Identification and Characterisation of a Novel Glutathione Synthetase Gene Family in the Plant Parasitic Nematode *Rotylenchulus reniformis*

Duqing Wu

Submitted in accordance with the requirements for the degree of Doctor of Philosophy (Ph.D.)

The University of Leeds

Faculty of Biological Sciences

School of Biology

September, 2019

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter Three of the thesis has appeared in publication as follows:

Lilley, C.J., Maqbool, A., **Wu, D**., Yusup, H.B., Jones, L.M., Birch, P.R., Banfield, M.J., Urwin, P.E. and Eves-van den Akker, S., 2018. Effector gene birth in plant parasitic nematodes: Neofunctionalization of a housekeeping glutathione synthetase gene. *PLoS genetics*, 14(4), p.e1007310.

Acknowledgements

I would like to thank my supervisor Prof. Peter Urwin for not only giving me chance to carry out this PhD in the UK, but also for a lot of support and help over the last four years.

I would also like to thank Dr. Catherine Lilley for her patience, wisdom and scientific guidance. Her professional advices and wise words always lighted me up in my project.

Many thanks also go to my secondary supervisor Dr. Sebastian Eves-van den Akker. He offered me a lot of suggestions for my project, without which it would have become complete mess.

In addition, thanks go to Dr Chi Trinh for his help in solving the crystal structures.

Finally, I will always remember all the lab members (past and present) in the Plant Nematology Group at University of Leeds. Thanks to everyone for the love of Friday cakes, teas and birthday cards. The friendship will become a treasure of life.

Abstract

The reniform nematode, *Rotylenchulus reniformis* Linford & Oliveira, is a sedentary species of plant parasitic nematode that is widely distributed in tropical and subtropical regions and causes significant economic loss. There has been little molecular characterisation of *R. reniformis*, particularly in relation to the function of its effectors. Recent genomic and transcriptomic resources have become available that provide evidence of the complex suite of effector genes in *R. reniformis.*

Expanded families of putative effector genes have been described for other plant parasitic nematodes. In particular it was noted that the *Globodera pallida* genome encoded a large number (30) of complete glutathione synthetase-like genes in comparison to the free-living nematode *C. elegans* which has a solitary glutathione synthetase (*gs*) gene. In this study, we have identified a profusion of 73 complete glutathione synthetase-like genes from the *R. reniformis* genome and transcriptomes. The phylogeny of *R. reniformis* GS-like genes divides this family into three major clades: Clade 1 contains only one sequence that is the likely ancestor of the *R. reniformis* GS gene family; Clades 2 and 3 represent two independent expansions that acquire their unique functions during evolution. In addition, most Clade 3 GS do carry a signal peptide for secretion while Clade 1 & 2 GS do not. Furthermore, most Clade 3 *gs* are most highly expressed in the parasitic female stage whereas Clade 1 & 2 *gs* are up-regulated in the non-parasitic stages. *In situ* analysis showed Clade 3 *gs* are expressed in the gland cell of *R. reniformis* which is a common site of nematode effector synthesis. In contrast, Clade 1 & 2 *gs* are expressed in the intestine tissues.

Glutathione synthetase is a key enzyme in the second step of glutathione biosynthesis. Biochemical analysis of GS from *R. reniformis* confirmed the functional diversity between each clade. Clade 1 GS exhibited the canonical GS enzyme activity which was all-but lacking in Clade 2 & 3 GSs. Crystallography was then exploited to investigate the structural differences between canonical and non-canonical GSs, indicating that an alternative substrate may be accepted by non-canonical GS.

This project also set out to investigate the functions of *R. reniformis* GS. None of the *R. reniformis* GS, including canonical GS could complement the Arabidopsis GS mutant *gsh2*. In addition, Arabidopsis overexpressing Clade 3 GS showed enhanced susceptibility to the cyst nematode *Heterodera schachtii*. In conclusion, this study

revealed evolved functional diversity of this expanded large GS family by phylogenetic, biochemical, structural and functional evidence.

Contents

Table of Figures

List of Tables

List of Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CBP	Cellulose-binding protein
CEP	C-TERMINALLY ENCODE PEPTIDE
CLE	CLAVATA3/ESR peptides
CN	Cyst nematode
CRT	Calreticulin
DHA	Dehydroascorbate
DIG	Digoxigenin
DMSO	Dimethyl sulfoxide
DOG	Dorsal gland promote
DTT	dithiothreitol
EST	Expressed sequence tags
GCL	Glutamate-cysteine ligase
GS	Glutathione synthetase
GSH	Reduced glutathione
GSSG	Oxidised glutathione
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMM	Hidden Markov Models
HPLC	High-performance liquid chromatography
IDA	INFLORESCENCE DEFICIENT IN ABSCISSION
IPTG	Isopropyl β-d-1-thiogalactopyranoside
JA	Jasmonate
MS	Murashige & Skoog
NADPH	Nicotinamide adenine dinucleotide phosphate
NGS	Next generation sequencing
NOS	Nopaline synthetase
PCR	Polymerase chain reaction
PDB	Protein data bank
PPN	Plant parasitic nematode
qRT-PCR	Quantitative Reverse-Transcriptase PCR
RKN	Root-knot nematode
RNAi	RNA interference
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE	Standard error
SUMO	Small Ubiquitin-like Modifier
TEMED	Tetramethylethylenediamine

UV-TPEF	Ultraviolet Two-Photon Excited Fluorescence
VAP	Venom allergen-like protein
γ-EC	γ-glutamylcysteine

Chapter 1

General Introduction

1 Introduction

1.1 Plant parasitic nematodes

Nematodes are found ubiquitously in nature - most of them are free-living (Basyoni and Rizk 2016). Nevertheless there are more than 4100 species of nematodes described as plant parasitic nematodes (PPNs) (Decraemer and Hunt 2006). PPNs are extremely important plant pathogens in crop production. Yield loss due to nematode infection is difficult to calculate as a lack of clear aboveground symptoms often contribute to an under-estimated and undetected threat until the crop losses become severe. However, it is estimated that PPNs cause over £157 billion in economic loss per year worldwide, posing a high threat to global food security (Nicol *et al.* 2011).

PPNs feed on the cytoplasmic contents of plant cells by means of their hollow stylets/ stylet analogues and display a wide variety of parasitic strategies (Seinhorst 1961). There are two distinctive categorical classifications, (a) ectoparasites that never penetrate into the host tissue but simply live in soil and use roots as a food resource when the ectoparasites encounter them and endoparasites that entirely entry into host tissue while feeding; (b) migratory species that leave the original parasitised tissue after a short feeding time to move to a new site and sedentary species that remain sedentary at a chosen feeding site (Seinhorst 1961). The most economically important and intensively studied are sedentary endoparasitic nematodes, particularly the root-knot nematodes and cyst nematodes (Jones *et al.* 2013).

1.1.1 Root-knot nematode

Root-knot nematode have more than 100 species and the most important species include *M. incognita*, *M. hapla*, *M. javanica*, *M graminicola* and *M. arenaria* (Escobar *et al.* 2015). Root-knot nematodes can parasitise almost every species of vascular plant including many important crops and vegetables (Taylor and Sasser 1978). Information about the overall economic loss caused by root-knot nematode is rare. However, there is growing evidences that suggests the problem of *Meloidogyne* spp. in most farms across the continent is a significant threat to crop production (Moens, Perry and Starr 2009; Onkendi *et al.* 2014). Recently, up to 70% of the total estimated economic losses inflicted by nematodes derive from rice alone (Kyndt, Fernandez and Gheysen 2014). *M*. *graminicola* is one of the most prevalent PPNs in rice agrosystems

and it is well-adapted in both upland (rainfed) and lowland (irrigated) conditions. Its short life cycle and wide host range make this species difficult to control (Mantelin, Bellafiore and Kyndt 2017).

All root-knot nematodes pass through from an embryonic stage, four juvenile stages (J1–J4) to an adult stage (Eisenback and Triantaphyllou 1991). Mature females lay eggs in a protective gelatinous matrix which forms an egg mass on the root surface. After the embryonic stage, the J1 which molts once in the egg hatches as infective J2 from the egg. These migratory nematodes penetrate directly into the host root and move intercellularly within the plant tissue to a preferred feeding site in the vascular cylinder (Chitwood and Perry 2009). The J2 then become sedentary and as it feeds on a specialised feeding site which consists of several giant cells, it becomes saccate and undergoes three moults to J3, J4 and reproductive adult stage, respectively. Root tissue around the nematode and the giant cells experience hyperplasia and hypertrophy resulting in the characteristic root gall. Galls usually develop 1-2 days after J2 penetration (Dropkin 1972). The J3 and J4 cannot feed due to a lack of stylet. The vermiform males then leave the roots and move freely in the soil without further feeding while the females continue to feed and enlarge to become saccate. Depending on the different reproductive strategies of particular species, amphimixis or parthenogenesis, the male may search for a female to mate, or remain in the soil and finally die (Eisenback and Triantaphyllou 1991).

Sedentary PPNs have evolved complex strategies to maintain their prolonged parasitism that can continue for many weeks. One of the most essential aspects is the manipulation of host cell genetic developments that results in formation of a novel cell type in host roots (Kyndt *et al.* 2013). Giant cells induced by root-knot nematodes are important for successful parasitism and are treated as food source for nematodes. Giant cells initiate from procambial cells and are expansions of single cells. Once nematodes reach a suitable site they manipulate the normal root cells to undergo drastic morphological and ultra-structural changes (Bird 1961). They often become more than 100 times larger than normal plant root cells. Their nuclei become hypertrophied in the absence of cytokinesis and the cytoplasm condenses with increasing numbers of mitochondria, plastids, ribosomes and structures of endoplasmic reticulum (Bird 1961; Jones 1981).

1.1.2 Cyst nematode

The cyst nematodes (mainly *Heterodera* & *Globodera* spp.) are obligate sedentary parasites of many important crops including soybean (*Heterodera glycines*), potato (*Globodera rostochiensis* & *G. pallida*), wheat (*Heterodera avenae* & *H. filipjevi*), and sugar beet (*H. schachtii*) (Jones *et al.* 2013). Compared with root-knot nematode, each species of cyst nematode has a much narrower host range (Stone 1986). Soybean cyst nematode has been reported to be responsible for over 1.5 billion dollar economic loss each year in USA alone (Bernard, Egnin and Bonsi 2017). A total of average 9% crop loss in potato production area worldwide (Turner and Rowe 2006) and estimated £50 million yield loss in the UK (Jones *et al.* 2017) is caused by potato cyst nematode.

Like root-knot nematodes, cyst nematodes also have four juvenile stages in addition to adult stage. A generalised life-cycle of cyst nematodes sees the J2 stage hatching from an egg stimulated by root exudates. The J2 locates a root and then invades the root primarily behind the root tip through mechanical use of the stylet and secretion of proteins. Cyst nematodes migrate through cells, which causes extensive necrosis of host cells (Turner and Rowe 2006). After intracellular migration to the inner cortex, J2 selects a suitable cell to form a unique feeding site termed syncytium as a source of nutrition and become sedentary near the vascular tissue (Sobczak and Golinowski 2011). After feeding, the nematode remains at this feeding site for several weeks, going through a further three moults to the adult stage (Jones *et al.* 2013). Sex is determined by environmental conditions, with the frequency of males increased in conditions of crowding or poor nutrition (Triantaphyllou 1973). Female cyst nematodes grow until their saccate bodies are visible at the root surface, whereas males revert to the vermiform body shape, leave the roots and follow sex pheromone gradients to find females. After fertilisation, the female cyst nematode dies and her body wall tans to form a cyst, which encloses the next generation of eggs (Sobczak and Golinowski 2011). Cyst nematodes remain dormant within the cyst, enabling them to persist for up to 20 years without a host.

Syncytia induced by cyst nematodes usually originate from a selected cortical, endodermal, or pericycle cell (Jones 1981). This cell undergo increased active metabolism, proliferation of mitochondria and plastids and dismantling of the central vacuole into several dispersed small secondary vacuoles (Golinowski, Grundler and Sobczak 1996). An increase in cytoplasmic organelles is then accompanied by cell

wall dissolution at the pit fields and fusion of neighbouring cell protoplasts, resulting in a large feeding cell with multiple enlarged nuclei, dense cytoplasm and thickened outer cell walls (Bohlmann and Sobczak 2014). Metabolic profiling analyses of syncytia described higher levels of starch and some specific sugar accumulation, and increased levels of many amino acids and phosphorylated metabolites in syncytia induced by *H. schachtii* (Hofmann *et al.* 2010).

1.1.3 Reniform nematode

The reniform nematode, *Rotylenchulus reniformis*, a sedentary semi-endoparasite with a large host range involving more than 350 plant species, is considered to be an important economic pathogen (Gaur and Perry 1991; Robinson *et al.* 1997). Evolutionary studies indicated that *R. reniformis* is most closely related to the cyst nematode genera and also shares common ancestry with migratory endoparasitic *Radopholus* spp (Holterman *et al.* 2009). *R. reniformis* is reported to suppress cotton yields to 40% of the yield potential. In some heavy infection areas, if no effective control measures were adopted, yield loss can increase to an estimated 100% (Westphal *et al.* 2004). A previous report indicated that *R. reniformis* has replaced root-knot nematode as the major pathogenic nematode of cotton in the mid-south region of United State (Robinson 2007). The estimated losses are still increasing due to several factors: (i) the lack of resistant cultivars, (ii) limited use of crop rotation in many areas, (iii) the lack of awareness of pathogenic nematodes as production constraints, especially the reniform nematode, (iv) the loss of highly effective, low-cost, fumigant nematicides (Starr *et al.* 2007). In 2014, an estimated loss of 74 million USD was caused by *R. reniformis* infection in the US cotton producing area (Lawrence *et al.* 2015).

Similar with other sedentary plant parasitic nematodes like root-knot nematode and cyst nematode, the above-ground symptoms of *R. reniformis* infection do not display any unique features merely common symptoms of nutrient deficiencies such as leaf loss, plant stunting and reduced crop production (Koenning *et al.* 2004). Although *R. reniformis* attacks host roots, unlike root-knot nematode it does not cause obvious phenotype changes to roots. It cannot be easily observed on the surface of roots like cyst nematodes. Root growth is usually reduced with limited secondary root development and root rot and necrosis can be seen in some plants (Jones *et al.* 2013). Disease complexes with other plant pathogens, such as *Fusarium oxysporum* (Neal

1954), Verticillium spp (Tchatchoua and Sikora 1983) and *Rhizoctonia solani* (Vadhera, Shukla and Bhatt 1995) lead to reduced shoot growth, wilt and further damage.

The *R. reniformis* life cycle begins when J2 hatch from eggs in the soil. However, unlike cyst and root-knot nematodes, *R. reniformis* J2 do not infect the host root but instead become immobile, assuming a crescent-like shape. Subsequently two moults through the J3 and J4 juvenile stages occur in the absence of feeding. This stage of the life cycle ends with the emergence of mobile non-infective vermiform males which remain in the soil and infective vermiform females (Ganji, Wubben and Jenkins 2013). Adult stage usually occurs 16 days after inoculation in susceptible cultivars (Ayala and Ramírez 1964). The females penetrate the host roots, inserting about one-third of the anterior body, and become sedentary, establishing feeding sites termed syncytia from endodermal and pericycle cells as their food source (Wyss 1997). The syncytium extends around the root as a single, curved cell layer (Jones and Dropkin 1975). As for the syncytia induced by cyst nematodes, *R. reniformis* syncytia also show significant cell wall dissolution, increased cytoplasmic density and nuclei with enlarged nucleoli (Rebois 1980). After feeding for around 10 days, the body of the female outside the root swells and assumes a kidney (i.e. reniform) shape. Within the subsequent 7-9 days under suitable conditions, the vulval glands produce a gelatinous matrix into which 40-200 eggs are laid (Sivakumar and Seshadri 1971). Reniform nematode gelatinous matrix is always completely outside the root, only the anterior end of the body is embedded in the root (Agudelo *et al.* 2004). The males do not feed and remain in the soil. The life cycle of reniform nematode is usually shorter than four weeks, but this depends on soil temperature (Jones *et al.* 2013). However, it can survive at least two years in the absence of a host in dry soil through anhydrobiosis, a survival mechanism that allows the J3 and J4 nematodes to enter an ametabolic state and live without water for extended periods of time (Radewald and Takeshita 1964). The whole life cycle of *R. reniformis* from egg to egg is from 22-29 days in susceptible cultivars (Ayala and Ramírez 1964).

The life cycle of *R. reniformis* is summarised in Figure 1.1.

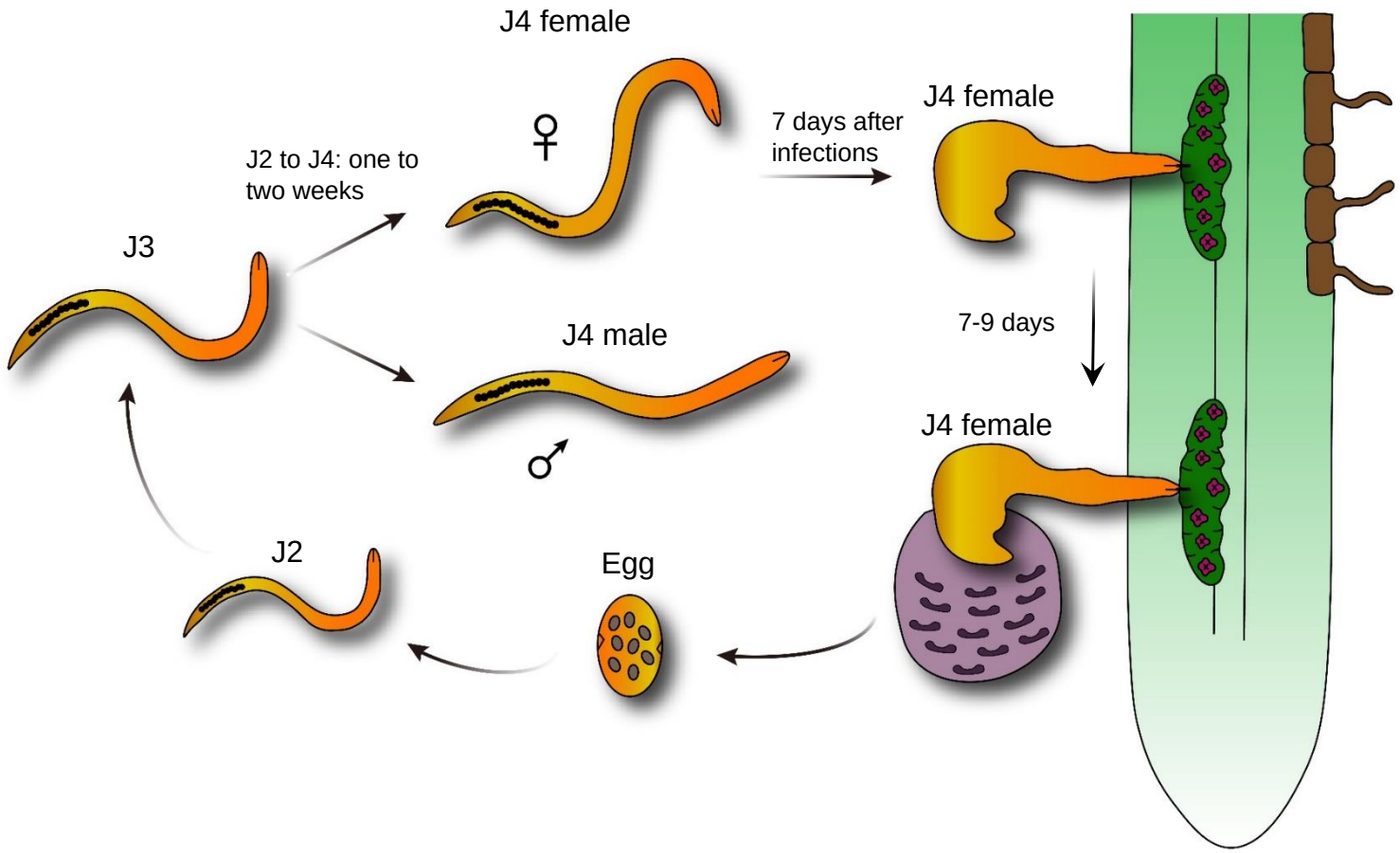


Figure 1.1 Life cycle of *R. reniformis*. The *R. reniformis* life cycle begins when J2 hatch from eggs in the soil, followed by subsequently two moults (J3 & J4) in absence of feeding. J4 stage ends with the emergence of mobile non-infective vermiform males which remain in the soil and infective vermiform females. The females penetrate the host roots, inserting about one-third of the anterior body, and become sedentary, establishing feeding sites termed syncytia from endodermal and pericycle cells as their food source. The syncytium extends around the root as a single, curved cell layer. After feeding for around 10 days, the body of the female outside the root swells and assumes a kidney (i.e. reniform) shape. Within the subsequent 7-9 days under suitable conditions, the vulval glands of the female produce a gelatinous matrix into which 40-200 eggs are laid. The whole life cycle of *R. reniformis* from egg to egg is from 22-29 days in susceptible cultivars.

1.1.4 Nematode gland cells

In order to initiate and maintain their interactions with the hosts, sedentary nematodes exploit their protrusible mouth structure termed a “stylet” to release secreted proteins termed effectors from their pharyngeal gland cells into the host roots through the stylet. (Davis *et al.* 2008; Hussey 1989). These gland cells are considered as the main source of effectors involved in plant parasitism (Hussey, Davis and Baum 2002). In highly evolved sedentary nematodes such as cyst nematodes and root-knot nematodes, there are three typical large secretory gland cells: one dorsal and two sub-ventral cells, which are the principal source of secretions that contain nematode effectors (Hussey 1989). Each gland cell contains a large nucleus with abundant Golgi complexes, rough endoplasmic reticulum and other organelles typical of secretory cells (Davis *et al.* 2000). The morphological changes between the dorsal and sub-ventral gland cells at different life stages suggest distinctive roles of these glands in the life cycle. During parasitism of sedentary nematodes, it is hypothesised that the sub-ventral glands function primarily but not exclusively in the root penetration and migration phases, while the dorsal gland plays a primary role in the subsequent formation and maintenance of the feeding cells (Mitchum *et al.* 2013). Very limited information on the gland cells of reniform nematode has been reported so far. Sedentary female reniform nematodes appear to have only one single dorsal gland which is more than one-half the stylet length and is posterior to the base of the stylet knobs (Dasgupta, Raski and Sher 2011).

1.2 Nematode effectors

Plant pathogens, such as fungi and bacteria, secrete a cocktail of proteins termed effectors into different cellular compartments of their hosts to modulate plant defence circuitry and enable their colonisation of plant tissue (Toruno, Stergiopoulos and Coaker 2016). By definition, effectors are parasite-produced proteins or small molecules that promote parasitism by suppression of host immunity and defences or by manipulation of the host cell biology (Hogenhout *et al.* 2009).

Like other plant pathogens, plant parasitic nematodes also secrete a wide range of effectors with multiple functions from promotion of movement in plant tissues to modification of host cells with the ultimate aims to exploit the host for nutrients. As such, effectors are considered to play a significant role in successful plant nematode parasitism (Mitchum *et al.* 2013). The nematode effector-containing secretions are

produced in several different organs, in addition to the previously mentioned pharyngeal gland cells, including the cuticle, amphids, the excretory/secretory system and the rectal glands (Rehman, Gupta and Goyal 2016).

1.2.1 Nematode effector identification

Recently, “next generation” sequencing technologies have emerged, making it affordable to sequence the transcriptome and whole genome of plant parasitic nematodes. The application of next-generation sequencing to PPNs has allowed a wide range of genome- or transcriptome-level comparisons, and undoubtedly, the identification of plant parasitic nematode effectors has benefited from advancements in these high-throughput assays and bioinformatic analysis (Ali *et al.* 2015).

Transcript data are available from a wide variety of nematodes. The first transcript database was for *M. incognita*, which was established in 2003 to analyse over 5700 expressed sequence tags (EST) from second-stage larvae (McCarter *et al.* 2003). Due to a rapid development of sequencing technologies, there are, in addition to EST databases, now a great many whole transcriptome datasets established for numerous economically important plant parasitic nematodes, such as *Ditylenchus africanus* (Haegeman *et al.* 2009), *D. destructor* (Peng *et al.* 2013), *M. graminicola* (Haegeman *et al.* 2013), *H. avenae* (Kumar *et al.* 2014), *Nacobbus aberrans* (Eves-van den Akker *et al.* 2014) and *H. schachtii* (Fosu-Nyarko *et al.* 2016).

By analysing these transcriptome datasets, not only were considerable numbers of PPN parasitism-related genes or effector genes revealed, but a lot of basic information concerning genes related to aspects of nematode biology such as behaviour was acquired as well. Furthermore, some studies focused on comparative transcript analysis from different life stages (e.g. parasitic stage and non-parasitic stage), providing greater insights into PPN effector expression and aiding identification. For example, a comparative EST study of different life stages (7, 14 and 30 days after infection) from *G. pallida* was carried out to identify over 50 secreted proteins up-regulated after the onset of parasitism and expression in pharyngeal gland cells was confirmed using *in situ* hybridization (Jones *et al.* 2009).

In addition to transcriptomes, continually improvements in genome sequencing and assembly have led to the recent production of draft genome assemblies for PPNs. The first whole nematode genome to be sequenced was that of *Caenorhabditis elegans*, a

free-living nematode which is now one of the most important model animals in biological research, especially in developmental biology and genetics (Consortium 1998). A breakthrough for plant parasitic nematode genome sequencing was that of the root-knot nematode *M. incognita* ten years after the *C. elegans* genome was published, reporting an 86 Mb genome size (Abad *et al.* 2008). Until now, several PPN genomes have been published successively, including *M. hapla* (Opperman *et al.* 2008), *Bursaphelenchus xylophilus* (Kikuchi *et al.* 2011), *Globodera pallida* (Cotton *et al.* 2014), *Pratylenchus coffeae* (Burke *et al.* 2015), *G. rostochiensis* (Eves-van den Akker *et al.* 2016a), *Ditylenchus destructor* (Zheng *et al.* 2016) and *H. glycines* (Masonbrink *et al.* 2019). The genome data from these plant parasitic nematodes has led to a large panel of putative nematode effectors being identified. By a BLAST search against other datasets, the genes corresponding to already known effectors can be predicted. At the same time, nematode effectors usually contain a signal peptide for secretion and have no transmembrane domain. Therefore, by further filtering the potential effector set for the presence of a signal peptide for secretion and absence of a transmembrane domain, pioneer effector genes are able to be identified. For example, by analysing the complete genome of *G. pallida* in association with transcriptomic data from most stages of the nematode life cycle, an enormous expansion of the SPRY domain protein family was described as a set of potential novel effectors (Cotton *et al.* 2014). A large number of orthologues of effectors from other nematodes as well as novel effector candidates were also identified (Thorpe *et al.* 2014).

What's more, as mentioned above, nematode gland cell/cells are believed to be the major sites of effector production. Therefore, direct examination and detection of the content of nematode gland cells coupled with transcriptomic analysis can open up the possibility of uncovering the plant parasitic nematodes effectors repertoires and the variability among different nematode genera, species and pathotypes. For example, a large number of potential parasitism genes that were expressed in gland cells during parasitic stages were identified from soybean cyst nematode *H. glycines* by creation of gland cell-specific cDNA libraries of various parasitic stages using cytoplasm microaspiration (Gao *et al.* 2003). Furthermore, secreted proteins from the root knot nematode *M. incognita* were directly examined using mass spectrometry, resulting in the identification of 486 possible effectors (Bellafiore *et al.* 2008). Recently, a more

advanced and direct technique was used to identify potential nematode effectors. Whole individual oesophageal gland cells from three plant-parasitic nematode species were separated and isolated to elucidate the transcriptomes of oesophageal glands (Maier *et al.* 2013).

Conserved protein motifs are often used for prediction of effector repertoires in many plant pathogens such as the RxLR motif (consensus sequence: Arg-Xaa-Leu-Arg) identified from oomycetes (Whisson *et al.* 2007) and the signal sequence characteristic of Type III secretion system from bacterial plant pathogens (Alfano and Collmer 2004). But for PPN, until recently, there were no reliable elements/motifs that could help to computationally predict effectors. Recently, by analysing the *G. rostochiensis* genome assembly, a dorsal gland promoter element motif (termed DOG Box with a consensus sequence ATGCCA) was identified in the promoter region of 77% of *G. rostochiensis* dorsal gland effectors and representatives from 26 out of 28 dorsal gland effector families. Dorsal gland effectors contained an average of 2.54 DOG boxes in their promoter regions, compared to 0.32 for all non-effectors (Eves-van den Akker *et al.* 2016a). In addition, a putative regulatory promoter motif 'STATAWAARS' associated with an expression profile in the pharyngeal gland cells from *B. xylophilus* was identified. This motif has the consensus sequence STATWWAWRS, and has six variable loci indicated by the DNA ambiguity code ([C|G]TAT[T|A][T|A]A[T|A][G|A][C|G]). 43% of STATAWAARS motif containing genes were found to encode a protein with a predicted signal peptide (n = 206), compared with 12.7% of all known genes in the *B. xylophilus* genome (Espada *et al.* 2018). Taken together, although not all effectors share such a motif and some non-effectors were found to have this motif in the promoter region, the presence of DOG or STATAWAARS promoter motif showed a large enrichment of effectors and can be used as a useful additional criterion to facilitate effector prediction.

1.2.2 Current status of *R. reniformis* effectors

Compared with studies of root-knot and cyst nematodes, there has been very little molecular characterisation of *R. reniformis*. Very few details about the molecular basis of interactions between *R. reniformis* and its host have been reported so far. A survey of *R. reniformis* ESTs that were sequenced from the sedentary parasitic female cDNA library indicated a number of putative effectors which shared high sequence similarity with those from other plant- or animal- nematodes (Wubben, Callahan and Scheffler

2010; Nyaku *et al.* 2013). Dorsal oesophageal gland-specific expression of the *R. reniformis* CLE homolog has been demonstrated by *in situ* hybridization (Wubben *et al.* 2015). Another diverse family of effectors called the C-TERMINALLY ENCODED PEPTIDE (CEP) plant peptide mimics was identified in *R. reniformis* (Eves-Van Den Akker *et al.* 2016b). The nematode encoded CEPs were first identified in root-knot nematode but not found in cyst nematode, although neither the expressional location nor biological activity of the peptides has been revealed (Bobay *et al.* 2013). The *R. reniformis* CEPs were hypothesised to increase host nitrate uptake and also regulate the size of the syncytial feeding site. In addition, several other *R. reniformis* effectors such as β-1,4-endoglucanses (Wubben, Ganji and Callahan 2010) and C-type lectins (Ganji, Jenkins and Wubben 2014) have also been identified, although limited details on their functions have been described.

1.2.3 Functional characterisation of nematode effectors

In the nematodes' migratory stage, many of the effectors secreted facilitate penetration and migration by degrading components of the plant cell wall, as well as enabling the nematode to suppress the plant's immune system (Smant and Jones 2011). In the nematodes' sedentary stage, formation of feeding cells is usually accompanied by alterations of plant hormone status and dramatic changes in gene expression associated with various aspects of plant growth and development (Mitchum *et al.* 2012). Obviously, effectors secreted during this stage play key roles in modifications of the host cell biology, inducing the formation of a metabolically highly active feeding cell as a nutrient source to sustain nematode growth and development, at the same time as regulating host defences. So far, most of the molecular work related to functional characterisation of PPN effectors has focused on cyst nematodes and root-knot nematodes (Vieira and Gleason 2019).

1.2.2.1 Cell wall architecture regulated by nematode effectors

The plant cell wall, which is primarily composed of a variety of polysaccharides, is the major obstacle for infecting plant parasitic nematodes during their migration within host roots. Cell wall modifying and degrading enzymes such as cellulases and pectate lyases that can depolymerize various structural polysaccharides of plant cell walls were the first nematode-secreted proteins to be localized in *planta* during infection (Jaouannet and Rosso 2013). The repertoire of cell wall modifying and degrading enzymes in different nematode genera vary dramatically, perhaps a reflection of the

diversity of plant cell wall components within different host plant species. Table 1.1 shows a comparison of the various putative cell wall degrading enzymes predicted to be encoded by different plant parasitic nematode published genomes.

In addition to cell wall modifying and degrading enzymes, plant parasitic nematodes also exploit other sets of sophisticated effector proteins released into feeding cells, which function in cell wall modification, of which cellulose-binding proteins (CBPs) are good examples. CBPs can bind to cellulose in *in vitro* assays (Gao *et al.* 2004), and were shown to have a direct strong interaction with *Arabidopsis* pectin methylesterase protein 3 by yeast two-hybrid assay, targeting and potentially activating this enzyme to aid nematode parasitism (Hewezi *et al.* 2008). Hence, we conclude that PPN could regulate cell wall modifications much more than we expect originally via an effector cocktail.

