

# Comparative Genomics of two Inbred Lines of the Potato Cyst Nematode *Globodera rostochiensis* reveals disparate Effector Family-specific Diversification Patterns

Joris J.M. van Steenbrugge (✉ [jorisvsteenbrugge@gmail.com](mailto:jorisvsteenbrugge@gmail.com))

Wageningen University and Research Wageningen Plant Research <https://orcid.org/0000-0002-9961-2538>

Sven van den Elsen

Wageningen University & Research

Martijn Holterman

Solynta <https://orcid.org/0000-0002-1134-6555>

Mark G. Sterken

Wageningen University & Research <https://orcid.org/0000-0001-7119-6213>

Peter Thorpe

University of St Andrews School of Medicine

Aska Goverse

Wageningen University & Research <https://orcid.org/0000-0002-7399-8743>

Geert Smant

Wageningen University & Research <https://orcid.org/0000-0001-6094-8686>

Johannes Helder

Wageningen University & Research <https://orcid.org/0000-0003-0408-3985>

---

## Research article

**Keywords:** heterozygosity, gland proteins, innate immune system, SPRYCSEC, CLE, effectoromics

**Posted Date:** May 11th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-509140/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on August 11th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07914-6>.

# Abstract

**Background:** Potato cyst nematodes belong to the most harmful pathogens in potato, and durable management of these soil-borne parasites largely depends on host-plant resistances. These resistances are pathotype specific. The current *Globodera rostochiensis* pathotype scheme that defines five pathotypes (Ro1 - Ro5) is for fundamental and practical reasons barely useful. As a result, resistant potato varieties are used worldwide in a non-informed manner.

**Results:** We generated two novel reference genomes of *G. rostochiensis* inbred lines derived from a Ro1 and a Ro5 population. These genome sequences comprise 173 and 189 scaffolds respectively, marking a » 24-fold reduction in fragmentation as compared to the current reference genome. We provide copy number variations for 18 effector families. Four dorsal gland effector families were investigated in more detail. SPRYSECs, known to be implicated in plant defence suppression, constitute by far the most diversified family with 60 and 99 variants in Ro1 and Ro5 distributed over 18 and 26 scaffolds. In contrast, CLEs, effectors involved in feeding site induction, show strong physical clustering. The 10 and 16 variants cluster on respectively 2 and 1 scaffolds. Given that pathotypes are defined by their effectoromes, we pinpoint the disparate nature of the contributing effector families in terms of sequence diversification and loss and gain of variants.

**Conclusion:** Two novel reference genomes allow for nearly complete inventories of effector diversification and physical organisation within and between pathotypes. Combined with insights we provide on effector family-specific diversification patterns, this constitutes a solid basis for an effectorome-based virulence scheme for this notorious pathogen.

## Background

Plant-parasitic nematodes have a significant impact on food and feed production worldwide. Every cultivated crop is parasitized by at least one nematode species, resulting in a net loss of over 70 billion US dollar annually [1]. From an economic point of view root-knot and cyst nematodes have the highest impact [2]. Whereas root-knot nematodes have a higher impact in warmer climate zones, cyst nematode problems mostly occur in the temperate regions. Unlike root-knot nematodes, most cyst nematodes have a defined center of origin. For example, soybean cyst nematodes originate from north-east Asia and have spread as a successful and highly harmful parasite to all major soybean-growing areas. Potato cyst nematodes diversified in the Andes in South America, and have now proliferated to all major potato production areas in the world (*e.g.* [3]). Outside of their centers of origin, cyst nematodes belong to the most harmful pathogens of the crops mentioned above.

One of the most widely applied control measures is the use of resistant host plants. Resistances against potato cyst nematodes tend to have a long agronomic life span due to cyst nematodes' unique biology. Potato cyst nematodes usually have only one generation per year through obligate sexual reproduction, go into diapause for months, and - once hatched - their motility is in the range of a few cm per day. These characteristics drastically slow down the process of selection and proliferation of virulent individuals. Potato breeders have introgressed the resistance gene *H1* from *Solanum tuberosum* ssp. *andigena* CPC 1674 into numerous potato cultivars from the 1960's onwards [4]. The *H1* gene confers resistance against *G. rostochiensis* pathotypes Ro1 and Ro4 [5], and this resistance gene is still effective in virtually all major potato producing countries.

Based on a number of differentials, pathotypes have been defined within the two potato cyst nematode species *G. rostochiensis* and *G. pallida*. Five pathotypes named Ro1 - Ro5 have been proposed for *G. rostochiensis*, whereas three pathotypes (Pa1 - Pa3) were discriminated within *G. pallida* [6]. Apart from being laborious and time-consuming, the current pathotype scheme has limited value as it lacks a solid genetic basis. The distinction between for instance the *G. pallida* pathotypes Pa2 and Pa3 is elusive [7]. For *G. rostochiensis*, genome-wide allele frequencies correlate with the geographical distribution of populations, regardless of pathotype [8, 9]. This indicates that the genetic basis of the pre-defined pathotypes is small. A robust pathotyping scheme for potato cyst nematodes is highly desirable because it would lead to far more efficient and durable use of the limited number of host plant resistances currently available. The availability of high-quality reference genome sequences from individual pathotypes would be an ideal starting point for pathotypes' molecular characterization.

Resistant plant species deploy R proteins as surveillance molecules that recognize specific effector molecules - or their activities - secreted by nematodes. Nematodes use a protrusible stylet to inject effector proteins into plant cells. Effectors are diverse and fulfil functions ranging from plant cell wall degradation to the induction of a feeding site and suppressing the plant's innate immune system [10]. The nematode produces effectors mainly in the subventral and the dorsal esophageal glands. Effectors are usually members of diversified gene families, and potato cyst nematode typically produces multiple variants per effector. An example is the SPRYSEC gene family that codes for a highly expanded set of proteins that act as activators and suppressors of plant defence [11]. One variant of this family, RBP-1, was shown to trigger the activation of the potato resistance gene *Gpa2* [12] resulting in local hypersensitive response. Effector proteins secreted by the cyst nematode parasite are most likely responsible for the activation of plant resistance proteins. However, this was demonstrated for only a small number of resistance genes (*Gpa2*; [12], *Cf-2*; [13]).

Sequencing the genome of plant-parasitic nematodes is more challenging than for other, larger organisms. With the currently available methods, it is practically impossible to isolate and sequence DNA from an individual nematode to gain enough coverage to generate a high-quality reference genome sequence—especially when isolating high molecular weight DNA required for long-read sequencing technologies. Reference genomes of plant-parasitic nematodes are therefore often based on the genetic material from a population. Consequently, the reference genome includes a substantial heterozygosity level, as the starting material includes a high degree of allelic variation. The current reference genome sequences of potato cyst nematodes *Globodera rostochiensis* [14] and *G. pallida* [15] were each generated using heterozygous starting material (selected field populations), and are relatively fragmented (respectively 4,377 and 6,873 scaffolds). In *G. rostochiensis*, Eves-van den Akker et al. (2016) [14] predicted 138 high confidence effector genes based on sequence similarity with previously described effector gene families. Furthermore, a third of these genes were identified to cluster on effector gene islands. Among these expanded gene families, sequence divergence between different pathotypes was estimated as well. While many single nucleotide polymorphisms and insertions/deletions were observed [14], the highly fragmented reference genome sequence made it challenging to distinguish between sequence and copy number variation. Less fragmentation in the genome sequence would similarly make it possible to display the degree of clustering of effector genes more accurately.

We generated a new set of reference genome sequences to allow for the accurate organization of effector genes and to compare copy number variation and sequence variation between the Ro1 and Ro5 pathotypes. A precise representation of these two sources of genetic variation is essential for developing molecular pathotyping methods in the future. The current reference genome sequence of *G. rostochiensis* shows a haploid genome size

of 95.9Mb in size [14], spread over eighteen diploid chromosomes [16]. Here we used two *G. rostochiensis* lines, one fully avirulent and one fully virulent with regard to the *H1* gene. The starting materials for these lines were two distinct field populations sampled from The Netherlands (Ro1-Mierenbos) and Germany (Ro5-Hamerz) [17]. The selection process started with a single cross between an individual male and a female. After multiple generations, fully avirulent Ro1 (Gr-line19) and fully virulent Ro5 (Gr-line22) lines regarding the *H1* resistance in potato were generated [17]. As a result, both Gr-Line19 and Gr-Line22 harbour limited genetic variation, with a theoretical maximum of 4 alleles per locus. For sexually reproducing species, this is the minimum level of heterozygosity that can be present in a population.

New genome assemblies were generated for each of the inbred lines based on PacBio long read-sequencing technology. Using these newly generated *G. rostochiensis* reference genome sequences with a substantially reduced number of scaffolds, we investigated the genomic organisation and the diversification of 18 effector families. A large number of differences in the number of paralogs and variation in sequence content were identified between the effector arsenals of the avirulent Gr-line19 and the virulent Gr-line22. Detailed insights in effector family diversification will form the basis for the generation of a molecular pathotype scheme for potato cyst nematodes.

## Results

### Genome Assemblies

Two inbred lines of the potato cyst nematode *G. rostochiensis* were initially derived from crossings between individuals from two populations, Ro1-Mierenbos and Ro5-Harmerz [17]. DNAs from Gr-Line19 and Gr-line22 were sequenced using PacBio sequencing technology with respectively 119X and 132X coverage, and assembled into two reference genome sequences (Table 1). Benefitting from this long read technology as well as the significantly smaller genetic background, the two newly generated *G. rostochiensis* genome assemblies are less fragmented than the first genome sequence that was published (nGr.v1.0) [14] while maintaining a comparable assembly size. The number of scaffolds in the new assemblies is about 24 fold lower than in the original *G. rostochiensis* reference genome sequence (Table 1). At the same time, the scaffold N50 increased about 20 fold from 0.085 to around 1.7 Mb.

Table 1  
Comparative genome statistics of three *G. rostochiensis* genome assemblies

<i>G. rostochiensis</i> population	Assembly ID	Size (Mb)	# scaffolds	Scaffold N50 (Mb)	BUSCO notation assessment results (in %)			
					Single	duplicated	fragmented	missing
Ro1 population from JHI PCN collection (JHI-Ro1)	nGr.v1.0 [14]	96	4377	0.085	82.8	1.0	8.3	7.9
Gr-Line19	Gr19v10	92	173	1.70	82.2	1.7	7.9	8.2
Gr-Line22	Gr22v10	101	189	1.80	81.5	1.3	8.3	8.9
BUSCO (Benchmarking Universal Single-Copy Orthologs) - eukaryota_odb10.								

We verified whether the scaffolds were truly derived from our target genomes, and not from associated pro- or eukaryotes. As shown in Figs S1 and S2, no obvious contamination was detected. Some scaffolds comprised regions with highest similarity to the Arthropoda (8.2% and 7.6% in Gr-Line19 and G2-Line22, respectively). This could be related to nuclear repeats as observed recently for *Heterorhabditis bacteriophora* [18], or other low-complexity regions. Regarding the assembly size and BUSCO score, the novel assemblies are comparable to the current reference. The assemblies of Gr-Line19 and Gr-Line22 harbor 2,733 and 6,572 gaps, respectively, covering in total 130 Kb and 150 Kb. As compared to the current *G. rostochiensis* reference [14], the number and lengths of gaps showed a 29-fold reduction.

The repeat content in both reference genome sequences was relatively low, 2.6% for Gr-Line19 and 1.6% for Gr-Line22. The GC content in repeat regions for Gr-Line19 (40.3%) was comparable to this genotype's overall GC content (39.1%). In Gr-Line22, the GC content in repeat regions (32.5%) was lower than the overall GC content (38.3%). In predicted protein-coding regions, the GC content is comparable between both reference genome sequences (Gr-Line19: 50.8%, Gr-Line22: 50.9%). Using Braker2 as a gene-prediction tool, 17,928 and 18,258 genes were predicted in the Gr-Line19 and Gr-Line22 genome assemblies, coding for 21,037 and 21,514 transcripts, respectively. The protein-coding regions take up approximately 33% (Gr-Line19) and 30% (Gr-Line22) of the genomes at an average density of 89.3 (Gr-Line19) and 86.6 (Gr-Line22) genes per Mb.

