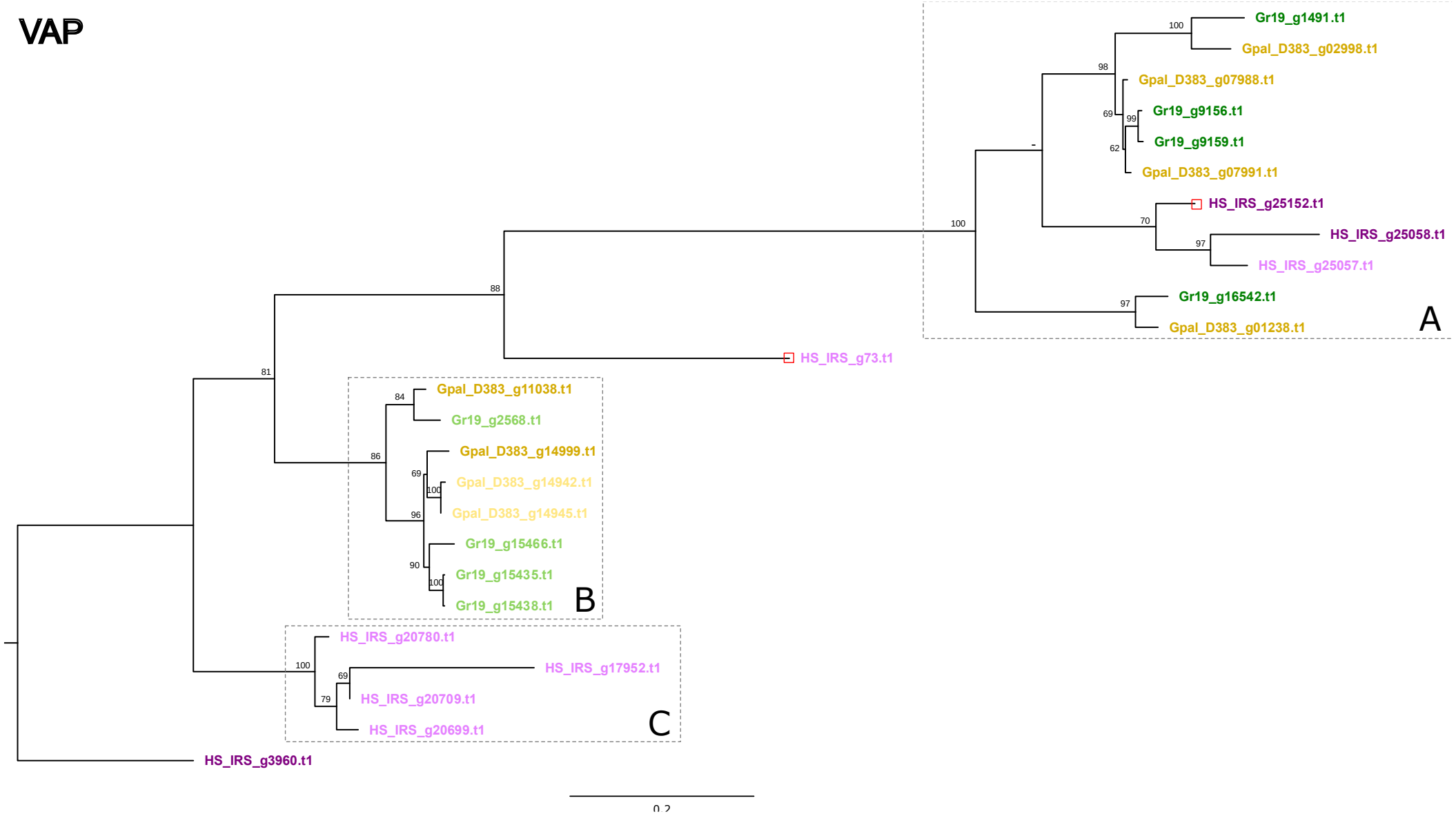

VAP
Gr19_g1491.t1
Gpal_D383_g02998.t1
Gpal_D383_g07988.t1
Gr19_g9156.t1
Gr19_g9159.t1
Gpal_D383_g07991.t1
HS_IRS_g25152.t1
HS_IRS_g25058.t1
HS_IRS_g25057.t1
Gr19_g16542.t1
Gpal_D383_g01238.t1
A
HS_IRS_g73.t1
Gpal_D383_g11038.t1
Gr19_g2568.t1
Gpal_D383_g14999.t1
Gpal_D383_g14942.t1
Gpal_D383_g14945.t1
Gr19_g15466.t1
Gr19_g15435.t1
Gr19_g15438.t1
B
HS_IRS_g20780.t1
HS_IRS_g17952.t1
HS_IRS_g20709.t1
HS_IRS_g20699.t1
C
HS_IRS_g3960.t1
0.2