Table 1.1: Comparison of predicted cell wall degrading enzymes from the sequenced genomes of different nematode species.

Nematode Species	Genome assembly length (Mb)	Cellulases	Xylanases	Arabinanases	Pectinases	Expansins	Total	Reference
M. incognita	86	21	6	2	32	20	81	(Abad *et al.* 2008)
M. hapla	53	6	1	2	24	6	39	(Opperman *et al.* 2008)
B. xylophilus	75	11	0	0	15	8	34	(Kikuchi *et al.* 2011)
G. pallida	125	16	0	1	0	9	26	(Cotton *et al.* 2014)
G. rostochiensis	96	11	0	1	3	7	22	(Eves-van den Akker *et al.* 2016a)
P. coffeae	19.7	1	2	2	1	3	9	(Burke *et al.* 2015)
D. destructor	113	3	0	1	1	0	5	(Zheng *et al.* 2016)
H. glycines	124	15	0	1	16	12	44	(Masonbrink *et al.* 2019)
C. elegans	100	0	0	0	0	0	0	(Consortium 1998)

1.2.2.2 Effectors mimicking plant proteins

Plant parasitic nematodes are able to secret small peptide effectors that mimic plant proteins or small ligands to promote parasitism (Hu and Hewezi 2018). One of the most striking examples in the nematode kingdom for mimicry of plant factors is the case of novel small gene families with a conserved C-terminal domain similar to that of the endogenous plant EMBRYO SURROUNDING REGION (ESR) (CLE)-related peptides. CLV3 is a member of the CLE family, the members of which can be identified by sequence similarity to CLV3 and the maize ESR gene products, which share a conserved 14 amino acids motif termed CLE box (Somssich et al. 2016). Plant CLE-related peptides were considered as intercellular signalling molecules that played a role in controlling the balance between meristem cell proliferation and differentiation (Fletcher *et al.* 1999; Sawa *et al.* 2006). Interestingly, CLE-like effectors have been identified in a range of sedentary endo-parasitic nematodes, such as *H. glycines* (Wang *et al.* 2005). The overexpression of a CLE-like effector from *H. glycines* and *H. schachtii* in *Arabidopsis* resulted in a wuschel-like phenotype that is very similar to reports of overexpression of plant CLEs (Wang *et al.* 2005; Wang *et al.* 2011). The nematode gene could also functionally complement the phenotype of the Arabidopsis mutant *clv3-1* (Lu *et al.* 2009). It has been reported that cyst nematode CLE proteins are delivered to the cytoplasm of syncytial cells, but ultimately function in the apoplast, which as expected, was similar to plant CLEs (Wang *et al.* 2010).

Secreted CLE peptides stimulate intracellular signalling through plasma membrane-localised receptors. Once secreted from plant stem cell, the CLV3 peptide is perceived by receptor-like kinases such as CLV1, CLV2/CORYNE (CRN) complex and receptor-like protein kinase 2 (RPK2) (Somssich *et al.* 2016). Recently, it has been shown that the CRN heterodimer receptor complex (Replogle *et al.* 2011), CLV1 and RPK2 (Guo *et al.* 2017) are required for the nematode CLE signalling network to facilitate nematode parasitism, suggesting a receptor kinase family protein may play a role in successful nematode-host interactions.

Another example is *Arabidopsis* INFLORESCENCE DEFICIENT IN ABSCISSION (IDA)-like peptide (Kim *et al.* 2018). IDA is a signalling peptide that regulates cell separation in *Arabidopsis* including floral organ abscission and lateral root emergence, and is highly conserved in flowering plant genomes. The *M. incognita* IDA-like genes, MiIDL1 and MiIDL2, encode a small protein with N-terminal signal peptide for secretion.

Exogenous treatment of synthetic MiIDL1 peptide rescued the abscission phenotype of the Arabidopsis *ida* mutant. At the same time, constitutive expression of MiIDL1 open reading frame with signal peptide complemented the Arabidopsis *ida* mutant to produce a wild-type phenotype. Furthermore, host-induced RNAi of MiIDL1 resulted in *Arabidopsis* plants with approximately 40% fewer galls on roots and reduced gall size. Taken together, MiIDL1 peptide was indicated to mimic plant IDA function and play a role in successful gall development (Kim *et al.* 2018).

1.2.2.3 Host plant hormone status manipulated by nematode effectors

As feeding sites develop, these selected cells undergo dramatic alteration of host metabolism, differentiation and reprogramming. A large number of studies associated with genetic and biochemical analyses have revealed that changes of plant hormone status are crucial to feeding cell development and this is directly controlled by nematode effectors (Gheysen and Mitchum 2011). Auxin is the main regulator involved in nematode-manipulated developmental reprogramming of their hosts. A nematode effector from *H. schachtii* called 19C07 interacted with the *Arabidopsis* auxin influx transporter in the plasma membrane and ectopic overexpression of this effector increased the rate of lateral root emergence and enhanced auxin influx (Lee *et al.* 2011). These effector-host interactions suggested a regulatory module in which a nematode effector manipulated the auxin flow into root cells adjacent to the initial feeding cells, thereby facilitating its contribution to syncytial development. Recently, an effector termed 10A07 from *H. schachtii* has been identified. Overexpression of Hs10A07 in *Arabidopsis thaliana* produced a hyper-susceptible phenotype in response to *H. schachtii* infection along with developmental changes reminiscent of auxin effects (Hewezi *et al.* 2015). Moreover, Hs10A07 was demonstrated to interact with an IAA16 transcription factor in the nucleus. IAA16 is an auxin-responsive protein that functions as a repressor of early auxin response at low auxin concentrations (Rinaldi *et al.* 2012). Hs10A07 was proposed to undermine the ability of IAA16 to regulate auxin response factors, triggering a down-regulation of auxin-dependent transcriptional programs required for syncytium initiation and formation (Hewezi *et al.* 2015).

The natural status of many other plant hormones, in addition to auxin, is significantly changed during PPN parasitism, such as cytokinin, Jasmonate (JA) and salicylate (SA) (Gheysen and Mitchum 2019). A nematode cytokinin-synthesising

isopentenyltransferase was showed to play a key role in syncytial formation of *H. schachtii* (Siddique *et al.* 2015). In addition, chorismate mutase effectors secreted by PPN have been indicated to disrupt SA levels of the host plant (Wang *et al.* 2018). Moreover, a tyrosinase-like protein secreted by *H. schachtii* was proved to increase the levels of auxin and ethylene precursors, two hormones involved in host susceptibility to cyst nematodes (Habash *et al.* 2017). Taken together, these studies support hypothesis that the manipulation of plant hormone pathways by nematode effectors contributes to successful parasitism.

1.2.2.4 Suppression of host defence responses

PPNs are exposed to plant defence responses all the time during parasitism, and must therefore exploit a suite of effectors to suppress these plant defence responses and mediate susceptibility (Goverse and Smant 2014). One of the best characterised examples is the venom allergen-like protein (VAP) family of cyst nematodes. VAPs are structurally conserved proteins present in secretions of both animal and plant parasitic nematodes studied to date (Wilbers *et al.* 2018). A VAP effector from *G. rostochiensis* termed GrVAP1 was demonstrated to interact with the extracellular cysteine protease $Rcr3^{pim}$ of tomato, which is required in the host resistance to nematodes (Lozano-Torres *et al.* 2012). Furthermore, GrVAP1 suppressed the activation of host defence responses mediated by surface-localized immune receptors during nematode migration and the interaction between GrVAP1 and $Rcr3^{pim}$ regulated defence-related programmed cell death (Lozano-Torres *et al.* 2014). Another good example is calreticulin (CRT). A CRT effector from *M. incognita* was indicated to play an important role in the suppression of plant innate defence during compatible interactions. In addition, expression of *M. incognita* CRT in *A. thaliana* suppressed the expression of defence marker genes as well as callose deposition (Jaouannet *et al.* 2013).

Reactive Oxygen Species (ROS) are a by-product of metabolism and can be destructive for cells. ROS burst is therefore considered as a significant chemical response involved in plant basal defence and induction of programmed cell death (Holbein, Grundler and Siddique 2016). The function of effectors in protecting the nematode from excessive oxidative stress was further studied using 10A06 of *H. schachtii*. 10A06 was revealed to interact specifically with *Arabidopsis* Spermidine Synthase 2, a key enzyme involved in spermidine biosynthesis. Ectopic

overexpression of 10A06 in *Arabidopsis* resulted in an increase in cellular spermidine that can function as a ROS scavenger by reducing free hydroxyl radicals. In this way, the nematode exploits 10A06 to inhibit plant basal defences by manipulating host ROS levels (Hewezi *et al.* 2010). Recently, a novel effector termed MjTTL5 isolated from *M. javanica* was shown to interact specifically with Arabidopsis ferredoxin: thioredoxin reductase catalytic subunit, a key component of the host antioxidant system, drastically increasing host ROS-scavenging activity, and hence suppressing plant basal defence and host resistance to the nematode infection (Lin *et al.* 2016).

Most of the times, plant ROS production is regulated by a wide range of enzyme families including the NADPH oxidases, encoded by Rboh genes. *Arabidopsis* encodes ten Rboh homologues (RbohA–RbohH) (Jiménez-Quesada, Traverso and Alché 2016). A recent study characterised the role of Rboh-mediated ROS production during a compatible interaction between Arabidopsis and *H. schachtii*. In this study, it was shown that *H. schachtii* infection activated the RbohD and RbohF to produce ROS, which suppressed host cell death and promoted syncytium formation to allow successful nematode parasitism (Siddique *et al.* 2014).

Furthermore, multiple antioxidant enzymes such as superoxide dismutases (Roze *et al.* 2008), peroxiredoxins (Dubreuil *et al.* 2011; Henkle-Duhrsen and Kampkotter 2001) and glutathione peroxidase (Jones *et al.* 2004) were found in the nematode secretions, which can scavenge the defensive ROS burst from the plant and minimize the effects of ROS. Recently, 52 glutathione synthetase genes were identified from *G. pallida* genome (Cotton *et al.* 2014). Glutathione synthetase is also a key enzyme involved in cellular redox status. Interestingly, about one-quarter of the nematode genes contained a signal peptide for secretion and these all showed a peak of expression in the early parasitic stages. Taken together, this suggested that the glutathione synthetase may function as effectors and play a role in ROS pathway manipulation.

Table 1.2: Recent evidence of plant-parasitic nematode effector function, published since 2015.

Gene name	Species	Function	Reference
Hs4E02	*H. schachtii*	Suppresses plant defence; targets and re-locates vacuolar papain-like cysteine protease RD21A.	(Pogorelko *et al.* 2019)
Hs30D08	*H. schachtii*	Interacts with a host auxiliary spliceosomal protein and alters expression of genes important for feeding site formation.	(Verma *et al.* 2018)
HsCLEB	*H. schachtii*	Encodes nematode B-type CLE peptides; regulates proliferation of vascular cells during feeding site formation.	(Guo *et al.* 2017)
Hs32E03	*H. schachtii*	Mediates host chromatin modifications to alter plant rRNA gene expression.	(Vijayapalani *et al.* 2018)
Hs25A01	*H. schachtii*	Has a role in nematode parasitism; interacts with F-box-containing protein, a chalcone synthase and the translation initiation factor eIF-2 β subunit.	(Pogorelko *et al.* 2016)
Hs10A07	*H. schachtii*	Undermines plant auxin family factor IAA6 and regulates host auxin response.	(Hewezi *et al.* 2015)
HgGLAND18	*H. glycines*	Suppression of both basal and hypersensitive cell death immune responses.	(Noon *et al.* 2016)
HaEXPB2	*H. avenae*	Encodes expansin-like protein; involved in host cell wall modification.	(Liu *et al.* 2016)
Ha18764	*H. avenae*	Suppresses programmed cell death triggered by BAX; Suppresses host defence responses.	(Yang *et al.* 2019)
GpSPRY-414-2	*G. pallida*	Encodes SPRYSEC effector; interacts with potato cytoplasmic linker protein-associated protein.	(Mei *et al.* 2018)
RrCEP1	*R. reniformis*	Increases host nitrate uptake and regulates the size of the syncytial feeding site.	(Eves-Van Den Akker *et al.* 2016b)
MiPFN3	*M. incognita*	Encodes profilin; binds to monomeric actin; expression in plant cells disrupts actin filaments.	(Leelarasamee, Zhang and Gleason 2018)
MiIDL1	*M. incognita*	Encodes plant IDA-like peptide; play a role in successful gall development.	(Kim *et al.* 2018)
MiSGCR1	*M. incognita*	Suppresses plant cell death induced by plant disease; Played a role in early stage of nematode infections.	(Nguyen *et al.* 2018)
MgGPP	*M. graminicola*	Suppresses plant defences; targets to the nuclei of giant cells.	(Chen *et al.* 2017)
MjTTL5	*M. javanica*	Encodes transthyretin-like protein; Interacts with *Arabidopsis* ferredoxin: thioredoxin reductase catalytic subunit; regulates host ROS-scavenging activity.	(Lin *et al.* 2016)
BxSapB1	*B. xylophilus*	Encodes saposin-like protein with a saposin B domain.	(Hu *et al.* 2019)

1.3 Glutathione synthetase

1.3.1 Glutathione biosynthesis

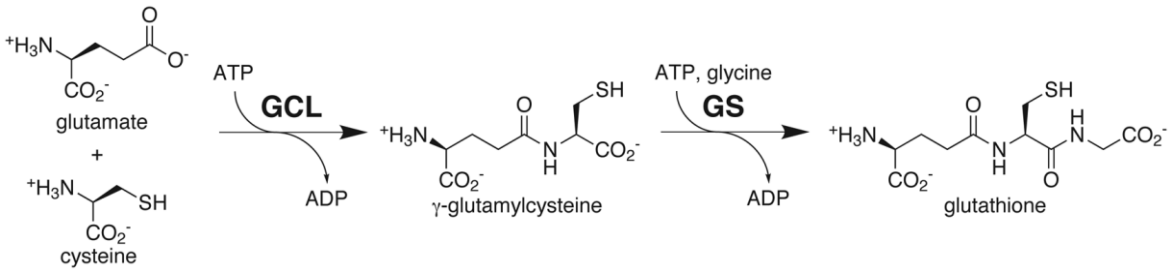
The tripeptide thiol glutathione (γ-L-glutamyl-L-cysteinyl-glycine or a functionally homologous thiol) is an essential small metabolite with multiple functions, such as preventing damage from reactive oxygen species and heavy metals (Meister 1995). Glutathione biosynthesis occurs through two conserved ATP-dependent steps in most organisms (Figure 1.2A). In the first reaction, glutamate-cysteine ligase (GCL; also known as γ-glutamylcysteine synthetase, EC 6.3.2.2) catalyses the formation of γ-glutamylcysteine (γ-EC) from cysteine and glutamate. In the second step, glutathione synthetase (GS; EC 6.3.2.3.) catalyses the addition of glycine to γ-EC to produce glutathione (Meister 1983). Taking *Arabidopsis* as an example, GCL and GS are each encoded by a single gene *GSH1* and *GSH2*, respectively (May and Leaver 1994; Wang and Oliver 1996). Reduced glutathione (GSH) is continuously oxidized to a disulphide form (GSSG) that is recycled to GSH by NADPH-dependent glutathione reductase in key organelles and the cytosol. Generally the ratio of GSH: GSSG in plant tissues such as leaves is maintained at 20:1 and the ratio may be varied in specific subcellular compartments (Noctor *et al.* 2012). Similarly in *C. elegans*, each of the synthetic enzymes is encoded by a single gene: GCS-1 and GSS-1 (Consortium 1998). GCS-1 was considered to play a role in worm resistance to arsenite (Luersen *et al.* 2013) and the oxidative stress response induced by infection of pathogenic bacteria (van der Hoeven *et al.* 2011). However, limited knowledge is known for biological functions of *C. elegans* GSS-1 so far.

Many factors affect the synthesis of glutathione, but the first step in the glutathione synthesis system is generally considered to be the rate-limiting step: GCL activity and cysteine availability are considered to be the most important factors (Noctor *et al.* 2012). Overexpression of *GSH1* or enzymes involved in cysteine biosynthesis in plants resulted in an increased glutathione content (Noctor *et al.* 1996; Harms *et al.* 2000), whereas overexpression of *GSH2* in *Arabidopsis* showed a relatively stable glutathione level (Strohm *et al.* 1995). Additionally, the subcellular localization of GCL and GS also plays a key role in the biosynthesis of glutathione. Immuno-electron microscopy of *Arabidopsis* leaf tissue showed that GCL is localized to the chloroplast and that GS is found within chloroplasts and the cytosol. The first step of glutathione

synthesis is plastidic while the second step is probably predominantly located in the cytosol (Galant *et al.* 2011).

Some plant taxa, particularly many legumes, contain glutathione homologues, in which the C-terminal residue is an amino acid other than glycine (Figure 1.2B). For example, in *Phaseolus coccineu*, homoglutathione (γ -Glu-Cys- β -Ala) was shown to be present instead of GSH (Klapheck 1988). In addition, cereals produce another GSH variant (hydroxymethylGSH; γ -Glu-Cys-Ser) through direct modification of GSH rather than alteration of the GSH biosynthesis pathway (Klapheck *et al.* 1992). In maize, exposure to cadmium activated the production of γ-glutamylcysteinylglutamate (γ -Glu-Cys-Glu) (Meuwly *et al.* 1995). Interestingly, gene duplication during evolution has resulted in the coexistence of different synthetases that produce GSH or glutathione homologues (Frendo *et al.* 2001). Novel GSH homologues in plants may remain to be discovered in the future. However, no similar situation has been described outside the plant kingdom so far.

A



B

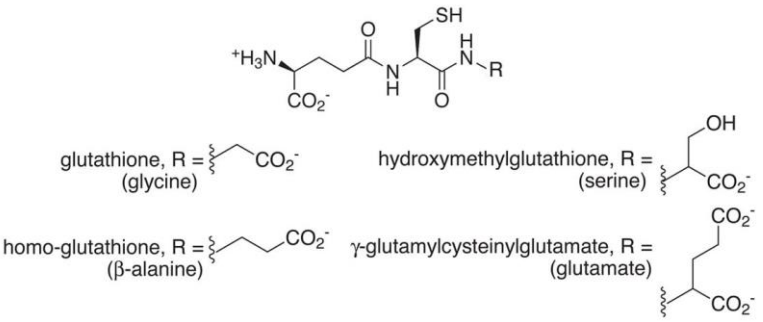


Figure 1.2 Glutathione biosynthesis. **(A)** Substrates and products of the reactions catalysed by glutamate-cysteine ligase (GCL) and glutathione synthetase (GS). **(B)** The chemical structures of glutathione analogs synthesized by various plants are shown. All share the core γ-glutamylcysteine structure with modifications to the third amino acid position as indicated.

1.3.2 Functions of glutathione during plant-pathogen interactions

Glutathione plays a wide range of roles in plants, from regulation of plant development to heavy metal detoxification, to tolerance to abiotic and biotic stress, and to metabolism of ROS and ascorbate (Rouhier, Lemaire and Jacquot 2008; Noctor *et al.* 2012). Glutathione has long been indicated to be linked to host defence response and it is now apparent that glutathione at least regulates the expression of stress defence genes and is involved in plant resistance to various plant pathogens. An important discovery had been shown by the analysis of glutathione-deficient mutants. For example, in the *Arabidopsis* mutant *rax1-1* that had >50% lowered foliar glutathione levels than wild-type, a wide set of defence-related genes and stress-responsive genes were shown to be responsive to changed glutathione metabolism within the hosts leaves infected with avirulent *Pseudomonas syringae* (Ball *et al.* 2004). Similar studies showed *Arabidopsis* mutant *pad2-1* that contained much lower amounts of GSH than wild-type displayed enhanced susceptibility to *P. brassicae* and *Spodoptera littoralis* (Parisy *et al.* 2007; Schlaeppi *et al.* 2008). In these studies, a certain level of glutathione was shown to be required for the synthesis of some plant defence-related molecules. In contrast to this, GSH metabolism was found to play a key role in nematode-induced root galls: depletion of GSH content in *Medicago truncatula* impaired nematode egg mass formation and modified the sex ratio of *M. incognita* (Baldacci-Cresp *et al.* 2012).

Moreover, programmed cell death is mainly controlled by perturbation in cellular redox balances through generation of different ROS such as hydrogen peroxide (H_2O_2) (van Doorn *et al.* 2011). It is noteworthy to underline that ROS-triggered cell death may be particularly effective against plant pathogens with biotrophic lifestyles, such as parasitic nematodes, due to the necessity of viable nutrition sources for these types of parasites (Lohar and Bird 2003). It has long been known that glutathione can interact with ROS and dehydroascorbate (DHA; the relatively stable oxidised form of ascorbate) and there is a close relationship between availability of H_2O_2 and glutathione status (Queval *et al.* 2007; Queval *et al.* 2009). Increased levels of GSH accumulation in tobacco and barley have been shown to occur during defence responses against biotrophic pathogens, protecting excess oxidative damage in the host cells surrounding Hypertensive Response area (Elzahaby, Gullner and Kiraly 1995; Fodor *et al.* 1997). A simple glutathione/ascorbate metabolic scheme was shown: glutathione

has the potential to act in detoxification and ascorbate could also be regenerated in the chloroplast by other mechanisms depending on ferredoxin or NADPH. In this pathway, GSH can be oxidized to GSSG by some ROS, such as H_2O_2, to allow regeneration of reduced ascorbate by providing electrons to diverse peroxidases. In addition GSH can also react with nitric oxide, the other major antioxidant in plant cells, to form S-nitrosoglutathione (Hogg, Singh and Kalyanaraman 1996).

In general, metabolic redox-dependent regulation of host cells plays a crucial role in plant responses to biotic stress. At the same time, the pathogens also exploit this mechanism to benefit themselves to promote parasitism.

1.4 Project overview

The reniform nematode, *R. reniformis*, is a devastating plant pathogen of global economic importance. Almost all the animals and plants investigated to date have only one gene-coding glutathione synthetase. However, a large number of GS-like sequences were found in the *R. reniformis* genome and transcriptome resources.

The aims of this project were to:

1. Identify the extent of the novel GS-like gene family from the *R. reniformis* genome assembly in association with transcriptome data using a computational approach.
2. Characterise the nature, structure and function of the GS gene family and analyse the expression profile and location of selected gene family members.
3. Solve representative GS crystal structures and understand how their active site conformations may influence their activity.
4. Investigate the roles of nematode GS-like genes in successful nematode parasitism.

Chapter 2

General Materials and Methods

2 General Materials and Methods

All routine chemicals and reagents were supplied from either Sigma Aldrich or Thermo Fisher Scientific unless specified otherwise.

2.1 Plant & bacterial growth media

2.1.1 Murashige &Skoog (½ MS10)

1 litre ½ MS10 liquid includes:

2.2 g MS medium including vitamins; 10 g sucrose; ELGA water.

Then pH was adjusted to around 5.7 by using KOH. Plant agar (Duchefa, UK) was used for flat plates at 2.2-2.4 g and for upright squares at 4 g per 400 ml ½ MS10 liquid. Then the media was autoclaved at 121 °C for 20 mins. If required, appropriate relevant antibiotics were supplemented into the agar media.

	Stock concentration	Final concentration
Ampicillin	50 mg/ml	50 µg/ml
Kanamycin	50 mg/ml	50 µg/ml
Rifampicin	50 mg/ml	50 µg/ml

2.1.2 Luria-Bertani (LB)

1 litre LB liquid included:

10 g Tryptone; 5 g Yeast extract; 10 g NaCl; ELGA water.

If LB agar media was needed, 1% bacteriological agar (w/v) was added into LB liquid. Then the media was autoclaved at 121 °C for 20 mins. If required, appropriate relevant antibiotics were supplemented into the agar media. For blue/white selection, 20 µl of 20 mg/ml X-gal were spread over the surface of the agar plate before use.

2.2 Biological materials

2.2.1 Maintenance of *Rotylenchulus reniformis*

Cotton seeds (Coker 201) were soaked in concentrated sulphuric acid for 30 seconds to clear the fibre from the seeds' surface and scarify the hard seed coat to promote germination. Seeds were then washed with running tap water four times for 2 min. The cleaned cotton seeds were placed on filter paper dampened with sterilised water in a Petri dish at room temperature. After 2 days, germinated seeds were sown in 9-cm-

diameter pots containing sand, loam and compost (Bailey's of Norfolk, UK) in a 1:1:2 ratio. Approximately 2 weeks after planting when the first set of true leaves were fully expanded, cotton seedlings were transplanted into sterilised silty loam:fine sand in a 2:1 ratio in 7" pots. The fresh sand:loam was mixed with soil and chopped infected roots from previously infected old plants (4-6 months old). The proportions depended on the infection rate of the old plants but typically approximately 1 part old soil containing roots: 5 parts new soil was used and mixed well. As the plants were in the pots for many months, slow release fertiliser granules were included at the recommended rate. The plants were grown in a glasshouse at 25-27 °C with a 16 h day length.

2.2.2 Collection of *R. reniformis* at different life-stages

A method for isolating eggs and parasitic stage feeding females of *R. reniformis* in sufficient quantities was carried out based on the protocol of (Ganji, Wubben and Jenkins 2013).

2.2.2.1 Egg collection and sterilisation

Infected cotton roots were cut into 2-3 cm pieces and then agitated in 1% sodium hypochlorite solution for 3 min. The liquid mixture was poured over nested 150, 63, 25 µm sieves and washed thoroughly with water. The above step can be repeated in fresh hypochlorite solution to extract more eggs. The eggs along with root debris on the 25 µm sieve were then washed into a 50 ml polypropylene tube in a total of 20 ml volume. The same volume of 70% sucrose solution was added into the tube and mixed well to suspend the eggs in 35% sucrose, followed by careful addition of 5 ml of water on the top of the sucrose-egg-debris mixture. The tube was then centrifuged at 1200 × *g* for 10 min. The root debris pelleted at the bottom of the tube while the eggs could be collected from the sucrose-water interface and transferred onto the 25 µm sieve. The eggs were then thoroughly washed with water to remove the sucrose, the eggs were finally concentrated and poured from the sieve in a small volume of water into a small glass beaker.

The eggs were transferred into a 1.5 ml micro-centrifuge tube for surface sterilisation. Water was removed after a brief centrifugation. 0.1% chlorhexidine digluconate; 0.5 mg/ml CTAB and 0.01% Tween 20 were added to the tube and this tube was then

incubated on a rotator at room temperature for 30 min. The sterilised eggs were thoroughly washed three times with sterilised water.

2.2.2.2 Collection of sedentary parasitic females

R. reniformis females were collected from infected cotton roots. Cleaned roots were cut into 2-3 cm pieces, and transferred to a blender (Waring, UK) in a small volume of water and disrupted with two 5-10 sec blends. The blended mixture was poured over nested 300, 150, 63 and 45 µm sieves and washed thoroughly with water. The mixture on the 150 and 63 µm sieves was then collected and transferred into a 50 ml polypropylene tube. Water was removed after centrifugation at 1500 *g* for 10 min and the nematode-debris pellet was re-suspended in 40 ml of 70% sucrose, followed by careful addition of 5 ml of water on the top of sucrose-nematode-debris mixture. Then the tube was centrifuged at 500 × *g* for 5 min. The root debris pelleted at the bottom of the tube while sedentary females could be collected from the sucrose-water interface into a beaker of water. The water was then poured over the 63 µm sieve and then the nematodes were thoroughly washed with water to remove sucrose. The mixture on the sieve was concentrated and washed into a small glass beaker with a small volume of water. The individual sedentary females were finally collected and separated from remaining debris under a stereo-binocular microscope.

2.2.3 Collection of *H. schachtii* cysts

The cysts of *H. schachtii* that had been propagated on cabbage plants were stored in damp 50:50 sand: loam mix at 4 °C. Cysts were collected by re-suspending the sand: loam mixture in three volumes of water. Once the heavy soil particles settled down to the bottom, the floating cysts were poured over and concentrated on a 300 µm sieve. Then the mixture on the sieve was washed onto a filter paper and the cysts were collected manually under a microscope.

2.2.4 Hatching of second-stage juveniles and sterilisation

The sterilised *R. reniformis* eggs were transferred to an autoclaved hatching jar containing a hatching ring with a 20 µm mesh. Sufficient sterile water was added to submerge the eggs. The jar was incubated at room temperature in the dark. Freshly hatched J2s were removed and the water replaced every few days.

H. schachtii J2 were hatched from cysts. 3 mM $ZnCl_2$ solution was used to replace sterile water as a stimulating hatching agent. Once *H. schachtii* J2 hatched, they were

collected into 15 ml centrifuge tube. A 0.1% chlorhexidine gluconate and 0.5 mg/ml CTAB solution was applied to sterile J2s for 30 min. The sterilised J2s were then thoroughly washed three times with sterilised water.

2.3 Molecular protocols

2.3.1 Polymerase Chain Reaction (PCR)

Primers were designed by Primer3Plus (available at http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesised by Integrated DNA Technologies. The ideal primer generally has the following characteristics:

1) The annealing temperature (Tm) between 55 and 65°C (usually corresponds to 45-55% G+C for a 20-mer). The annealing temperature of the primers was determined at NEB Tm Calculator (https://tmcalculator.neb.com/#!/main).

2) Absence of dimerization capability.

3) Absence of significant hairpin formation (usually >3 bp).

4) Lack of secondary priming sites in the template.

5) Low specific binding at the 3' end, to avoid mispriming.

PCR was carried out for sequences of interest with relevant primer pairs.

For cloning purpose, Phusion proof-reading enzyme (New England BioLabs, UK) was required in the PCR reaction. Each PCR reaction contained 5 µl 5x buffer, 1 µl 10 mM dNTPs, 1 µl of 10 µM each relevant primer, 20-50 ng of DNA template, 0.5 µl Phusion enzyme and ddH_2O to make a 25 µl final volume. The typical PCR cycling conditions were: 98 °C for 30 sec, followed by 30-35 cycles of 98 °C for 10 sec, the specific annealing temperature for 30 sec (depending on the primer sequences), 72 °C for 30-120 sec (depending on the length of the target), ending with an extension at 72 °C for 10 min.

For colony screening purpose, MyTaq Red Mix (Bioline, UK) was used in the PCR reaction. Colonies were screened for presence of desired gene by PCR. Each PCR reaction contained 10 µl 2x MyTaq Red Mix, 1 µl of 10 µM each relevant primer and ddH_2O to make a 20 µl final volume. A single colony was touched by a P200 tip and this tip was then inserted into the PCR reaction mixture and mixed by pipetting up and down. Or one microlitre of grown bacterial culture was used as DNA template. The

typical PCR cycling conditions were: 95 °C for 60 sec, followed by 25 cycles of 95 °C for 15 sec, the specific annealing temperature for 15 sec and 72 °C for 30 sec, ending with an extension at 72 °C for 5 min.

If downstream experiments required purified DNA, the PCR reaction was cleaned up using the QIAquick PCR Purification Kit (Qiagen, Germany) based on the manufacturer's instructions.

2.3.2 Agarose gel electrophoresis and gel extraction

PCR products were analysed by agarose gel electrophoresis. Approximately 1% w/v agarose was added into TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) and completely dissolved by microwaving for 1.5 min. DNA was visualised by addition of GelRed (Cambridge Bioscience, UK) into the molten agarose at a concentration of 1: 20000. The gel was typically electrophoresed at 100 volts for around 30-40 min. If required, DNA bands of interest were extracted from the gel using a QIAquick Gel Extraction Kit (Qiagen, Germany) based on the manufacturer's instructions.

2.3.3 Restriction enzyme digestion of DNA

Restriction enzyme digestions were usually carried out in 20 µl volume. Each digestion included: a final concentration of 1 × NEBuffer (New England Biolabs, UK); Approximately 1 µg DNA; the relevant restriction enzymes. The buffer should ensure 100% enzyme activity. The mixture was incubated at relevant temperature (usually 37 °C) for three hours. Agarose gel electrophoresis was used to confirm complete enzyme digestion, and to determine the sizes of DNA fragments produced.

2.3.4 DNA ligation

Phusion polymerase does not produce A-overhangs for subsequent T/A cloning of the PCR products. In this case, 1 µl Taq DNA polymerase as well as 1 µl 10 mM dATP and sufficient ThermoPolTM Superscript Reaction buffer was added to the purified DNA fragment and the reaction incubated at 72 °C for 10 min to allow cloning into pGEM-T Easy vector (Promega) after gel extraction.