Synteny between the newly generated genomes and the current reference genome [14] was evaluated using a progressive genome alignment. Homologous regions larger than 3 kb and their genomic organization are presented in Fig. 1. A broad span of regions in the JHI-Ro1 reference assembly shows homology to both new assemblies (respectively 67%, 72%, and 61% of the total assembly sizes for JHI-Ro1, Gr-Line19 and Gr-Line22). While the total numbers of base pairs that are covered in a homologous region are roughly within a 10% range of each other, both new assemblies show substantially larger continuous and so far uncovered regions. Seventy homologous regions shared exclusively between the Gr-Line19 and Gr-Line22 genomes were identified (Fig. 1 represented by yellow arches). The size of these regions ranges from 3,048 to 23Kb. These regions comprise 97 genes, and all genes are on the JHI-Ro1 genome with identity levels varying between 23% and 100% (Supplemental Table S1).

## **Heterozygosity and structural variation between the two *Globodera rostochiensis* genomes.**

The inbred lines Gr-Line19 and Gr-Line22 originate from single crossings of individuals, and, as a result, the genetic variation is expected to be smaller than for field populations. To pinpoint the effect of this genetic bottleneck caused by a single crossing, a comparison was made between the proportion of heterozygous and homozygous single nucleotide variants between Gr-Line19 (Ro1) and JHI-Ro1, the selected field population used to generate the current *G. rostochiensis* reference genome [14] while using the Gr-Line22 (Ro5) genome as a reference. Among the called variants that passed the quality filter (JHI-Ro1 n = 584,145; Gr-Line19 n = 716,491), 37% of the JHI-Ro1 loci were homozygous, as compared to 47% of the variants in Gr-Line19 (Fig. 2A). The increased level of homozygosity in Gr-Line19 reflects the relatively narrow genetic basis of this inbred line.

Secondly, structural variation (e.g. insertions, deletions, inversions) of approximately 1kb or larger within the individual lines and between Gr-Line19 and Gr-Line22 was determined. The proportions of heterozygous and homozygous structural variants with fragment sizes > 1kb were compared (Fig. 2B). The structural variation within Gr-Line19 and Gr-Line22 was minimal (Fig. 2B). This observation confirms the low level of structural intra-population heterozygosity. The proportion of homozygous variants was nearly identical while comparing Gr-Line19 with Gr-Line22 and *vice versa* (Gr-Line22 *versus* Gr-Line19: 85.09% & Gr-Line19 on Gr-Line22 85.06%).

## **Expansion of Effector gene Families**

We identified homologs of 18 known effector gene families from which at least one member was shown to be expressed in the subventral (n = 6) or the dorsal (n = 11) oesophageal gland cells, or in the amphids (n = 1) (Table 2, Fig. 3A).

Table 2  
Effector families mapped in genomes of *G. rostochiensis* lines Gr-line19 and Gr-line22.

Expression	Effector family	Functionality / similarity		Reference
Subventral esophageal glands	GH5*	Beta 1,4 endoglucanase	CWDE	[69]
	GH30	xylanase, glucosylceramidase, <i>etc.</i>		[70]
	GH43	candidate arabinanase		[14]
	GH53	candidate arabinogalactanase		[14]
	PL3**	Pectate lyase		[71]
	Hg-GLAND 10	cellulose binding protein		[72, 73]
Dorsal esophageal gland	<b>SPRYSEC</b>	Suppression and activation of plant innate immunity	Immune	[26]
	GSS	glutathione synthetase-like effectors involved in redox regulation	Feeding site	[44]
	<b>CLE</b>	CLAVATA3/ESR-related peptides, mimic plant CLEs	Feeding site	[29]
	<b>1106</b>	PTI and ETI suppressor	Immune	[20]
	Hg16B09	Suppression plant innate immunity		[74]
	Hg-GLAND1	ETI suppressor		[73] [19]
	<b>Hg-GLAND5</b>	PTI suppressor		[73] [19]
	Hg-GLAND6 (4D06)	PTI suppressor		[73] [19]
	Hg-GLAND 12	Pioneer (function unknown)	Feeding site	[73]
	Hg-GLAND 13	Invertase ( <i>Rhizobium</i> )		[73]
Hg-GLAND 14	Endopeptidase ( <i>Ascaris suum</i> )		[73]	

\* GH, Glycoside Hydrolases; \*\*PL, Polysaccharide Lyases, Family numbering according to CAZy (<http://www.cazy.org>)

'Expression' refers to nematode organs in which at least one effector family member was shown to be expressed. Subventral esophageal glands of potato cyst nematode are mainly active during migration to the host plant, and host plant penetration. The dorsal esophageal gland shows highest activity during feeding site induction and maintenance. Amphids are chemosensory organs located at the head region of the nematode. For effector families in bold diversification and physical distribution are investigated in detail. CWDE: cell wall-degrading enzymes, Immune: effector families for which at least one member is known to affect the plant innate immune system, Feeding site: effector families for which at least one member is known to be involved in feeding site induction or maintenance.

Expression	Effector family	Functionality / similarity	Reference
Amphids	HYP	hyper-variable extracellular effector (function unknown)	[75]
* GH, Glycoside Hydrolases; **PL, Polysaccharide Lyases, Family numbering according to CAZy ( <a href="http://www.cazy.org">http://www.cazy.org</a> )			
'Expression' refers to nematode organs in which at least one effector family member was shown to be expressed. Subventral esophageal glands of potato cyst nematode are mainly active during migration to the host plant, and host plant penetration. The dorsal esophageal gland shows highest activity during feeding site induction and maintenance. Amphids are chemosensory organs located at the head region of the nematode. For effector families in bold diversification and physical distribution are investigated in detail. CWDE: cell wall-degrading enzymes, Immune: effector families for which at least one member is known to affect the plant innate immune system, Feeding site: effector families for which at least one member is known to be involved in feeding site induction or maintenance.			

For each of these gene families, the copy number differences between Gr-Line19 and Gr-Line22 were determined. The number of paralogs per effector families varied from 99 SPRYSEC variants in Gr-Line22 to a single Hg-GLAND14 gene with signal peptide in the same line. Among the 18 effector families, six have a lower number of paralogs in Gr-Line22, six have an equal number of paralogs, whereas six have a higher number of variants in Gr-Line22 (Fig. 3A).

Four effector families show a relatively large difference in the number of paralogs between Gr-Line19 and Gr-Line22. SPRYSEC is by far the most speciose effector family in both lines, but Gr-Line22 harbor 36 more variants with signal peptide than Gr-Line19. Similarly, 11 Hg-GLAND5 homologs were present in Gr-Line19, while Gr-Line22 comprised 18 paralogs with a signal peptide. The reverse was also observed for Glutathione synthetase (GSS). Five copies were found in Gr-Line19, whereas this effector family is not present in Gr-Line22. Though less radical, a drastic decrease was observed for the subventral gland effector family GH30. Whereas six variants with signal peptide were identified in Gr-Line19, only two were found in Gr-Line22. It is noted that the GH30 family harbors various glycoside hydrolases that were previously categorized as GH5.

## Genomic organisation of effector genes

To characterise the genomic organisation of effector genes, the shortest distance between each gene and the closest adjacent gene was calculated (either at the 3' or 5'-end of the full genomic sequence). This was done for effector genes, as well as for known non-effector genes (i.e. BUSCO gene set). The distances based on the full set of predicted genes ranged from extremely gene sparse to extremely gene dense regions (Fig. 3B). BUSCO genes are generally located in regions that are more gene dense than expected at random (Wilcoxon Rank Sum test  $P < 0.0001$ ). Effector genes expressed in either the dorsal or the subventral esophageal gland cells are often located in more gene sparse regions both as compared to non-effector genes and to any random gene (Wilcoxon Rank Sum test  $P < 0.0001$ ).

Furthermore, the spatial organization and diversification between two pathotypes of *G. rostochiensis* lines is presented for four selected effector families that are expressed in the dorsal esophageal glands during parasitic life stages. Hg-GLAND5 effectors are known as pathogen-associated molecular pattern (PAMP) triggered immunity (PTI) suppressors [19]. Members of the effector family 1106 were demonstrated to suppress both PTI and effector-triggered immunity (ETI) [20]. The highly speciose SPRYSEC family was shown to be involved in both the suppression and the activation of the plant immune system [10]. CLE-like effectors were demonstrated to

be involved in feeding site induction by mimicking the functionality of endogenous host-plant CLE peptides [21]. Concentrating on the distribution of individual family members over the relevant scaffolds, large differences in the level of clustering per family are observed (Fig. 4). While the 60 and 99 SPRYSEC variants are distributed over respectively 18 and 26 scaffolds, the moderately diversified CLE family is concentrated on two scaffolds in case of Gr-Line19, and on a single scaffold for Gr-Line22.

## A. Hg-GLAND5

Hg-GLAND5, also referred to as 'putative gland protein G11A06' [22], has first been discovered in soybean cyst nematode *Heterodera glycines*. This effector is expressed in the dorsal gland during a range of parasitic life stages, and it functions as a PTI suppressor [19]. In a transcriptional analysis of two *H. glycines* races, the expression level of Hg4J4-CT26, a GLAND5 family member, was shown to be highly race-dependent [23]. Searches in public genome database revealed that both PCN species harbour homologs of HG-GLAND5 [24].

In *G. rostochiensis*, the GLAND5 effector family comprises 13 and 23 members in Gr-Line19 and Gr-Line22, respectively. Among these variants, three and five are unlikely to be involved in parasitism as the corresponding protein sequences are not preceded by a signal peptide (SP) for secretion (Fig. 3A). A phylogenetic analysis of the GLAND5 family based on the coding sequences using RaxML revealed an initial split between four effectors without an SP on the one hand, and all functional GLAND5 effectors on the other (Fig. 5). Four clusters with mainly secreted GLAND5 variants could be discerned. Local expansion of effector variants in Gr-Line22 is present in three clusters. Most notable is the diversification in Box I, where eight related GLAND5 representatives from Gr-Line22 surround a single Gr-Line19 variant. Box III and Box IV illustrate expansion in Gr-Line22 as well, although less extreme. Alternatively, Box II shows expansion in Gr-Line19, where two representatives of Gr-Line19 are present and a single variant of Gr-Line22.

## B. 1106

The 1106 gene family encodes mainly secreted proteins, and members were demonstrated to suppress both PTI and ETI responses [20]. In this study, a conserved region of 1106 variants was shown to hybridize in the dorsal gland of infective juveniles of *G. rostochiensis*. Gr-Line19 contains ten paralogs, whereas 14 1106 paralogs were found in Gr-Line22. In terms of organization, the genes in Gr-Line22 and Gr-Line19 show a comparable degree of physical clustering (Fig. 4). To investigate the diversification of the effector family 1106, the phylogenetic relationship between the variants identified in Gr-Line 19 and Line 22 was examined (Fig. 6). In many cases, a 1106 variant in Gr-line 19 had a single, orthologous equivalent in Gr-line22 (see e.g. Gros19\_g2102 and Gros22\_g4744, and Gros19\_g2104 and Gros22\_g4746). Notably, the relationship between the small clusters of 1106 variants was largely unresolved. Within cluster I in Fig. 6, four Gr-Line22 variants were present, and only one representative from Gr-Line19. Two variants, Gros22\_g4703 and Gros22\_g4696, deviate substantially from the other 1106 family members. It is noted that these variants are not preceded by a signal peptide for secretion and thus are unlikely to act as effectors. Cluster II is highlighted as it represents a local expansion of this effector family in Gr-Line22.