Fig. 6

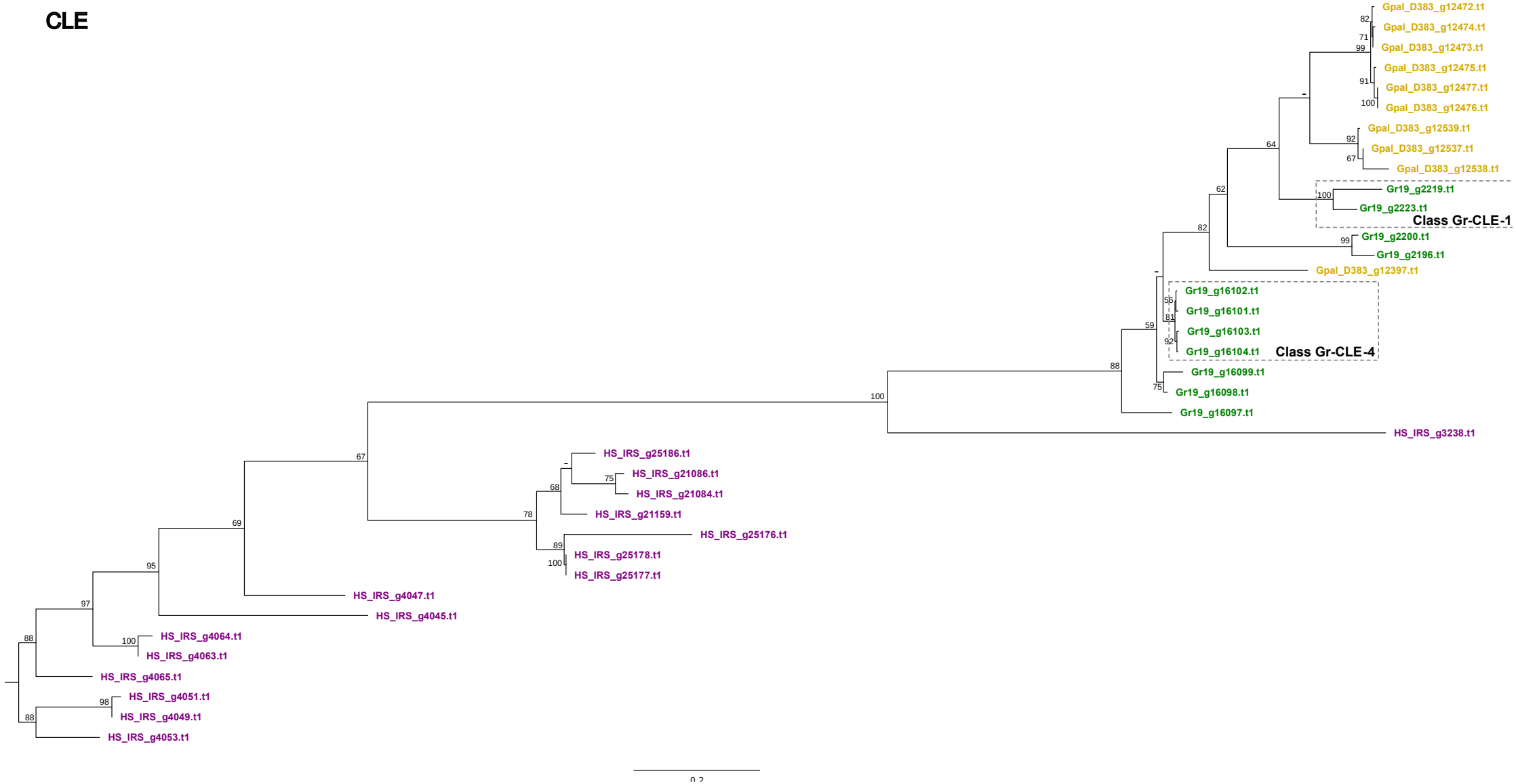

CLE
Gpal_D383_g12472.t1
Gpal_D383_g12474.t1
Gpal_D383_g12473.t1
Gpal_D383_g12475.t1
Gpal_D383_g12477.t1
Gpal_D383_g12476.t1
Gpal_D383_g12539.t1
Gpal_D383_g12537.t1
Gpal_D383_g12538.t1
Gr19_g2219.t1
Gr19_g2223.t1
Class Gr-CLE-1
Gr19_g2200.t1
Gr19_g2196.t1
Gpal_D383_g12397.t1
Gr19_g16102.t1
Gr19_g16101.t1
Gr19_g16103.t1
Gr19_g16104.t1
Class Gr-CLE-4
Gr19_g16099.t1
Gr19_g16098.t1
Gr19_g16097.t1
HS_IRS_g3238.t1
HS_IRS_g25186.t1
HS_IRS_g21086.t1
HS_IRS_g21084.t1
HS_IRS_g21159.t1
HS_IRS_g25176.t1
HS_IRS_g25178.t1
HS_IRS_g25177.t1
HS_IRS_g4047.t1
HS_IRS_g4045.t1
HS_IRS_g4064.t1
HS_IRS_g4063.t1
HS_IRS_g4065.t1
HS_IRS_g4051.t1
HS_IRS_g4049.t1
HS_IRS_g4053.t1
0.2