Linear DNA insert fragments and relevant linear vector were then combined at a 3: 1 ratio. T4 DNA ligase, relevant buffer (final concentration 1X) and ddH_2O were added to make a final 10 µl reaction volume, which was incubated at room temperature for 30-60 min or at 4 °C overnight.

2.3.5 Method for making *E. coli* ultra-competent cells

10-12 large *E. coli* DH5α colonies were picked up and incubated in 250 ml LB medium in a 1 L flask at 19 °C with 200 rpm shaking until the OD_{600} value reached 0.5 (normally takes 24-36 hours). Once the desired OD_{600} value was reached, the cultures were cooled down on ice for 10 min, followed by centrifugation at 4000 rpm for 10 min at 4 °C. The cell pellets were gently resuspended in 80 ml ice-cold TB (10 mM PIPES, 15 mM $CaCl_2$, 250 mM KCl) and then stored on ice for 10 min. The cells were centrifuged again at 4000 rpm for 10 min at 4 °C, followed by resuspending in 20 ml ice-cold TB with the addition of 1.4 ml DMSO (the DMSO needs to be stored at -20 °C overnight before use). 100 to 200 µl of cells was aliquoted into individual 1.5 ml microfuge tube and then quickly frozen in liquid nitrogen and store at -80 °C.

2.3.6 Transformation of competent *E. coli*

LB agar plates with relevant antibiotics were pre-warmed and dried at 37 °C, followed by thawing competent *E. coli* DH5α cells on ice. The ligation was added into the cells and left on ice for 5 minutes. The cell mix was pipetted directly onto the pre-warmed plates and gently spread, and then incubated at 37 °C overnight.

2.3.7 Plasmid DNA extraction from *E. coli*

Single colonies were picked out into 5 ml LB liquid medium containing relevant antibiotics and then incubated at 37 °C with 200 rpm shaking overnight. The cultures that produced a PCR product with an expected size were used for plasmid extraction using a Qiaprep Spin MiniKit (Qiagen, Germany) based on the manufacturer's instructions.

2.3.8 DNA sequencing

5 µl 30-100 ng/µl purified plasmids were submitted for sequencing. The DNA sequencing service was provided by GeneWiz.

2.4 RNA extraction and cDNA synthesis

Total RNA was extracted from liquid nitrogen frozen tissue sample using an RNeasy kit (Qiagen, Germany) according to the relevant protocol and including an on-column DNase I digestion. For extraction from nematode samples, the manufacturer's instructions for animal tissues were followed. For extraction from plant samples, the manufacturer's instructions for plant tissues were followed.

Reverse transcription was then carried out from 100 ng to1 µg total RNA to make first strand cDNA by using Superscript II Reverse transcriptase (Invitrogen, UK) according to the manufacturer guidelines. The DNA/RNA concentrations were measured by NanoDrop spectrophotometers (Thermo Fisher Scientific, UK).

2.5 General primers

M13F: TGTAAAACGACGGCCAGT

M13R: CAGGAAACAGCTATGAC

2.6 Methods of statistical analysis

Statistical analyses were carried out using a Student’s t-test assuming a two-tailed distribution with an unequal variance. Error bars presented on all graphs illustrate the Standard Error (SE) of the Mean.

Chapter 3

Identification of a Glutathione Synthetase Gene Family in *R. reniformis*

3 Identification of a Glutathione Synthetase-like Gene Family in *R. reniformis*

3.1 Introduction

3.1.1 Genomic resources for *R. reniformis*

Rotylenchulus reniformis is a major agricultural pest (Robinson *et al.* 1997) but until very recently there were limited genomic resources available for this nematode and more detailed knowledge of its parasitic mechanism is still required. The estimated size of the *R. reniformis* genome, based on flow cytometry in two separate studies, is around 190 Mb (Ganji *et al*. 2013; Showmaker *et al.* 2019) which is considered as one of the largest PPN genome investigated to date. Genomic characteristics such as genome rearrangements, transpositions, tandem repeats and segmental duplications are often features of large genomes (Tang 2007), potentially making assembly of the *R. reniformis* genome more challenging.

In 2014, the first genome draft for *R. reniformis* by shotgun sequencing was reported, with the authors indicating the identification of a range of genes associated with core biological processes, and highlighting a number of genes in categories such as detoxification, carbohydrate-active enzymes and 'parasitism genes' (Nyaku *et al.* 2014). Interestingly, within the category of antioxidant genes, 8 contigs were found to encode glutathione synthetases, although the overall homology to their best protein matches was generally low. However, the quality of this genomic resource is a barrier to progress. Over 1.2 million genomic reads were generated by 454 sequencing from whole-genome amplified DNA pooled from four female nematodes. This represented about 380 Mb of sequences, providing only 2-fold coverage of the genome and the assembled contigs covered only 37 Mb. In addition, 89% of the 67,317 contigs were shorter than 1000 bp so it is perhaps not surprising that a relatively small number of GS genes were represented in this assembly. A novel draft genome assembly of *R. reniformis* using both small- and large-insert libraries to provide >70 Gb of Illumina sequence data in total, was reported this year (Showmaker *et al.* 2019). This higher quality assembly is 314 Mb and contains genes encoding 86% of the core eukaryotic proteins. The larger assembly size than indicated from flow cytometry may be due to unresolved haplotypes within the heterogeneous population of *R. reniformis* that provided the starting material. With the help of this genomic resource, numerous

R. reniformis homologues of known plant-parasitic nematode effector molecules were identified, such as chorismate mutase, CEP, CLE peptides, ubiquitin extension protein and venom allergen-like protein as well as many cyst nematode pioneer effectors (Showmaker *et al.* 2019). Access to this genome assembly was made available to us prior to publication and, due to its better quality, all the *R. reniformis* genomic analysis in this study were based on the described dataset (RREN1.0, GCA_001026735.1 under BioProject No. PRJNA214681).

In addition to genome data, there are also a number of transcript-based sequence resources for *R. reniformis*. An expressed sequence tag (EST) analysis of parasitic females of *R. reniformis* was performed in 2010, which represented a small portion of the entire *R. reniformis* transcriptome but nevertheless provided a starting point for studying *R. reniformis* from a functional genomic perspective (Wubben, Callahan and Scheffler 2010). An RNAseq approach using 454 sequencing of egg and J2 RNA produced 20,596 contigs, although the average length of these was only 231 bp (Nyaku *et al.* 2013). Another life cycle stage specific transcriptomic resource for J2 and parasitic J4 female *R. reniformis* were provided recently (Eves-Van Den Akker *et al.* 2016b), which was exploited for the initial identification of *R. reniformis* GS members in this study. Most recently, and most comprehensively, the transcriptome assemblies of five life stages (eggs, J2, J3, vermiform adult and sedentary female) of *R. reniformis* were presented (Showmaker *et al.* 2018). Completeness assessment of these assemblies using CEGMA ranged from 81.45% to 83.06%. In this thesis, the transcriptome assembly and transcripts containing GS-like domains were identified in the *R. reniformis* next-generation sequencing (NGS) data (ERA PRJEB8325 and SRR949271) as described by Eves-Van Den Akker *et al.* 2016b. In addition, the expression analysis was based on the datasets under BioProject no. PRJNA286314.

3.1.2 Glutathione synthetase genes in other species

Glutathione synthetase in general is present in a broad diversity of eukaryotic and prokaryotic organisms (Mooz and Meister 1967). Despite relatively high sequence similarity within each main group (~30%-40%), there is little similarity between eukaryotic and prokaryotic GS genes. Previous phylogenetic analysis showed that the eukaryotic GS did not evolve directly from the bacterial GS and it is uncertain whether these proteins are homologous or arose by convergent evolution (Copley and Dhillon 2002). Taking *E. coli* GS as a representative of the prokaryotic GS family, it has 316

coding amino acids containing an ATP-grasp domain and the molecular weight calculated from the predicted amino acid sequence is around 35 kDa (Gushima *et al.* 1984), which is smaller than a typical eukaryotic GS.

The first characterised mammalian GS was isolated from rat kidney. Rat kidney GS has 474 amino acids which showed no significant similarity to the enzyme from *E. coli* (Huang *et al.* 1995). Human GS was found to have the same amino acid length with a molecular mass of 52 kDa. Southern blots of human genomic DNA hybridized with the GS cDNA revealed a relatively simple pattern of strongly hybridising fragments, indicating the absence of a gene family and suggesting that there is only one GS gene copy in the human genome (Gali and Board 1995). Similarly, only one GS gene was discovered in the genomes of *C. elegans* (Consortium 1998; Li *et al.* 2004) and many other eukaryotic organisms. Interestingly, the genomes of some plant species were found to contain more than one GS gene. For example, the soybean genome contains two GS and two hGS genes, with each pair sharing 87 and 93% sequence identity, respectively (Frendo *et al.* 2001; Schmutz *et al.* 2010). Also, three GS genes were isolated from the rice genome and all of the encoded proteins displayed GS enzyme activity, whereas only one of them had hGS enzyme activity (Yamazaki, Ochiai and Matoh 2019). It is therefore hypothesised that atypical GS likely arose from canonical GS by divergent evolution after the first duplication event because these plant genomes have undergone several rounds of genome duplication (Galant *et al.* 2011).

A large expansion of glutathione synthetase genes has been recently demonstrated in many plant parasitic nematodes (Cotton *et al.* 2014; Lilley *et al.* 2018), including *R. reniformis*. All animals and most plants investigated previously possess only one gene coding for GS. Given the fact that glutathione deficiency impaired root-knot nematode development in *M. truncatula* (Baldacci-Cresp *et al.* 2012), it was hypothesised that this unexpected expansion of GS genes in plant parasitic nematodes may be associated with successful nematode parasitism (Cotton *et al.* 2014).

3.1.3 The GS domain as a computational tool to predict GS-like genes

Proteins generally have one or more functional regions, which are commonly termed 'domains'. Today, Pfam has become the most popular database for identification of conserved domains within protein. Pfam is a database of protein families that includes

their annotations and multiple sequence alignments generated using Hidden Markov Models (HMM) (El-Gebali *et al.* 2019). HMM are probabilistic models used for the statistical inference of homology built from an aligned set of curator-defined family representative sequences (Krogh *et al.* 1994). In Pfam, the HMM search is exploited on a large sequence collection to discover all homologues of a diverse superfamily.

Previous studies indicated that all eukaryotic GS enzymes have similar domains and all belong to the ATP-grasp superfamily that contains an ATP-grasp fold (Copley and Dhillon 2002). The ATP-grasp fold is conserved within the ATP-grasp superfamily and is characterized by two alpha helices and beta sheets that hold onto the ATP molecule between them (Fawaz, Topper and Firestine 2011). Therefore, members of the ATP-grasp superfamily typically have an overall structural design containing three common conserved focal domains. In addition, by analysing the structures of eukaryotic GS, a substrate-binding domain was identified. This domain has a 3-layer alpha/beta/alpha structure (Polekhina *et al.* 1999). Taken together, Pfam domain GSH_synth_ATP (PF03917) and GSH_synthase (PF03199) which represent the GS ATP-grasp fold and GS substrate-binding domains respectively can be exploited to predict GS-like genes from the genome and transcriptome resource of *R. reniformis*.

3.2 Aims

1. To identify the complement of GS-like sequences in the *R. reniformis* genome and transcriptome.
2. To define the phylogenetic relationship between the *R. reniformis* GS-like genes and those of other nematodes.
3. To analyse the spatial and temporal expression profiles of *R. reniformis* GS-like genes as a basis for understanding their likely roles.

3.3 Materials and methods

3.3.1 Computational prediction of GS-like genes

3.3.1.1 GS-like sequence search

GS-like PFAM domain-containing sequences (PF03917 & PF03199) were identified in the genome assembly (RREN1.0, GCA_001026735.1 under BioProject No. PRJNA214681) (Showmaker *et al.* 2019) and the J2 and parasitic J4 female life-specific transcriptome resources (ERA PRJEB8325 and SRR949271) (Eves-Van Den Akker *et al.* 2016b) using hidden Markov models SEARCH v 3.1b2 (El-Gebali *et al.* 2019). Additional GS-like sequences were identified in the genome and transcriptome by sequence similarity searches with BLAST v 2.4.0 (Li *et al.* 2015) using all 52 full-length *G. pallida* GS amino acid sequences (Lilley *et al*. 2018) as queries.

The results of these two identification pipelines were merged, to produce a single list of unique GS-like genes present in either/both the genome and transcriptome. Several of the GS-like sequences identified were clearly partial (short sequences that were lacking either the 5’ or 3’ end, or both a start and stop codon). This can be the result of insufficient sequencing depth in the transcriptome, assembly artefacts, or incorrect gene calls using the genomic information. To highlight additional genes that were likely truncated or misassembled/predicted, the predicted proteins encoded by all the GS-like sequences from the genome and transcriptome database searches were aligned with Muscle 3.8.31 (Edgar 2004), and visualised in Jalview 2.9.0b2 (Waterhouse *et al.* 2009).

For any given apparently incomplete *R. reniformis* GS-like gene prediction from the genomic information, an attempt was first made to manually identify the sequence information missing from the 5’ and/or 3’ regions using the following procedure:

1. A related, apparently full length, GS-like sequence was selected.
2. In genomic regions 1 kb adjacent upstream and/or downstream of the gene with missing information, sequence similar to the apparently full length reference GS was identified using BLASTn.
3. If regions of high similarity were identified that co-incided with canonical intron exon boundaries, they were added to the original gene model and the new coding sequence was put back into the GS list. All the amino acid sequences were realigned and then manually checked for congruence in an iterative approach.

3.3.1.2 Rules to remove GS sequences

Having corrected as many apparent misprediction/assembly artefacts as possible, a series of rational criteria were designed to remove redundant sequences, and those that could not be corrected.

1. Sequences which shared 100% amino acid identity with other GS-like sequences were removed, to leave one representative.
2. Gene predictions that were incomplete by lacking either the highly conserved N or C termini, and for which the missing sequence could not be found in the genome assembly were removed.
3. GS-like sequences apparently missing internal exons or with introns potentially retained in the gene model (as determined by multiple alignment) were amplified from cDNA, cloned and sequenced as described in the General methods section to provide a high confidence sequence for analysis.
4. The sequences that remained incomplete, were missing many nucleotides within the gene and for which cloning subsequently failed, were then removed.
5. When fixing partial genes, if more than one partial sequence mapped to the same location in the genome as reference, only a single complete sequence was left in the final list, the others were removed.

Taken together, we were sufficiently confident that the remaining corrected, non-redundant, and likely full length GS-like sequences could be treated as individual genes for further analysis. Primers for amplification and cloning were designed in the 5‘and 3’ untranslated regions. All primers used for the amplification of GS-like coding regions from *R. reniformis* cDNA are listed in Table 3.1.

3.3.2 Phylogenetic analyses

A phylogeny of GS-like sequences from plant parasitic nematodes *Bursaphelenchus xylophilus*, *Longidorus elongatus*, *Pratylenchus penetrans*, *M. incognita*, *Nacobbus aberrans*, *R. reniformis*, *G. rostochiensis*, *G. pallida*, *H. schachtii* and *H. avenae*, and free-living nematodes and animal parasitic nematodes *C. elegans*, *C. briggsae*, *C. remanei*, *C. nigoni*, *C. brenneri*, *Strongyloides ratti*, *Brugia malayi*, *Loa loa*, *Trichinella spiralis*, *T. suis*, *T. native*, *T. patagoniensis*, *T. pseudospiralis*, *T. muris*, *Pristionchus pacificus*, *Ancylostoma ceylanicum*, *Diploscapter pachys*, *Toxocara canis*, *Onchocerca flexuosa*, *Ascaris suum* and *Wuchereria bancrofti* (termed the ‘all

nematode GS sequences' in this thesis), and a separate phylogeny of *R. reniformis* GS sequences with *C. elegans* GS as an outgroup were built. The deduced amino acid sequences of corrected, non-redundant, and likely full length GS-like sequences were aligned using Muscle 3.8.31. A Bayesian phylogenetic tree derived from this alignment was then generated using TOPAli V2.4, using the WAG + Gamma model. Bayesian inferences of all nematode and *R. reniformis* GS phylogeny were run for 2,500,000 generations with 25% burn-in value and 1,000,000 generations with 25% burn-in value, respectively. The phylogenetic tree was re-rooted by the known outgroup GS Clade containing the single *C. elegans* sequence in FigTree V1.4.3 (available at http://tree.bio.ed.ac.uk/software).

3.3.3 Signal peptide prediction

The presence/absence of N-terminal signal peptides in the *R. reniformis* GS proteins was predicted using the SignalP 4.1 Server (Petersen *et al.* 2011).

3.3.4 Expression profiling of GS genes across the *R. reniformis* life-cycle

The *R. reniformis* raw RNAseq reads from five life stages: egg, J2, J3, vermiform adult and sedentary female, were downloaded from NCBI under BioProject no. PRJNA286314 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP059368). The RNAseq pipeline was carried out at https://usegalaxy.org/ (Afgan *et al.* 2018). The raw RNAseq reads were assembled and normalised using Trinity (Grabherr *et al.* 2011). The assembled sequences were subsequently trimmed and filtered for adapters and low-quality base calls with Trimmomatic (Bolger, Lohse and Usadel 2014). The trimmed reads were then mapped back to all the full-length GS-like nucleotide sequences by BLASTn. The transcript expressions were counted as Transcripts Per Million (TPM) values using Salmon (Patro *et al.* 2017). Transcript abundance data and relative expression for each GS-like sequence was calculated as the average TPM of each life stage. Morpheus (https://software.broadinstitute.org/morpheus/) was used for the generation of an expression heatmap. The relative expression values were then calculated based on the TPM values of each genes from different life stages.

3.3.5 *In situ* hybridisation

3.3.5.1 Preparation of DIG-labelled DNA probes

A 200-250 bp fragment of selected, cloned GS genes of interest was amplified from plasmid DNA using Phusion proof-reading enzyme, ensuring that the sequence was

specific for the gene of interest. Following gel electrophoresis and excision of the amplified fragment from the gel, asymmetric PCR was carried out to incorporate digoxigenin (DIG) labelled dUTP into two single-stranded DNA probes using the following reagents and reaction conditions. All primers used in preparation of *in situ* hybridization probes are summarised in Table 3.2.

Reagents: 2 µl 10 × buffer; 4 µl 5 µM forward or reverse primer (the sense probes amplified with forward primers were used as negative controls); 0.5 µl Biotaq polymerase; 1.5 µl DIG DNA labelling Mix (Roche); 50 ng purified PCR product as template and RNase-free water to a final volume of 20 µl.

Asymmetric PCR was carried out using either forward or reverse primer only by incubating at 94 °C for 2 minutes, followed by 35 cycles of 94 °C for 15 seconds, annealing temperature for 30 seconds, and 72 °C extension for 90 seconds.

Two microlitres of each probe were analysed on an agarose gel alongside 1.5 µl of unlabelled template DNA fragment. The molecular mass of the labelled product should be larger due to the incorporated DIG. The DIG labelled probes were stored at -20 °C until required.

3.3.5.2 Fixation of nematodes

Fixation of J2 stage nematodes.

Eggs of *R. reniformis*, extracted from roots according to 2.2.2.1 were incubated in sterile tap water at 25 °C to allow hatching of J2 stage nematodes. J2s were collected into a 1.5 ml microcentrifuge tube (Non-stick, RNase free) and pelleted by brief centrifugation. The J2s were resuspended in 1 ml of fixative (2% paraformaldehyde in RNase-free M9 buffer) and the tube was placed on its side at 4 °C for 18 hours and then fixed at room temperature for an additional 12 hours.

Fixation of nematodes from infected plants.

The roots from an infected cotton plant were washed and cut into around 2 cm sections and then blended briefly (5 sec) in a volume of tap water sufficient to cover the root segments. The root and water were transferred to a 500 ml glass beaker and the volume increased to 300 ml, followed by addition of 100 ml formaldehyde (around 37%) to give a final concentration of approximately 10% formaldehyde. The beaker was covered with foil and left in a fume hood. Three days later, the roots were tipped onto

a 63 µm pore sieve and washed briefly with tap water. Then the roots were transferred to a blender and blended with tap water for 5-10 sec. Nematodes were collected on a tower of sieves: 300 µm, 150 µm, 63 µm, 25 µm. The roots and worms from the 63 µm and 150 µm sieves were collected into separate 50 ml centrifuge tubes, followed by centrifugation at 2455 *g* for 5 min. Supernatant was removed and the pellet was re-suspended in 40 % sucrose. Five millilitres of tap water was carefully added to the tube to form an upper layer and then spun at 1500 × *g* for 10 min. The white nematode layer at the sucrose: water interface was removed with a glass pipette and placed into a beaker of tap water. Then, the nematodes were collected into a watch glass containing a small volume of sterile tap water and a stereo-binocular microscope was used to facilitate removal of debris and excess water from the watch glass with a pipette.

3.3.5.3 Hybridisation and detection of probe

In situ hybridization was carried out based on the method of de Boer *et al*. (1998) with minor modifications as follows:

The clean nematodes were concentrated into a 1.5ml microcentrifuge tube and then re-suspended in 150-200 µl RNase free M9 buffer containing 10% fixative per mm of nematode pellet. Around 100 µl of the nematode suspension were pipetted as an elongated drop on a clean microscope slide. The nematodes were cut on the slide using a single edge razor blade until over 50% of the nematodes were chopped.

The cut nematodes were incubated in proteinase-K solution. J2 nematodes were incubated in 0.5 mg/ml concentration of proteinase-K solution in 1 ml M9 buffer at room temperature for 30 min. Parasitic stage sedentary female nematodes from infected plants were incubated in 2 mg/ml concentration of proteinase-K solution in 1 ml M9 buffer at 37 °C for 90 min.

Hybridisation was performed overnight at 50 °C and an estimated probe concentration was 300 ng/ml.

Table 3.1: Primers used for amplification of full-length *R. reniformis* GS-like coding regions.

Primer name	Seq 5'-3'	TM (°C)
GS4-F	TTTCACGATCCTGAGACAA	58
GS4-R	ATCAAATTGACCAATCACG	
GS5-F	ATGTCATCGGCATTCAAAATGA	63
GS5-R	TGATCACTTCCCATTGATCATC	
GS18-F	AATGAAAAATTTCTTTTATCTAAGG	56
GS18-R	AAAACAAGAAAAAGTTCAATAAAGAT	
GS20-F	AATATTTCTTTCTTCTAACGCTTTT	56
GS20-R	TAGAAAAATACGGATAATAAAAATCT	
GS27-F	GTGCAAATGTCGATATTTT	55
GS27-R	GCATAATGGAATAGGAATAG	
GS36-F	TTCCAATGATTTTCATGCA	57
GS36-R	ATCCGAATTTTACAAGCCA	
GS44-F	AATGATGAAATTGGTGCAA	56
GS44-R	ATCCAACAATGATAATAGCA	
GS49-F	TGTCTTAAACCCGGATTTTC	60
GS49-R	CATCATCATCATCATCGCATA	
GS50-F	AATATTTTCATGGCATCGA	57
GS50-R	AGAACAGGTATGGCGAGTC	
GS51-F	TAATGAAGCATTCTGTGAA	55
GS51-R	TCATATTATCATGAACCCA	
GS55-F	TACAATAATGTTCGTCCAAA	57
GS55-R	GCAAAATGCTAATAACCAAA	
GS59-F	CGAACGACAACAAATAATGT	59
GS59-R	GGGGTTCTTAATACAGGAAA	
GS64-F	CCCTATCCTCGCCAACTGT	62
GS64-R	TCGTCAAATTCCAAATGCC	
GS65-F	TAAAACTAGAATGGAATTGC	55
GS65-R	TATGGTTATTATCTTCTCGG	
GS66-F	TATTCTTTGCTTGCTTCCCA	62
GS66-F	TTCAATCGCTCCGAACAAAT	
GS67-F	ATGGCTATTTTGCTGAATAT	57
GS67-R	TACATCATTTCCCATAGGTT	
GS72-F	TCTTCTGCAACTACCGATA	56
GS72-R	TTGGGTGAAAACTTGATAT	

Table 3.2: Primers used in *in situ* hybridisation.

Primer name	Seq 5'-3'	TM (°C)	Length of probe
In situ-GS1-F	CCAACCGCAATTTGAGCTCAA	64	246
In situ-GS1-R	TTCTGGTTCACCTCACCGATG	65	
In situ-GS2-F	GAACCAACGGAAGCGTACATG	64	223
In situ-GS2-R	TCCATGGCCTGGTAGAACAAC	65	
In situ-GS11-F	CAATTCCTATGCCATTGCGGG	65	229
In situ-GS11-R	CAACTCGTTGAGTGCCTGTTG	65	
In situ-GS14-F	TGGGAGGTGGAGCAGATGAC	67	212
In situ-GS14-R	GCGCTGGAACTATGGATTTT	61	
In situ-GS23-F	GACATTGTTCCCGTCCAAAT	61	206
In situ-GS23-R	TCTGCTGTCGGTATCCCTCT	65	
In situ-GS36-F	CCCTGAACTTGTTGTATTGGC	62	223
In situ-GS36-R	TCATTGTTCCCTTCGGCTTG	63	
In situ-GS49-F	TGAACTGTTCCACCAAGCAG	65	202
In situ-GS49-F	TCGTTGGAATACCATGCTGA	64	
In situ-GS55-F	AAGAGGCAATGACCCTGTTG	63	234
In situ-GS55-F	CATGATGTAGCTGGCCTTCA	63	
In situ-GS67-F	CAATTGGGCTGATGATGATG	59	210
In situ-GS67-F	GGGTGTCAGTTGCATTGTTG	63	

3.4 Results

3.4.1 Discovery of a large group of GS-like genes from *R. reniformis*

A computational approach described in the Section 3.3.1 combining both transcriptome and genome information was exploited to identify all GS-like gene family members in *R. reniformis*. The discovery pipeline is illustrated in Figure 3.1. 184 sequences were identified to contain the ATP-grasp domain of a glutathione synthetase (PF03917) while 92 sequences contained a GS substrate-binding domain (PF03199) in the genome. After merging these two groups, 188 GS-like sequences were identified using the Pfam domain search. At the same time, 107 sequences were identified from the *R. reniformis* genome by similarity searches with *G. pallida* GS amino acid sequences. After merging these two groups, 189 GS-like sequences were identified in the genome. In addition, 71 GS-like sequences were identified using the same method from the transcriptome assemblies. After merging these two groups, a total of 260 GS-like sequences were found in the *R. reniformis* genome and transcriptome. Where possible, partial, incomplete, and mis-predicted sequences were manually refined by either searching upstream and downstream regions of the genome or by amplification of coding regions from cDNA, followed by cloning and sequencing as detailed in the Section 2.3.1.

A typical example of the process to search in the genome assembly for the missing sequence information from a partial GS gene is shown in Figure 3.2: *Rre-gs68* was identified from the transcriptome assembly and was considered to be a partial sequence because of a lack of a start codon in the predicted protein (Figure 3.2 A). By analysing and comparing the original *Rre-gs68* transcriptome sequence and the *R. reniformis* genome data, the missing N-terminal region was identified in the genome using the predicted gene g32685.t1 as reference (Figure 3.2 B). Although there are some minor differences between the original *Rre-gs68* sequence and g32685.t1, the missing N terminal sequence was directly added before the original *Rre-gs68* sequence to form a new correctly fixed GS-like protein (Figure 3.2 C).

An example of the process of PCR to fix a partial sequence is shown in Figure 3.3. The original genomic sequence corresponding to *Rre-gs59* was amplified by PCR using gene specific primers and the relevant *R. reniformis* life-stage cDNA as template (Figure 3.3A). In this case, *R. reniformis* female cDNA was utilised as *Rre-gs59*

sequence reads were more abundant in the female transcriptome dataset. A clear band of 1500 bp was amplified (Figure 3.3A). The purified, amplified products were then cloned into pGEM-T Easy vector and up to 6 representative clones for each gene were sequenced (Figure 3.3B). The missing information of the partial sequences was fixed based on the sequencing results (Figure 3.3C). In addition, some cloned cDNAs were found to differ slightly from the expected sequence in the genome and or transcriptome resources and were subsequently used to replace the original sequence.

Despite many attempts to fix all the partial or mispredicted GS sequences, there remained 186 sequences that were still apparently incomplete. These sequences were removed from the final GS list according to a series of criteria described in the Methods section. A total of 73 corrected, non-redundant, and likely full length sequences remained for further study. All these 73 *R. reniformis* GS-like sequences are listed in the Appendix 1.

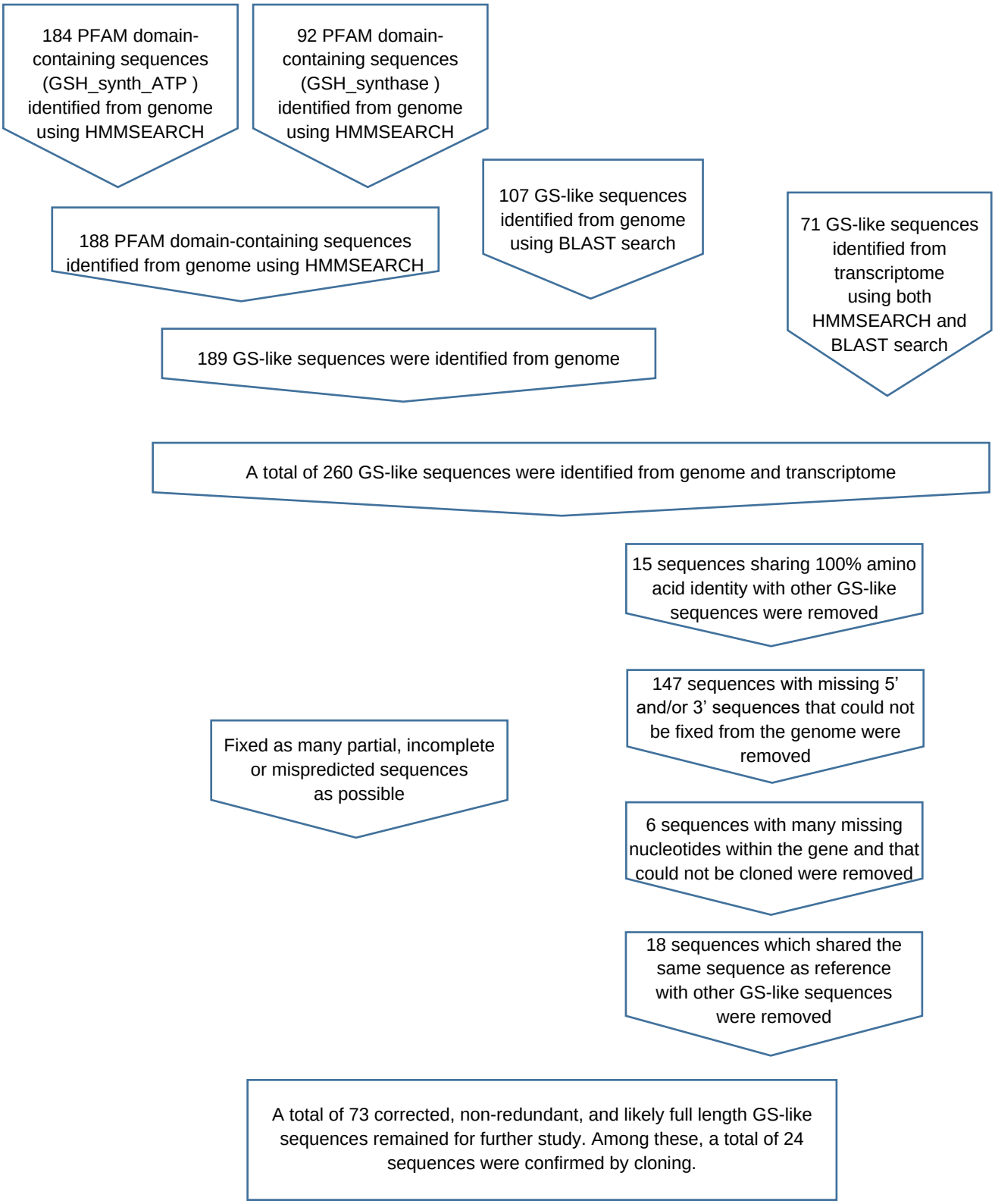


Figure 3.1: Overview of the identification pipeline for the GS gene family in *R. reniformis*. In brief, based on both transcriptome and genome sequences, computational homology analysis associated with manual confirmation was used to identify GS gene family members in *R. reniformis*. Some of the truncated sequences identified may be genuine GS-like genes that have arisen during genome expansion but may have lost their functions, however, these sequences were not included in the final list.