## C. SPRYSEC

The SPRYSEC gene family encodes for secreted proteins that contain an SPLa and RYanodine receptor. SPRYSECs are produced in the dorsal esophageal glands, and this effector family is thus far the most expanded one among the plant-parasitic nematodes [11]. Several SPRYSECs from *G. rostochiensis* were shown to be

implicated in the suppression and the activation of defence-related cell death [25]. Suppression was demonstrated for the variants SPRYSEC-4, -5, -8, -15, -18, and -19, whereas only SPRYSEC15 elicited a defence response in tobacco [10]. In the closely related cyst nematode species *G. pallida*, a single SPRYSEC variant - RBP-1 - was shown to be responsible for the evasion of the potato resistance gene *Gpa2*, thus preventing a local HR [12]. No direct ortholog of RBP-1 could be found among the *G. rostochiensis* SPRYSECs (identity lower than 50%), with the used filtering criteria.

The diversification of the SPRYSEC-like variants in Gr-Line19 and Gr-Line22 was investigated by analysing the phylogenetic relationships. Although the number of SPRYSECs in Gr-Line19 ( $n = 60$ ) was already higher than for any other effector family, Gr-Line22 was shown to harbour even more members of this effector family ( $n = 99$ ) (Fig. 3A). Maximum-likelihood-based inference revealed several SPRYSEC clusters (Fig. 7). Due to the poor backbone resolution, no statements can be made about the relationship between these clusters. It is noted that the support values for the more distal parts of the SPRYSEC tree are substantially higher than the support values for most of the more proximal bifurcations. Three large (A, B, and D) and two smaller (C, E) SPRYSEC clusters could be identified. The majority of Gr-Line19 variants have a single orthologous equivalent in Gr-Line22, while Gr-Line22 shows expansion in each of the clusters.

As compared to B and D, cluster A shows the highest level of diversification. Both types of asymmetric SPRYSEC expansion were found in this cluster. Box I in Cluster A comprises a single Gr-Line22 and three Gr-Line19 SPRYSECs. Box II exemplifies Gr-Line22 expansion, where four closely related Gr-Line22 SPRYSECs surround a single Gr-Line19 variant.

Cluster B harbours three of the SPRYSEC variants described in [10] (SPRYSEC-4, -5 and -8), all of which seem to be represented by a single orthologous pair. As the orthologs in both lines are highly similar, we assume that these SPRYSECs are actively suppressing the host plant immune response in both lines.

Cluster C is characterized by a set of genes homologous to SPRYSEC-15 that are considerably expanded in Gr-Line22. It is noted that *Gros19\_g2329* also is the closest match of SPRYSEC-18, however only with 60% identity.

Cluster D unites SPRYSEC variants with a low degree of diversification. Although most Gr-Line19 variants have a single equivalent in Gr-Line22, there are a few examples of further diversification in Gr-Line22. Box III shows a notable example of a diversification event where a single Gr-Line19 variant has five closely related equivalents in Gr-Line22.

SPRYSEC-19, a variant that was demonstrated to suppresses programmed cell death mediated by several immune receptors [25], localized in cluster E (*Gros19\_g12677*). SPRYSEC-19 was first identified in a *G. rostochiensis* Ro1 Mierenbos population [26], which is the population Gr-Line19 was originally derived from. Cluster E shows the Gr-line 22 equivalent of SPRYSEC-19 (*Gros22\_g7323*).

## D. CLE-like

The CLE-like gene family is an unusual effector family coding for prepropeptides that are delivered via the stylet of the infective J2 to the syncytial cell. For CLEs from the related cyst nematode species *Heterodera glycines* with domain structures similar to *G. rostochiensis*, it was shown that the mature propeptide comprised a nematode-specific translocation signal that facilitated the export from the developing syncytium to the apoplast [27].

Subsequently, the protein is cleaved outside the plant cell, and bioactive CLEs are released [28]. Two classes of CLE-like proteins were found to be expressed in the dorsal gland of *G. rostochiensis*. Members of the *Gr-CLE-1* class showed moderate ( $\approx 10$  fold) upregulation in early parasitic life stages (peak in parasitic J-3), whereas *Gr-CLE-4* representatives showed an over 1,000 fold in later parasitic stages (at 21 dpi) [29].

The two *G. rostochiensis* lines 19 and 22 harbour 10 and 16 CLE variants, and both lines comprise members of the *Gr-CLE-1* and the *Gr-CLE-4* class. As shown in the phylogenetic analysis (Fig. 8), members of class *Gr-CLE-4* show little variation among each other, while *Gr-CLE-1s* show a higher level of diversification. As compared to Gr-Line19, the number of *Gr-CLE-4* variants had doubled from four to eight in Gr-Line22. On the contrary, each of the Gr-Line19 representatives of *Gr-CLE-1* had a single homolog in Gr-Line22. In Fig. 8, Box I shows an example of a homologous gene pair, with a tentative duplication in Gr-Line22.

In addition to sequence similarity within the Gr-CLE function classes, there is also a high degree of physical clustering (Fig. 9). The Gr-CLE-4 variants are all located adjacent to each other, and not interspersed by any other gene. Based on this remarkable physical organization, we hypothesize that one or more duplication events in this region underlies the copy number difference of Gr-CLE-4 effectors ( $n = 4$  in Gr-Line19;  $n = 8$  in Gr-Line22).

## Discussion

Due to specific biological characteristics of plant-parasitic nematodes, host plant resistances tend to be a remarkably durable means to manage this category of soil-borne pathogens. The main challenge is the actual developing and breeding resistant host-plant varieties. As the genetic basis for virulence in plant-parasitic nematodes is unknown, breeding for resistance can only be done on a trial-and-error basis. The whole process is, therefore, inefficient, and thus time consuming and expensive. The availability of molecular-based pathotyping methods of plant-parasitic nematode populations would allow for the design of more targeted resistance. Here we concentrated on two *Globodera rostochiensis* inbred lines, Gr-Line19 and Gr-Line22, with distinct pathotypes [17]. Resulting from a single male-female crossing by Janssen et al. (1990) [17], each of these lines' genomic background is small, with a maximum of 4 haplotypes per locus. These small genomic backgrounds significantly simplify the generation of high-quality reference genomes, which has been a challenge for sexually reproducing plant-parasitic nematodes in the past. Therefore, we expect that the reference genome sequences of Gr-Line19 and Gr-Line22 are a more accurate representation of the *G. rostochiensis* genome, bringing the concept of molecular pathotyping a step closer. Furthermore, long-read sequencing technology allowed us to generate reference genomes about 24 fold less fragmented than the current reference genome [14]. This higher continuity made it possible to pinpoint the physical distribution and the diversification in a way that was not possible with the highly fragmented JHI-Ro1 reference genome sequence.

One of the main technical challenges we tried to overcome was generating high-quality reference genome sequences of a highly heterozygous nematode species. The assembly sizes of the novel reference genomes are comparable to the *G. rostochiensis* JHI-Ro1 genome [14] as well as to the estimated *G. rostochiensis* genome size [30]. Therefore, it is likely that the high levels of heterozygosity did not negatively impact the assembly by the presence of haplotigs [31]. To further reduce the number of scaffolds, possibly to a chromosome level, it might be advantageous in the future to supplement long-read sequencing with other techniques such as optical mapping [32, 33].

We furthermore assessed the effect of generating a genome assembly of a highly inbred line instead of a field population. A comparison was made between SNPs' zygosity called on the basis of short-read data and we observed that Gr-Line19 had a 10% higher proportion of homozygous SNPs than the JHI-Ro1 population. Since more than 50% of the called SNPs in Gr-Line19 are still heterozygous, it seems reasonable to assume that the currently measured heterozygosity levels provide a more realistic picture of the heterozygosity that is present in an individual. This observation is in line with previous findings for the cyst nematode *Heterodera glycines* by Ste-Croix et al. (2021) [34] showing that individuals from a single population can be homozygous or heterozygous vis-à-vis putative virulence genes.

A more detailed analysis of four effector families for which at least a subset of members are known to be expressed in the dorsal gland of nematodes during feeding site induction or maintenance revealed dozens of novel potentially virulence-associated variants.

To some extent, our starting point was comparable to the approach taken by Bekal et al. (2008) [35]. Within the soybean cyst nematode *Heterodera glycines* subsets of populations with comparable pathogenicity have been defined based on their multiplication characteristics on a set of seven soybean indicator lines. Populations that shared multiplication characteristics were coined 'HG types' [36]. Subsequently, [35] used two inbred lines that were either avirulent ('TN10'; HG type 0) or virulent ('TN20'; HG type 1, 2, 3, 4, 5, 6, 7). 454 micro-bead sequencing of these indicator lines resulted in the generation of tens of millions of short reads (110–120 bp), which allowed for whole-genome comparative analysis. These efforts resulted in 239 homozygous SNPs between TN20 and TN10 [35]. Although the relationship between these SNPs and pathogenicity is unclear, these SNPs could be considered one of the first molecular markers for pathogenicity in cyst nematodes. Here we took it one step further, by identifying copy number variation that might serve as potential pathotype-specific molecular markers. Copy number variation is relevant, as it has been linked to virulence in various pathogens [37, 38] including plant-parasitic nematodes [39].

For potato cyst nematodes, Folkertsma et al. (1996) [40] used AFLP assays [41] to characterize pathotypes of the potato cyst nematodes *G. rostochiensis* and *G. pallida*. Almost 1,000 marker loci were employed to genotype populations of both potato cyst nematode species. These analyses revealed genetic markers that can distinguish between the *G. rostochiensis* pathotypes Ro1, Ro3 and Ro4, while such loci appeared to be absent for the *G. pallida* pathotypes Pa2 and Pa3. In a more extensive approach focussing on *G. rostochiensis* only, Mimeo et al. (2015) [8] employed a restriction enzyme-based genotyping-by-sequencing approach. The genotypic characterization of 23 populations, covering all five pathotypes, revealed a clear distinction between pathotypes Ro1 and Ro2 on the one hand, and Ro3, Ro4, and Ro5 on the other. Moreover, their analyses seemed to demonstrate intra-pathotype variation within Ro1. However, it is noted that with 14 populations from 9 different countries, Ro1 was substantially overrepresented in this research.

The first reference genome for *G. rostochiensis* was published by Eves-van den Aker (2016) [14], and - in conjunction with this - the intra-species variation regarding members of known effector families was mapped. All five *G. rostochiensis* pathotypes were represented in this study, and whereas homozygous molecular markers to discriminate between Ro4 and Ro5 could be identified, this was not possible for the remaining three pathotypes. Moreover, this research confirmed the large genotypic diversity of populations that are all labeled Ro1, indicating that there are many possible genotypes that yield a similar Ro1-like virulence. Here it might be mentioned that Ro1 and Ro4 share the inability to parasitize potato genotypes that harbor the *H1* resistance gene from *Solanum*

*tuberosum* ssp. *andigena* CPC 1673. Moreover, the H1 resistance genes have been introgressed in most commercial potato varieties, and potato cyst nematode populations worldwide have been exposed to these resistance genes likely including the ones characterized by Eves-van den Akker (2016) [14]. So, although these pathotypes share their avirulence with regard to the *H1* gene, they belong to another *G. rostochiensis* genotype and differ significantly in intra-pathotype variation.

We used pathotypically characterized inbred lines as a starting point from which we generated new reference genome sequences. On this basis, complete effector families could be mapped and compared. In essence, the make-up of effector families in lines with distinct pathogenic characteristics could vary because of (1) non-synonymous variants in sequence in a given set of effector genes and/or (2) effector gene loss or gain. The balance between these two (dependent) sources of variation varies in a pathogen-dependent manner. The genome-wide comparison of three *Microbotryum* species parasitizing distinct Caryophyllaceae allowed Beckerson et al. (2019) [42] to define the secretomes of the individual species. Their analyses revealed that host specificity was explained by rapid changes in effector genes rather than by variation in the effector copy numbers. With a similar underlying question, Qutob et al. (2009) [43] investigated two effector genes families of *Phytophthora sojae*, Avr1a and Avr3a in a range of races. The presence of multiple copies of nearly identical genes on the Avr1a and the Avr3a locus was suggested to contribute to the fitness of these races, and races with distinct pathogenicities were characterized by variations in effector gene numbers. These examples demonstrate that both sources of variation can generate differences in pathogenicity among plant pathogens. Here we specifically focussed on effector gene loss and gain effects, and observed that both events happen in the avirulent Gr-Line19 as well as the virulent Gr-Line22. Previous studies show that, at least in potato cyst nematodes, single nucleotide polymorphisms are also related to virulence (e.g. [12]), which indicates that both types of genomic variance are relatable with virulence.