Fig. 7

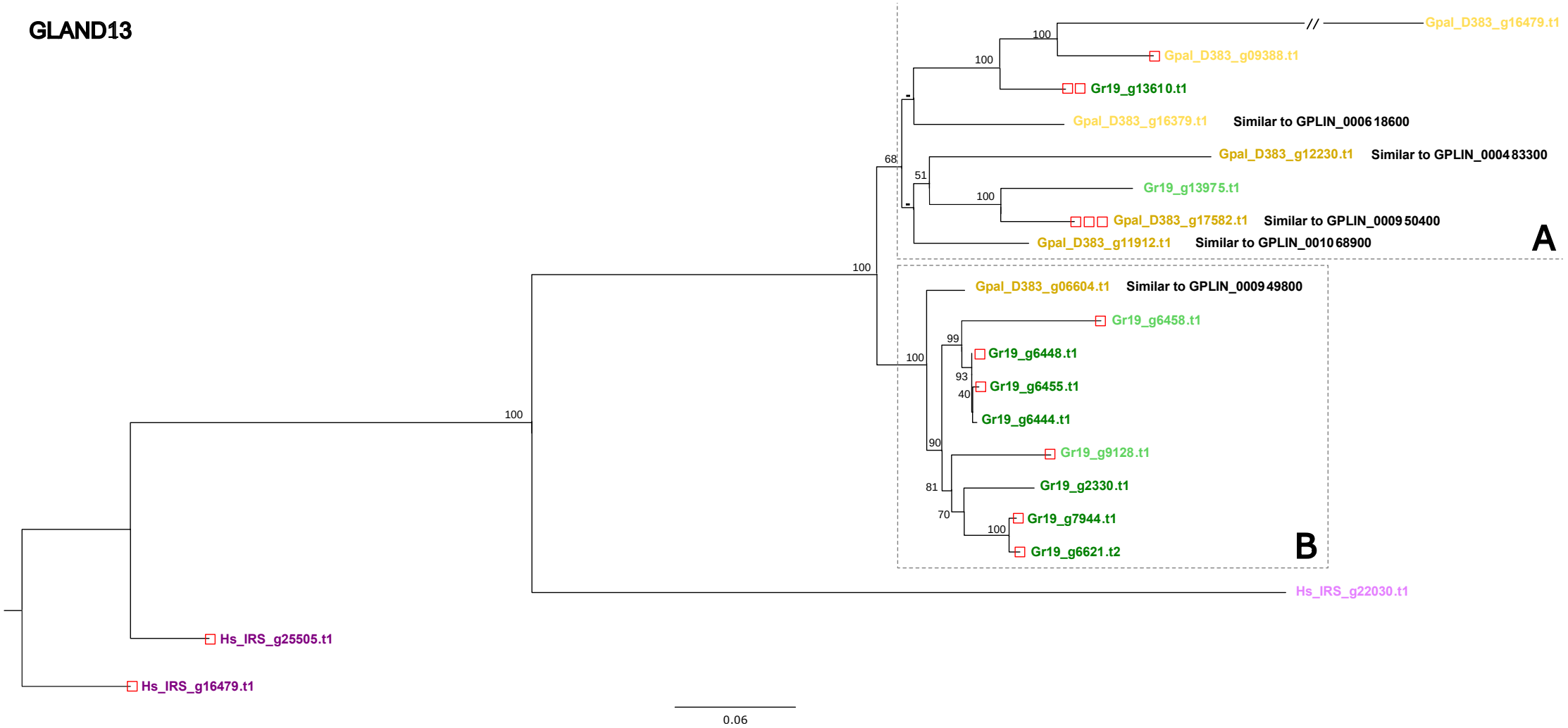

GLAND13
Gpal_D383_g16479.t1
Gpal_D383_g09388.t1
Gr19_g13610.t1
Gpal_D383_g16379.t1
Similar to GPLIN_000618600
Gpal_D383_g12230.t1
Similar to GPLIN_000483300
Gr19_g13975.t1
Gpal_D383_g17582.t1
Similar to GPLIN_000950400
Gpal_D383_g11912.t1
Similar to GPLIN_001068900
A
Gpal_D383_g06604.t1
Similar to GPLIN_000949800
Gr19_g6458.t1
Gr19_g6448.t1
Gr19_g6455.t1
Gr19_g6444.t1
Gr19_g9128.t1
Gr19_g2330.t1
Gr19_g7944.t1
Gr19_g6621.t2
B
Hs_IRS_g22030.t1
Hs_IRS_g25505.t1
Hs_IRS_g16479.t1
0.06