A

>Original *Rre-gs68*

NICIACCCFYCVFGETSGQQEIDVQVLVEDALDYGHYVGLIHRAKDHLKSSDLSEVSAMALFPSPFPRQVFEDANNVQEALAELYFRVANDYEFLMNAYREVRKVDKTVDKL
MNLLEDIRKKGIHQPIGLMMMRADYMANMNEQNSESPYEIKQIEVNIGAVGGATCEKATLVHRRVLAKAGMTSVVLPDNNATDTLAMGMYQAWKAFNNEQAIIVTIIGKLGQ
KTQYEMRKAEYKATELSGGKIRTVCMNLTEANEKLTLDDNFNLRLDDQIVAVVNYRLARNIPEKFLTDEKIDVWTKMEVSTAIKSPPLNYEIACTKKMQQVLAEKDVVEKFF
LEPKDAKKVAAIRKFQARMWSLDHNDEKTQAVIQDAIEHPDRYVLKPNKDGGGNNLWEEEMKIKLETLKPEERSQYILMQRIRPFVGKNFLKRPLEQARYEDQVVTELSIFG
ALLGNQENGKILHNKGGGHMMRSKPKHVNEGGLEMGAGFYDSPLLI-

B

```
      g32685.t1       MAILLNICIACCCYYCVFGETSGQQEIDVQVLVEDALDYGHYVGLIHRAKDHLKSSDLSE
Origin Rre-gs68       -----NICIACCCFYCVFGETSGQQEIDVQVLVEDALDYGHYVGLIHRAKDHLKSSDLSE
                           ********:**************************************************

      g32685.t1       VSAMALFPSPFPRQVFEDANNVQEALAELYFRVANDYEFLMNAYREVRKVDKTVDKLMNL
Origin Rre-gs68       VSAMALFPSPFPRQVFEDANNVQEALAELYFRVANDYEFLMNAYREVRKVDKTVDKLMNL
                      ************************************************************

      g32685.t1       LEDIRKKGIHQPIGLMMMRADYMANQNEQNPDSPYEIKQIEVNIGAVGGATCEKATLVHR
Origin Rre-gs68       LEDIRKKGIHQPIGLMMMRADYMANMNEQNSESPYEIKQIEVNIGAVGGATCEKATLVHR
                      ************************ ****.:****************************

      g32685.t1       RVLAKAGITSVVLPDNHATNTLAMGMYQAWKAFGNENAIIVTIIGKLGQKTQYEMRKAEY
Origin Rre-gs68       RVLAKAGMTSVVLPDNNATDTLAMGMYQAWKAFNNEQAIIVTIIGKLGQKTQYEMRKAEY
                      *******:********:**:*************.**:************************

      g32685.t1       KATELSGDKIRTVCMNLTEANEKLTLDDDFNLRLDDQIVAVVNYRLARNIPEKFLTDEKM
Origin Rre-gs68       KATELSGGKIRTVCMNLTEANEKLTLDDNFNLRLDDQIVAVVNYRLARNIPEKFLTDEKI
                      *******.********************:*******************************:

      g32685.t1       EVWTKMEVSTAIKSPPLNYEIACTKKMQQVLAEKDVVEKFFPEPKDAKKVAAIRKFQARM
Origin Rre-gs68       DVWTKMEVSTAIKSPPLNYEIACTKKMQQVLAEKDVVEKFFLEPKDAKKVAAIRKFQARM
                      :************************************* ******************

      g32685.t1       WSLDHNDEKTQAVIQ---------VIK---------------------------------
Origin Rre-gs68       WSLDHNDEKTQAVIQDAIEHPDRYVLKPNKDGGGNNLWEEEMKIKLETLKPEERSQYILM
                      **************         *:*

      g32685.t1       ------------------------------------------------------------
Origin Rre-gs68       QRIRPFVGKNFLKRPLEQARYEDQVVTELSIFGALLGNQENGKILHNKGGGHMMRSKPKH


      g32685.t1       -------------------
Origin Rre-gs68       VNEGGLEMGAGFYDSPLLI
```

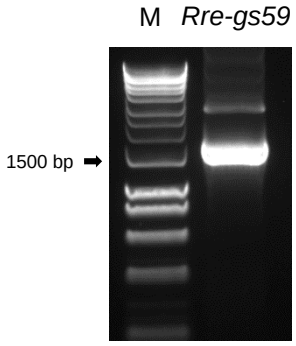
C

>Fixed *Rre-gss68*

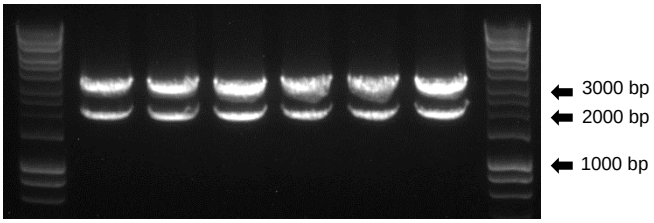
MAILLNICIACCCFYCVFGETSGQQEIDVQVLVEDALDYGHYVGLIHRAK
DHLKSSDLSEVSAMALFPSPFPRQVFEDANNVQEALAELYFRVANDYEFL
MNAYREVRKVDKTVDKLMNLLEDIRKKGIHQPIGLMMMRADYMANMNEQN
SESPYEIKQIEVNIGAVGGATCEKATLVHRRVLAKAGMTSVVLPDNNATD
TLAMGMYQAWKAFNNEQAIIVTIIGKLGQKTQYEMRKAEYKATELSGGKI
RTVCMNLTEANEKLTLDDNFNLRLDDQIVAVVNYRLARNIPEKFLTDEKI
DVWTKMEVSTAIKSPPLNYEIACTKKMQQVLAEKDVVEKFFLEPKDAKKV
AAIRKFQARMWSLDHNDEKTQAVIQDAIEHPDRYVLKPNKDGGGNNLWEE
EMKIKLETLKPEERSQYILMQRIRPFVGKNFLKRPLEQARYEDQVVTELS
IFGALLGNQENGKILHNKGGGHMMRSKPKHVNEGGLEMGAGFYDSPLLI-

Figure 3.2. The process of finding missing 5' or 3' sequence information using *Rre-gs68* as an example. (A) The original sequence of *Rre-gs68*. (B) *Rre-gs68* was fixed using the adjacent 1 kb sequence in the genome. The missing 5' end information was identified in the genome using g32685.t1 as reference. (C) The missing N terminal information (red underlined), including the predicted start codon, was added to the original sequence of *Rre-gs68* to form the new correctly fixed GS gene.

A



B



C

```
g27519.t1       MSKFQPIFVTFIVVLLRCYGDSTPDSDKDAINVSSIEDEVDLKILADDAIDFAQNNGLII
 Rre-GS59       MSKFQPIFVTFIVVLLRCYGDSTPDSDKDAINVSSIEDEVDLKILADDAIDFAQNNGLII
                ************************************************************

g27519.t1       RTNDHPTESDISAFAAFTLFPTQFPRKQFHQAYDVQEAMSLLYFRISRDYDFLIKIASEI
 Rre-GS59       RTNDHPTESDISAFAAFTLFPTQFPRKQFHQAYDVQEAMSLLYFRISRDYDFLVKIASEI
                ************************************************:***********

g27519.t1       TKNDYAVEKMLEIVQTIHEEAKLGKINQPISLVLQRS-----------------------
 Rre-GS59       TKNDYAVEKMLEIVQKIHEEAKLGKINQPISLVLQRSDYMCHMNPKAQGKEDQYQLKQIE
                ***************.********************

g27519.t1       -------------------------------------------GIYLAWQQFKNPNAIV
 Rre-GS59       VNNGPIGAILVERVRKLHQRMLAKANMDGGSMLPENRPFNTIAEGIYLAWQQFKNPNAIV
                                                           ***************

g27519.t1       VTIIGSKRNRFRFEQAQLEYELERISGNKIKNIVYMNMNEAHESLRLAKDNSLMLGDRVV
 Rre-GS59       VTIIGSKRNRFRFEQAQLEYELERISGNKIKNIVYMNMNEAHESLRLAKDNSLMLGDRVV
                ************************************************************

g27519.t1       GVVYFRRGFLIKPHPLADQQFVTRLLIERSTAIKSPTVALELASMKKIQQVLAKPNMVEQ
 Rre-GS59       GVVYFRRGFLIKPHPLADQQFVTRLLIERSTAIKSPTVALELASMKKIQQVLAKPNMVEQ
                ************************************************************

g27519.t1       FFPDPKDADKVAVIRATFANLWGLEKEDEETEAVIQDAIAHPENYVLKPNREGGGHNYWG
 Rre-GS59       FFPDPKDADKVAVIRATFANLWGLEKEDEETEAVIQDAIAHPENYVLKPNREGGGHNYWG
                ************************************************************

g27519.t1       HEISEKLSAFSMTDRKEHILMERLRPFVAQNYPIRAGGDVRLENIVTEFSTYGYLVGNIQ
 Rre-GS59       HEISEKLSAFSMTDRKEHILMERLRPFVAQNYPIRAGGDVRLENIVTEFSTYGYLVGNIQ
                ************************************************************

g27519.t1       DGEVLYNKGHGHLMRTKIESVTEGGILEGSGFYDSPYLID
 Rre-GS59       DGEVLYNKGHGHLMRTKIESVTEGGILEGSGFYDSPYLID
                ****************************************
```

Figure 3.3: Typical procedure for fixing mis-predicted GS genes by cloning using *Rre-GS59* as an example. (A) PCR product amplified from cDNA by gene specific primers. M: DNA ladder. A clear band was shown at 1500 bp position, which was purified and subcloned into pGEM-T vector. (B) Plasmid DNA for 6 independent clones digested with EcoR I to release the insert and confirm correct cloning. (C) Alignment of the original sequence identified from the genome database (g27519.t1) and the sequencing result. The missing information of the partial sequences was fixed based on the sequencing results.

3.4.2 Phylogenetic tree of the *R. reniformis* GS family

First of all, a phylogenetic tree was constructed based on an amino acid alignment of 241 GS-like sequences from 31 nematode species (Figure 3.4). All these 241 nematode GS-like sequences are listed in the Appendix 2.

The single GS-like sequences from each free-living nematode and animal parasitic nematode (black) were limited to a single clade that also contained only one sequence from each plant parasitic nematode species except *M. incognita* (Figure 3.4). The polyploid genome of *M. incognita* and *S. ratti* contributed two genes to this Clade. Furthermore, it is noteworthy that GS-like sequences from the majority of PPNs (red and green) experienced at least two large gene expansions. GS-like genes in the first expansion were only present in the nematodes belonging to the order Tylenchida, and vary in number from 2 (*P. penetrans* and *M. incognita*) to 12 (*R. renifomis*) genes per species. Like animal parasitic and free-living nematodes, *B. xylophilus* in the order Aphelenchida and *L. elongatus* in the order Dorylaimida do not display any expansion of GS-like genes. The second and larger expansion of GS-like genes was present only in the cyst and reniform nematodes, which both induce syncytial feeding sites. In addition, *R. reniformis* (red in Figure 3.4) has the largest number of GS gene family members, which may be due to its large genomic size and polyploidy (Sommer and Streit 2011). Also, *R. reniformis gs* sequences in this clade were generally clustered together in sub-clades within the phylogeny, whereas the sequences from the different cyst nematode species are more evenly dispersed and inter-mixed.

In order to understand the specific evolutionary relationship amongst the *R. reniformis gs* genes, we focused on the *R. reniformis* phylogeny. An amino acid alignment was made between the 73 likely full length, non-redundant GS-like sequences from *R. reniformis* and the single *gs* gene from *C. elegans* (*cel-gss1*). A Bayesian phylogenetic tree was constructed based on this alignment. As shown in Figure 3.5, the *R. reniformis* GS family was clearly divided into three major clades, reflecting the same overall structure as for the larger nematode phylogeny. Clade 1 (red) contained only one GS-like sequence (named *Rre-gs1*) from *R. reniformis*, together with the *C. elegans* GS, while Clade 2 (blue) and Clade 3 (yellow, orange and green) which can be split into three sub-clades represented the first and the second expansions of GS genes in *R. reniformis*, respectively. The location of *R. reniformis* GS1 indicated this gene was the ancestral gene that was the origin of the *R. reniformis* GS family.

3.4.3 Additional sequence analysis of GS-like genes

The presence of a signal peptide for secretion was predicted for each GS-like sequence (Figure 3.6 shows examples of negative and positive signal peptide predictions; Figure 3.7). Interestingly, none of Clade 1 and 2 *gs* genes were predicted to encode a signal peptide for secretion, whereas most of Clade 3 GS-like genes contain a signal peptide at the N-terminal. Given that the presence of a signal peptide is a key feature for nematode effector (Mitchum *et al.* 2013), Clade 3 GS-like genes are likely to encode secreted proteins which play a significant role in nematode parasitism. In addition, we found that Clade 2 GS-like sequences shared a short and somewhat variable C-terminal extension of the approximate consensus sequence P[A|S]SE[F|L][Q|H] with unknown functions yet (Figure 3.8), which were also identified in Clade 2 *gs* sequences from other plant parasitic nematode species.

3.4.4 Expression profiles

To facilitate functional classification of individual GS clades, the heatmap of the GS transcript abundance data from five life-specific stages (egg, J2, J3, adult vermiform and female) was plotted (Figure 3.7). The *R. reniformis* raw RNAseq reads from five life stages: egg, J2, J3, vermiform adult and sedentary female, were downloaded from NCBI under BioProject no. PRJNA286314 (Showmaker *et al.* 2019). As shown in Figure 3.6, both Clade 1 and 2 GS-like genes were highly expressed at the non-parasitic stages (egg, J2, J3 and vermiform adult). Given the fact that *R. reniformis gs1* gene is genetically closest to *cel-gss1* compared to the rest of *R. reniformis* GS-like sequences, *R. reniformis gs1* was considered as a typical housekeeping *gs* gene and played a similar role with *cel-gss1* involved in glutathione biosynthesis in nematode cells. By contrast, most of Clade 3 GS-like genes were significantly up-regulated at the parasitic female stage, indicating Clade 3 GS may play a role in plant nematode parasitism and function as ‘effector’ during plant-nematode interactions. Interestingly, several Clade 3 *gs* genes do not fit the overall trend (*e.g*. Clade 3 sequences that do not have signal peptide or those up-regulated in non-parasitic stages but with a predicted signal peptide). Most of these abnormal Clade 3 *gs* genes are contained in the sub-clade 2 of the tree, suggesting that this sub-clade may share a special function.

3.4.5 Cloning of GS-like genes from *R. reniformis* cDNAs

In order to characterise their enzymatic activity and roles in plant-nematode interactions in more detail, 24 GS-like genes distributed across all three clades, with a range of expression profiles including those Clade 3 genes with and without signal peptides were selected for further study. Primers were designed to amplify the complete predicted coding regions from cDNA of the appropriate life-stage and a number of resulting clones were sequenced for each gene. The cloned genes were *Rre-gs1* representing Clade 1, *Rre-gs2*, *Rre-gs4*, *Rre-gs5*, *Rre-gs11* representing Clade 2, *Rre-gs14*, *Rre-gs18 Rre-gs20*, *Rre-gs23*, *Rre-gs27*, *Rre-gs36*, *Rre-gs44*, *Rre-gs49*, *Rre-gs50*, *Rre-gs51*, *Rre-gs55*, *Rre-gs57*, *Rre-gs59*, *Rre-gs61*, *Rre-gs64*, *Rre-gs65*, *Rre-gs66*, *Rre-gs67* and *Rre-gs72* representing Clade 3. All these sequences are listed in Supplementary figure 2.

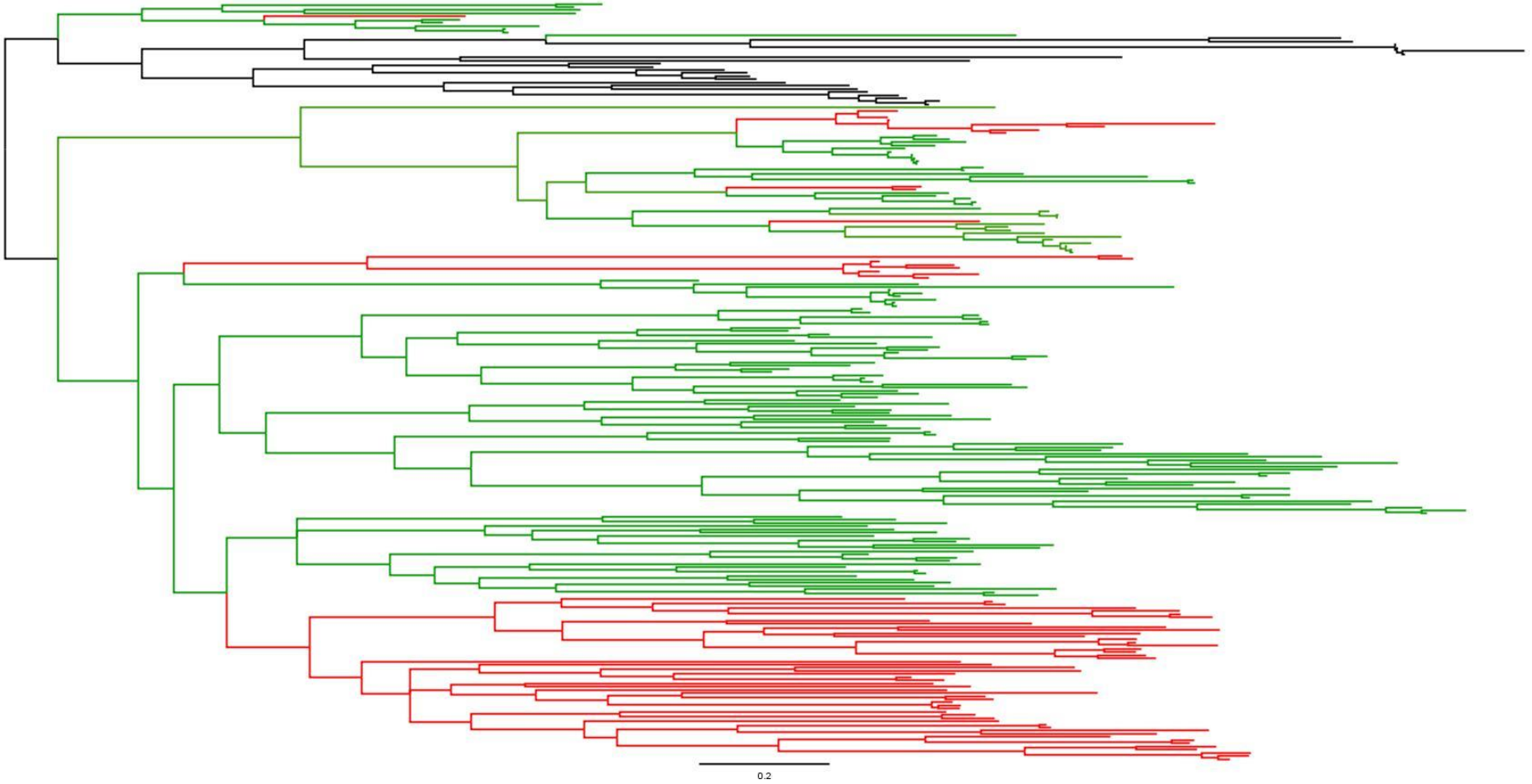


Figure 3.4: Phylogenetic tree to understand nematode GS evolution. A Bayesian phylogenetic tree based on a GS protein alignment from 31 nematode species, where free-living nematodes and animal parasitic nematodes are in black, reniform nematodes are in red and other plant parasitic nematodes are in green. Branch line width is scaled by support.

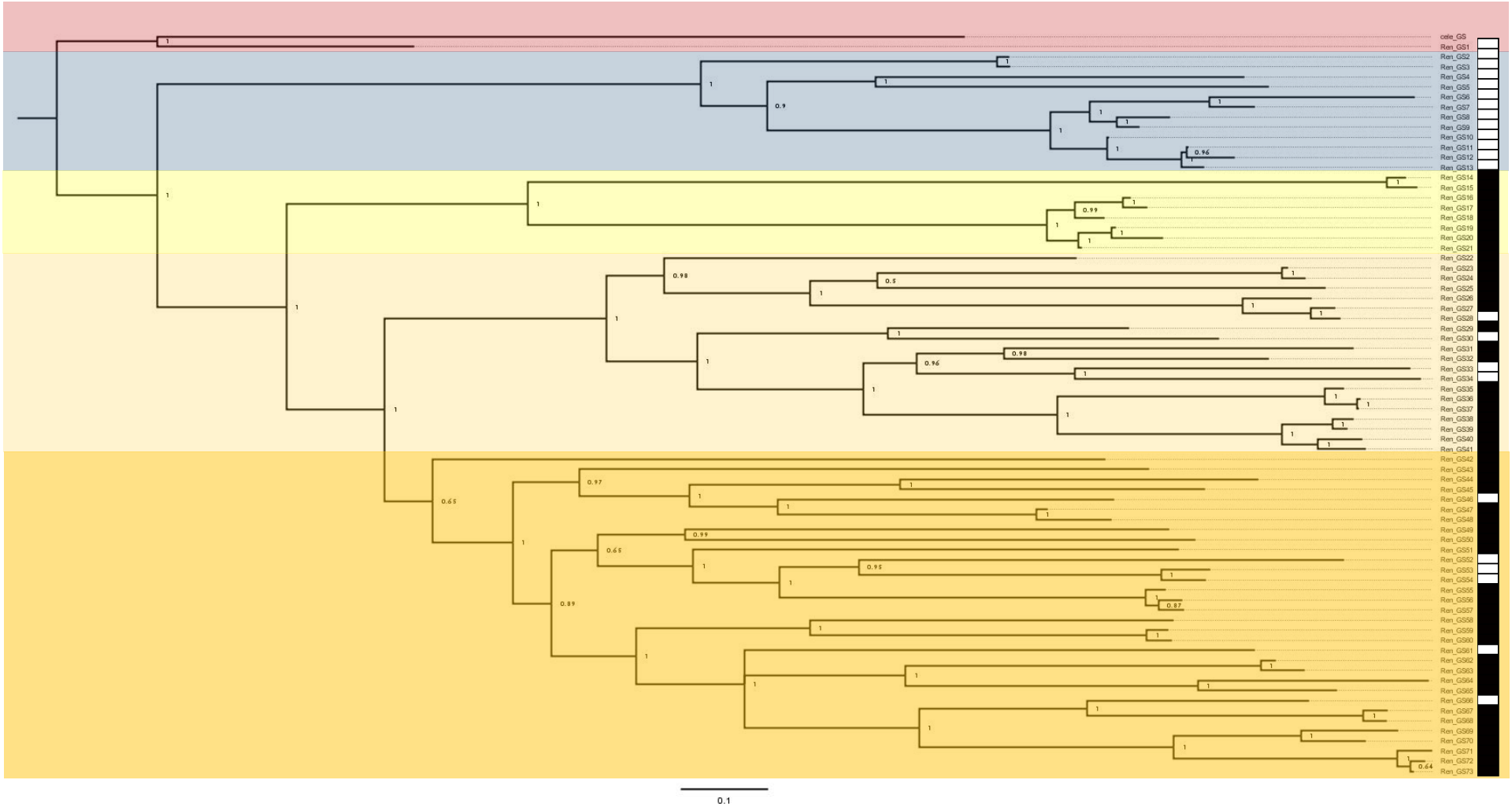


Figure 3.5: Phylogenetic tree of GS genes of *R. reniformis*. A Bayesian phylogenetic tree was made using a protein alignment of 73 GS sequences from *R. reniformis* and the single GS gene from *C. elegans*. Bootstrap support values for 1 million iterations are shown as node labels. Branch line width is scaled by support. These 74 sequences are broadly divided into three major clades. Red: Clade 1; Blue: Clade 2; Yellow: Clade 3, Subclade 1; Light orange: Clade 3, Subclade 2; Dark orange: Clade 3, Subclade 3. The bars alongside the phylogeny represent the prediction of signal peptide. Black bars indicate the presence of a signal peptide for secretion within a particular sequence while white ones represent the absence of a signal peptide within that sequence.

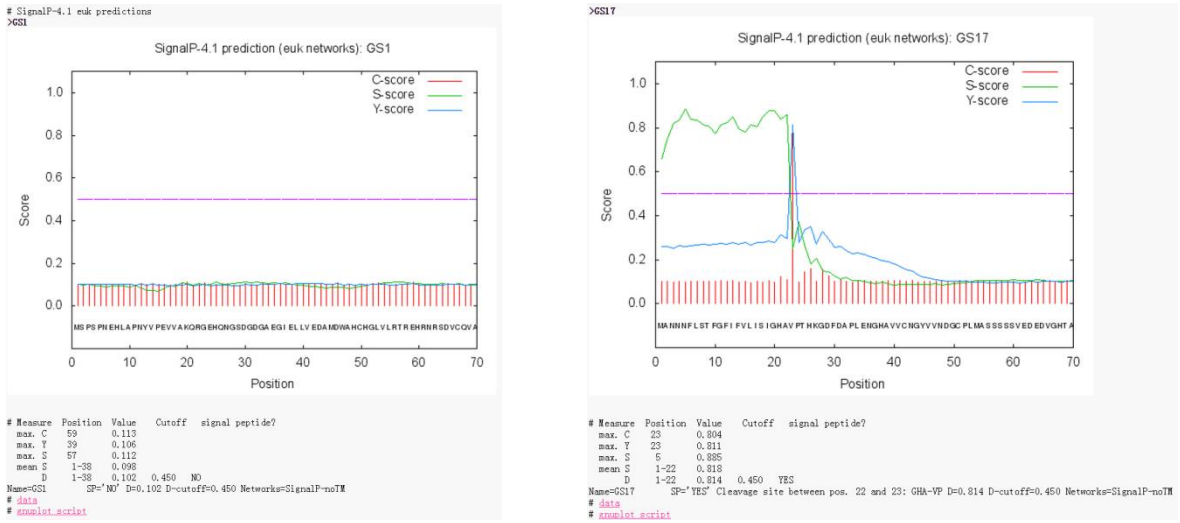


Figure 3.6 Example of the negative and positive outputs for signal peptide prediction using SignalP V4.1. The probability that the provided protein sequence contains a signal peptide and the position of the predicted signal peptide are indicated. In this case, the sequence shown on the left, *R. reniformis* GS1, does not have a signal peptide while that on the right, *R. reniformis* GS17 has a signal peptide at the amino acid positions 1 to 22.

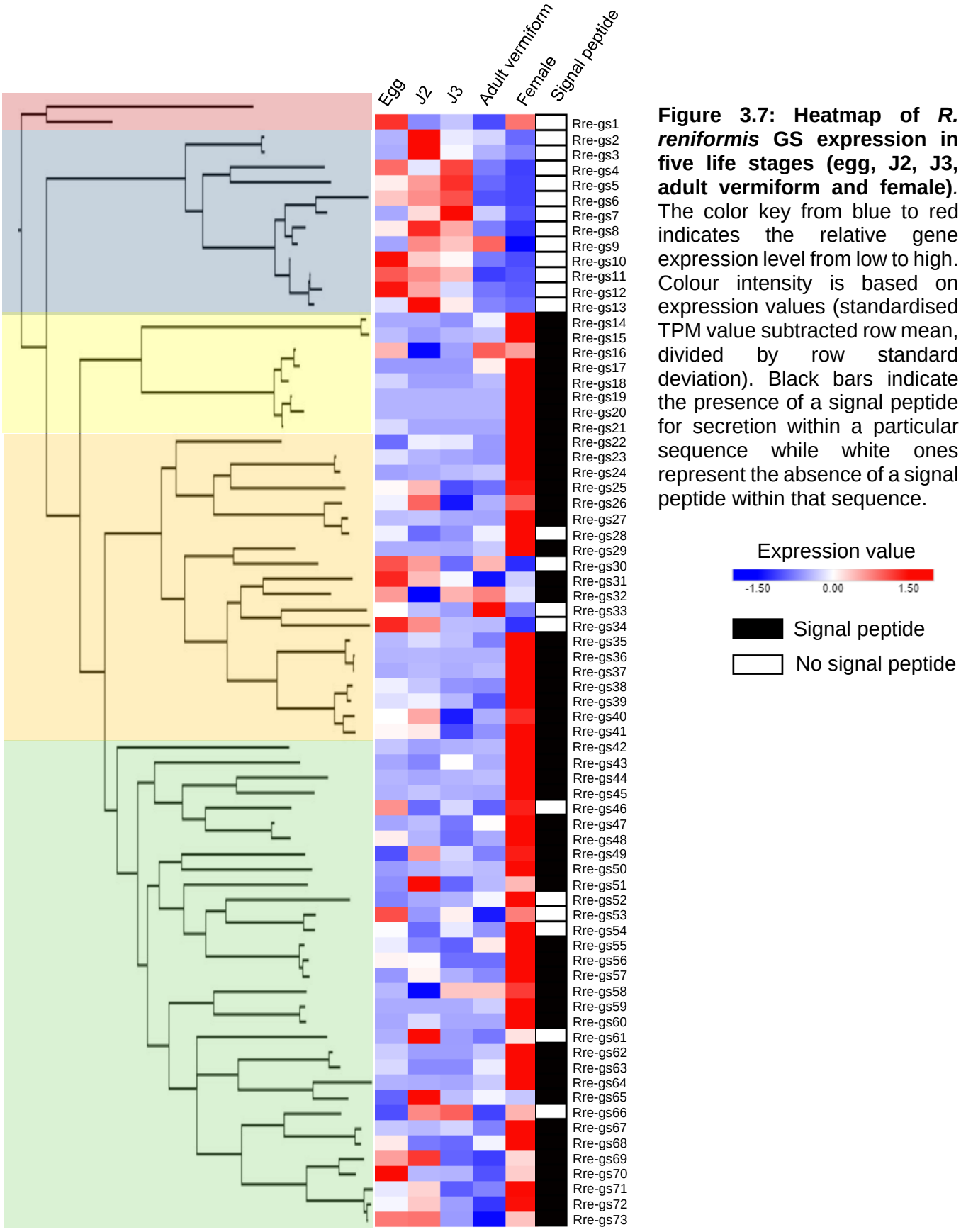
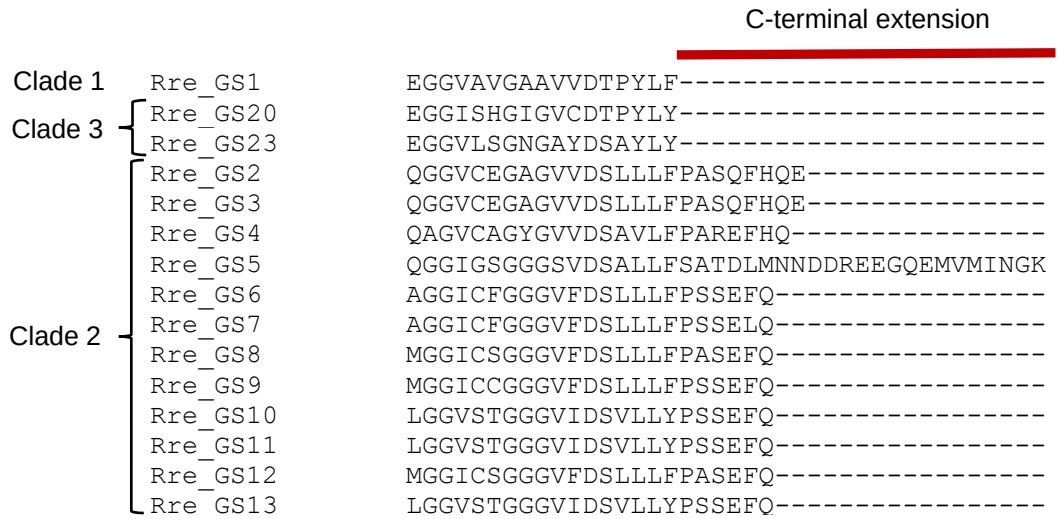


Figure 3.7: Heatmap of *R. reniformis* GS expression in five life stages (egg, J2, J3, adult vermiform and female). The color key from blue to red indicates the relative gene expression level from low to high. Colour intensity is based on expression values (standardised TPM value subtracted row mean, divided by row standard deviation). Black bars indicate the presence of a signal peptide for secretion within a particular sequence while white ones represent the absence of a signal peptide within that sequence.

A



B

Figure 3.8 C-terminal extension associated with *R. reniformis* Clade 2 GS that is absent from all Clade 1 and Clade 3 GS. (A) Alignment of a short C-terminal extension associated with Clade 2 GS sequences. * indicates consensus residues. (B) The consensus sequence of the six amino acids C-terminal extension of unknown significance.