In case of the tropical root-knot nematode *Meloidogyne incognita*, [39] tried to pinpoint the genetic basis of avirulence and virulence with regard to the tomato resistance gene *Mi-1.2*. Genome-wide characterization of two pairs of avirulent and virulent lines revealed 20 gene families that all showed a lower number of copies in both virulent *M. incognita* lines. It is noted that the 20 families included pioneers and household genes, and not known effector families. Hence, although a lower copy number per gene family was associated with virulence, this research did not identify gene loss events that could be causally related to virulence.

We separately considered the dorsal esophageal gland-expressed effectors that are thought to be involved in immune response suppression and feeding site induction, and the subventral gland-expressed effectors that are active during plant penetration. Concerning dorsal gland-expressed effector families, *G. rostochiensis* Gr-Line19 harboured on average 14 genes per effector family, while on average, 19 members were identified per effector family in Gr-Line22, a homozygous virulent line with regard to the H1 resistance gene. Two exceptions to this overall increase in effector copy number should be mentioned; GSS appeared to be absent in Gr-Line22, while lower copy numbers were observed for Hg-GLAND6. GSS is a peculiar example as this was the first example of a cyst nematode effector family that arose from neofunctionalization of a household gene, glutathione synthetase [44]. Two expansion waves of GS-like genes were described, and only in the second wave GS-like genes with a signal peptide for secretion arose. The resulting clade 3 comprised cyst nematode effectors. It should be noted that GSS genes in this clade share only 36% protein identity. The thresholds used in this study to identify GSS family members were presumably too strict to identify more GS-like members.

The effector families expressed in the subventral glands that are included in this study showed less expansion than the dorsal gland specific families, with only small differences in copy numbers between the two lines. Strikingly, a substantial number of genes belonging to this category lack a signal peptide presence. Since many of these genes (e.g., glycoside hydrolases, pectate lyases) code for cell wall-degrading or modifying enzymes, the proteins would have to be secreted to make physical contact with the plant in order to perform their function. One hypothesis could be that these genes are, in fact, pseudogenes. However, this seems unlikely as manual inspection showed that most of these SP lacking genes show a RNAseq signal (results not shown). Whereas ample RNAseq data allowed for an accurate prediction of the intron-exon structure, the transcription start site is more difficult to predict without additional experimental data. If transcription start sites were misplaced, we could have missed a preceding signal peptide. Alternatively, it could be that this cyst nematode genuinely harbours effector variants without apparent signal peptide similar to the invertases identified in *Meloidogyne incognita*. These effectors were suggested to be acquired at a late stage during cyst nematode evolution [45].

Phylogenetic analysis of effector families as presented here takes along both effector diversification and effector loss and gain. These data clearly demonstrate that the balance between both sources of variation differs per effector family. Whereas effector family 1106 showed overall little copy number variation, SPRYSECs showed a 65% increase in copy number in Gr-Line22, and in case of the GLAND5 family significant diversification was accompanied by a sharp increase of the copy number in Gr-Line22. Other population genetic studies on plant-pathogenic fungi and oomycetes showed exclusively low [46] or high [47] levels of diversification between effector genes. We are not aware of other plant pathogens for which such drastic contrasts in diversification pattern between effector families were described.

Because of its extreme level of physical clustering of CLE effectors in both *G. rostochiensis* inbred lines we investigated its genomic organisation in more detail. Potato cyst nematodes produce and secrete mimics of plant CLEs. Plant CLEs are signalling components that were shown to be conserved in both Arabidopsis and potato roots [29]. Among *Globodera* CLE genes two functional classes are distinguished, CLE1 and CLE4. The main difference between these classes is the composition of CLE peptides that are present as small cleavable units separated by small spacers at the protein's C terminus. Mitchum et al. (2012) [21] described a single CLE1 representative, and here a second potential CLE1 variant is identified in both *G. rostochiensis* lines (Fig. 8a). This second variant has a domain structure similar to GrCLE1 (Supplemental Fig. 3), and the conservation of the domain structure makes it plausible this variant has a CLE1-like function. Notable is the putative duplication event of Gr-CLE4 genes in Gr-Line22. Gr-CLE-4 genes are highly conserved, even between pathotypes and we assume that this duplication event might result in a higher production of GrCLE4 peptides. A dose effect for a nematode effector was previously reported for the 32E03 effector of the beet cyst nematode *Heterodera schachtii* [48]. So our finding might suggest that the virulent *G. rostochiensis* line 22 might exert a stronger CLE4 peptide-based effects on its host.

## Conclusions

Molecular pathotyping is an essential element in durable disease management. After all, this will allow breeders to use host plant resistances in a targeted way, and it allows farmers to make a more informed decision which potato variety to grow in the field. The existing pathotyping system for *G. rostochiensis* classifies populations into five pathotypes (Ro1-Ro5) on the basis of their relative multiplication rates on a number of *Solanum* differentials, and this systematic was used as starting point for the generation of new pathotyping platform. By generating

high quality reference genomes from two pathotypically-distinct inbred lines, we were able to generate broad overviews of effector families including their diversification and spatial organisation. On the basis of a selection of four effector families, dozens of effector variants could be pinpointed that were unique for either of the two inbred lines Gr-Line19 (Ro1) and Gr-line22 (Ro5). Once these data are supplemented by re-sequencing data from well-characterized *G. rostochiensis* field populations, comparative effectomics would be within reach. Comparative effectomics will provide a fundament for our understanding of compatible and incompatible host-nematode interactions as well as for a new, biologically insightful pathotyping scheme as a basis for the durable use of host plant resistances.

## Methods

### Experimental design

This study was designed to generate new reference genomes for two *G. rostochiensis* inbred lines that differ in their responses to the dominant potato resistance gene *H1* (Gr-Line 19: fully avirulent, Gr-Line 22: fully virulent). After the development of a gentle nematode lysis protocol, PacBio long-read technology was used to sequence these two lines. Assembled reference genomes were used to study the positioning and the diversification of a wide range of effector families that are expressed in early parasitic stages (pre-parasitic J2) in the subventral esophageal glands or in later developmental stages (parasitic J2, J3, J4) in the dorsal esophageal glands. Via a protruded stylet nematode injected effectors produced in these glands into their host plant. Upon the identification of effector family members, phylogenetic relationships were established both within and between the two distinct inbred lines, and their spatial organisations were established and compared.

### DNA isolation and sequencing

Cysts from two *G. rostochiensis* lines that were selected for being fully avirulent Ro1 (Gr-line19) or fully virulent Ro5 (Gr-line22) with regard to the *H1* gene [17] were used as starting material for the collection of pre-parasitic second-stage juveniles (J2). J2 nematodes were concentrated, and sucrose centrifugation was used to purify the nematode suspension [49]. After multiple rounds of washing of the purified nematode suspension in 0.1 M NaCl, nematodes were resuspended in sterilized 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 [50]. 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). DNAs of both inbred lines (each 10–20 µg) were sequenced using Pacific Biosciences SMRT sequencing technology at Bioscience (Wageningen Research, Wageningen, The Netherlands) Gr-line19 was sequenced to a depth of approximately 119X with an average read length of 5,641 bp, whereas Gr-line22 was sequenced 132X with an average read length of 7,469 bp. Depth was calculated based on the assembly sizes. In parallel, a 2x 250 bp Illumina NovaSeq run resulted in 188x coverage of paired-end reads per line used to polish the initial assemblies. The raw sequencing reads and the genome assemblies are available under NCBI accession PRJNA695196.

### Genome assembly

Raw PacBio reads were first corrected by merging haplotypes with the correction mode of Canu v1.8 [51], allowing a corrected error rate of 15% and a corrected coverage of 200. Using long-read assembler wtdgb2 v2.3 [52],

approximately one hundred assemblies were generated per inbred line, optimizing the parameters minimal read length, k-mer size, and minimal read depth. The quality of the initial assemblies was assessed based on whether the assembly size was close to the genome size estimate [14]. Completeness of the genome was assessed using BUSCO v3 [53] using the standard library of eukaryotic single copy genes. Based on the criteria mentioned before, the most optimal assembly was then selected for each line and used for post-assembly processing.

After determining the most optimal assembly, remaining unmerged haplotigs were filtered from the assembly using Purge Haplotigs v1.0.4 [31]. The assembly was then tested for contamination using the blobtools pipeline v.1.0.1 (Supplemental Figs. 1 and 2) [54]. Contigs were scaffolded with PacBio reads using SSPACE-Longread [55]. The remaining gaps in the scaffolds were then filled using a consensus alignment approach [56] NovaSeq data were used to polish the resulting assemblies using three iterations of Arrow v2.3.3 (<https://github.com/PacificBiosciences/GenomicConsensus>) and five iterations of Pilon v1.23 [57] each. Repeat regions were soft masked using RepeatModeler v1.0.11 (<https://github.com/Dfam-consortium/RepeatModeler>) and RepeatMasker v4.0.9 (<http://www.repeatmasker.org/RepeatMasker/>). Gene annotations were predicted for both assemblies using BRAKER v2.1.2 [58] aided by RNAseq datasets of different *G. rostochiensis* life stages that were mapped using Hisat v2.2.0 [59] (NCBI BioProject accessions: PRJEB12075, PRJNA274143, PRJNA274143).

## Genome Synteny

Genome synteny was determined between the genome assembly of Line19, Line22 and the previous Ro1 reference genome (NCBI BioProject PRJEB13504) through a progressive genome alignment using Mauve v2.4.0. The alignment was then visualized in Circos v0.69-9 [60], showing only syntenic regions of 3kb and larger.

## Estimating Heterozygosity levels and structural variation

Heterozygosity levels were estimated based on the frequency of heterozygous *versus* homozygous variants (SNPs and small indels). A comparison was made between Gr-Line19 and the JHI-Ro1 population, using the Gr-Line22 genome assembly as a reference. Illumina reads were mapped with Burrows-Wheeler Aligner [61] using default settings. For Gr-Line19, a library of Illumina NovaSeq reads (accessions: SRR13560389, SRR13560388) was used, and for JHI-Ro1, Illumina HiSeq reads (accessions: ERR114519) were mapped against the reference. Variants were called with bcftools v.1.9 [62] with multiallelic variant calling enabled, at a maximum depth of 1,000 reads.

The structural variation between the newly generated assemblies of Gr-Line19 and Gr-Line22 was estimated by the frequency of heterozygous *versus* homozygous structural variants (SVs) with fragment size > 1kb. Raw Pacbio reads of Gr-Line19 were mapped against the Gr-Line22 (and *vice versa*) using NGMLR v0.2.7 [63] with default settings. SVs were then called using Sniffles v1.0.10 running the standard settings [63].

## Identification of Effector Homologs

Effector genes were identified in both line Gr-line19 and Gr-line22 based on the proteomes predicted by BRAKER2 [58]. Phobius [64] was used to check for the presence of a signal peptide for secretion. Homologs for glycoside hydrolase (GH) families 5, 30, 43, 53, Pectate lyase 3, Glutathione Synthetase were identified with HMMER v3.2.1 [65] based on pre-calculated profile HMMs in the PFAM database [66] (entries PF00150, PF02055, PF04616 and PF07745, PF03211, PF02955 respectively). SPRYSEC homologs were identified by testing protein sequences for a SPRY domain (hmm profile PF00622). Arabinogalactan galactosidase homologs were identified with a custom profile HMM-based on UniProt sequences (entries O07012, Q65CX5, Q65CX4, D9SM34, P48841, O31529, Q8X168,

Q5B153, O07013, P83692, P48842, P83691, Q4WJ80, B0XPR3, A1D3T4, Q2UN61, Q0CTQ7, A2RB93, Q9Y7F8, B8NNI2, Q76FP5). CLE-like homologs were identified with a custom profile HMM-based on UniProt sequences (D1FNJ7, D1FNK5, D1FNJ9, D1FNK2, D1FNK8, D1FNK3, D1FNK0, D1FNK4). GenBank peptide sequences JQ912480 to JQ912513 were used to generate a custom profile HMM for the effector family 1106. Based on GenBank entries KM206198 to KM206272, a custom profile HMM was made for the HYP effector family. Homologs of the *Heterodera glycines* effector families Hg16B09 (GenBank: AAO85454) and GLAND1-18 (GenBank: KJ825712 to KJ825729) were identified with blastp, with the following cut-offs: an identity score higher than 35%, a query coverage of at least 50%, and an E-value lower than 0.0001.