Fig. 8A

Bar-graphs – effector family members / distribution DOG boxes

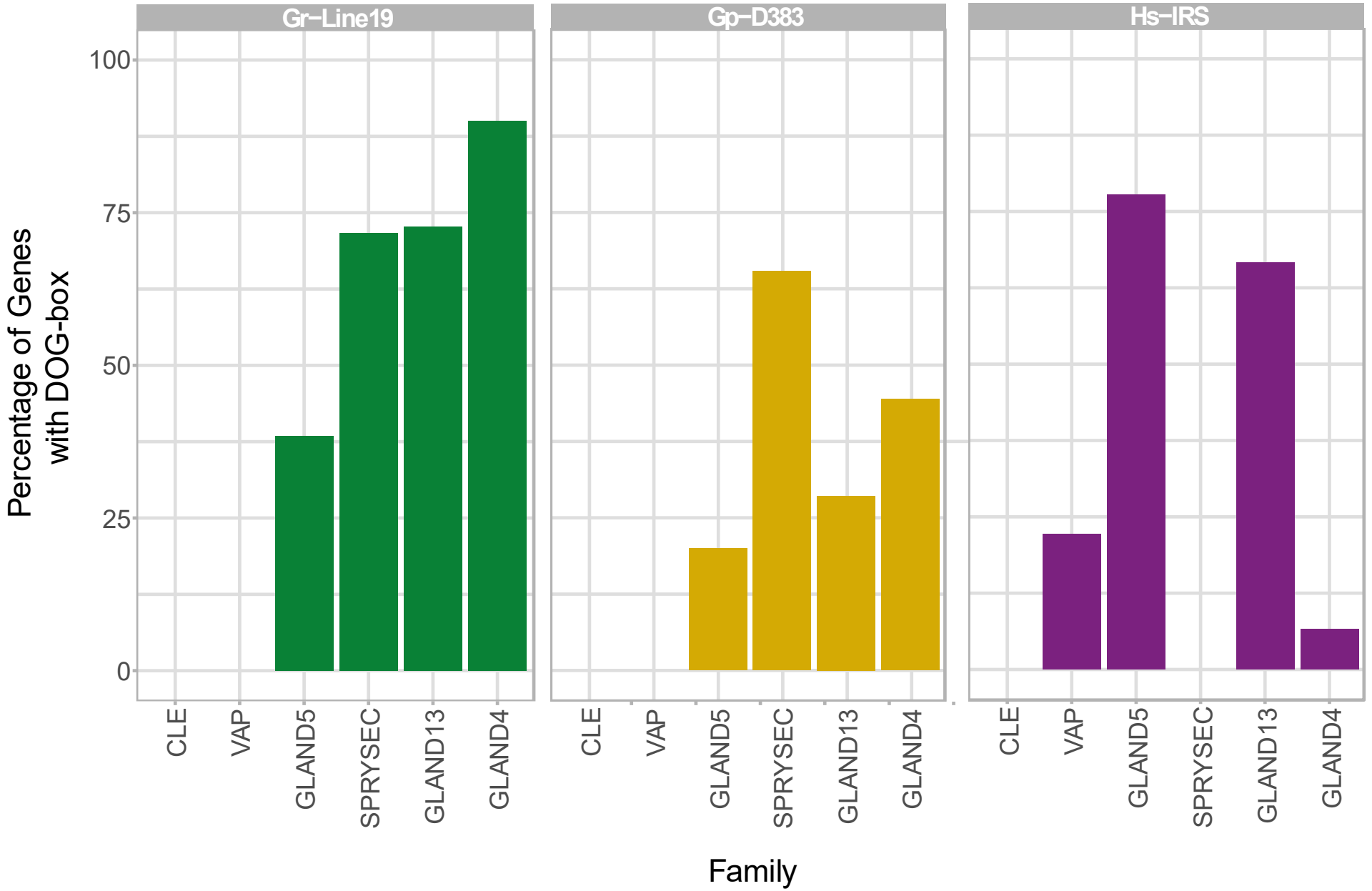
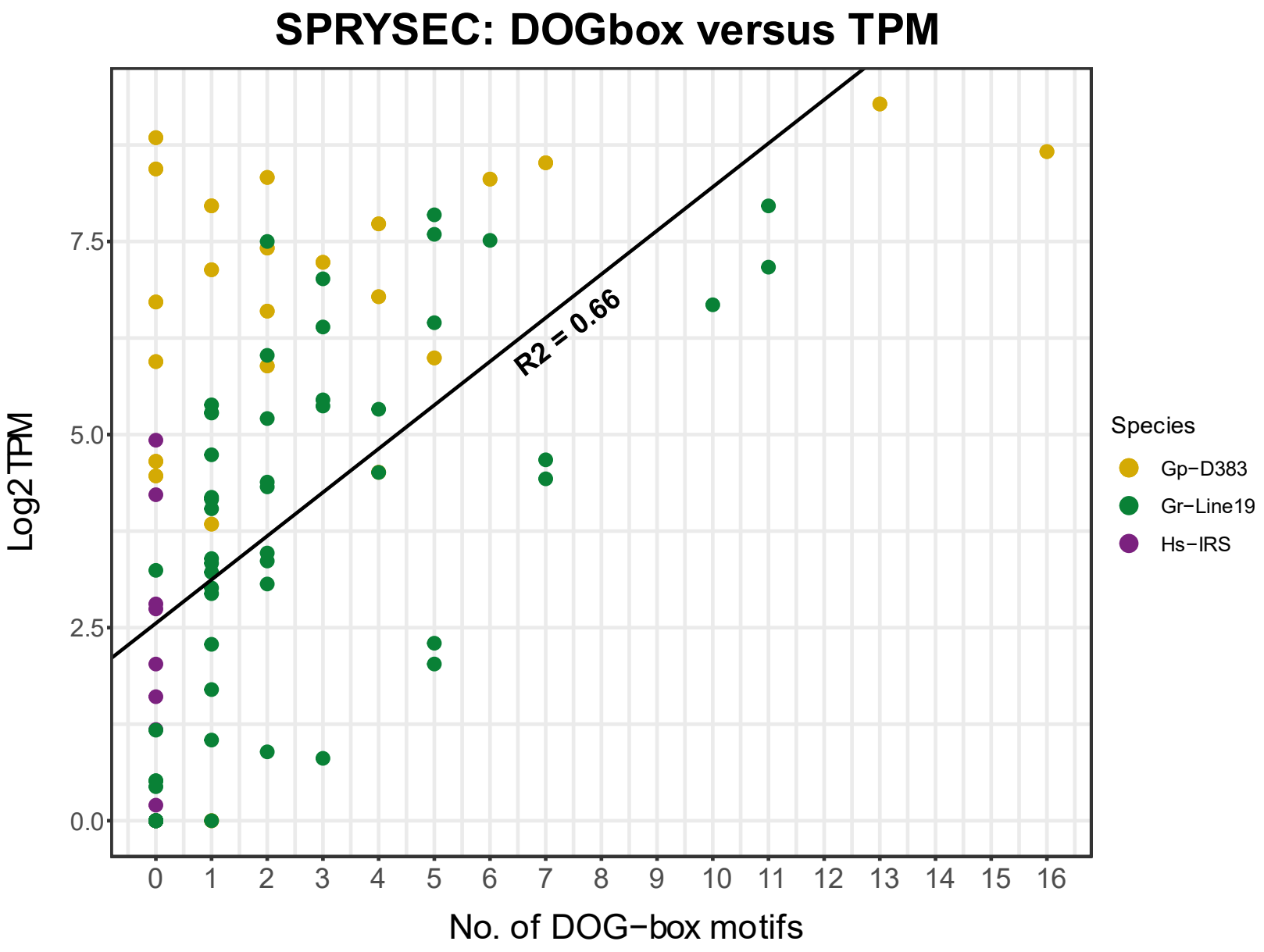


Fig. 8B


SPRYSEC: DOGbox versus TPM
Log2 TPM
7.5
5.0
2.5
0.0
R2 = 0.66
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
No. of DOG−box motifs
Species
Gp−D383
Gr−Line19
Hs−IRS