3.4.6 Spatial expression of *R. reniformis* GS-like genes within nematode

As shown in Figure 3.6, a large number of the *R. reniformis* GS-like genes in Clade 3 encode a protein with a N-terminal signal peptide, whereas all genes in Clade 1 and 2 do not. In order to support the hypothesis that GS-like genes in Clade 3 may be considered as “effectors”, *in situ* hybridization was carried out to indicate the spatial expression of GS-like genes of different life stage within nematodes. A range of GS-like genes were selected representing those that were expressed at either the J2 or female stage to include members from each clade for the *in situ* hybridisation assay.

Complementary and non-complementary DNA probes were made by asymmetric PCR using reverse and forward primers respectively. Figure 3.9A shows agarose gel electrophoresis of two ~200 bp probes used in *in situ* hybridization. When an aliquot of each probe was run alongside an aliquot of the corresponding unlabelled template DNA on agarose gel, an increase in molecular mass of the labelled product was observed due to successful incorporation of Digoxigenin (DIG).

Using a complementary DIG-labelled DNA probe for *in situ* hybridisation, the transcripts of *Rre-gs1* which comes from Clade 1 as well as *Rre-gs2*, *Rre-gs4* and *Rre-gs11* which come from Clade 2 were localised in the intestine of the non-parasitic J2 nematodes (Figure 3.8B-E). None of these genes are predicted to encode a signal peptide. On the other hand, the transcripts of *Rre-gs14*, *Rre-gs23*, *Rre-gs36*, *Rre-gs49*, *Rre-gs55* and *Rre-gs67* which come from Clade 3 and do encode a GS with a signal peptide, are expressed specifically in the single large secretory pharyngeal gland cell of the adult female (Figure 3.9). No such staining patterns were observed with the non-complementary sense probes used as negative controls (Figure 3.9F-I, Figure 3.10G-L). This indicates that the proteins encoded by *Rre-gs14*, *Rre-gs23*, *Rre-gs36*, *Rre-gs49*, *Rre-gs55* and *Rre-gs67* are likely to be secreted *in planta* during nematode parasitism.

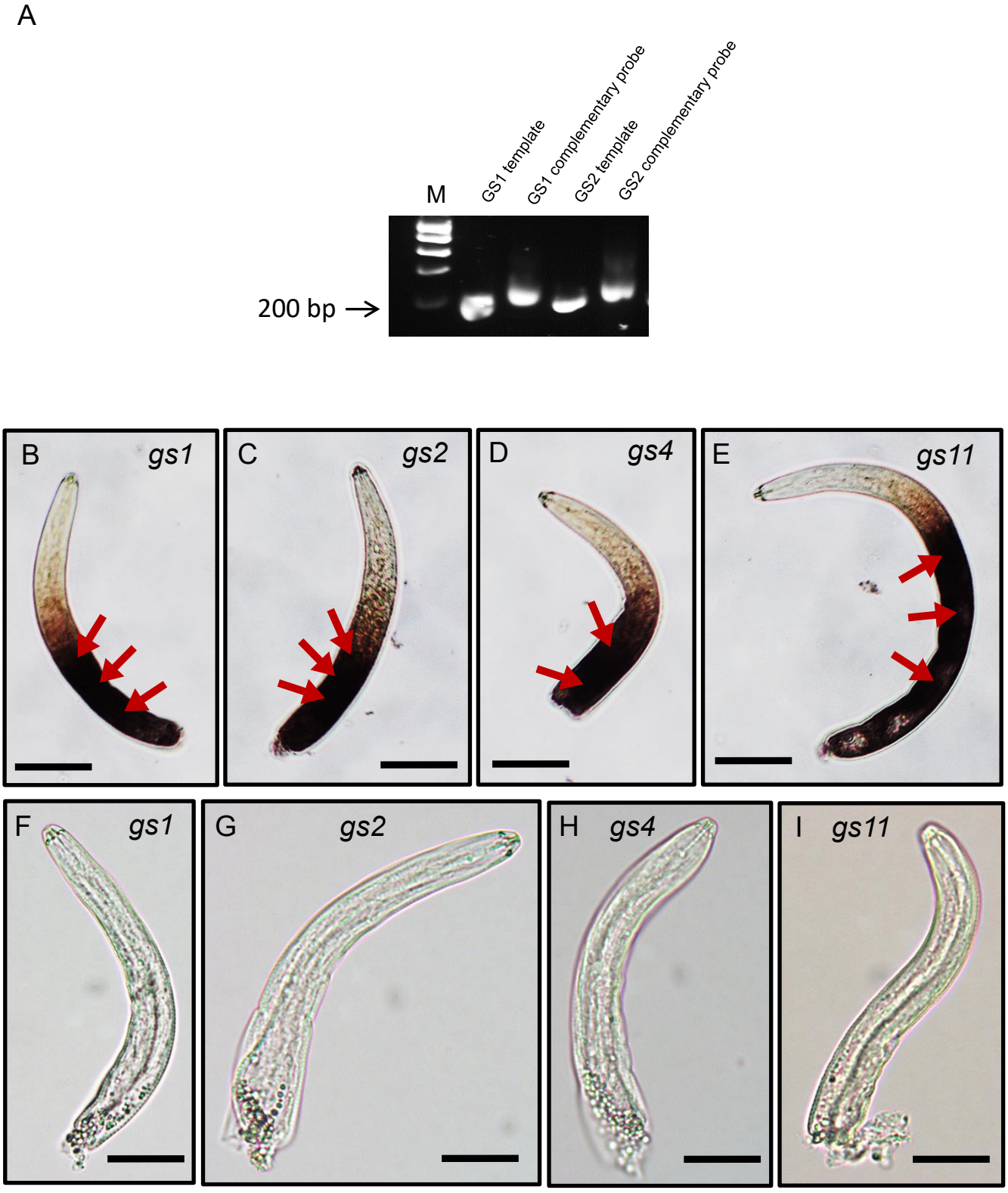


Figure 3.9: *In situ* hybridization of *R. reniformis* GS-like gene members of Clade 1 & 2 in J2s. (A) Example of probes used for *in situ* hybridisation. M: DNA ladder. (B) - (E) *Rre-gs1* (Clade 1), *Rre-gs2*, *Rre-gs4* and *Rre-gs11* (all are Clade 2) are expressed in the intestine of J2 nematodes (red arrowheads). Dark staining represents where the genes are expressed. (F) - (I) No such staining patterns were seen with negative control sense probes. Scale bars = 50 µm.

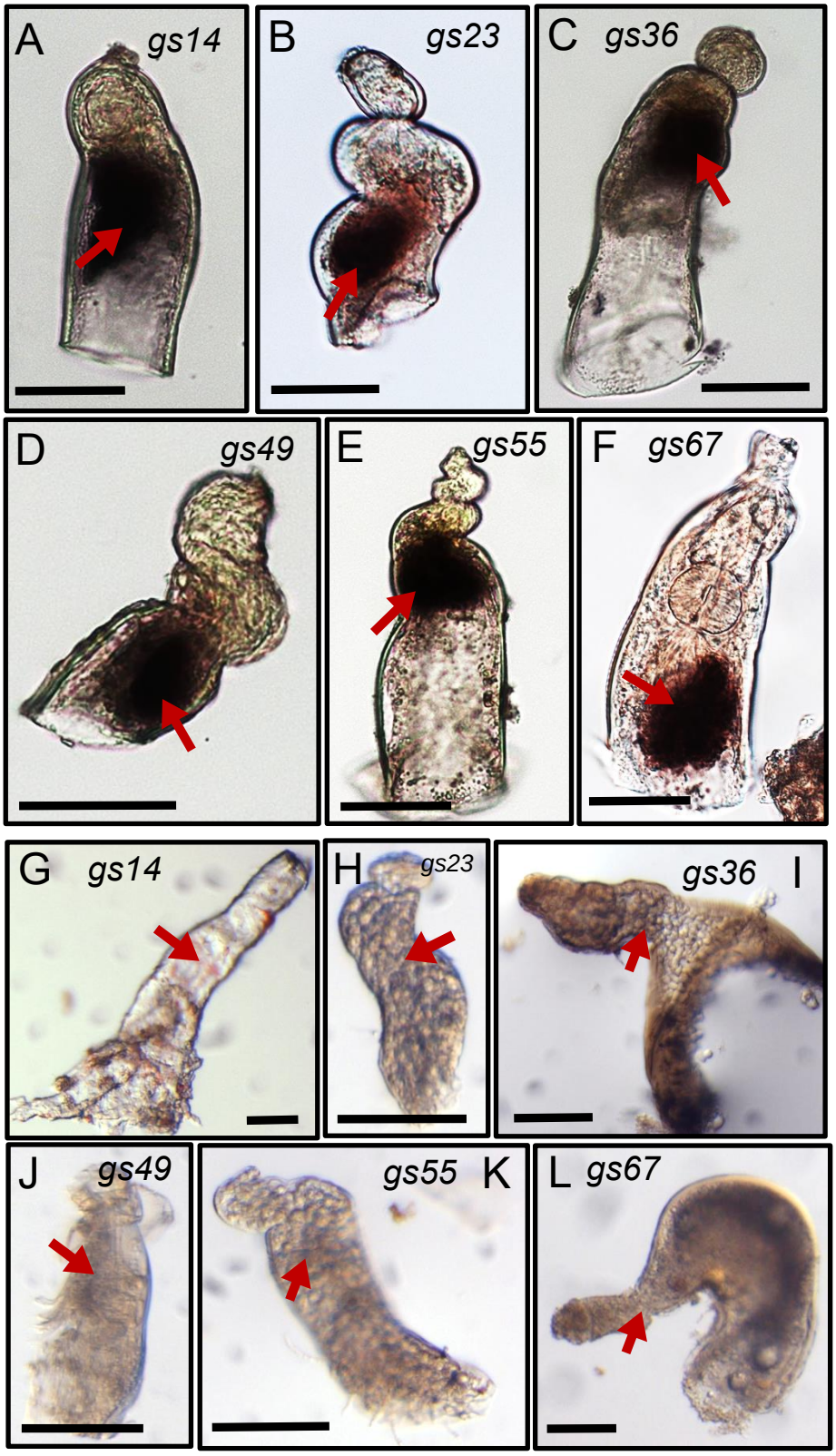


Figure 3.10: *In situ* hybridization of *R. reniformis* GS-like gene members of Clade 3 in parasitic females. Dark staining represents where the genes are expressed. (A) - (F) *Rre-gs14*, *Rre-gs23*, *Rre-gs36*, *Rre-gs49*, *Rre-gs55* and *Rre-gs67* from Clade 3 of the GS phylogeny are expressed in the pharyngeal gland cell (red arrowhead). (G) – (L) No such staining patterns (red arrowhead) were seen with the negative control sense probes. Scale bars = 50 µm.

3.5 Discussion

3.5.1 A large group of GS-like genes were identified from *R. reniformis*

Sequencing the genome of *R. reniformis* represents a key step in identifying genes underlying the plant-nematode interaction and for studying the evolution of parasitism. In this study, the draft genome assembly (Showmaker *et al.* 2019) together with five life stage-specific transcriptome assemblies including parasitic stage and non-parasitic stages (Showmaker *et al.* 2018) of *R. reniformis* were exploited to identify and classify GS-like sequences. From the genome and transcriptome resources, a total of 260 GS-like sequences including 189 sequences from the genome and 71 sequences from the transcriptome assemblies were identified using the GS domain BLAST and *G. pallida* GS gene similarity search, followed by a manual refinement of the list of GS-like genes. In this way, a large number of GS-like sequences were grouped into the *R. reniformis* GS-like gene list. All the obviously truncated sequences had to be removed from the final list although some of them may be genuine GS-like genes that have arisen during genome expansion and likely lost their function, since lots of these GS-like sequences (over 100 members) looked partial and incomplete and could not be fixed.

The discovery of such a large number of GS-like sequences in a plant parasitic nematode is unprecedented even when compared to the around 50 GS-like sequences in *G. pallida* (Cotton et al. 2014). Despite a lot of GS-like sequences identified in *R. reniformis*, most of them were shown as obviously incomplete sequences which lack necessary information and cannot be refined and fixed. This may be due to the poor quality of the genome assembly of *R. reniformis* as a result of unresolved haplotypes stemming from heterogeneity within the *R. reniformis* population used for DNA extraction (Leach, Agudelo and Lawton-Rauh 2012). Therefore, one of the methods to improve *R. reniformis* genome quality is to utilise more inbred population. In addition, some genome assemblies have to be artificially large because of the difficulty of assembling repetitive sequences (Kikuchi, Eves-van den Akker and Jones 2017). Often, genomic repeats and transposons are associated with gene duplication events and gene family expansions. They may similarly be involved in the unprecedented expansion of GS-like sequences identified in *R. reniformis*. Therefore, by the nature of how they are formed, they may be difficult to assemble properly. Long read genome assembly can span stretches of repetitive regions and thus produce a

more contiguous reconstruction of the genome (Jung *et al.* 2019). Currently, the two most important third-generation DNA sequencing technologies, Pacific Biosciences and Oxford Nanopore, are able to produce long reads with average fragment lengths of over 10,000 base-pairs that can be advantageously used to improve the genome assembly (Del Angel *et al.* 2018).

Assembly refinements of *M. incognita* polyploid genome can be considered as a good example for the improvement of a highly polymorphic but relatively fragmented genome assembly. Root-knot nematodes have very complex origins involving the mixing of several parental genomes by hybridisation. *M. incognita* was first sequenced to acquire a 86 Mb genome assembly with 19, 212 predicted genes (Abad et al. 2008). A recent re-sequencing of *M. incognita* reported a 184 Mb genome assembly with 45, 351 predicted genes (Blanc-Mathieu *et al.* 2017). One possible reason of a better genome assembly may be due to the multi-pass assembler MIRA to generate contigs from the 454 genomic libraries as it is optimised for highly heterozygous genomes (Chevreux 2007). Moreover, Sanger reads of the *M. incognita* first draft genome sequence (Abad et al. 2008) were used to separate a maximum of repeats and heterozygous regions. Also, Illumina data was used to correct the homopolymer errors of the 454 contigs.

After removing those partial sequences, a total of 73 GS-like sequences remained in the final *R. reniformis* GS family. Some of the partial GS-like sequences are probably real GS members, however, without better genomic resources it is very difficult to find them back. Considering *G. pallida* contains around 50 GS-like genes (Lilley *et al.* 2018) and a 124 Mb genome (Cotton et al. 2014), the actual number of *R. reniformis* GS-like genes is predicted to be around 100 based on the 314 Mb genome size estimated by flow cytometric analysis (Nyaku *et al.* 2014). Furthermore, one possible way to improve the GS annotation is to exploit the latest transcriptomic resources. Initial *gs* gene identifications at transcript levels were performed using J2 and J4 female transcriptomic databases (Eves-Van Den Akker *et al.* 2016b). Although most GS-like genes that we know of already are expressed at the sedentary female stage, that may not be the case for all GS-like genes. A very recent transcriptomic resource (Showmaker *et al.* 2018) with five life-specific stages may provide more useful information on the GS annotation.

As introduced in Chapter 1, most eukaryotic organisms own only a single gene coding GS. However, it is not rare that multiple *gs* genes were revealed in the genome sequences. For example, three rice GS homologs were isolated and all of them showed typical GS activity with one of them able to catalyse the synthesis of hydroxymethyl-glutathione from γ-EC L-serine in an ATP-dependent manner (Yamazaki, Ochiai and Matoh 2019). It is hypothesised that the rice genome which has experienced large scale genome duplications was responsive for an expansion of *gs* genes (Yamazaki, Ochiai and Matoh 2019). Similarly, *R. reniformis* also witnessed an unexpectedly larger *gs* gene expansion from the genome and transcriptome datasets. Phylogenetic analyses showed that nematodes have evolved to be parasites on up to 18 separate occasions in their evolutionary history (Blaxter *et al.* 1998), suggesting GS family expansions may occur during multiple gene duplication events.

3.5.2 Three major Clades shown in the GS family

The discovery in the plant parasitic nematode *R. reniformis* of a large expansion of *gs* genes leads us to explore their evolutionary relationship. The overall phylogeny of all nematode *gs* genes divides the GS family into three major clades. The Clade 1 represents the ancestral GS clade which contained only one sequence from each investiged nematode except *M. incognita* and *S. ratti* due to their polyploid genome. In addition, the Clade 1 *gs* sequences from plant parasitic nematodes except *L. elongates* are limited into a sub-clade of Clade 1 while those from animal parasitic nematodes and free-living nematodes are limited into another sub-clade, indicating that Clade 1 *gs* genes from plant parasitic nematodes and non-plant parasitic nematodes appear to have evolved independently.

Interestingly, two gene expansions were shown to only present in PPN, which were represented by Clade 2 and Clade 3, respectively (Figure 3.3). However, these gene expansions in plant parasitic nematode species exclude sequences from the migratory ectoparasite *L. elongatus* and the non Tylenchid migratory endoparasite *B. xylophilus*. Furthermore, Clade 3 represents a larger expansion but contains a narrower species (only present in syncytia-forming cyst and reniform nematodes), indicating Clade 3 GS may be involved in formation of syncytia. Generally, the sequence identity between Clade 2 and Clade 3 is around 30%-40%. However, the sequences in Clade 3 share around 35% identity, which are much lower than those of Clade 2 (~57%). In addition, unlike *R. reniformis* Clade 2 *gs* sequences which distribute within this clade evenly,

Clade 3 *gs* sequences were grouped into a few subclades. Taken together, these results suggest Clade 3 *gs* sequences have undergone more duplication and diversification events during evolutionary history.

3.5.3 Diverse sub-functionalisation within the large *R. reniformis* GS gene family

To understand functional classification of individual *R. reniformis* GS clades, the transcript abundance data from five life-specific stages (egg, J2, J3, adult vermiform and parasitic female) and canonical signal peptide prediction were exploited (Figure3.6). Some *gs* genes in the Clade 1 such as *C. elegans* GS have been well-studied (Buzie and Enjuakwei 2007). They are typical GS that can catalyse the addition of glycine to γ-EC and are considered as a 'housekeeping' gene. Consistent with the *C. elegans gs* gene, the Clade 1 *R. reniformis gs* has no predicted signal peptide for secretion and has a relatively stable expression level in all the life stages, which supports the assumption that *R. reniformis* GS1 functions as a typical GS enzyme in nematode.

All Clade 2 *gs* genes lack a signal peptide for secretion and have a higher expression level in the non-parasitic stages (from egg to vermiform adult), suggesting Clade 2 GS are not secreted proteins and not involved in nematode parasitism. Considering Clade 2 GS-like genes had a high absolute expression level at parasitic female stage, even though the relative expression was higher in the non-parasitic stages, these non-secreted GS may be needed in many nematode tissues rather than the single gland cell and are likely to play their roles in nematode development.

In the Clade 3, the presence of a signal peptide for secretion was indicated to be strongly correlated with the corresponding gene being up-regulated in the parasitic female stage, indicating that these genes function during the parasitism process and may be considered as 'effector GS'. Interestingly, a few GS-like genes in Clade 3 do not encode a protein with a signal peptide for secretion but are highly expressed at parasitic female stage. Previous reports introduced signal peptide is important for effector but not always necessary. Several effector candidates released from nematode stylets without a signal peptide have been reported (Bellafiore *et al.* 2008). To investigate this, *Rre-gs44* which is highly expressed at the parasitic female stage but lacks a signal peptide has also been tested in *in situ* hybridisation assay. However,

no staining was observed in the adult females and so this hypothesis would need to be tested more rigorously across Clade 3.

In addition, there are some Clade 3 GS-like genes highly expressed in the non-parasitic stages but possess signal peptide for secretion. Similar phenomenon has been described before. For example, two glutathione peroxidases were identified from *G. rostochiensis* (Jones *et al.* 2004). One protein has a signal peptide for secretion while the other is predicted to be intracellular. Both genes are expressed in all parasite stages tested and the secreted one was shown to function at the surface of nematodes (Jones *et al.* 2004). Given that nematodes are exposed to the hostile environment all the time rather than only at the parasitic stage, some secreted proteins that protect nematodes themselves are likely to not be restricted at the parasitic stage but also function at the non-parasitic stage.

Given the hypothesis that the functions of GS gene family were diversified by clades, the expressional locations within nematodes were further examined. Ten GS genes, with representatives from each of the three clades, were analysed by *in situ* hybridisation to elucidate their spatial expression. In this study, *Rre-gs1* from Clade 1 together with *Rre-gs2*, *Rre-gs4* and *Rre-gs11* from Clade 2 were shown to be expressed within the intestine at the J2 stage, which was consistent with our assumption that Clade 1 and 2 GS function intracellularly. By contrast, *Rre-gs14*, *Rre-gs23*, *Rre-gs36*, *Rre-gs49*, *Rre-gs55* and *Rre-gs67* from Clade 3 were found to be expressed in the secretory pharyngeal gland cell of the parasitic stage, which a common site of effector production (Davis, Hussey and Baum 2004). In conclusion, these *in situ* hybridisation results, together with the signal peptide prediction and temporal expression data strengthened our hypothesis of sub-functionalisation within *R. reniformis* GS family and Clade 3 GS-like genes may play a role as nematode effectors during biotrophic interactions.

3.6 Summary

1. A large group of glutathione synthetase genes were identified from the *R. reniformis* genome and transcriptome assemblies using bioinformatic approaches.
2. The *R. reniformis* GS-like gene family was divided into three major clades. Clade 1 had only one sequence and Clade 2 and 3 represented two large gene family expansions.
3. GS-like genes in Clade 1 and 2 are expressed more highly in the non-parasitic stages and do not encode a signal peptide for secretion, whereas most of the GS genes in Clade 3 are expressed more highly in the parasitic stage and encode a signal peptide.
4. *In situ* hybridisation revealed that Clade 1 and 2 GSs are expressed in the nematode intestine or their whole body, and Clade 3 GSs are expressed specifically in gland cells and are predicted to be ‘effector’ GS.

Chapter 4

Biochemical characterisation of *R. reniformis* glutathione synthetases

4 Biochemical characterisation of *R. reniformis* glutathione synthetases

4.1 Introduction

4.1.1 Biochemical characterisation of GS from other species

Glutathione is present in the majority of living cells and is also the most abundant intracellular thiol. Glutathione synthetase is a key enzyme in the second step of the glutathione biosynthesis pathway. It catalyses the addition of glycine to gamma-glutamylcysteine, to produce glutathione (Meister 1983). GS have been found in a large number of species including Gram-negative bacteria, yeast, mammals, and plants. Despite catalysing the same reaction, prokaryotic and eukaryotic GS genes shared very low sequence similarity (Copley and Dhillon 2002). Both types are, however, members of the ATP-grasp fold superfamily. The biochemical characteristics of several representatives in the overall GS family are introduced below.

4.1.1.1 Prokaryotic GS

Escherichia coli GS is the most well characterised representative of the prokaryotic GS family, which act as homotetrameric enzymes. Previous studies showed the glutathione synthetic activity of *E. coli* GS to be 15-650 pmol min^{-1} μg^{-1} (Watanabe *et al.* 1986). The K_m values of *E. coli* GS for γ-EC, glycine, and ATP were 0.24 mM, 0.91 mM, and 240 μM, respectively (Tanaka *et al.* 1992). It has been reported that glutathione disulfide (GSSG) is an inhibitor of *E. coli* GS, whereas GSH is almost ineffective (Gushima *et al.* 1983). In addition, expression of both *E. coli* GS in the yeast *Saccharomyces cerevisiae* resulted in an unchanged glutathione level. However, co-expression of *E. coli* GCL and GS in *S. cerevisiae* caused a significant increase in glutathione content (Ohtake *et al.* 1989). Taken together, these results supported it is GCL rather than GS that is the rate-limiting enzyme in glutathione biosynthesis. Furthermore, Lys18, Arg86, Asn283, Ser286, Thr288 and Glu292 of *E. coli* GS were shown by X-ray crystallography and affinity labelling studies to be key residues in binding of the γ-EC substrate (Yamaguchi *et al.* 1993; Hibi *et al.* 1993). Site-directed mutagenesis of these residues and kinetic measurements of the mutant enzymes were applied to analyse their roles in γ-EC binding (Hara *et al.* 1995). This study indicated that Arg86 was not only critical for γ-EC binding but also had a role in maintaining the structural integrity of the enzyme.

4.1.1.2 Eukaryotic GS

Currently, the human form of the enzyme, which has only 10% sequence identity with *E. coli* GS, is the most studied eukaryotic GS. Kinetic study of human GS revealed K_m values for γ-glutamyl-aminobutyrate (a non-thiol analog of γ-EC), ATP, and glycine were 0.65 mM, 220 μM and 1.34 mM, respectively (Njalsson *et al.* 2001). In addition, human GS was shown to be an allosteric enzyme and exhibited an unusual kinetic behaviour for the binding of γ-EC substrate. Within hyperbolic saturation of ATP and glycine, the K_m value for γ-glutamyl-aminobutyrate became much lower (164 μM) (Njalsson *et al.* 2001). Taken together, these results suggested that there is a close catalytic dependence between the two substrates of the enzyme, generating a negative cooperativity for binding of γ-EC substrate. In this type of allosteric regulation, the binding of γ-EC at one active site significantly reduces substrate affinity at another active site (Ingle 2015). As a result, human GS was considered as an ideal model for exploring the role of protein-protein interactions in allosteric communications as the obligate homodimer (Ingle 2015). Val44, Val45 and Asp458 were shown to play a role in modulation of this allosteric communication, and are the only three residues known to modulate allostery in GS to date (Ingle *et al.* 2019). All the mutations in these residues led to reduced enzyme activity, decreased γ-EC binding cooperativity, and lower thermal stability (Slavens *et al.* 2011; Brown *et al.* 2011). Furthermore, a variety of structural alignment methods were exploited and four highly conserved residues of human GS (Glu-144, Asn-146, Lys-305, and Lys-364) were identified as the ATP binding sites (Polekhina *et al.* 1999). Experimental and computational site-directed mutagenesis revealed that residue mutations showed no major changes to overall enzyme structure. However, the ligand binding was significantly affected by these mutations, suggesting that these residues played an essential role in GS enzyme activity (Dinescu *et al.* 2004).

Plant GS share around 30-40% sequence identity with human GS and kinetic parameters of GS from some plant species have been described (Frendo *et al.* 2001; Jez and Cahoon 2004; Yamazaki, Ochiai and Matoh 2019). Taking *A. thaliana* GS as an example, the enzyme rate reached around 7910 pmol min^{-1} μg^{-1}, which is a little higher than human GS but much higher than reported prokaryotic GS. The K_m values of *A. thaliana* GS for γ-EC, glycine, and ATP were 39 μM, 1.51 mM and 57 μM, respectively, which were similar to those of other eukaryotic GS (Jez and Cahoon

2004). Compared with the *E. coli* GS (Tanaka *et al.* 1992), the K_m values of *A. thaliana* GS for both γ-EC and ATP were 10-fold lower. The kinetic mechanism of *A. thaliana* GS was investigated using initial velocity analysis and product inhibition, suggesting the equation for a random Ter-reactant model was best fit to the observed data. In this equation, the binding of either ATP or γ-EC increased the binding affinity of the other substrate to *A. thaliana* GS. Likewise, the binding of ATP or glycine increased binding affinity for the other ligand. In contrast, binding of either glycine or γ-EC resulted in a reduced binding affinity for the second molecule. Taken together, this model indicated that binding of either ATP or γ-EC was preferred first followed by addition of glycine (Jez and Cahoon 2004). Moreover, based on the crystal structures of the human and yeast GS, twelve amino acid residues involved in binding of γ-EC and ATP were determined. Site-directed mutagenesis of these residues was performed to examine the effect of these mutations on steady-state kinetics, ATP binding, pH-dependence of catalysis, and solvent kinetic isotope effects. Numerous important effects associated with particular residues were identified in this study. For example, mutation of Arg-132 and Arg-454, which are positioned at the interface of the two substrate-binding sites, affected the enzymatic activity (Herrera *et al.* 2007).

Several GS from nematodes, including the model nematode *C. elegans,* have been studied. The enzyme activity of *C. elegans* GS was determined to be around 1860 pmol min^{-1} μg^{-1} at an optimum pH of 7.0, which is three times lower than that of human GS. The lower activity of *C. elegans* GS might be due to the substitution of the bulky valine residue for Ala386 (Njalsson *et al.* 2001). In addition, K_m values for γ-EC, ATP and glycine were calculated to be around 196 μM, 250 μM and 2.04 mM, respectively, which were in close range to those reported for GS from other eukaryotes (Njalsson *et al.* 2000; Meierjohann, Walter and Muller 2002). Furthermore, many GS-like genes have been identified from some plant parasitic nematodes, such as *G. pallida* (Cotton *et al.* 2014). In a parallel study to this work, different members of the *G. pallida* GS family exhibited very different biochemical characteristics associated with the different family clades (Lilley *et al.* 2018). However, the mechanism of these distinctive biochemical characteristics still remains unclear.

In conclusion, although both prokaryotic and eukaryotic GS share the same functions, they have to be grouped into separate subfamilies due to their low sequencel identity.

In general, the enzyme rates of eukaryotic GS are higher than those of prokaryotic members.

Table 4.1 Comparison of kinetic parameters of GS enzyme representatives

	V_{max} (pmol min^{-1} μg^{-1})	K_m [γ-EC] (mM)	K_m [ATP] (μM)	K_m [glycine] (mM)
C. elegans	1860	0.196	250	2.04
H. sapiens	6010	0.65	220	1.34
P. falciparum	5240	0.107	59	5.04
R. norvegicus	11300	0.042	37	0.913
A. thaliana	7910	0.039	57	1.51
E. coli	15-650	0.24	240	0.91

4.2 Aims

1. To determine the catalytic activities of *R. reniformis* GS-like enzymes.

2. To reveal the kinetic mechanism of *R. reniformis* GS-like enzymes.

3. To discover conserved and functionally important residues.

4.3 Materials and methods

4.3.1 Enzymology of *R. reniformis* glutathione synthetases

4.3.1.1 Constructs for protein expression in *E.coli*

All *R. reniformis* GS-like coding regions of interest and the *Arabidopsis* GS (At5g27380) were cloned (without their predicted signal peptide if appropriate) into the pOPIN S3C vector (Bird 2011). The target protein was expressed with a HIS tag, a chaperone, and a 3C protease cleavage site as an N-terminal fusion to the protein of interest (HIS6-SUMO-3C-POI). Vector constructions were based on the In-Fusion HD Cloning Kit (ClonTech, UK) according to the In-Fusion HD Cloning Kit User Manual. A typical cloning procedure involved the following:

1) Around 1 µg linearized pOPIN S3C vector was generated by Kpn I and Hind III digestion at 37 °C for 3 hours and purified from the gel.

2) The target fragment was amplified by PCR from the existing p-GEM clone using gene specific primers with a 15 bp extension homologous to the vector ends. Primers used in pOPIN S3C vector constructions are shown in Table 4.2.

3) After isolation and purification of the amplified target fragment from the gel, the In-Fusion Cloning reaction was set up in a 5 µl reaction system containing 1 µl 5× In-Fusion HD Enzyme Premix, 50-100 ng linearized vector and appropriate volume of target fragment to make the vector: insert molar ratios be at around 1:3.

4) The reaction was incubated at 50 °C for 15 min and then placed on ice.

5) The reaction was then transformed into *E. coli* strain SHuffle (Lobstein *et al.* 2012), followed by incubation on the pre-warmed LB plates with 50 µg/ml Ampicillin at 30 °C overnight. The method for the preparation of competent SHuffle strain was similar with the general method for the normal *E. coli* DH5α competent strain with minor modification. The growth temperature of SHuffle strain was 30 °C rather than 37 °C.

6) The plasmid DNA was extracted from *E. coli* and was then sent to sequence with the vector primers.

Pop-detect-F: TAG CCT GCG CTT TCT GTA TGA

Pop-detect-R: CAA GGG GCT TCA TGA TGT CC

7) The sequencing results were aligned with the original predicted gene sequences using Muscle to check that the N-terminal fusion was in frame and no errors had been introduced during amplification. The correct plasmids were then stored at -20 °C for further studies.