## Phylogeny

Multiple Sequence Alignments were generated based on the coding sequences of the orthologs per effector family, using Muscle v3.8.1551 [67] using standard options. Phylogenetic trees were then generated with RaxML v8.2.12 [68] running a GTRGAMMA model with 100 bootstrap replicates. The resulting trees were visualized and organized in Figtree (v. 1.4.4).

## Declarations

## Ethics approval

Ethics approval was not needed for this study.

## Consent for publications

Consent for publication was not needed for this study.

## Availability of data and materials

The datasets generated and/or analysed during the current study are available under NCBI Bioproject accession PRJNA695196. Scripts used for the generation of genome assemblies, and result figures are available open-source under the MIT license on Github ([https://github.com/Jorisvansteenbrugge/GROS\\_genomes](https://github.com/Jorisvansteenbrugge/GROS_genomes)).

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was funded as part of a grant by the Netherlands Organization for Scientific Research (NWO) as part of the Applied and Technical Science domain (TTW) under grant no. 14708. PT: bioinformatics and computational biology analyses were supported by the University of St Andrews Bioinformatics Unit (AMD3BIOINF), funded by Wellcome Trust ISSF award 105621/Z/14/Z.

## Acknowledgements

We thank John T. Jones for sharing his expertise in the assembly of plant-parasitic nematode genome sequences.

## Contributions

SvdE performed the DNA extraction and library preparations. JvS generated the genome assemblies. JvS performed the comparative genomics/effectomics and phylogenetic analyses. JvS and JH wrote the manuscript; all other authors co-commented on the manuscript.

## References

1. Nicol JM, Turner SJ, Coyne DL, den Nijs L, Hockland S, Maafi ZT: **Current Nematode Threats to World Agriculture.** In *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Edited by Jones J, Gheysen G, Fenoll C. Dordrecht (The Netherlands): Springer; 2011: 21–43
2. Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, Kikuchi T, Manzanilla-López R, Palomares-Rius JE, Wesemael WML, Perry RN: **Top 10 plant-parasitic nematodes in molecular plant pathology.***Molecular Plant Pathology* 2013, **14**:946-961.
3. Plantard O, Picard D, Valette S, Scurrah M, Grenier E, Mugniéry D: **Origin and genetic diversity of Western European populations of the potato cyst nematode (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci.***Molecular Ecology* 2008, **17**:2208-2218.
4. Fuller JM, Howard HW: **Breeding for resistance to the white potato cyst-nematode, *Heterodera pallida*.***Annals of Applied Biology* 1974, **77**:121 - 112.
5. Toxopeus HJ, Huijsman CA: **Genotypical background of resistance to *Heterodera rostochiensis* in *Solanum tuberosum*, var. *andigenum* [1].***Nature* 1952, **170**:1016.
6. Kort J, Ross H, Rumpfenhorst HJ, Stone AR: **An International Scheme for Identifying and Classifying Pathotypes of Potato Cyst-Nematodes *Globodera rostochiensis* and *G. pallida*.***Nematologica* 1977, **23**:333-339.
7. Phillips MS, Trudgill DL: **Variations in the ability of *Globodera pallida* to produce females on potato clones bred from *Solanum vernei* or *S. tuberosum* ssp. *andigena* CPC 2802.***Nematologica* 1983, **29**:217-226.
8. Mimee B, Duceppe MO, Véronneau PY, Lafond-Lapalme J, Jean M, Belzile F, Bélair G: **A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis.***Molecular Ecology Resources* 2015, **15**:1356-1365.
9. Thevenoux R, Folcher L, Esquibet M, Fouville D, Montarry J, Grenier E: **The hidden diversity of the potato cyst nematode *Globodera pallida* in the south of Peru.***Evolutionary Applications* 2020, **13**:727-737.
10. Ali S, Magne M, Chen S, Obradovic N, Jamshaid L, Wang X, Bélair G, Moffett P: **Analysis of *Globodera rostochiensis* effectors reveals conserved functions of SPRYSEC proteins in suppressing and eliciting plant immune responses.***Frontiers in Plant Science* 2015, **6**:623.
11. Diaz-Granados A, Petrescu AJ, Goverse A, Smart G: **SPRYSEC Effectors: A Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity.***Front Plant Sci* 2016, **7**:1575.
12. Sacco MA, Koropacka K, Grenier E, Jaubert MJ, Blanchard A, Goverse A, Smart G, Moffett P: **The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death.***PLoS Pathogens*

2009, **5**:e1000564.

13. Lozano-Torres JL, Wilbers RHP, Gawronski P, Boshoven JC, Finkers-Tomczak A, Cordewener JHG, America AHP, Overmars HA, Van 't Klooster JW, Baranowski L, et al: **Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode.***Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:10119-10124.
14. Eves-van den Akker S, Laetsch DR, Thorpe P, Lilley CJ, Danchin EGJ, Da Rocha M, Rancurel C, Holroyd NE, Cotton JA, Szitenberg A, et al: **The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence.***Genome Biology* 2016, **17**:124.
15. Cotton JA, Lilley CJ, Jones LM, Kikuchi T, Reid AJ, Thorpe P, Tsai IJ, Beasley H, Blok V, Cock PJA, et al: **The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode.***Genome biology* 2014, **15**:R43.
16. Rouppe van der Voort JNAM, Van Enckevort ELJG, Pijnacker LP, Helder J, Gommers FJ, Bakker J: **Chromosome number of the potato cyst nematode *Globodera rostochiensis*.***Nombre de chromosomes du nematode a kyste de la pomme de terre *Globodera rostochiensis** 1996, **19**:369-374.
17. Janssen R, Bakker J, Gommers FJ: **Selection of virulent and avirulent lines of *Globodera rostochiensis* for the H1 resistance gene in *Solanum tuberosum* ssp. *andigena* CPC 1673.***Revue de Nématologie* 1990, **13**:265-268.
18. McLean F, Berger D, Laetsch DR, Schwartz HT, Blaxter M: **Improving the annotation of the *Heterorhabditis bacteriophora* genome.***GigaScience* 2018, **7**.
19. Pogorelko G, Wang J, Juvale PS, Mitchum MG, Baum TJ: **Screening soybean cyst nematode effectors for their ability to suppress plant immunity.***Molecular Plant Pathology* 2020, **21**:1240-1247.
20. Finkers-Tomczak A: **Co-evolution between *Globodera rostochiensis* and potato driving sequence diversity of NB-LRR resistance loci and nematode suppressors of plant immunity.** Wageningen University, Laboratory of Nematology; 2011.
21. Mitchum MG, Wang X, Wang J, Davis EL: **Role of nematode peptides and other small molecules in plant parasitism.** In *Annual Review of Phytopathology*, vol. 50. pp. 175-195; 2012:175-195.
22. Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS: **The parasitome of the phytonematode *Heterodera glycines*.***Molecular Plant-Microbe Interactions* 2003, **16**:720-726.
23. Wang G, Peng D, Gao B, Huang W, Kong L, Long H, Peng H, Jian H: **Comparative transcriptome analysis of two races of *Heterodera glycines* at different developmental stages.***PLoS ONE* 2014, **9**:e91634.
24. Yang S, Pan L, Chen Y, Yang D, Liu Q, Jian H: ***Heterodera avenae* GLAND5 Effector Interacts With Pyruvate Dehydrogenase Subunit of Plant to Promote Nematode Parasitism.***Front Microbiol* 2019, **10**:1241.
25. Postma WJ, Sloopweg EJ, Rehman S, Finkers-Tomczak A, Tytgat TOG, van Gelderen K, Lozano-Torres JL, Roosien J, Pomp R, van Schaik C, et al: **The effector SPRYSEC-19 of *Globodera rostochiensis* suppresses CC-NB-LRR-mediated disease resistance in plants.***Plant Physiology* 2012, **160**:944-954.
26. Rehman S, Postma W, Tytgat T, Prins P, Qin L, Overmars H, Vossen J, Spiridon LN, Petrescu AJ, Govere A, et al: **A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR protein from a susceptible tomato.***Molecular Plant-Microbe Interactions* 2009, **22**:330-340.
27. Wang J, Dhroso A, Liu X, Baum TJ, Hussey RS, Davis EL, Wang X, Korkein D, Mitchum MG: **Phytonematode peptide effectors exploit a host post-translational trafficking mechanism to the ER using a novel**

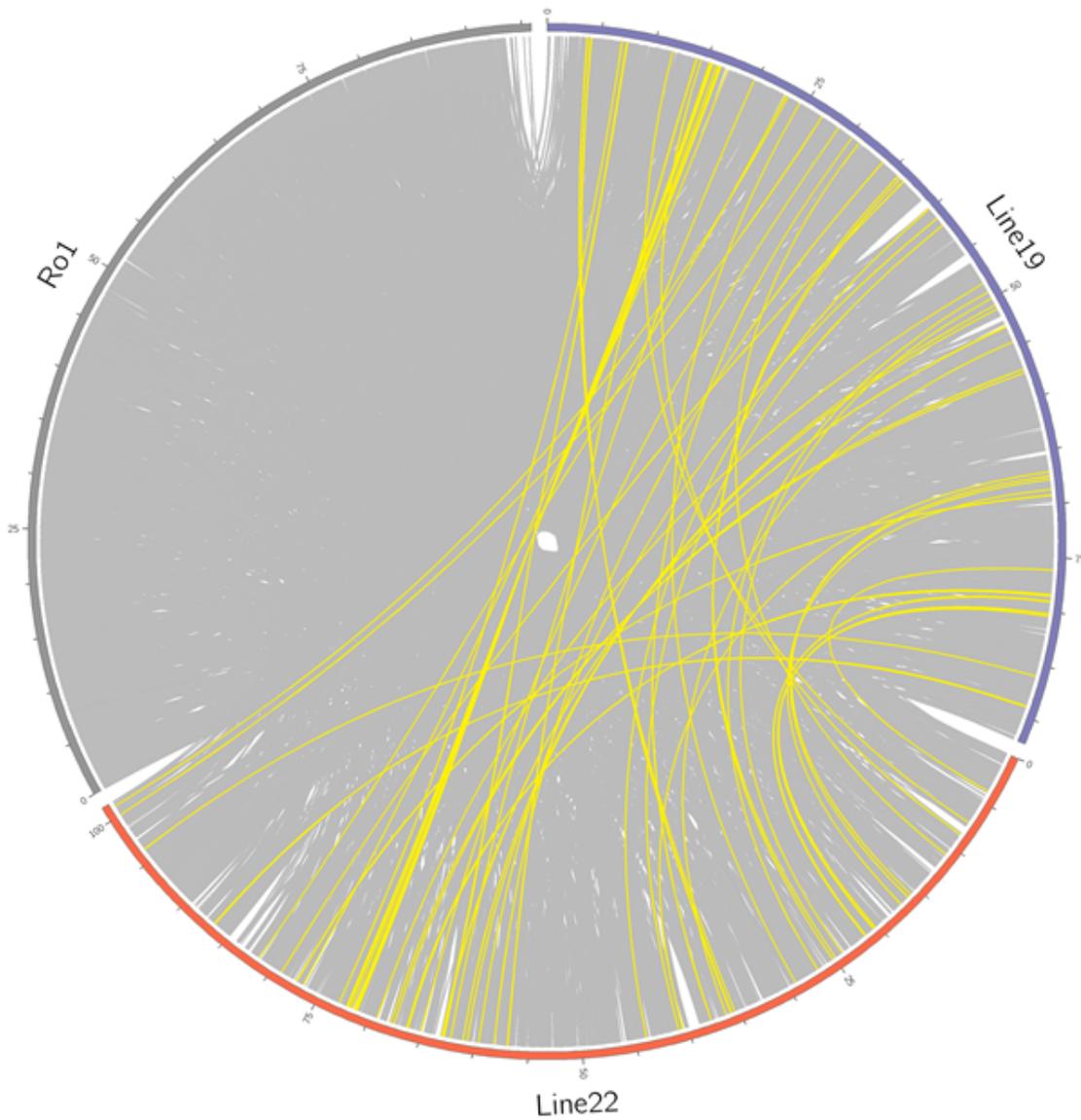
**translocation signal.***New Phytologist* 2020.

28. Gheysen G, Mitchum MG: **Phytoparasitic nematode control of plant hormone pathways.***Plant Physiology* 2019, **179**:1212-1226.
29. Lu SW, Chen S, Wang J, Yu H, Chronis D, Mitchum MG, Wang X: **Structural and functional diversity of CLAVATA3/ESR (CLE)-like genes from the potato cyst nematode *Globodera rostochiensis*.***Molecular Plant-Microbe Interactions* 2009, **22**:1128-1142.
30. Grisi E, Burrows PR, Perry RN, Hominick WM: **The genome size and chromosome complement of the potato cyst nematode *Globodera pallida*.***Fundam Appl Nematol* 1995, **18**:67–70.
31. Roach MJ, Schmidt SA, Borneman AR: **Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies.***BMC Bioinformatics* 2018, **19**:460.
32. Deschamps S, Zhang Y, Llaca V, Ye L, Sanyal A, King M, May G, Lin H: **A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping.***Nature Communications* 2018, **9**:4844.
33. Field MA, Rosen BD, Dudchenko O, Chan EKF, Minoche AE, Edwards RJ, Barton K, Lyons RJ, Tuipulotu DE, Hayes VM, et al: **Canfam-GSD: De novo chromosome-length genome assembly of the German Shepherd Dog (*Canis lupus familiaris*) using a combination of long reads, optical mapping, and Hi-C.***GigaScience* 2020, **9**:giaa027.
34. Ste-Croix DT, Gendron St-Marseille A-F, Lord E, Bélanger RR, Brodeur J, Mimee B: **Genomic Profiling of Virulence in the Soybean Cyst Nematode Using Single-Nematode Sequencing.***Phytopathology* 2021, **111**:137-148.
35. Bekal S, Craig JP, Hudson ME, Niblack TL, Domier LL, Lambert KN: **Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing.***Molecular Genetics and Genomics* 2008, **279**:535-543.
36. Niblack TL, Arelli PR, Noel GR, Opperman CH, Orf JH, Schmitt DP, Shannon JG, Tylka GL: **A revised classification scheme for genetically diverse populations of *Heterodera glycines*.***Journal of Nematology* 2002, **34**:279-288.
37. Brynildsrud O, Gulla S, Feil EJ, Nørstebø SF, Rhodes LD: **Identifying copy number variation of the dominant virulence factors msa and p22 within genomes of the fish pathogen *Renibacterium salmoninarum*.***Microbial genomics* 2016, **2**:e000055.
38. Zhao S, Gibbons JG: **A population genomic characterization of copy number variation in the opportunistic fungal pathogen *Aspergillus fumigatus*.***PLoS ONE* 2018, **13**:e0201611.
39. Castagnone-Sereno P, Mulet K, Danchin EGJ, Koutsovoulos GD, Karaulic M, Da Rocha M, Bailly-Bechet M, Pratz L, Perfus-Barbeoch L, Abad P: **Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita*.***Molecular Ecology* 2019, **28**:2559-2572.
40. Folkertsma RT, Rouppe Van Der Voort JNAM, De Groot KE, Van Zandvoort PM, Schots A, Gommers FJ, Helder J, Bakker J: **Gene pool similarities of potato cyst nematode populations assessed by AFLP analysis.***Molecular Plant-Microbe Interactions* 1996, **9**:47-54.
41. Vos P, Hogers R, Bleeker M, Reijans M, Vandeleee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M: **AFLP - a New Technique for DNA-Fingerprinting.***Nucleic Acids Research* 1995, **23**:4407-4414.
42. Beckerson WC, Rodríguez De La Vega RC, Hartmann FE, Duhamel M, Giraud T, Perlin MH: **Cause and effectors: Whole-genome comparisons reveal shared but rapidly evolving effector sets among host-specific plant-castrating fungi.***mBio* 2019, **10**:e02391.

43. Qutob D, Tedman-Jones J, Dong S, Kuflu K, Pham H, Wang Y, Dou D, Kale SD, Arredondo FD, Tyler BM, Gijzen M: **Copy number variation and transcriptional polymorphisms of *Phytophthora sojae* RXLR effector genes Avr1a and Avr3a.***PLoS ONE* 2009, **4**:e5066.
44. Lilley CJ, Maqbool A, Wu D, Yusup HB, Jones LM, Birch PRJ, Banfield MJ, Urwin PE, Eves-van den Akker S: **Effector gene birth in plant parasitic nematodes: Neofunctionalization of a housekeeping glutathione synthetase gene.***PLoS Genetics* 2018, **14**:e1007310.
45. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, et al: **Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*.***Nature Biotechnology* 2008, **26**:909-915.
46. Talas F, McDonald BA: **Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination.***BMC Genomics* 2015, **16**:996.
47. Flier WG, Grünwald NJ, Kroon LPNM, Sturbaum AK, Van Den Bosch TBM, Garay-Serrano E, Lozoya-Saldaña H, Fry WE, Turkensteen LJ: **The population structure of *Phytophthora infestans* from the Toluca Valley of central Mexico suggests genetic differentiation between populations from cultivated potato and wild *Solanum* spp.***Phytopathology* 2003, **93**:382-390.
48. Vijayapalani P, Hewezi T, Pontvianne F, Baum TJ: **An effector from the cyst nematode *Heterodera schachtii* derepresses host rRNA genes by altering histone acetylation.***Plant Cell* 2018, **30**:2795-2812.
49. Jenkins WR: **A rapid centrifugal-flotation technique for separating nematodes from soil.***Plant Disease Reporter* 1964, **48**:48.
50. Holterman M, van der Wurff A, van den Elsen S, van Megen H, Bongers T, Holovachov O, Bakker J, Helder J: **Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades.***Molecular Biology and Evolution* 2006, **23**:1792-1800.
51. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM: **Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.***Genome Research* 2017, **27**:722-736.
52. Ruan J, Li H: **Fast and accurate long-read assembly with wtdbg2.***Nature Methods* 2020, **17**:155-158.
53. Seppey M, Manni M, Zdobnov EM: **BUSCO: Assessing genome assembly and annotation completeness.** In *Methods in Molecular Biology*, vol. 1962. pp. 227-245; 2019:227-245.
54. Laetsch DR, Blaxter ML: **BlobTools: Interrogation of genome assemblies.***F1000Research* 2017, **6**:1287.
55. Boetzer M, Pirovano W: **SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information.***BMC Bioinformatics* 2014, **15**:211.
56. van Steenbrugge JJM: **Jorisvansteenbrugge/GapFiller: GROS Assembly version.***Zenodo* 2021.
57. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM: **Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement.***PLoS ONE* 2014, **9**:e112963.
58. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M: **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.***NAR Genomics and Bioinformatics* 2021, **3**:108.
59. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.***Nature Biotechnology* 2019, **37**:907-915.

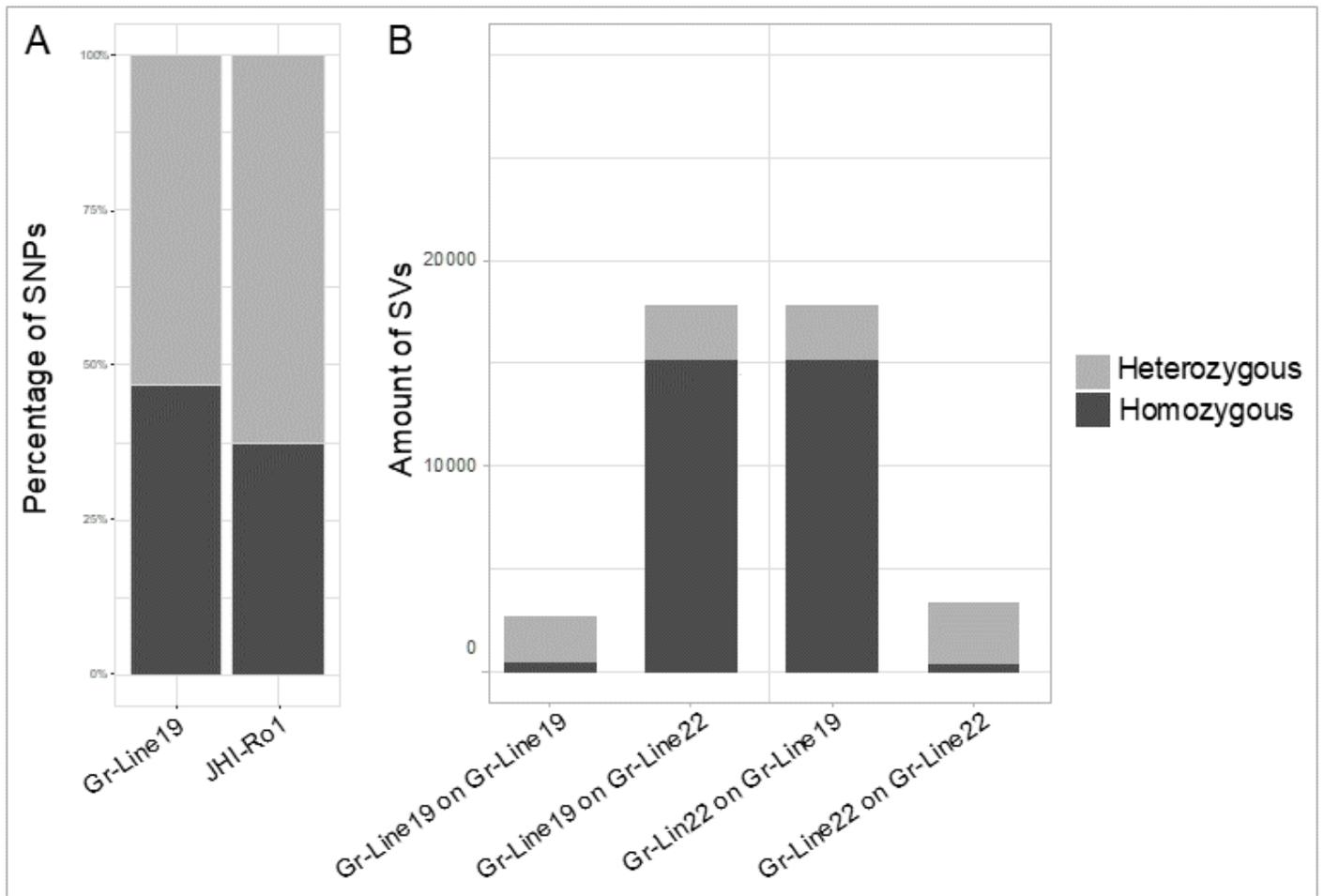
60. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: An information aesthetic for comparative genomics.***Genome Research* 2009, **19**:1639-1645.
61. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.***Bioinformatics* 2009, **25**:1754-1760.
62. Li H: **A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.***Bioinformatics* 2011, **27**:2987-2993.
63. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, Schatz MC: **Accurate detection of complex structural variations using single-molecule sequencing.***Nature Methods* 2018, **15**:461-468.
64. Käll L, Krogh A, Sonnhammer ELL: **A combined transmembrane topology and signal peptide prediction method.***Journal of Molecular Biology* 2004, **338**:1027-1036.
65. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M: **Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions.***Nucleic Acids Research* 2013, **41**:e121.
66. Mistry J, Finn R: **Pfam: A domain-centric method for analyzing proteins and proteomes.** In *Methods in Molecular Biology*, vol. 396. pp. 43-58; 2007:43-58.
67. Edgar RC: **MUSCLE: Multiple sequence alignment with high accuracy and high throughput.***Nucleic Acids Research* 2004, **32**:1792-1797.
68. Stamatakis A: **RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies.***Bioinformatics* 2014, **30**:1312-1313.
69. Smant G, Stokkermans J, Yan YT, de Boer JM, Baum TJ, Wang XH, Hussey RS, Gommers FJ, Henrissat B, Davis EL, et al: **Endogenous cellulases in animals: Isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes.***Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:4906-4911.
70. Mitreva-Dautova M, Roze E, Overmars H, De Graaff L, Schots A, Helder J, Goverse A, Bakker J, Smant G: **A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*.***Molecular Plant-Microbe Interactions* 2006, **19**:521-529.
71. Popeijus H, Overmars H, Jones J, Blok V, Goverse A, Helder J, Schots A, Bakker J, Smant G: **Enzymology: Degradation of plant cell walls by a nematode.***Nature* 2000, **406**:36-37.
72. Hewezi T, Howe P, Maier TR, Hussey RS, Mitchum MG, Davis EL, Baum TJ: **Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with *Arabidopsis* pectin methylesterase: Cooperative cell wall modification during parasitism.***Plant Cell* 2008, **20**:3080-3093.
73. Noon JB, Hewezi T, Maier TR, Simmons C, Wei JZ, Wu G, Llaca V, Deschamps S, Davis EL, Mitchum MG, et al: **Eighteen new candidate effectors of the phytonematode *Heterodera glycines* produced specifically in the secretory esophageal gland cells during parasitism.***Phytopathology* 2015, **105**:1362-1372.
74. Hu Y, You J, Li C, Pan F, Wang C: **The *Heterodera glycines* effector Hg16B09 is required for nematode parasitism and suppresses plant defense response.***Plant Science* 2019, **289**.
75. Eves-van den Akker S, Lilley CJ, Jones JT, Urwin PE: **Identification and Characterisation of a Hyper-Variable Apoplastic Effector Gene Family of the Potato Cyst Nematodes.***PLoS Pathogens* 2014, **10**:e1004391.