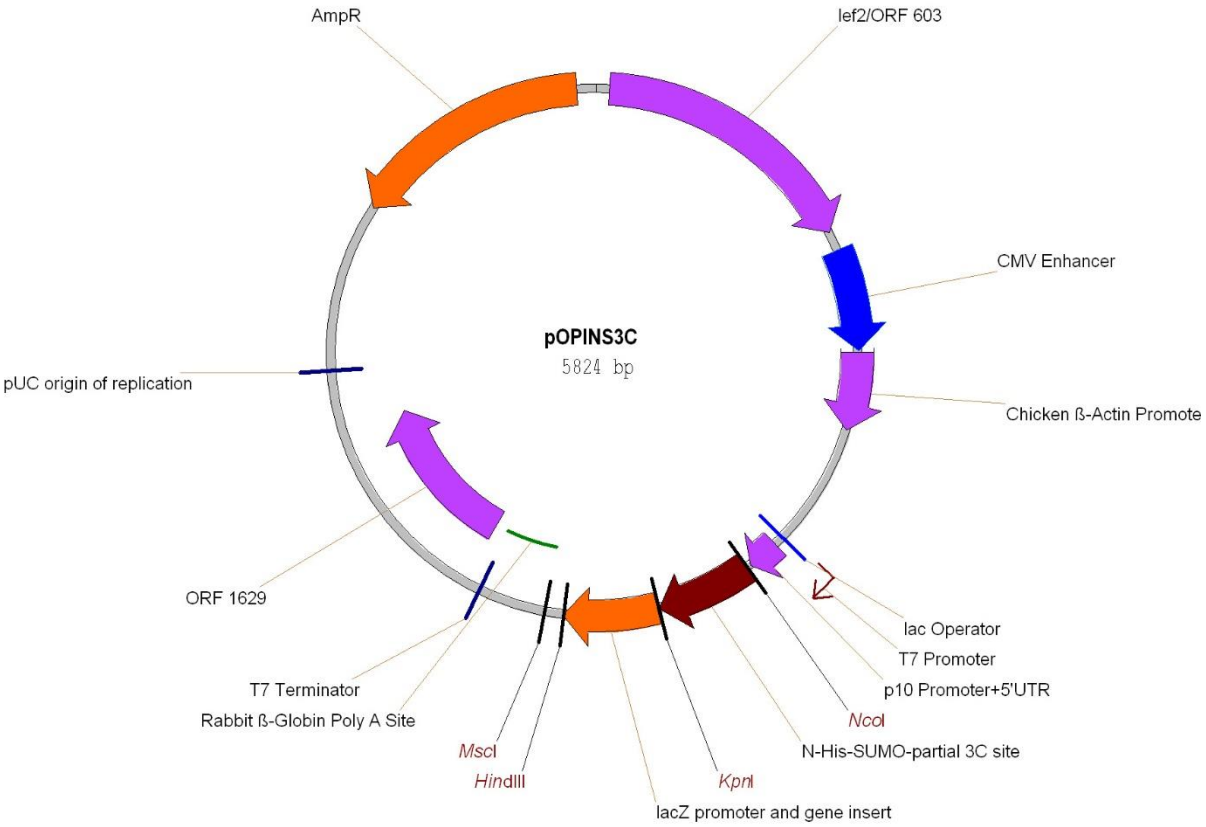


Figure 4.1: Diagram of pOPIN S3C vector (Addgene) with coding GS insert.

Table 4.2: Primers used to clone GS genes into pOPIN S3C for protein expression. The extensions homologous to the vector ends and restriction enzyme sites are in lowercase.

Primer name	Seq 5'-3'	Annealing temperature (°C)
pOPIN-GS1-F	aagttctgtttcagggcccgTCCCCATCACCGAACGAA	64
pOPIN-GS1-R	atggtctagaaagctttaCTAGTGATTTACAGCAACTCCTC	
pOPIN-GS2-F	aagttctgtttcagggcccgGTGGTGACACTCCCTCCCAA	67
pOPIN-GS2-R	atggtctagaaagctttaTCATTCTTGGTGAAATTGGCTGG	
pOPIN-GS11-F	aagttctgtttcagggcccgACATCGATCAGCAACGGACA	65
pOPIN-GS11-R	atggtctagaaagctttaTCACTGAAACTCGCTAGACG	
pOPIN-GS14-F	aagttctgtttcagggcccgGCCCATATTCCGGAAGGTAA	61
pOPIN-GS14-R	atggtctagaaagctttaCTACACCAGGAAAGGCGAGT	
pOPIN-GS20-F	aagttctgtttcagggcccgGAAGCTGATGCCGAAATAACT	62
pOPIN-GS20-R	atggtctagaaagctttaCTAGTACAAGTACGGAGTGTC	
pOPIN-GS23-F	aagttctgtttcagggcccgGGGCCTGTCGATGAAAATG	61
pOPIN-GS23-R	atggtctagaaagctttaCTAATACAGGTATGCACTATCG	
pOPIN-GS49-F	aagttctgtttcagggcccgGTGCCAACCCACAAGGGG	65
pOPIN-GS49-R	atggtctagaaagctttaCTAGACCACCAGGTATGGCG	
pOPIN-GS55-F	aagttctgtttcagggcccgACTGAAGATGCTTCTACTGA	61
pOPIN-GS55-R	atggtctagaaagctttaCTACACAAGCAATGGTGAAT	
pOPIN-AtGS-F	aagttctgtttcagggcccgGGCAGTGGCTGCTCTTC	64
pOPIN-AtGS-R	atggtctagaaagctttaTCAAATCAGATATATGCTGTCCAAGA	

4.3.2.2 Small scale protein expression and purification

A single *E. coli* colony harbouring the expression construct of interest was inoculated into 5 ml LB medium containing 50 µg/ml ampicillin and grown overnight at 30 °C with shaking at 200 rpm. 5 µl of this bacterial culture was used to inoculate a fresh 5 ml LB medium containing 50 µg/ml ampicillin at 30 °C with shaking at 200 rpm until the OD_{600} reached 0.5-0.8. The shaking incubator was then cooled to 18 °C and 1 mM final concentration IPTG was added into the culture. Protein expression was then induced overnight. To extract the His-tagged protein, the 5 ml culture was centrifuged at 13,000 × rpm for 2 minutes. The pelleted cells were then re-suspended in 1 ml lysis buffer (50 mM Tris-HCl (pH8), 500 mM NaCl, 50 mM glycine, 5% (v/v) glycerol and 20 mM imidazole), followed by sonication (10 cycles of 30s on and 10s off in an ice bath) until lysed. The cell debris was then pelleted by centrifugation at 13,000 × rpm for 2 minutes and the supernatant was transferred to a new tube containing 50 µl of Ni-NTA resin (Qiagen, Germany). The mixture was incubated at room temperature with rotating for 10 minutes and the resin was pelleted by centrifugation at 13,000 × rpm for 2 minutes. The resin was then washed three times with 200 µl wash buffer (20 mM Tris (pH7.6), 250 mM NaCl and 20 mM imidazole). The His-tagged protein was eluted off the resin with 50-100 µl of elution buffer (20 mM Tris (pH7.6), 250 mM NaCl and 500 mM imidazole). After the resin was pelleted by centrifugation, the supernatant containing purified protein was transferred to a fresh tube for analysis by SDS-PAGE.

4.3.2.3 Large scale protein expression and purification

Having identified bacterial clones expressing the protein of interest successfully, a single colony was added into 50 ml LB medium containing 50 µg/ml ampicillin, and incubated at 30 °C with shaking at 200 rpm overnight. 5 ml of the culture was transferred to 1 L fresh LB medium containing 50 µg/ml ampicillin and incubated at 30 °C with 200 rpm shaking. In order to acquire sufficient protein for downstream experiments, a total of 6-8 L bacterial culture was usually needed for each expression construct. Once the OD_{600} value reached 0.5-0.8, IPTG was added to a final concentration of 1 mM and the culture was incubated at 18 °C with 200 rpm shaking for 16 hours. The induced culture was cooled on ice for 10 minutes and then centrifuged at 12000 × *g* for 10 minutes. The bacterial pellets were either stored at -80 °C or used directly for protein purification.

The cell pellets were re-suspended in lysis buffer (50 mM Tris-HCl (pH8), 500 mM NaCl, 50 mM glycine, 5% (v/v) glycerol and 20 mM imidazole with the addition of one EDTA-free protease inhibitor tablet per 50 ml solution (Roche, Switzerland)), and were then lysed by sonication (10 cycles of 30s on and 10s off in an ice bath). The cell lysates were clarified by 15,000 × *g* centrifugation for 40 minutes at 4 °C in a pre-cooled centrifuge. The clear supernatant was transferred to a clean glass bottle and stored on ice prior to further purification.

A 1 ml HIS-trap Fast Flow Ni^{2+}-NTA column (GE Healthcare, UK) was applied to an AKTA Xpress (GE Healthcare, UK), and pre-equilibrated with wash buffer (20 mM Tris (pH7.6), 250 mM NaCl and 20 mM imidazole). The soluble total protein extract was passed through the column at a flow rate of 1 ml/minute. His-tagged proteins were eluted using a gradient of increasing imidazole by altering the elution buffer (20 mM Tris (pH7.6), 250 mM NaCl and 500 mM imidazole) from 0% to 100%. Real time absorbance at A_{280} and A_{230} were collected to indicate the presence of proteins in the fractions during the elution. Pooled peak fractions containing the eluted protein were then collected and imidazole was removed by buffer exchange into A4 buffer (20 mM HEPES pH 7.5 + 0.15 M NaCl) using PD MiniTrap G-25 columns (GE Healthcare, UK). The proteins after buffer switching were subsequently cleaved to remove the His+SUMO tag by overnight digestion with 3C protease (2B Scientific, UK) at 4 °C at a ratio of 100:1 (protein: protease). Complete digestion was confirmed by SDS-PAGE. Mature proteins were separated from the His+SUMO tag by passing the solution over a 1 ml Ni^{2+}-NTA column manually, followed by washes of the column with 10 ml A4 buffer. The unbound mature proteins in wash buffer were then concentrated to 10-20 mg/ml using a 30, 000 molecular weight cutoff protein concentrator (Thermo Fisher Scientific, UK). The concentration of each protein was measured using Quick Start™ Bradford dye reagent (Bio-rad, UK) using the associated Quick Start™ Bovine Serum Albumin Standard Set (Bio-rad, UK) to generate a standard curve. The purified proteins were stored in small aliquots at -80 °C until needed.

4.3.2.4 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

A 10% SDS-PAGE gel was prepared in two sections. A 10 ml resolving gel was first prepared by combining 4 ml 30% acrylamide/bis-acrylamide (Severn Biotech Ltd, UK), 2.6 ml 1.5 M Tris (pH8.8), 100 µl 10% (w/v) ammonium persulfate, 100 µl 10% (w/v) SDS, 10 µl TEMED and 3.19 ml ELGA water and adding this gel mix into the lower

75% of the gel casting rig. Once set, a stacking gel was prepared on top of the resolving gel after combining 1.34 ml 30% acrylamide/bis-acrylamide (Severn Biotech Ltd, UK), 2.5 ml 1.5 M Tris (pH6.8), 100 µl 10% (w/v) ammonium persulfate, 100 µl 10% (w/v) SDS, 10 µl TEMED and 5.95 ml ELGA water. The gel comb to form the wells was inserted into the stacking gel before it set. When both gels had set, the protein samples were mixed with 6 × sample loading buffer (the final concentration contained 62.5 mM Tris-HCl (pH 6.8), 2% (w/v) SDS, 0.1 % Bromophenol Blue, 10 % glycerol and 5% β-mercaptoethanol) and were then heated at 99 °C for 10 minutes. The denatured samples were loaded into each well for separation. The gel was electrophoresed at 100 volts in running buffer (3 g Tris base, 14.4 g glycine and 1 g SDS in 1 L water) with BenchMark™ Protein Ladder (Thermo Fisher Scientific, UK) until the dye reached the end of the gel. The gel was stained in staining buffer (2 g Coomassie Blue, 100 ml acetic acid, 400 ml methanol and 500 ml water) at room temperature with slight shaking for at least two hours. The proteins were then visualised after de-staining in wash buffer (100 ml acetic acid, 400 ml methanol and 500 ml water) until bands were clearly visible.

4.3.2.5 Enzymology of recombinant GS proteins

As introduced above, GS is an ATP-dependent enzyme that releases free phosphate during the reaction. The initial rate of glutathione formation was assumed to be equimolar to the rate of phosphate release. Consequently, in this study the GS enzymatic activity was calculated by measuring inorganic free phosphate in aqueous solution based on the malachite green assay protocol. Under acidic conditions, malachite green molybdate can combine with phosphate to form a green molybdophosphoric acid complex (Geladopoulos, Sotiroudis and Evangelopoulos 1991).

The malachite green assay solution contained: 1 M HCl, 1 mM malachite green and 8.5 mM ammonium molybdate supplemented with 0.1 % Triton N-101 shortly before use. Once mixing all the reagents, the malachite green mixture was stirred at room temperature for 30 minutes, followed by filtering through a 0.45 µm filter using a syringe and filter disc. The assay solution was then stored in a plastic container wrapped in aluminium foil, to protect from the light.

A standard curve of absorbance at 630 nm was produced using 0 to 100 µM KH_2PO_4 solution as standards. 100 µl volumes of standard solutions were added into 700 µL of malachite green assay solution, followed by 20 minutes incubation at room temperature. Absorbance values at 630 nm were then read using a EL_X800 Microplate Reader (Bio-Tek Instruments, Inc., UK).

All enzymatic assays were performed with purified recombinant GS proteins. A typical reaction mixture (final volume 100 µl) contained 100 mM HEPES (pH 7.5), 20 mM $MgCl_2$, 5 mM dithiothreitol, 1 mM γ-EC, 2.5 mM ATP, 100 mM glycine and ~5 µg GS protein. The reaction mixtures were then incubated at 30 °C for a suitable time course. 15 µl reaction mixture was then taken out at intervals and added to 105 µl malachite green assay solution in a 96 well plate. The absorbance values at 630 nm were recorded after 20 minutes incubation at room temperature. The reactions with all substrates but without enzyme were used as controls while the reactions with enzyme but without γ-EC were used as blanks. At least four technical repeats were performed per reaction. The initial enzyme rate of GS was calculated following subtraction of the blank reading.

4.3.2.6 Kinetic analysis of GS enzyme activity

Kinetic assays were conducted in the same manner, with concentrations of either γ-EC varied from 0 to 10 mM, or ATP from 0 to 50 mM, or glycine from 0 to 50 mM. Initial velocity kinetic studies were performed by varying two of the three substrates while holding the third substrate constant under identical reaction conditions. The Sigma Plot software was used to determine kinetic parameters. Curve-fitting was carried out with software Origin 2018b (www.originlab.com).

4.3.2.7 Site-directed mutagenesis of *R. reniformis* GS sequences

Directed sequence changes were made to selected Clade 2 and 3 *R. reniformis* GS clones, in order to analyse the role of particular amino acid residues. The GS genes selected for site-directed mutagenesis were *Rre-gs11* to represent Clade 2 GS and *Rre-gs14*, *-gs23*, *-gs55* and *-gs72*, to represent Clade 3 GS. Site-directed mutagenesis of GS cDNA sequences cloned in pGEM-TEasy was performed using a Q5 Site-Directed Mutagenesis Kit (New England BioLabs, UK). Primers were designed using the NEB online software: NEBaseChanger.neb.com. Primers used in site-directed mutagenesis are shown in Table 4.3.

A typical PCR reaction mixture contained 12.5 µl Q5 Hot Start High-Fidelity 2× Master Mix, 0.5 µM forward primer, 0.5 µM reverse primer, 1-25 ng pGEM vector containing the gene of interest as template DNA and nuclease-free water to make a final volume of 25 µl. The PCR cycling conditions were: 98 °C for 30 sec, followed by 25 cycles of 98 °C for 10 sec, the specific annealing temperature for 30 sec and 72 °C for 2 min, ending with an extension at 72 °C for 2 min.

1 µl PCR product was then assembled with 5 µl 2× KLD Reaction Buffer, 1 µl 10× KLD Enzyme Mix and 3 µl nuclease-free water. The reaction mixture was incubated at room temperature for 5 min, followed by transformation of *E. coli* DH5α cells and sequencing of cloned inserts as described in the General methods section. The correct mutated sequences was subsequently cloned into pOPIN S3C vector for protein expression as described in 4.3.1.1.

Table 4.3: Primers used for site-directed mutagenesis. The sequences homologous to the original GS sequence are in uppercase. The sequences homologous to the desired mutated GS sequence are in lowercase. For each GS gene, three regions were selected for mutagenesis. P1 and P2 represent regions associated with glutamic acid binding while P3 represents the glycine binding sites.

Primer name	Seq 5'-3'	Annealing temperature (°C)
GS2-p1-F	gtttgagcagTTGCAGTTCACCACCTTC	63
GS2-p1-R	aggttgaactcGTCAGGCTGGTTGACGAA	
GS2-p2-F	cggcTACATCCCGGAACACTACC	65
GS2-p2-R	tagcgCATGTGGCACAGCGCAAC	
GS2-p3-F	GGGCGGGGTGgcagctGGTGCCGGAG	65
GS2-p3-R	TGGTTCACATTGGCCGGC	
GS11-p1-F	aacatgtacgatcaaCTCCAGTTTGTGATGTTC	57
GS11-p1-R	gcgttcttctggctgATTAACGTACAGGATCATC	
GS11-p2-F	AGTTCACATGagaTATGGGTATTTGCCAG	60
GS11-p2-R	AAGCCAACACGTTTGGTG	
GS11-p3-F	TGGAGGGGTCgcagctGGCGGTGGCGTG	66
GS11-p3-R	AGATTCTGGGAGGCGGCT	
GS14-p1-F	acgcaacatgTACGACCAGCGCCAGGTG	69
GS14-p1-R	tcttctggctgCTCCACCACAATGATGATTGCTTCC	
GS14-p2-F	AAGAACCATTtatACACCTGGCAGC	62
GS14-p2-R	TGGTACACAACGGCAACT	
GS14-p3-F	CAATGATGCCgctGGGGACACAAGTGC	70
GS14-p3-R	TCTGCCGGCGACTTGGCC	
GS20-p1-F	aatatgtatgatcaaGAGCGAATTTTGGCTAGC	62
GS20-p1-R	acgctcctcctgttgTCGGACAATGAGCATGGC	
GS20-p2-F	ggatatTCATCCCGGGCTCTTCGA	66
GS20-p2-R	gtcgcgGAGGTACACAACCGCAATTTTG	
GS20-p3-F	GGGCGGGATCgcagctGGCATCGGAG	62
GS20-p3-R	TCCTTTTGATGCGACCAC	
GS23-p1-F	caacatgtacgatCAGTGGGATCTGGAGGAG	65
GS23-p1-R	cgttcttctggctgGGCCAGGATCACCATGAT	
GS23-p2-F	CCGGTACCCGtatGATCCCAATGATCCG	61
GS23-p2-R	TGGAACACAATGGCCACT	
GS23-p3-F	GGGCGGAGTAgcagctGGGAACGGGG	66
GS23-p3-R	TCGTTGCTCTTCACGCTC	
GS55-p1-F	aactattttgcacaaCATAAAATAGCCCAAGAGC	59
GS55-p1-R	caattcaccaccctgTGGTTCTACAACAGCCAC	
GS55-p2-F	AGAGTGACCTaTTTGAGTCCAATTG	60

GS55-p2-R	GTAAAAAATCACGGCCAC	
GS55-p3-F	GGGTGGAATTgcagctGGCACCGGAGTATAC	63
GS55-p3-R	TCGTTGGCCCAAACCCAT	

4.4 Results

4.4.1 Expression and purification of *R. reniformis* GS from *E.coli*

A range of *R. reniformis* GS genes were selected for characterisation of their enzymatic activity *in vitro* and to determine if the members of the expanded gene family retained canonical GS function, despite their sequence divergence. The 16 selected genes were *Rre-gs1* representing Clade 1, *Rre-gs2*, *-gs4* and *-gs11* representing Clade 2, and *Rre-gs14*, *-gs18*, *-gs20*, *-gs23*, *-gs27*, *-gs36*, *-gs44*, *-gs49*, *-gs55*, *-gs57*, *-gs61* and *-gs67* representing Clade 3. For comparison, the only GS enzyme from the model plant *Arabidopsis* was also chosen. All genes of interest were sub-cloned into the pOPIN S3C^{TM} vector without their signal peptide if appropriate. This generated translational fusions at the N-terminus of the GS proteins to add both a 6x His tag for purification and a small ubiquitin-like modifier (SUMO) tag to enhance protein stability and solubility. Each gene was heterologously expressed in an *E. coli* system, and their products were purified using nickel affinity chromatography. All selected GS proteins were first expressed in small scale cultures to identify whether the bacteria expressed the protein of interest successfully (Figure 4.2). Samples were analysed from cultures at four stages: pre-IPTG induction total extract, post-IPTG induction total extract, induced total soluble protein and eluted purified protein (Figure 4.2A-D) to screen the target protein expression. Based on the amino acid sequences, the *R. reniformis* GS proteins were predicted to range from 50-70 kDa. The predicted size of each GS is shown in Table 4.4. In addition, the size of the His-SUMO tag is around 13 kDa. Therefore, the expected band of the fusion proteins should be near the 80 kDa position.

All GS of interest except Rre-GS27 and Rre-GS61 were expressed successfully in the *E. coli* small scale cultures. Figure 4.2 shows an example of small scale expression and purification of some GS. In Figure 4.2D, strong bands of GS11, GS49, GS20 and GS18 at appropriate positions were detected on SDS-PAGE gel following purification, suggesting that these GS proteins were expressed successfully in the bacteria. Moreover, in addition to the expected bands, many additional protein bands were still detected on the gel, which may be due to the less stringent washing of the Ni-NTA resin in these small-scale batch purifications.

Subsequently, successfully expressing *E. coli* colonies were used for large scale GS production. Sufficient protein for the downstream experiments (>1000 µg) was produced for Rre-GS1, -GS2, -GS11, -GS14, -GS20, -GS23, -GS49 and -GS55. Here we take Rre-GS14 as an example. The predicted molecular mass of Rre-GS14 protein is approximately 59 kDa. Figure 4.3 shows SDS-PAGE analysis of protein samples from six stages of the large-scale purification: pre-IPTG induction whole cell, post-IPTG induction whole cell, total soluble fraction, first purification using HIS column, after cleavage of the HIS-SUMO tag and after the second HIS column purification to remove the cleaved tag (from left to right). A strong band detected at around 80 kDa after the first purification on the HIS column indicated that GS14 with HIS and SUMO tag was successfully expressed. The following two lanes showed a drop in molecular mass of approximately 13 kDa corresponding to the loss of the His and SUMO tags. A clear band of approx. 22 kDa observed at the bottom of the gel after cleavage of the HIS-SUMO tag is the His-tagged 3C protease. This was absent after the 2nd HIS-Trap purification indicating that it, together with the cleaved HIS-SUMO tag, was successfully removed from the purified Rre-GS14 protein.

Table 4.4 The predicted size of *R. reniformis* GS proteins (without starting code and signal peptide if appropriate)

Proteins	Expected molecular mass (kDa)	Proteins	Expected molecular mass (kDa)
GS1	56.3	GS50	55.3
GS2	64.8	GS51	57.4
GS4	63.5	GS55	55.5
GS5	63.4	GS57	55.7
GS11	60.1	GS59	57.0
GS14	56.1	GS61	54.9
GS18	55.8	GS64	55.3
GS20	55.4	GS65	55.0
GS23	58.0	GS66	55.7
GS27	58.3	GS67	54.1
GS36	58.6	GS72	59.4
GS44	58.7	AtGS	60.2
GS49	62.3		

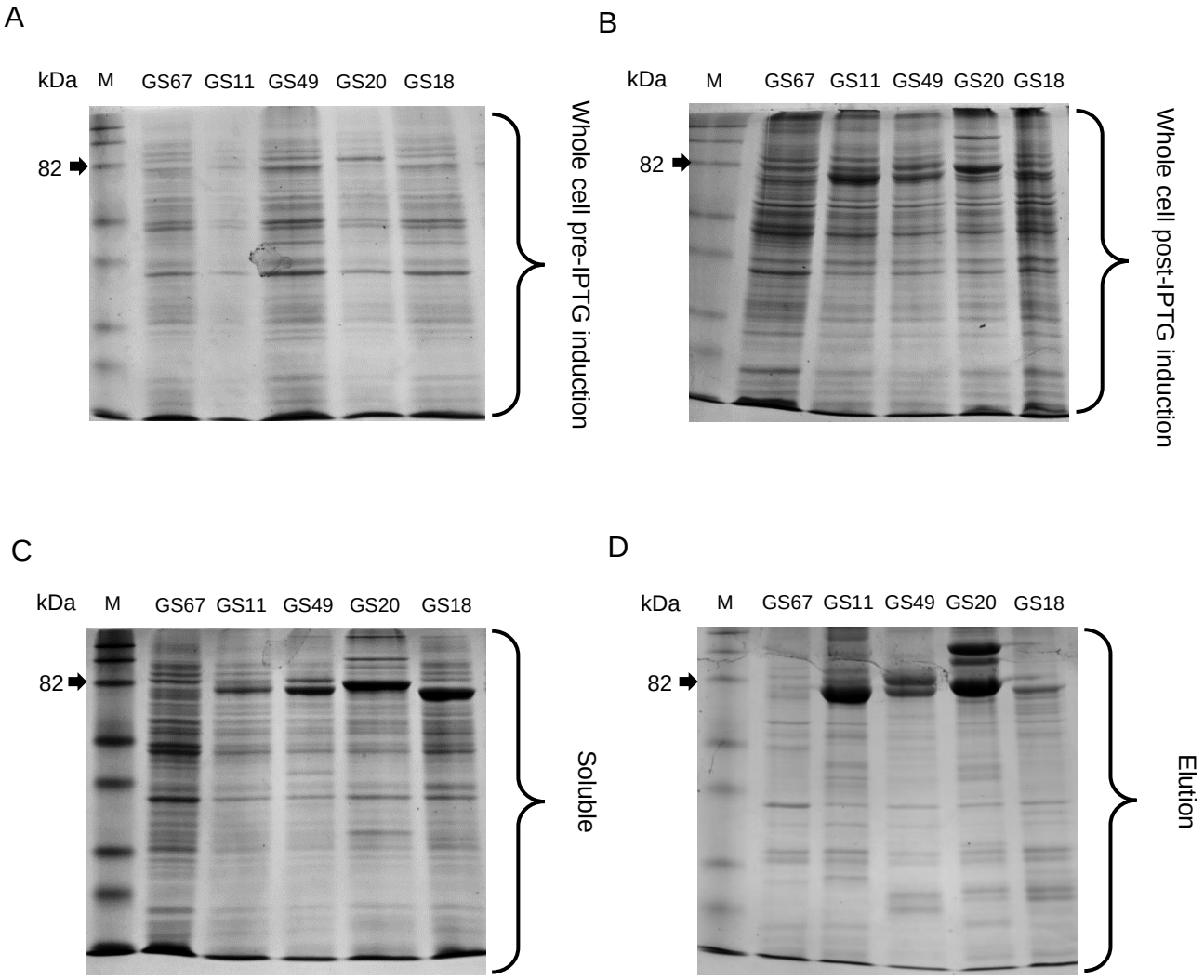


Figure 4.2: Examples of expression and purification of *R. reniformis* in small scale. Protein samples were prepared for (A) pre-IPTG induction control, (B) post-IPTG induction control, (C) soluble protein control and (D) target proteins after elution. Strong bands of GS11, GS49 GS20 and GS18 at around 80 kDa were detected, suggesting that these GS proteins were expressed successfully in the bacteria

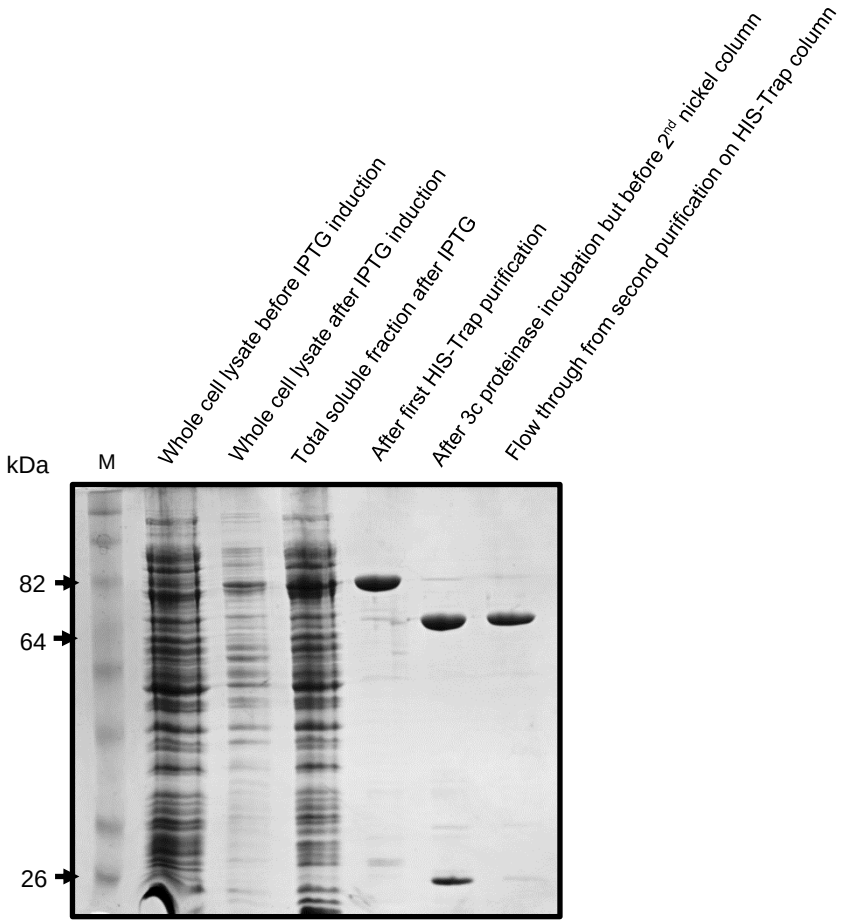


Figure 4.3: Large scale expression and purification of *R. reniformis* GS14. SDS-PAGE analysis of recombinantly expressed Rre-GS14 protein. M: Molecular mass marker. In the lane of first purification of HIS column stage, a strong band detected at around 80 kDa position, indicating Rre-GS14 with HIS and SUMO tag was successfully expressed. A clear band with a drop in molecular mass of approximately 13 kDa was shown after 3C proteases overnight incubation. At the bottom of this lane, the band shows the 3C proteases. After the second purification, only one single band was shown on the gel, demonstrating that mature GS14 proteins were purified.

4.4.2 Enzymatic activity of *R. reniformis* GS

4.4.2.1 Time course analysis of *R. reniformis* GS enzyme activity

Prior to examination of initial enzyme rate, full time course analysis of Rre-GS glutathione synthetic activities was performed to determine the most suitable time points for the experiments. Rre-GS1 representing Clade 1, Rre-GS2 and -GS11 representing Clade 2, Rre-GS14 and Rre-GS23 representing Clade 3 were selected for these experiments. Figure 4.4 indicates that the absorbance at 630 nm, due to phosphate release from ATP, became maximum after 40 min for Rre-GS1, whereas the absorbance at 630 nm for the Clade 2 Rre-GS enzymes and Clade 3 Rre-GS enzymes reached the peak after 60-80 min and 120 min, respectively.

4.4.2.2 Rre-GS initial enzyme rates

Initial enzyme rates for the proteins used in the time course analysis plus AtGS and Rre-GS20, Rre-GS49 and Rre-GS55 from Clade 3 were determined by measuring phosphate release from ATP in the presence of canonical substrates (γ-EC, glycine and ATP) using standard curve of absorbance provided by 0 to 100 µM KH_2PO_4 solution. Initial enzyme rates were determined by measuring phosphate release from ATP in the presence of canonical substrates (γ-EC, glycine and ATP). Standard curve of absorbance was draw using 0 to 100 µM KH_2PO_4 solution as standard solutions (data now shown). Initial enzyme rates of GS proteins were calculated based on the absorbance difference between GS reactions in the presence of all substrates and the blank reaction that lacked γ-EC in order to discount background ATP hydrolysis (Figure 4.5). Using this approach, the initial rate of *Arabidopsis* GS (AtGS) phosphate release was 7476 (± 132) pmol min^{-1} $µg^{-1}$, which was consistent with a previous report for *Arabidopsis* GS of around 7500 pmol min^{-1} $µg^{-1}$ (Jez and Cahoon 2004). This validated the assay system so providing confidence in the data obtained for the previously uncharacterised *R. reniformis* GS.

Interestingly, the initial enzyme rate of the Rre-GS from different Clades exhibited a significant diversity (Figure 4.5). The initial rate of the canonical Rre-GS1 reached 1028 (± 119) pmol min^{-1} $µg^{-1}$, which was slightly lower than the reported enzyme rate of the canonical GS from the non-parasitic nematode *C. elegans* (Buzie and Enjuakwei 2007). By contrast, the initial rates of Rre-GS from both Clade 2 & 3 were extremely low and varied from 6.5 ± 0.3 (Rre-GS20) to 17.6 ± 0.4 (Rre-GS2) pmol min^{-1} $µg^{-1}$.

There was no apparent difference in glutathione synthetic activity between the two Clades.