## Figures



**Figure 1**

Synteny between Gr-Line19, Gr-Line22, and JHI-Ro1 based on a progressive genome alignment in Mauve. Only syntenic regions larger than 3 kb are shown. Yellow lines represent regions that are syntenic between Gr-Line19 and Gr-Line22.

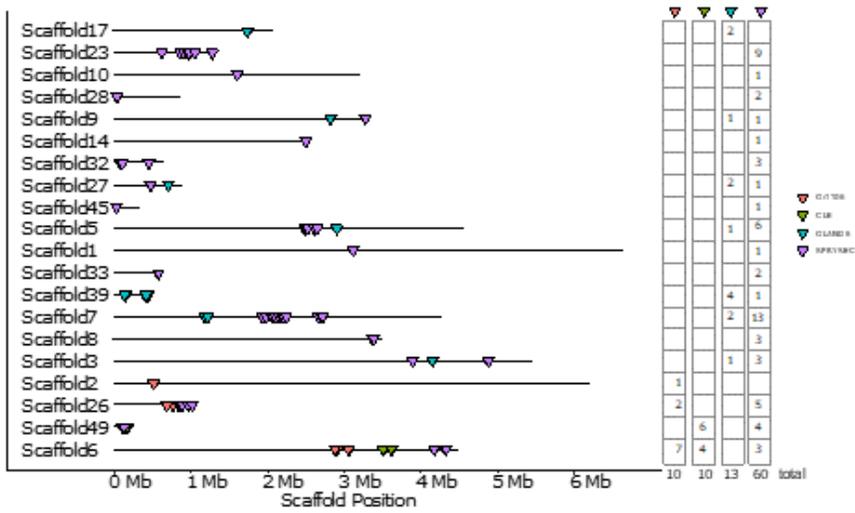


**Figure 2**

(A) Relatively level of SNP heterozygosity of Gr-Line19 and JHI-Ro1 (JHI – James Hutton Institute, Scotland, UK). Both lines were compared to Gr-Line22. JHI Ro1 was used as the current *G. rostochiensis* reference genome. (B) Comparison overall genomic constitution of *G. rostochiensis* inbred lines 19 and 22. In this overview structural variants are shown. Structural variants are DNA region of approximately 1 kb and larger in size, and can include inversions, insertions, deletions.



Gr-Line 19



Gr-Line 22

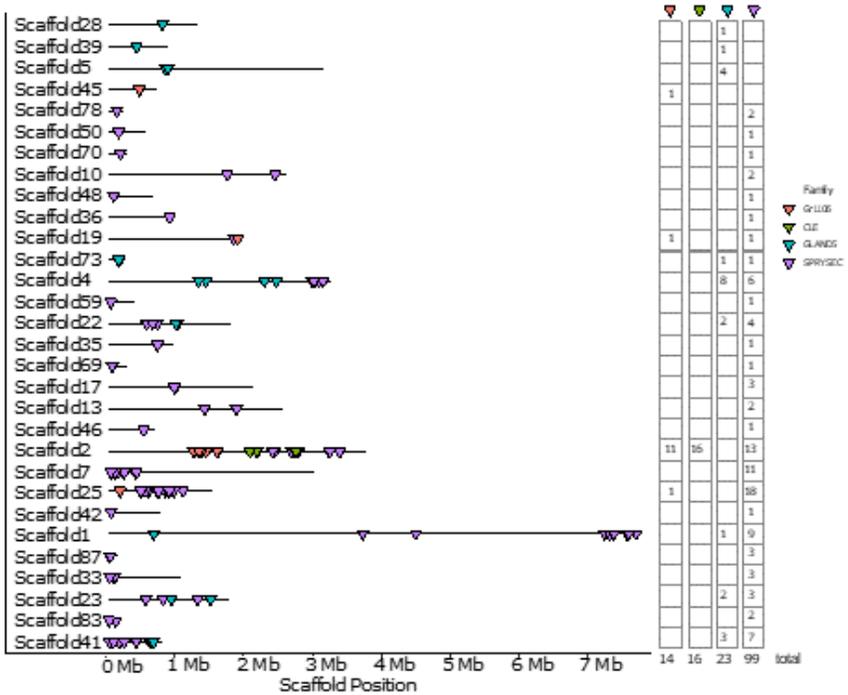
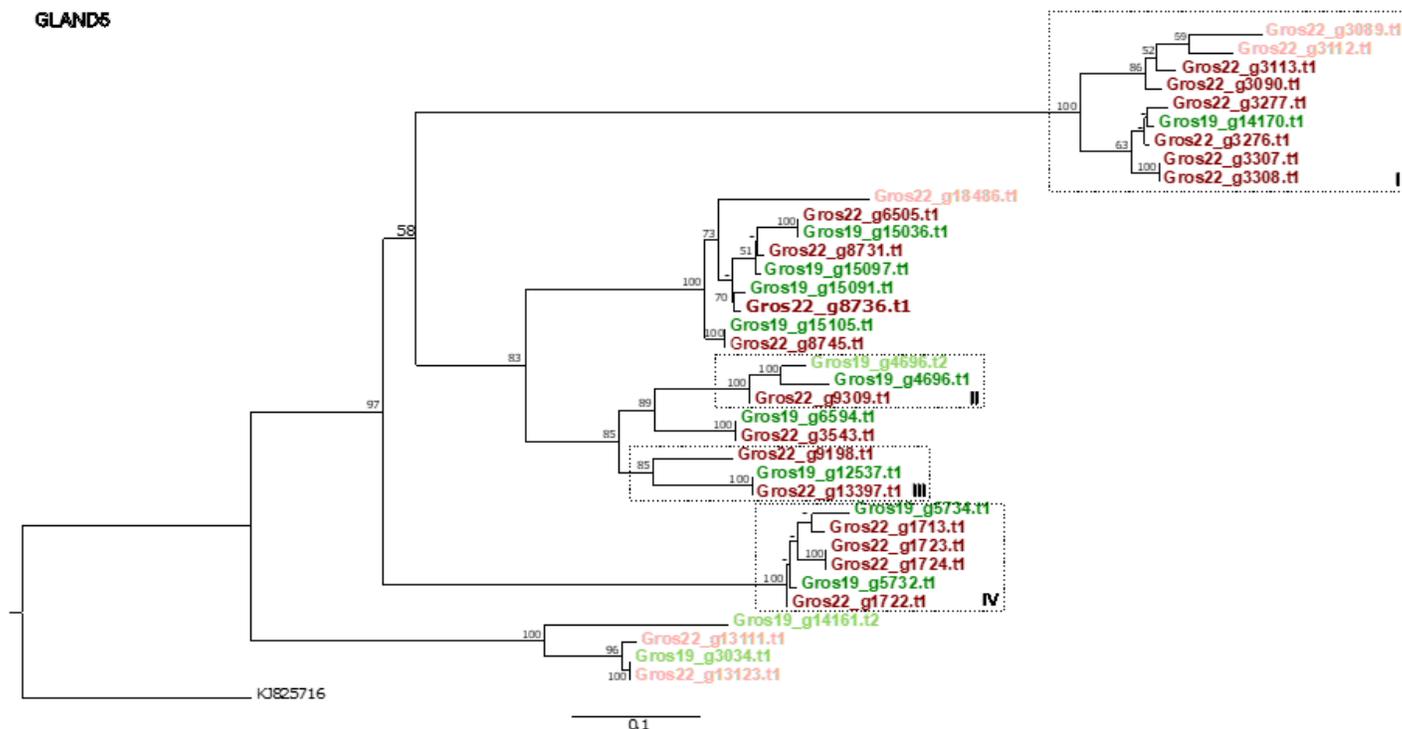


Figure 4

Spatial distribution of genes belonging to the effector families Gr1106 (Red), CLE (Green), GLAND5 (blue), and SPRYSEC (purple). Each triangle indicates the genomic position of a single gene. At the right, the number of variants per effector family are given for each scaffold.

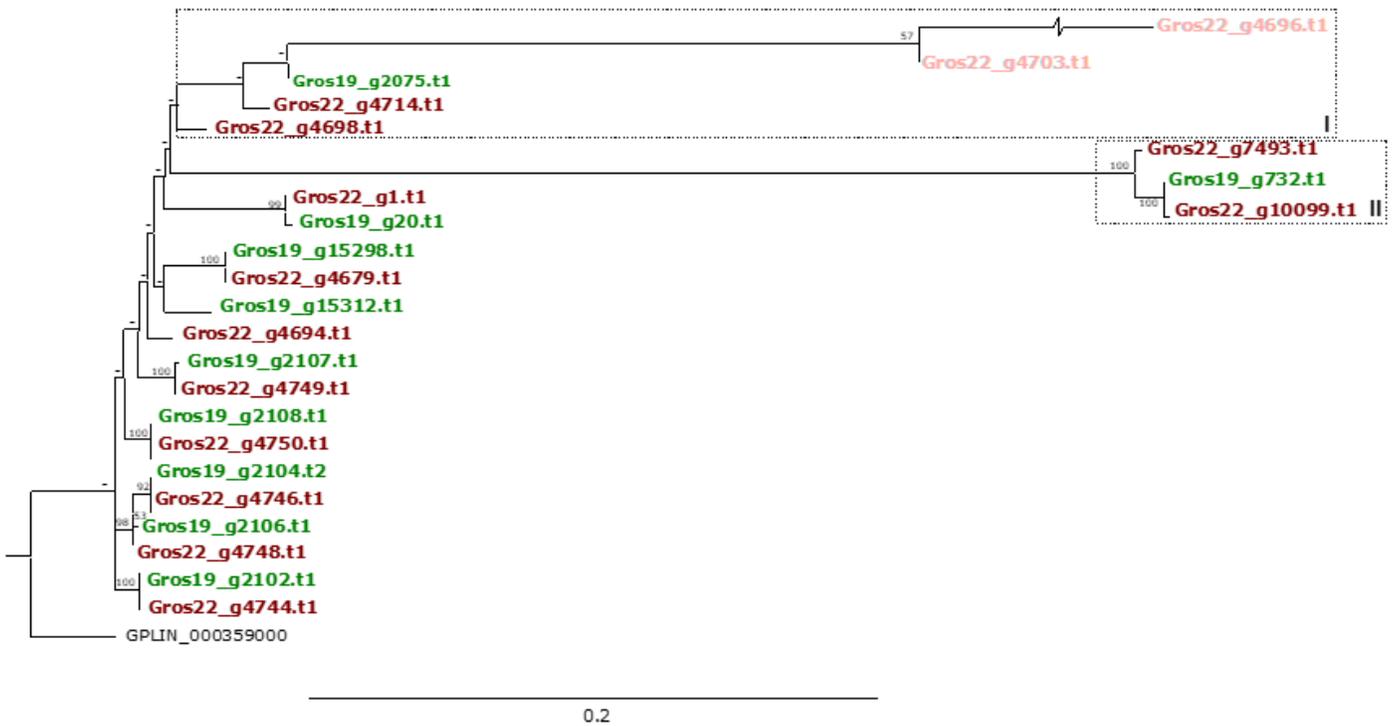
GLAND5



**Figure 5**

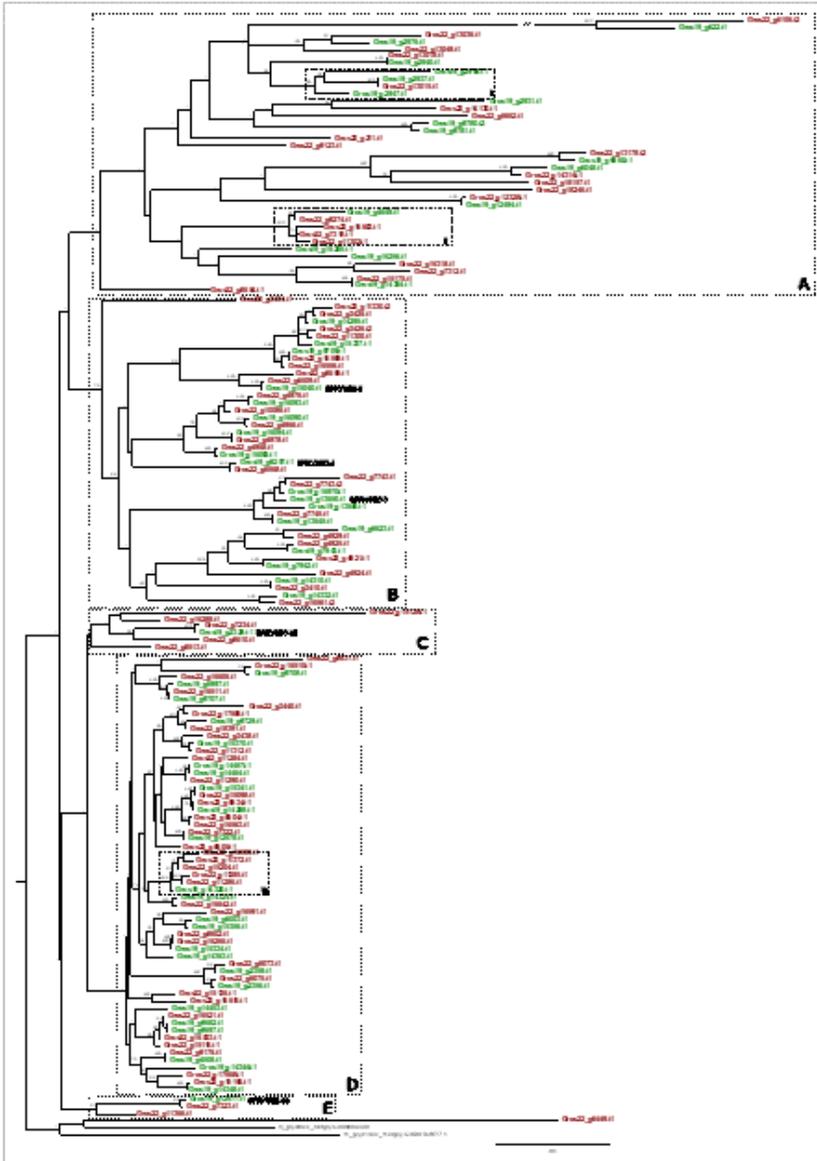
Phylogeny of GLAND5 effector genes of both Gr-Line19 (Green) and Gr-Line22 (dark red). A multiple sequence alignment was made using MUSCLE on the coding sequence. A phylogenetic tree was made using RAxML using a GTRGAMMA model, validated by 100 bootstrap replicates. Bootstrap values < 50% are indicated by a “-”. Lighter shades of green or red indicate effector variants that lack a signal peptide for secretion. Boxed clusters (Roman I – IV) highlight asymmetric expansions in one of the two *G. rostochiensis* lines.

## Effector family 1106



**Figure 6**

Phylogeny of Gr1106 effector genes of both Gr-Line19 (green) and Gr-Line22 (dark red). 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 "-". Lighter shades of green or red indicate effector variants that lack a signal peptide for secretion. Boxed clusters (Roman I - II) highlight asymmetric expansions in *G. rostochiensis* line Gr-Line22.



**Figure 7**

Phylogeny of SPRYSEC effector genes of both Gr-Line19 (green) and Gr-Line22 (dark red). Only SPRY proteins with a signal peptide for secretion are included. 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 "-". Closest homologs to the functionally described SPRYSEC-4, SPRYSEC-5, SPRYSEC-8, SPRYSEC-15, and SPRYSEC-19 [10] are shown. Clusters of SPRYSEC variants are boxed (A-E). Boxed clusters (Roman I – III) highlight asymmetric expansions in one of the two *G. rostochiensis* lines.

Effector family CLE

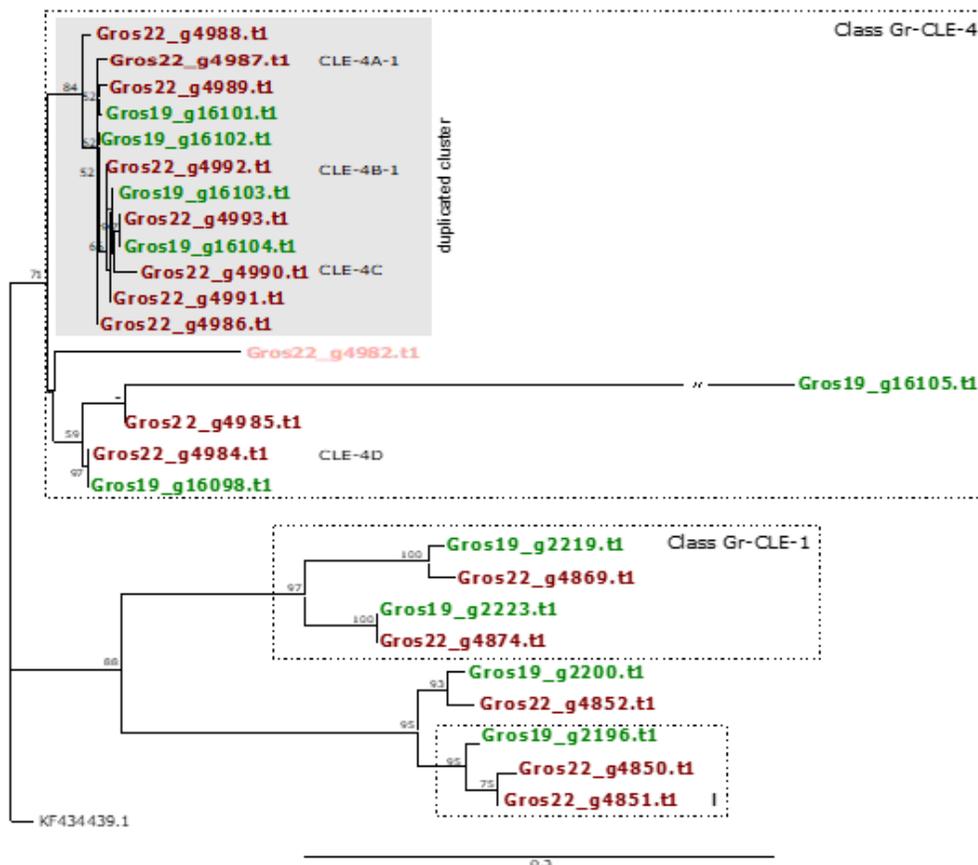
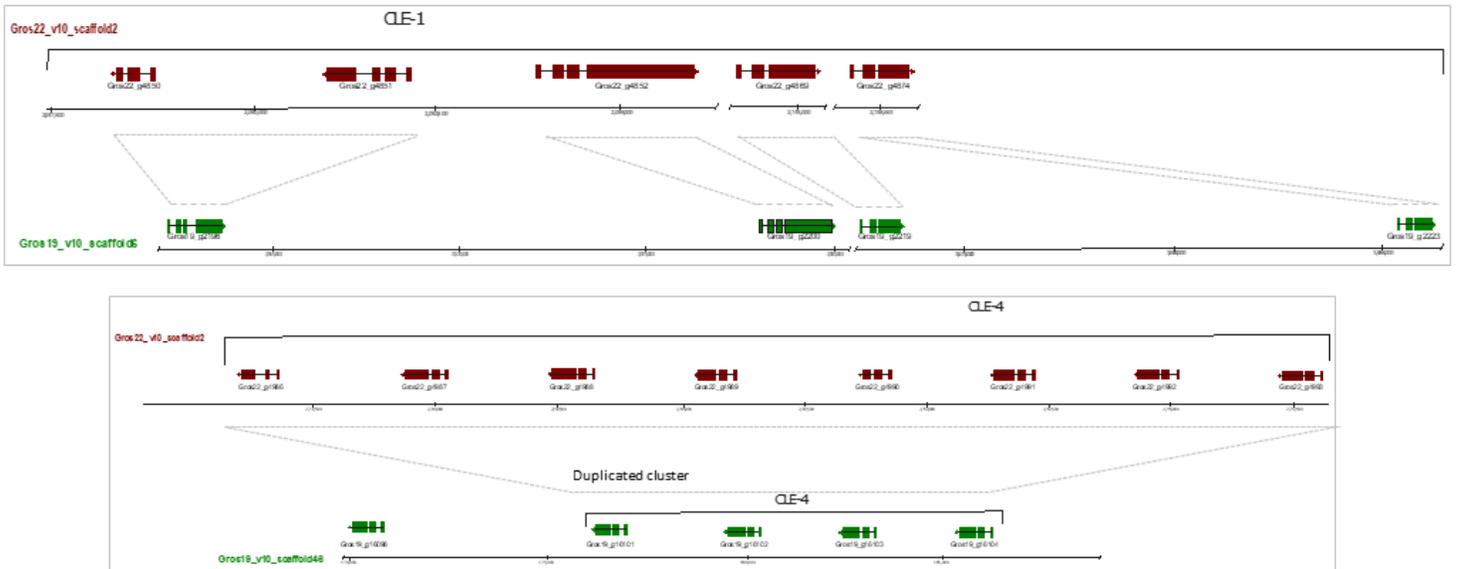


Figure 8

Phylogeny of CLE effector genes of both Gr-Line19 (Green) and Gr-Line22 (dark red). 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 "-". Lighter shades of green or red indicate effector variants that lack a signal peptide for secretion. Genes belonging to the functional classes GR-CLE-1 and Gr-CLE-4 are labeled with dashed boxes. A boxed cluster (Roman I) highlights the asymmetric expansions in one of the two *G. rostochiensis* lines.



**Figure 9**

Spatial organization of CLE functional classes CLE-1 and CLE-4 for both Gr-Line19 (Green) and Gr-Line22 (Dark red).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementalfigures.pptx](#)
- [SupplementalTable1.xlsx](#)