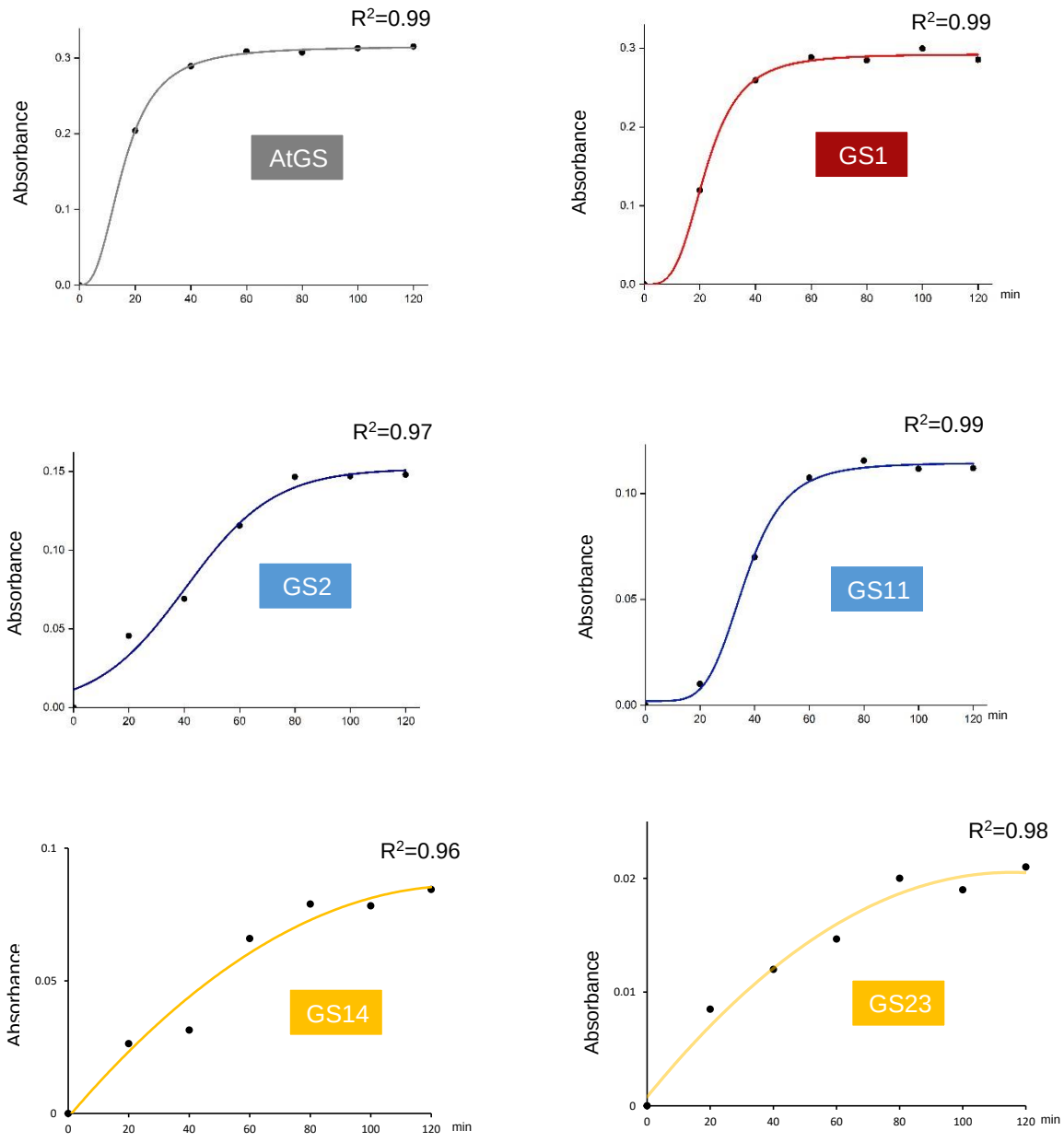


Figure 4.4: Time course analysis of *R. reniformis* GS enzyme activity. *Arabidopsis* GS (grey), *R. renifomis* GS1 representing Clade 1 (red), GS2 and GS11 representing Clade 2 (blue), GS14 and GS23 representing Clade 3 (yellow) were used for time course analysis. The absorbance at 630 nm of *Arabidopsis* GS and GS1 became maximum after 40 min, whereas the absorbance at 630 nm of Clade 2 GS enzymes and Clade 3 GS enzymes reached the peak after 60 to 80 min and 120 min, respectively.

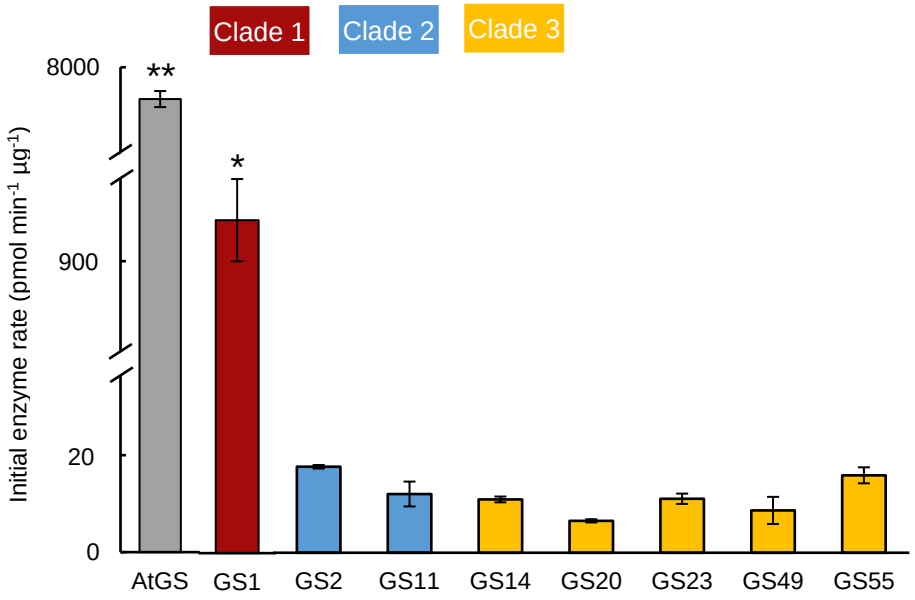
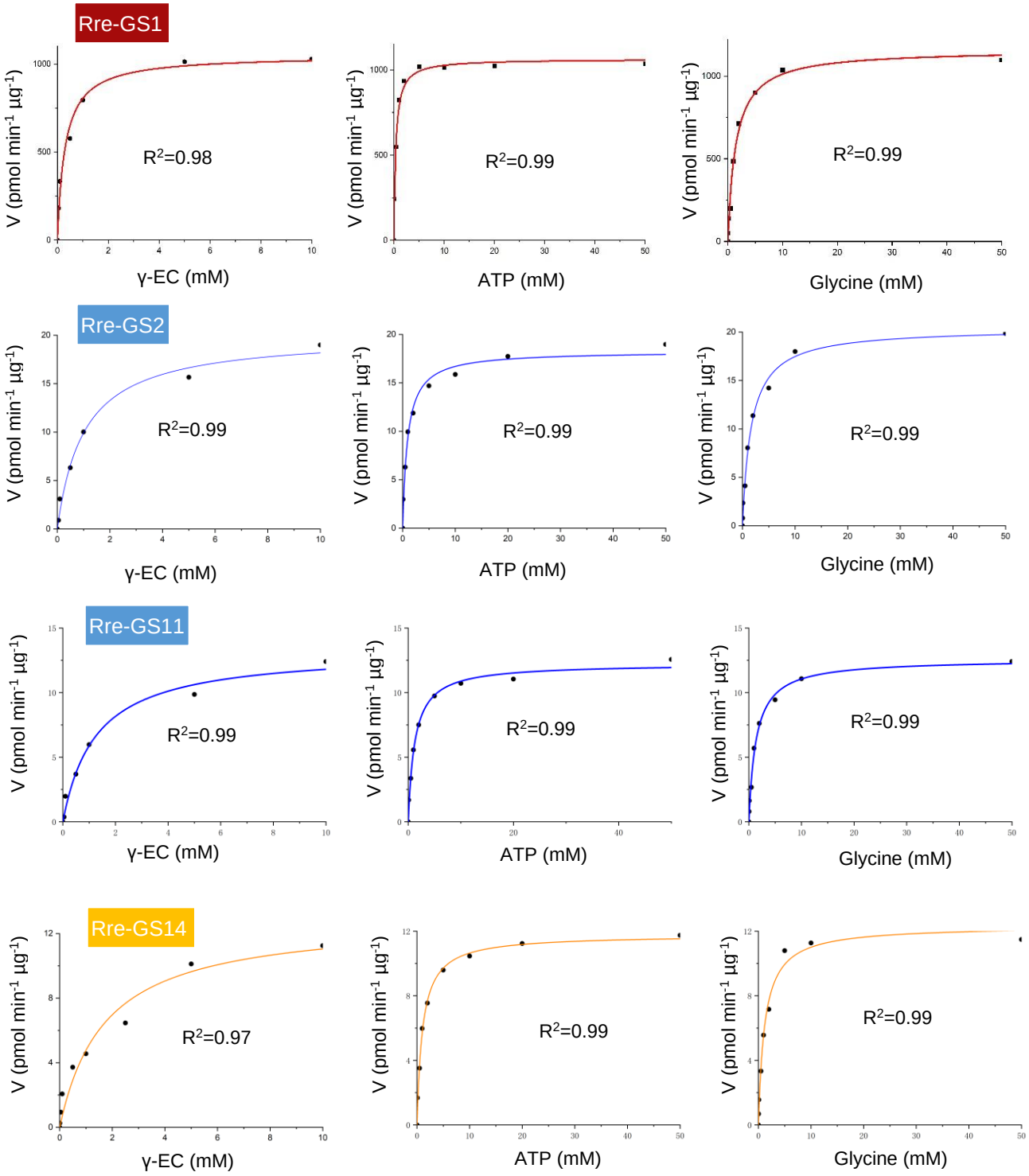


Figure 4.5: Initial enzyme rates of *R. reniformis* GS proteins. Purified protein for *Arabidopsis* GS (AtGS), *R. reniformis* GS1, GS2, GS11, GS14, GS20, GS23, GS49 and GS55 were tested for glutathione synthetase activity by measuring phosphate release from ATP in the presence of canonical substrates (γ-EC, glycine and ATP). The mean values were calculated from 4 independent experiments. Error bars indicate standard error of the mean. * indicates a statistically significant difference (One-way ANOVA, P<0.05).

4.4.2.3 Kinetic analysis of *R. reniformis* GS

To understand the mechanistic information of *R. reniformis* GS activity, a kinetic analysis was carried out using initial velocity data collected as a function of varying substrate concentrations. Figure 4.6 shows saturation curves of Rre-GS1 -GS2, -GS11, -GS14, -GS20, -GS23, -GS49 and -GS55 for γ-EC, ATP and glycine, respectively. All displayed data fit the Michaelis Menten equation. Steady-state kinetic parameters (K_m) for γ-EC, ATP and glycine were determined for Rre-GS1, -GS2, -GS11, -GS14, -GS20, -GS23, -GS49 and -GS55 (Table 4.5).

Purified recombinant *R. reniformis* GS1 enzyme displayed K_m values (K_m [γ-EC] =0.305 mM, K_m [ATP] = 364 µM, K_m [glycine] = 1.44 mM) (Table 4.5) which are in close range to those reported for the GS from other eukaryotes (Buzie and Enjuakwei 2007), indicating *R. reniformis* GS1 enzyme has similar biochemical characteristics with some typical eukaryotic GS enzymes. Compared with Rre-GS1, Clade 2 & 3 GS enzymes displayed nearly 10-fold and 5-fold higher K_m values for γ-EC and ATP, respectively, whereas they showed similar K_m values for glycine (Table 4.5), suggesting that Rre-GS1 has stronger affinity for γ-EC and ATP than all the Clade 2 & 3 GS enzymes. Given the fact that Clade 2 & 3 GS lack canonical enzyme activity, this study emphasised the hypothesis that Clade 2 & 3 GS may accept an alternative substrate instead of γ-EC whereas glycine may still be a substrate for non-canonical GS enzymes.


Rre-GS1
V (pmol min-1 μg-1)
R2=0.98
γ-EC (mM)
R2=0.99
ATP (mM)
R2=0.99
Glycine (mM)
Rre-GS2
R2=0.99
γ-EC (mM)
R2=0.99
ATP (mM)
R2=0.99
Glycine (mM)
Rre-GS11
R2=0.99
γ-EC (mM)
R2=0.99
ATP (mM)
R2=0.99
Glycine (mM)
Rre-GS14
R2=0.97
γ-EC (mM)
R2=0.99
ATP (mM)
R2=0.99
Glycine (mM)

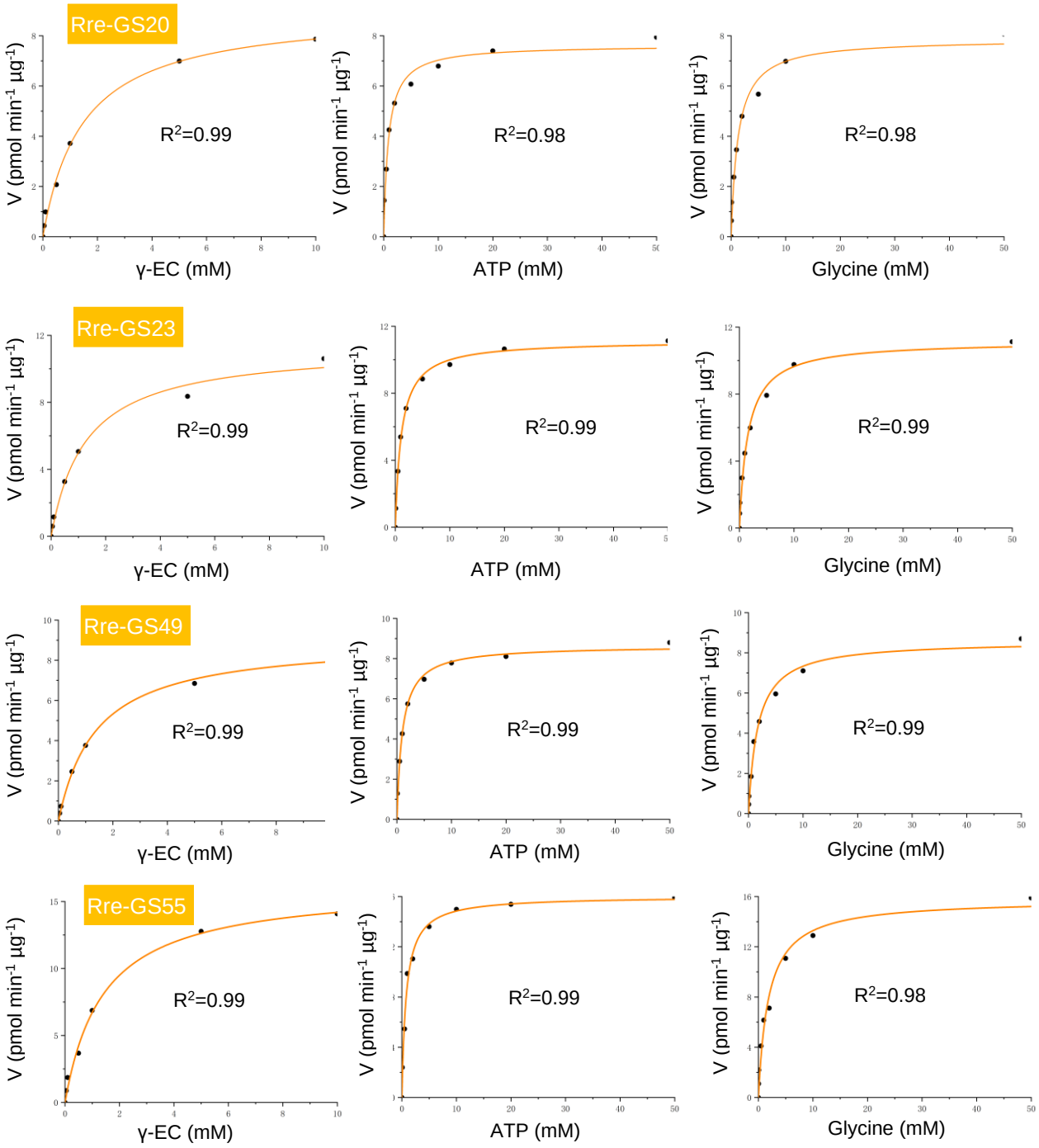


Figure 4.6: Saturation curves for Rre-GS1, Rre-GS2, Rre-GS11, Rre-GS14, Rre-GS20, Rre-GS23, Rre-GS49 and Rre-GS55 with varying concentrations of different canonical GS substrates. The enzymatic assays were carried out by varying the concentration of one substrate while the other two were maintained at saturating concentrations. The data were fitted by nonlinear least-squares regression analysis, and the lines of best fit to the Michaelis-Menten equation are shown. The square of the correlation coefficient (R^2) for the global fit of all displayed data are shown.

Table 4.5 Key kinetic parameters of the recombinant *R. reniformis* GS enzymes. All K_m and V_{max} values are expressed as mean ± standard error, for an n = 3

	GS1	GS2	GS11	GS14	GS20	GS23	GS49	GS55
K_m [γ-EC] (mM)	0.305 ± 0.08	1.02 ± 0.17	1.26 ± 0.25	1.71 ± 0.23	1.46 ± 0.16	1.28 ± 0.21	1.39 ± 0.1	1.41 ± 0.18
K_m [ATP] (μM)	364 ± 32.2	946 ± 124.1	1243 ± 132.7	1061 ± 91.4	857 ± 122.5	1137 ± 62.5	1007 ± 88.0	791 ± 81.5
K_m [glycine] (mM)	1.44 ± 0.15	1.68 ± 0.19	1.38 ± 0.17	1.20 ± 0.19	1.27 ± 0.22	1.57 ± 0.22	1.72 ± 0.22	1.87 ± 0.34
V_{max} (pmol min^{-1} μg^{-1})	1050 ± 40.4	20.0 ± 0.93	13.3 ± 0.77	12.9 ± 1.2	9.0 ± 0.29	11.4 ± 0.53	9.0 ± 0.19	16.2 ± 0.61

4.4.3 Sequence analysis of active site residues in *R. reniformis* GS

To elucidate the reason for the loss of canonical GS enzyme activity in *R. reniformis* Clade 2 and 3 GS-like enzymes, the amino acid sequences of all 23 cloned *R. reniformis* GS-like genes were aligned with those of the structurally solved potato GS (StGS), a non-canonical *G. pallida* GS protein (Gpa-GSS22) (Lilley *et al.* 2018) and human GS (Dinescu *et al.* 2004). As introduced above, three important substrates (ATP, γ-EC and glycine) are required for glutathione production catalysed by GS. Conserved active residues were therefore examined based on the binding pockets for each substrate.

Residues in the ATP-binding pocket of Gpa-GSS22 were highly conserved in sequence and in position with canonical GS from other eukaryotes (Lilley *et al.* 2018) and were similarly conserved across all the *R. reniformis* GS-like proteins (Figure 4.8A; Table 4.6). Given the fact that *R. reniformis* Clade 2 & 3 GS-like enzymes lacked the typical GS activity and have an extremely low rate of ATP turnover when provided with the canonical substrates, it can be hypothesised that *R. reniformis* Clade 2 & 3 GS-like enzymes still belong to the ATP-grasp subfamily but may accept alternative substrates, which is responsible for their distinct catalytic activities.

The binding of γ-EC in canonical GS enzymes is coordinated at both the glutamate and the cysteine residue. Three coordinating residues for cysteine were identified in the structurally-solved human GS and potato GS (Polekhina *et al.* 1999; Lilley *et al.* 2018). The alignment of amino acid sequences between these two GS and all cloned *R. reniformis* GS indicated that residues in the cysteine binding pocket of *R. reniformis* GS were relatively conserved, in Clade 2 and 3 enzymes as well as in the canonical Rre-GS1 (Figure 4.7; Figure 4.8B; Table 4.6). Perfect conservation was revealed in the first catalytic residue arginine, while the other two coordinating residues in the cysteine binding pocket, which interact with the C-alpha backbone of cysteine, were not always the same but were largely conserved and were preferably uncharged.

In contrast, the glutamate of γ-EC was only coordinated by interactions with charged side chains of six residues in the binding pocket of potato StGS (Lilley *et al.* 2018). What's more, all these six positions in the glutamic acid binding pocket are highly conserved among the GS sequences that show canonical catalytic activity (plant GS, human GS, *C. elegans* GS and *R. reniformis* Clade 1 GS) (Figure 4.7; Figure 4.9A),

whereas these demonstrated a significant diversification among the GS sequences which did not display typical enzyme activity (*R. reniformis* Clade 2 & 3 GS) (Figure 4.9B). Among 19 cloned *R. reniformis* Clade 2 & 3 GS sequences, there are 49 different amino acid compositions in these 6 positions, none of them is consensus. Interestingly, at the fifth position of the glutamic acid binding pocket, a 100% conserved and a fairly conserved arginine were shown in the canonical and non-canonical GS enzymes, respectively, indicating that this position may not be responsible for the loss of canonical enzyme activity. Similarly, residues in the glycine binding pocket are highly conserved in sequences which have canonical GS activity (Figure 4.7), but significantly diversified in those which do not have canonical GS activity (Figure 4.10).

Taken together, none of the residues of non-canonical GS enzymes in the glutamic acid binding pocket and glycine binding pocket are conserved, suggesting that there is a possible change in γ-EC and/or glycine specificity in these enzymes and a novel substrate may be accepted into *R. reniformis* Clade 2 & 3 GS.

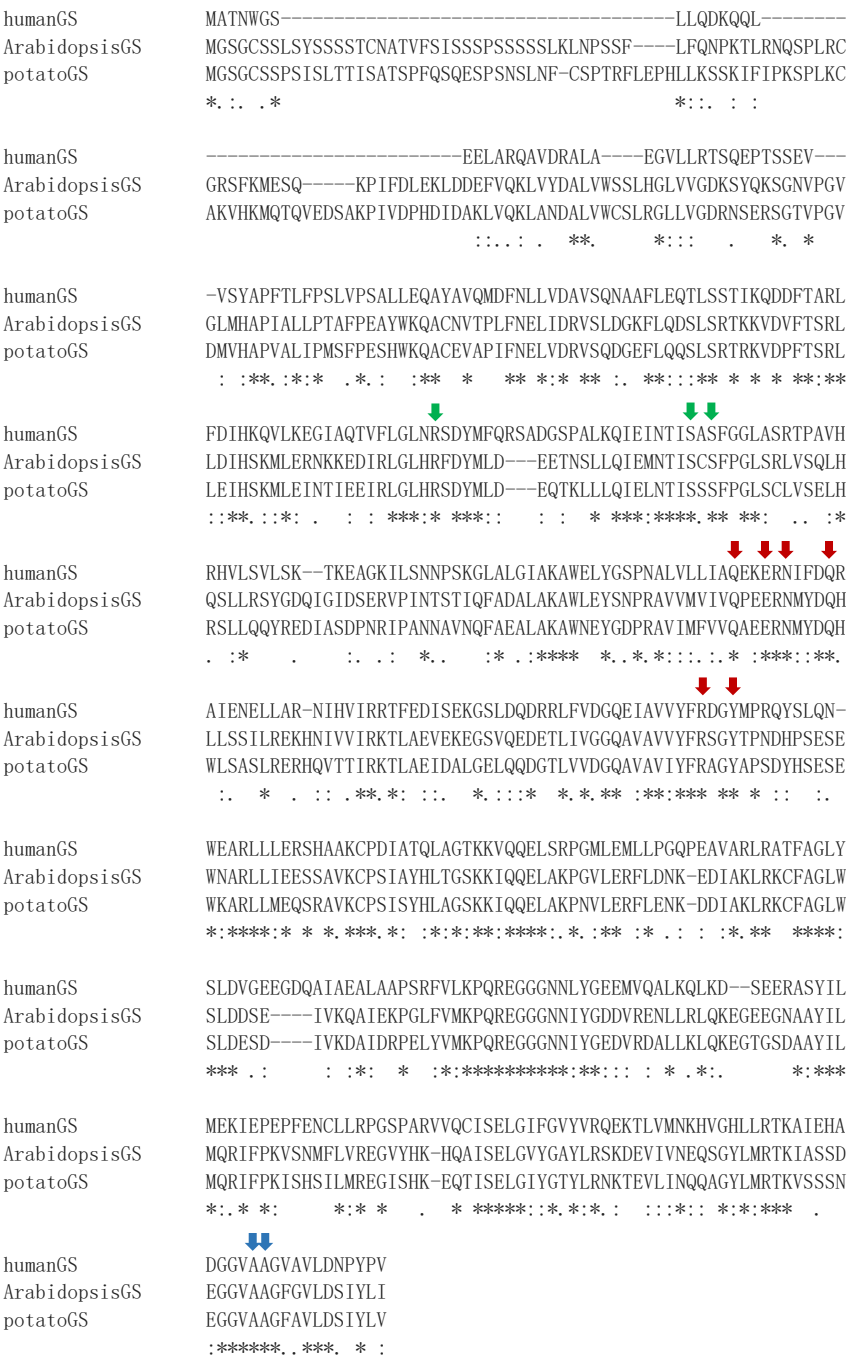
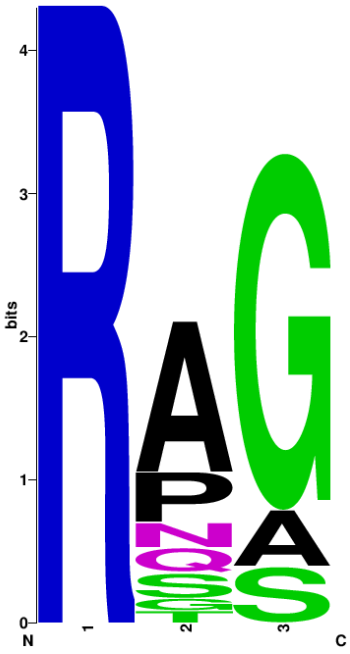


Figure 4.7: An alignment of canonical GS indicating conserved active residues. Green arrows: cysteine binding residues. Red arrows: glutamic acid binding residues. Blue arrows: glycine binding residues.

A

Residue conservations in the ATP binding pocket

B



Residue conservations in the cysteine binding pocket

Figure 4.8: (A) Residues in ATP binding pocket. (B) Residues in cysteine binding pocket. An alignment of amino acid sequences of all 23 cloned *R. reniformis* GS-like genes, potato GS, a non-canonical *G. pallida* GS and human GS was exploited to investigate the relevant active residues. In plant GS, the cysteine of the di-peptide substrate (γ-EC) is coordinated by the side chain of an arginine, and the backbone of two serines. The arginine is conserved among all cloned *R. reniformis* GS. The two serines are not 100% conserved but the equivalent residues are preferentially small and uncharged amino acids.

A

Residue conservation in the glutamic acid binding pocket of GS sequences which have canonical GS enzyme activity.

B

Residue variation in the glutamic acid binding pocket of GS sequences which lack canonical GS enzyme activity

Figure 4.9: Residues in the glutamic acid binding pocket. (A) Residues in the glutamic acid binding pocket of canonical GS enzymes. In canonical GS enzymes including human GS, Arabidopsis GS, potato GS, *C. elegans* GS and *R. reniformis* GS1, the residues involved in glutamic acid binding were highly conserved. (B) Residues in the glutamic acid binding pocket of non-canonical GS enzymes. In the non-canonical *R. reniformis* Clade 2 and 3 GS enzymes, the residues involved in glutamic acid binding show a high level of variability.

A

Residue conservation in the glycine binding pocket of GS sequences which have canonical GS enzyme activity.

B

Residue variation in the glycine binding pocket of GS sequences which lack canonical GS enzyme activity

Figure 4.10: Residues in the glycine binding pocket. (A) Residues in the glycine binding pocket of canonical GS enzymes. In canonical GS enzymes including human GS, Arabidopsis GS, potato GS, *C. elegans* GS and *R. reniformis* GS1, the residues involved in glycine binding (Red rectangle) were highly conserved. (B) Residues in the glycine binding pocket of non-canonical GS enzymes. In non-canonical GS enzymes including *R. reniformis* Clade 2 and 3 GS, the residues involved in glycine binding (Red rectangle) are highly variable.

Table 4.6: Structure-guided comparison of key residues in the ATP binding pocket and substrate binding pocket of *R .reniformis* GS.

Clade		1	2				3																		
Protein	StGS	Rre_GS1	Rre_GS2	Rre_GS4	Rre_GS5	Rre_GS11	Rre_GS14	Rre_GS18	Rre_GS20	Rre_GS23	Rre_GS27	Rre_GS36	Rre_GS44	Rre_GS49	Rre_GS50	Rre_GS51	Rre_GS55	Rre_GS57	Rre_GS59	Rre_GS61	Rre_GS64	Rre_GS65	Rre_GS66	Rre_GS67	Rre_GS72
ATP binding residues	M (141)	M	M	L	L	L	V	M	M	M	M	M	M	M	F	M	M	M	M	M	F	F	M	M	M
	I (152)	I	V	V	V	V	I	-	-	I	I	I	I	I	I	I	I	I	I	I	I	V	I	I	V
	E (153)	E	E	E	E	E	D	Q	Q	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
	N (155)	N	N	N	N	N	S	S	A	N	N	N	N	N	N	N	N	S	N	N	N	N	N	N	N
	K (318)	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
	V (370)	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
	K (372)	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
	N (381)	N	N	N	N	I	E	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	Y (383)	F	F	F	F	F	H	F	F	F	F	Y	W	W	W	W	W	W	W	W	W	W	W	W	W
	M (408)	M	Q	Q	Q	M	M	M	M	M	M	Q	M	M	M	M	M	M	M	M	M	M	M	M	M
	Q (409)	Q	Q	Q	Q	K	E	E	E	Q	Q	Q	Q	Q	Q	Q	Q	Q	E	Q	Q	Q	Q	Q	Q
	I (411)	I	I	I	I	I	I	L	L	L	L	I	L	L	L	I	L	L	L	L	L	L	L	I	L
	E (434)	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
	R (459)	R	R	K	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
	K (461)	K	K	R	K	K	R	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
Cysteine binding residues	R (137)	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
	S (158)	A	Q	P	P	Q	N	-	-	G	N	A	P	A	A	A	A	A	P	A	T	A	A	A	-
	S (160)	S	G	G	G	G	A	-	-	G	G	G	G	A	A	G	G	G	G	G	G	G	G	G	G
Glutamic acid binding residues	Q (222)	G	Q	N	S	Q	M	A	G	G	K	R	Q	S	S	P	P	P	G	P	N	R	G	G	N
	E (225)	N	P	P	N	P	N	-	-	K	R	K	T	S	S	I	S	S	N	P	D	D	N	Q	D
	N (227)	N	C	C	T	C	L	-	-	T	H	P	A	N	T	T	T	T	F	N	F	L	T	T	A
	Q (231)	Q	Q	Q	Q	Q	Q	-	-	L	T	S	Q	Q	Q	D	D	D	Q	-	E	E	M	M	M
	R (279)	R	A	T	A	A	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	S
	Y (282)	Y	Y	N	N	Y	-	Y	Y	-	G	P	F	H	F	L	F	F	F	R	R	R	R	R	R
Glycine binding residues	A (471)	A	C	C	G	S	A	N	S	L	W	L	L	S	I	R	M	M	L	S	T	V	Q	E	F
	A (472)	V	E	A	S	T	P	H	H	S	E	N	K	C	E	I	K	K	E	R	I	I	N	M	H

4.4.4 Site-directed mutagenesis of *R. reniformis* non-canonical GS

As hypothesised above, the sequence variability in the glutamic acid binding sites and/or glycine binding sites may play a role in loss of canonical GS enzyme activity in the *R. reniformis* Clade 2 and 3 GS. To analyse the roles of the selected residues in the loss of canonical GS activity, site-directed mutagenesis of either the predicted glutamic acid binding residues or predicted glycine binding residues in non-canonical GS representatives Rre-GS11, -GS14, -GS20, -GS23, -GS55 and -GS72 was carried out and the enzyme rates of these mutants were then measured.

As introduced above, six key residues involved in glutamate binding and two amino acids associated with specificity of glycine binding were identified in canonical GS enzymes. In potato GS, the six residues in the glutamate binding pocket are Q222, E225, N227, Q231, R279 and Y282, and the two residues in the glycine binding pocket are A471 and A472 (Figure 4.11) (Lilley *et al.* 2018). Considering the relatively large distances between some of these residues in the primary sequence, the decision was taken to carry out the mutagenesis on three separate regions. Region 1 represented the first four coordinating residues of the glutamate binding pocket (Q, E, N and Q), Region 2 covered the remaining two residues for coordination of glutamate (R and Y) and Region 3 altered the two key residues of the glycine binding pocket (A and A). The range of mutants made by site-directed mutagenesis is shown in Table 4.7. In addition, short amino acid inserts were shown in the Region 1 of *R. reniformis* Clade 2 & 3 GS, which contributed to bad alignments in this region of these sequences. Therefore, the whole sequences in the Region 1 of these non-canonical GS were substituted by the same region of potato GS or the inserts were deleted (Table 4.7).

Most of the mutant variants did not show any significant change in initial enzyme rate when provided with canonical GS substrates (Figure 4.12A). However, one of the Rre-GS55 mutants (R241Q, S244E, T246N, and D250Q) displayed an approximate 2.5-fold increased enzyme rate, although the other two Rre-GS55 mutants with a change in the 2nd region of the glutamate bind pocket (F303Y) or with altered glycine binding residues (M501A, K502A) retained similar enzyme activity to native Rre-GS55 (Figure 4.12A). K_m values were further calculated. Rre-GS55 variant 1 (R241Q, S244E, T246N, and D250Q) showed a stronger affinity to both γ-EC and ATP but similar affinity to glycine (Figure 4.12B, Table 4.8). Considering that four residues in the glutamic acid

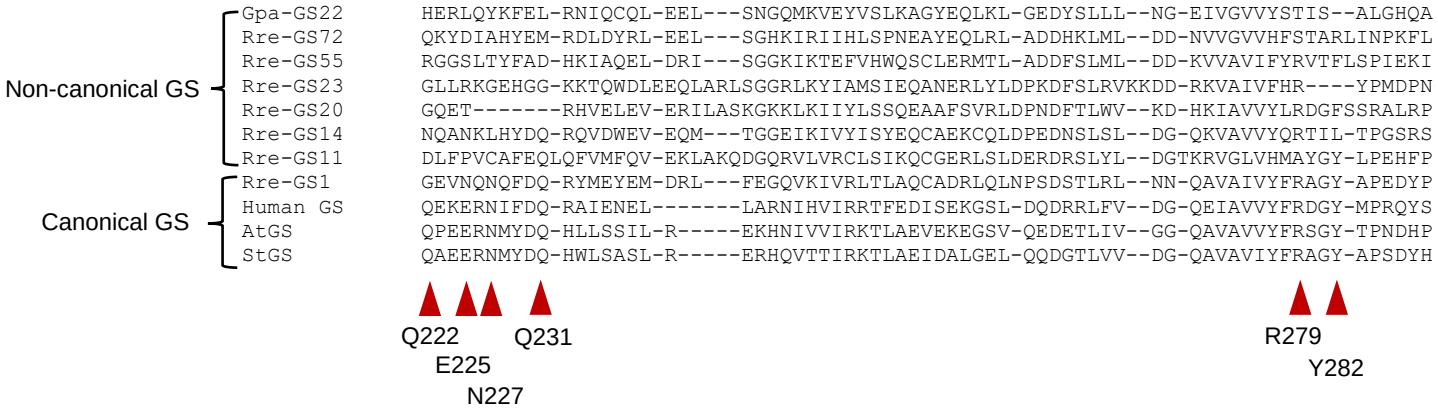
binding pocket were altered in Rre-GS55 variant 1, these four positions may play a role in canonical GS activity.

In addition, sequence alignment between canonical and non-canonical GS enzymes identified some short additional stretches of amino acids in this region in many of the *R. reniformis* non-canonical GS enzymes. Although the deletion of these short insertions in non-canonical GS enzymes did not rescue the activity, they are considered to potentially affect the canonical GS activity as they make the space of the glutamic acid binding pocket smaller and narrower.

Polarity of amino acids is very important in the formation of different non-covalent bonds between amino acids and ligands (Radzicka and Wolfenden 1988). In the sixth position of the glutamic acid binding pocket in canonical GS enzymes, an 100% conserved polar tyrosine is shown. However, a nonpolar phenylalanine was found at the same position of Rre-GS55. The variant F303Y did not recover the canonical GS enzymatic activity, indicating a single substitution at one position of the glutamic acid binding pocket is not sufficient to endow GS55 with canonical activity in the context of the rest of the protein.

Considering that Rre-GS55 variant 1 is the only one where just the four amino acid changed, with no deletions or changes of other amino acids, another possible reason for the partial recovery of the canonical GS activity is the whole region contributes to the structure of the binding pocket rather than some individual amino acids.

A



B

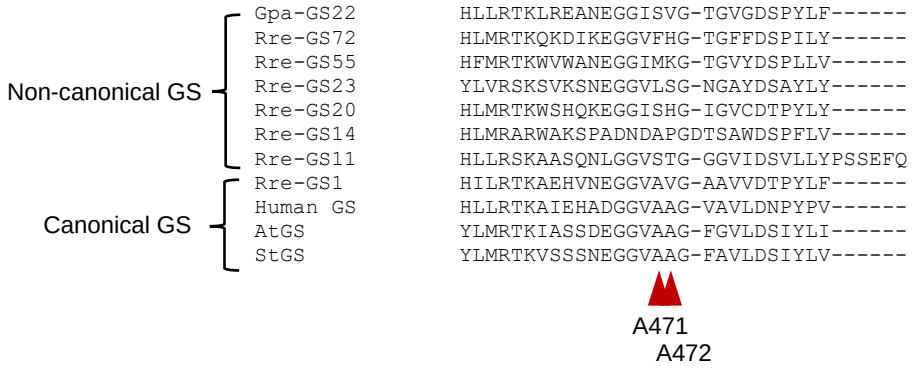
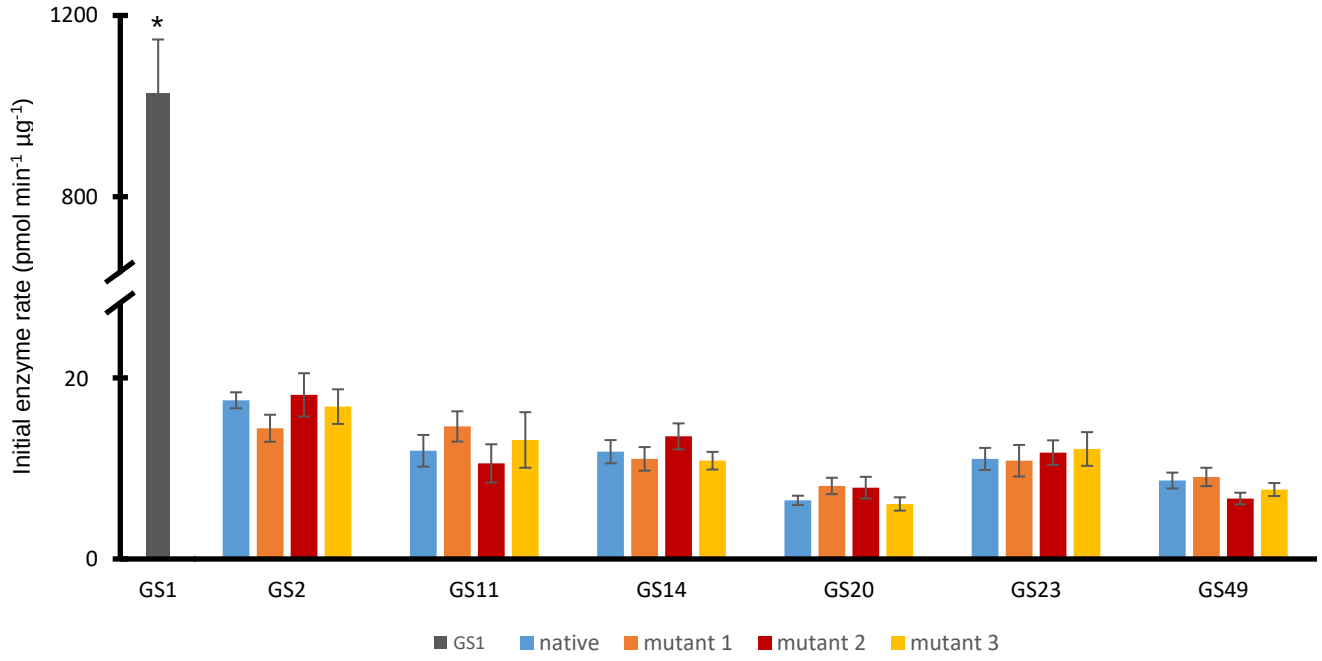


Figure 4.11: Residue selection for mutagenic strategy. Alignments of multiple GS amino acid sequences indicating key residues mutated in this study. (A) Residues in the glutamic acid binding pocket (red arrows). (B) Residues in the glycine binding pocket (red arrows). The amino acid numbering is based on the potato GS (StGS) sequence.

Table 4.7: Mutated residues of selected non-canonical GS enzymes. The amino acid numbering is based on each GS sequence. NC represents no change required in this position.

	Region 1				Region 2		Region 3	
StGS	Q	E	N	Q	R	Y	A	A
GS2	NC	L295E	P297N, C299, H300 were deleted	NC	A358R	NC	C556A	E557A
GS11	QPDLFPVCAFEQ (254 to 265) was substituted by 'QAEERNMYDQ'				A321R	NC	S519A	T520A
GS14	M244Q	Q247E	L251, H252 were deleted	NC	NC	L310Y	NC	P508A
GS20	G224Q	T247E	HVELEV (229 to 234) was substituted by 'NMYDQ'		NC	F284Y	S480A	H481A
GS23	R265Q, L260, A261, G262 were deleted	NC	G270N	T274Q	NC	M331Y	L525A	S526A
GS49	SKIGS (305 to 309) was substituted by 'QAEE'		NC	NC	NC	H368Y	S565A	C566A
GS55	R241Q	S244E	T246N	D250Q	NC	F303Y	M501A	K502A

A



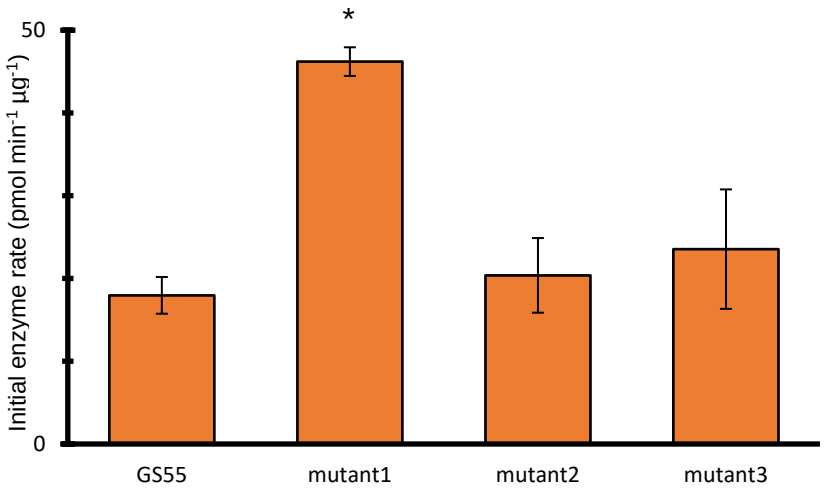


Figure 4.12 Effects of mutants in substrate binding sites. (A) Initial enzyme rates of native Rre-GS1, Rre-GS2, Rre-GS11, Rre-GS14, Rre-GS20, Rre-GS23, Rre-GS49, and corresponding mutants. * indicates significant difference between variants (One-way ANOVA, n=4). (B) Initial enzyme rates of Rre-GS55, GS55 Variant 1 (R241Q, S244E, T246N, D250Q), GS55 Variant 2 (F303Y) and GS55 Variant 3 (M501A, K502A). Error bars indicate the Standard Error of the Mean. * indicates significant difference between variants (One-way ANOVA, n=4).

Table 4.8: Kinetic parameters of Rre-GS55 and corresponding Rre-GS55 mutants for γ-EC, ATP and glycine, respectively. The amino acid numbering is based on the Rre-GS55 sequence.

Enzyme	Substrate	K_m
GS55	γ-EC	1.41 ± 0.18 mM
	ATP	791 ± 91.4 µM
	Glycine	1.87 ± 0.31 mM
Variant 1 (R241Q, S244E, T246N, D250Q)	γ-EC	0.62 ± 0.11 mM
	ATP	383 ± 43.5 µM
	Glycine	1.91 ± 0.21 mM
Variant 2 (F303Y)	γ-EC	1.33 ± 0.22 mM
	ATP	756 ± 65.8 µM
	Glycine	1.82 ± 0.28 mM
Variant 3 (M501A, K502A)	γ-EC	1.23 ± 0.21 mM
	ATP	695 ± 51.7 µM
	Glycine	1.63 ± 0.26 mM

4.5 Discussion

4.5.1 *R. reniformis* GS displayed distinctive biochemical activity.

As described above, phylogeny, temporal and spatial expression patterns of *R. reniformis* GS family members suggested a functional diversity. Here, distinct biochemical activities were also indicated among each *R. reniformis* GS clades.

First of all, *R. reniformis* GS1, the only sequence in Clade 1, displayed 1028 (± 349) pmol min^{-1} μg^{-1} initial enzyme rate by calculating the phosphate release speed, which is consistent with a previous report for *C. elegans* GS of 1860 pmol min^{-1} μg^{-1} (Buzie and Enjuakwei 2007) and several other eukaryotic GS (Meierjohann, Walter and Muller 2002). By contrast, the Clade 2 & 3 GS witnessed at least a 60-fold reduction in GS synthetic activity (Figure 4.5). Given the hypothesis that Clade 2 & 3 GS may play a different role than GS1, we speculate the Clade 2 & 3 GS probably gain a non-canonical function to produce alternative products. In addition, even though GS1 is considered to be a typical GS enzyme, the initial enzyme rate of GS1 is much lower than human GS (6010 pmol min^{-1} μg^{-1}) (Dinescu *et al.* 2004) and *A. thaliana* GS (7910 pmol min^{-1} μg^{-1}) (Jez and Cahoon 2004). Moreover, *G. pallida* Clade 2 GS exhibited much higher initial rate than *R. reniformis* forms and also stronger affinity to γ-EC (Lilley *et al.* 2018). Given that only one *G. pallida* Clade 2 GS and two *R. reniformis* ones were investigated, this huge difference between the activity of *G. pallida* and *R. reniformis* Clade 2 GS may be due to the untypical Clade 2 GS selected in this study. Otherwise, this suggests Clade 2 GS still maintain partial GS activity.

Kinetic parameters also support the diversity in GS function. K_m values of the Rre-GS1 enzyme are in close agreement to those of *C. elegans* GS (Buzie and Enjuakwei 2007) but are much higher than those of some plant GS such as Arabidopsis GS (Jez and Cahoon 2004) and potato GS (Lilley *et al.* 2018) for γ-EC and ATP, indicating that *R. reniformis* GS1 has similar biochemical characteristics with *C. elegans* GS rather than plant GS. Compared with Rre-GS1, K_m values of Clade 2 & 3 GS enzymes for γ-EC and ATP showed a nearly 10-fold and 5-fold increase, respectively. However, the K_m values for glycine among all *R. reniformis* GS were similar (Table 4.2). Taken together, this indicated GS1 has a stronger affinity to γ-EC and ATP than all the Clade 2 & 3 GS enzymes. Additionally, given the fact that Clade 2 & 3 GS lack canonical enzyme activity, this also suggested all the Clade 2 & 3 GS enzymes are not sensitive

to glycine concentration at the saturating γ-EC and ATP conditions. In other words, Clade 2 & 3 GS enzymes may accept alternative substrate instead of γ-EC.

Product release from enzyme active sites is often reversible and rebinding is common in many enzyme systems (Cao and De La Cruz 2013). Liberated product(s) can effectively compete with substrate binding to enzyme active sites and inhibit enzyme cycling. Human GS is one such good example. Human GS is considered as an ideal model to study allosteric regulation. The enzyme is negatively cooperative towards γ-EC. In this way, when the first γ-EC substrate binds and glutathione forms, the substrate affinity of the second subunit of human GS decreases (Oppenheimer *et al.* 1979). However, the negative cooperative binding effect of γ-EC observed for human GS enzyme was not found in *C. elegans* GS (Buzie and Enjuakwei 2007).

4.5.2 An alternative substrate may be accepted by non-canonical GS

Such biochemical diversity is highly unusual among eukaryotic GS enzymes. As introduced above, kinetic analysis suggested an alternative substrate may be accepted in non-canonical *R. reniformis* GS enzymes, which is responsible for the lost typical GS activity.

The ATP-grasp superfamily is a highly variable protein family where ATP binding is conserved but insertion of secondary structure elements with different functions permits distinctive substrates binding (Lee, Redfern and Orengo 2007). The canonical GS enzymes are typical members of the ATP-grasp superfamily. Interestingly, most residues in the binding pocket of GS sequences including both canonical and non-canonical GS were highly conserved (Figure 4.8A), suggesting the non-canonical GS enzymes still belong to the ATP-grasp superfamily despite the extremely low rates of ATP turnover when *R. reniformis* Clade 2 & 3 GS enzymes were provided with normal substrates.

The canonical product glutathione consists of three amino acids: cysteine, glutamic acid and glycine. Residues in the cysteine binding pocket of *R. reniformis* GS were relatively conserved (Figure 4.8B). The first catalytic residue is a conserved arginine which plays an important role in interactions between the enzyme and the cysteinyl moiety of the substrate by a hydrogen bond (Polekhina *et al.* 1999). The other two corresponding residues are varied but prefer to be preferentially neutral. Furthermore, in the canonical GS enzymes, all the positions in the glutamic acid binding pocket were

highly conserved. These residues are preferably large and polar amino acids which form hydrogen bonds with the glutamyl moiety of GSH (Fyfe, Alphey and Hunter 2010). However, in the non-canonical GS enzymes, these positions of the glutamic acid binding pocket witnessed a significant diversification, in which some small and hydrophobic amino acids occupied the native positions. In addition, several short amino acid insertions were present in this region, which is likely to contribute to a failure of γ-EC to come into the substrate binding pocket. Similarly, two conserved, small, neutral and nonpolar amino acids were identified in the glycine binding pocket of canonical GS enzymes whereas varied amino acids were shown in these positions of non-canonical GS enzymes. Taken together, this supported the hypothesis that γ-EC, especially its glutamate portion, and/or glycine may not be accepted into *R. reniformis* Clade 2 & 3 GS enzymes.

Site-directed mutagenesis also supported this hypothesis. The effect of mutating residues in the first four glutamic acid binding pocket revealed varied contributions to substrate binding and catalysis. GS55 mutant (R241Q, S244E, T246N, D250Q) displayed a significant increased initial enzyme rate while the other two GS55 mutants (F303Y; M501A, K502A) showed no changed enzyme activity, indicating that glutamic acid binding residues may play more important roles in the recovery of the canonical GS activity. Also, this suggested an alternative substrate other than γ-EC may be accepted in the binding pocket of non-canonical GS enzymes. Another possible explanation for the partial recovery of the canonical GS activity is due to the ‘better’ whole region that contributes to the structure of the binding pocket as Rre-GS55 variant 1 is the only one where just the four amino acid changed, with no deletions or changes of other amino acids. Accordingly, to test the functional significance of these positions, more substitutions at each position and corresponding multiple mutants should be generated in the future.

As introduced above, it is usual for some plant GS enzymes to exploit varied substrates. However, in these cases, the γ-EC carbon backbone is still normally used as a scaffold and the variation is restricted to the terminal amino acid (Skipsey, Davis and Edwards 2005). For example, the homo-glutathione synthetase of soybean catalyses the addition of β-alanine instead of glycine to γ-EC to produce homoglutathione. The two sequential alanine were previously identified as active residues in the glycine binding pocket of GS. However, in the homo-glutathione

synthetase of soybean, they are replaced by Leu487 and Pro488. A double mutation (L487A/P488A) can convert the substrate preference of hGS from β-alanine to glycine. In addition, structural comparison of hGS and human GS revealed that the Ala-rich loop in hGS which contains Leu487 and Pro488 is shifted to accommodate the longer alanine into hGSH (Galant *et al.* 2009).

In the *R. reniformis* non-canonical GS enzymes, the substitutions of these two sequential alanine in the glycine binding positions have also been demonstrated. However, the double mutants of *R. reniformis* non-canonical GS which replaced corresponding amino acids to alanine did not improves catalytic efficiency using glycine. In addition, the non-canonical *G. pallida* GS do not exhibit any preference to β-alanine and other terminal amino acids (Lilley *et al.* 2018). Taken together, these results indicate that the lack of canonical activity in the Clade 2 and 3 GS is not solely due to the use of an alternative terminal amino acid substrate.

Considering that the coordinating residues in the cysteine binding pocket of all the *R. reniformis* GS enzymes are relatively conserved, the alternative substrate(s) of non-canonical GS is likely to maintain a sulfydryl group. Using the mass spectrometric approach, around 300 sulfur metabolites were identified in *Arabidopsis* seedlings (Glaser *et al.* 2014). However, most of them remain uncharacterised and many of these could be small molecule thiols, providing a wealth of potential substrates but also making it extremely challenging to predict likely substrates of the non-canonical *R. reniformis* GS. High performance liquid chromatography analysis of thiols in poplar overexpressing a bacterial GS revealed two novel peaks, in addition to GSH. The peaks were particularly abundant in conditions in which leaf glycine contents were depleted (Noctor *et al.* 2012). Furthermore, a series of small molecular weight thiols with only small portion of glutathione were shown in the syncytia of *G. pallida* by HPLC analysis (Lilley *et al.* 2018). Taken together, these results suggest the novel substrate can be investigated via either generating transgenic plants expressing non-canonical GS genes or analysing thiol content in the syncytial material for *R. reniformis*.

4.6 Summary

1. *R. reniformis* GS1 from Clade 1 showed canonical GS activity (catalysing the formation of glutathione from γ-EC and glycine), whereas the other two clades GS exhibited extremely lower canonical GS activity.
2. Key kinetic parameters of *R. reniformis* GS for different substrates were calculated respectively.
3. By analysing all the *R. reniformis* GS amino acid sequences with some structurally solved GS sequences, active residues involving in ATP binding pocket, cysteine binding pocket, glutamic acid binding pocket and glycine binding pocket were identified.
4. Site-directed mutagenesis suggested it is first four glutamic acid binding residues that may be responsive for the lacking canonical GS activity.

Chapter 5

Crystal structural analysis of *R. reniformis* glutathione synthetases

5 Crystal structural analysis of *R. reniformis* glutathione synthetases

5.1 Introduction

5.1.1 An overview of protein crystallography

Over the last six decades, structural biology has provided a wealth of information that has contributed to a better understanding of biological structures and relevant functions (Shoemaker and Ando 2018). X-ray crystallography, electron microscopy and nuclear magnetic resonance are routinely exploited to solve the structure of macromolecules. As of June 28, 2019, there were 153328 total entries in the Protein Data Bank (PDB), of which 89.3% were determined by X-ray crystallography, 8.3% by nuclear magnetic resonance, and 2.2% by electron microscopy. However, there is no 'all-purpose' method as each technique has their unique advantages and disadvantages which are summarised in the Table 5.1.

Interpreting the X-ray diffraction data from many identical molecules in an ordered arrays like crystal is the most common experimental methods of obtaining a structural model of a protein macromolecule, which allows a great resolution of individual atoms (Rhodes 2010). Like small molecules, proteins can be crystallised, for structural determination by X-ray crystallography. When the incident X-ray beam bounces off identical crystal atoms, the scattered beams are collected on the detector, producing a diffraction pattern. As the wavelength of X-ray is already known and the crystal is gradually rotated, the angle and intensity of these scattered beams are able to be measured and the clouds of electrons (or the electron density map) in the molecules of the crystal is therefore interpreted. Based on this map, the average position of all the atoms in the crystal, chemical bonds, the angle and length of the bonds and other relative information can be determined (Giacovazzo *et al.* 2002). A typical protein crystallography pipeline includes crystals production, X-ray diffraction data collections and interpretations, phases determination, protein model building, structure refinement and a final model production (Shi 2014).

Since the first determination of the myoglobin crystal structure in 1957 (Kendrew *et al.* 1958), thousands of protein crystal structural models were built. As shown above, the estimated molecular weight of *R. renifomis* GS proteins are between 50-70 kDa, which is not suitable for electron microscopy and nuclear magnetic resonance (Table 5.1).

Therefore, X-ray crystallography is exploited for determination of the structures of *R. renifomis* GS proteins in this study.

Table 5.1 The comparison of X-ray crystallography, NMR and Cryo-EM

	X-ray crystallography	Electron microscopy	Nuclear magnetic resonance
Advantages	1. High resolution (1-3 Å) 2. Broad molecular weight range 3. Easy for model visualising and interpreting	1. Easy sample preparation 2. Showing the structure in native state 3. Small sample size	1. High resolution 2. Can provide information for secondary structure, dynamic study and identify side-chain motion
Disadvantages	1. Protein has to form a stable crystal that diffract well 2. Difficult and time-consuming for crystal production	1. Expensive 2. The resolution of Cryo-EM map is not high enough (>2.8 Å) 3. Applicable to samples of high molecular weights only	1. Need for concentrated solution that is difficult to prepare 2. Currently limited to small proteins 3. Difficult for model interpreting
Objects	1. Crystallisable samples 2. Soluble proteins, membrane proteins, ribosomes, DNA/RNA and protein complexes	1. >150 kDa 2. Virions, membrane proteins, large proteins, ribosomes, complex compounds	1. < 40–50 kDa 2. Water soluble samples

5.1.2 Structural overview of GS family

The crystal structures of a number of GS proteins from different species have been solved to date (Lilley *et al.* 2018). Figure 5.1 shows a structural comparison of some GS representatives. The GS family consists of two major groups: prokaryotic GS and eukaryotic GS. The first solved crystal structure of GS came from *E. coli* B at 2.0 Å under pH 6.0 condition (PDB: 2GLT), which can represent prokaryotic GS and be considered as the first member of the ATP-grasp superfamily (Yamaguchi *et al.* 1993). The crystal structure of *E.coli* GS showed that *E.coli* GS is a tetramer with four identical subunits and consisted of three major domains: the N-terminal, the central and C-terminal. The N-terminal domain mainly consisted of a six-stranded β-sheet sandwiched between two α-helices. The central domain consisted of a four-stranded anti-parallel β-sheet and two α-helices which were located on the same side of the β-sheet. The C-terminal consisted of a five-stranded anti-parallel β-sheet which was surrounded by three α-helices. In addition, the ATP binding pocket located in the cleft between the central and C-terminal and the ATP binding residues were surrounded by two set of motif which consisted of an anti-parallel β-sheet and a glycine-rich loop (Yamaguchi *et al.* 1993). As GS shows optimal catalytic activity at pH 7.5, a refined crystal structure *E. coli* GS was determined under the biochemically optimal condition (PDB: 1GSH) (Matsuda *et al.* 1996). The significant structural difference of this model is a ~0.35 Å movement of the central domain towards the N-terminal domain. As a result of this spatial movement, several new polar interactions between domains and subunits formed, contributing to a tighter dimer (Matsuda *et al.* 1996).

There is barely detectable sequence and structural similarity between the eukaryotic GS proteins and their bacterial counterparts although they catalyse the same reaction. Moreover, unlike tetrameric *E. coli* GS, the GS from eukaryotes such as human, yeast, plant, plant parasitic nematode and animal parasitic nematode is dimeric (Polekhina *et al.* 1999; Gogos and Shapiro 2002; Galant *et al.* 2009; Lilley *et al.* 2018; Fyfe, Alphey and Hunter 2010).

The crystal structure of human GS (PDB: 2HGS) was determined in complex with ADP, two magnesium ions, a sulfate ion and glutathione. The human GS monomer is a compact molecule with the shape of a flat, equilateral triangle. The main structural units were an eight-stranded β-sheet together with α-helices packing on either side of the sheet. In addition, a domain named the lid because of its role in providing access

to the ATP-binding sites, consisted of a four-stranded anti-parallel β-sheet with three α-helices packing on one side (Polekhina *et al.* 1999).The ligands including ADP, magnesium ions, sulfate ion and glutathione are bound to a central cavity on one side of the molecule. This cavity is surrounded by three loops, with the first loop (S-loop) playing a role in binding with glutathione and the other two named as the glycine-rich loop (G-loop) and the alanine-rich loop (A-loop) due to their amino acid composition (Polekhina *et al.* 1999).

As described above, various plant species produce glutathione homologs in which the terminal Glycine is substituted with a different amino acid. To understand the structural evolution and biochemical diversity of homoglutathione synthetase (hGS) from glutathione synthetase, the crystal structures of soybean hGS were solved at three separate states: the apoenzyme in an open active site conformation (PDB: 3KAJ); bound to γ-EC (PDB: 3KAK); and a closed form with hGSH, ADP, one sulfate ion and three magnesium ions bound in the active site (PDB: 3KAL) (Figure 5.1 C-E) (Galant *et al.* 2009). Similar with human GS, the overall structure of soybean hGS was also in a triangle shape. It mainly consisted of a smaller lid domain, a G-loop and an A-loop. The lid domain was formed by an anti-parallel β-sheet packing with two α-helices around the sheet. Furthermore, the crystal structures of soybean hGS under similar conditions in either the presence or absence of ligands indicated a domain movement and rearrangement of active site loop (Galant *et al.* 2009), supporting the hypothesis that enclosure of the active sites may prevent hydrolysis of the reactive acylphosphate intermediate (Herrera *et al.* 2007).

Except the GS proteins mentioned above, there are several another eukaryotic GS proteins that have been structurally determined, such as *S. cerevisiae* GS (PDB: 1M0T) (Gogos and Shapiro 2002), *Trypanosoma brucei* GS (PDB: 2WYO) (Fyfe, Alphey and Hunter 2010), *Solanum tuberosum* GS (PDB: 5OES) and *G. pallida* non-canonical GS (PDB: 5OEV, 5OEU, 5OET) (Lilley *et al.* 2018). In general, all these eukaryotic GS proteins shared structural similarities, as well as differences between eukaryotic and prokaryotic GS. All the eukaryotic GS proteins investigated to date contain at least the lid domain, the G-loop and the A-loop.

5.1.3 Structure of the G-loop and A-loop

Although all the eukaryotic GS share very low sequence identity (~10%-40%) between them, they have the same basic fold pattern and belong to the ATP-grasp superfamily of proteins. A common and defining feature of this family is possession of a very flexible glycine-rich loop that forms part of the ATP binding pocket (Galperin and Koonin 1997). The G-loop was indicated to be the most strictly conserved region in eukaryotic GS and play a key role in glycine and ATP binding (Dinescu, Anderson and Cundari 2007). Take human GS as an example, the main-chain amide of two residues in the G-loop was shown to interact with the phosphate oxygens of ATP. In addition, the main-chain nitrogen of Gly369 from the G-loop was indicated to be contacted with one sulfate ion (Polekhina *et al.* 1999).

The A-loop provides a cover over the active site cleft so that the loop will move to allow the substrates to come in to the active sites. At the same time, the A-loop closely contacts with the glycyl end of glutathione and interacts using main chain functional groups. In *T. brucei* GS, the amides of Val541 and Met542 of the A-loop interacted with glycyl end of glutathione (Fyfe, Alphey and Hunter 2010). In human GS, a similar interaction pattern was demonstrated between glutathione and the A-loop, although the residues concerned were Val461 and Ala462 (Polekhina *et al.* 1999).

5.1.4 Structure of substrate binding loop (S-loop)

Both A-loop and G-loop exhibited large catalytic loop motion during the catalytic cycle of GS, which manipulated access to the substrate binding pocket. Not as flexible as the G-loop and A-loop, the S-loop is relatively stable and rigid (Dinescu, Anderson and Cundari 2007). Residues of the S-loop in human GS were shown to form a wall of the active sites (F266-R267-D268-G269-Y270-M271-P272-R273-Q274-Y275-S276) and were considered to bind the substrate γ-EC (Ingle *et al.* 2019). In human GS, the γ-glutamyl moiety of glutathione formed a salt bridge with Arg267 and interacted with the N atom Arg267 by hydrogen bond. Moreover, the aromatic side chain of Tyr270 was in a position to form a hydrophobic face against the thiol moiety of glutathione. In addition, the main-chain oxygen of the cysteinyl moiety of glutathione contacted with the amide nitrogen of Ser151 and the side chain of Arg125 via hydrogen bonds, while the amide group of the cysteinyl moiety bonded with the main-chain of Ser149 (Polekhina *et al.* 1999). Similar situations have been described for other GS structures (Gogos and Shapiro 2002; Fyfe, Alphey and Hunter 2010). The structures of the S-

loop are highly conserved although the specific amino acids in this loop varies. In summary, the S-loop residues play a dominant role in γ-EC binding and the mutations in the S-loop impaired γ-EC binding significantly.

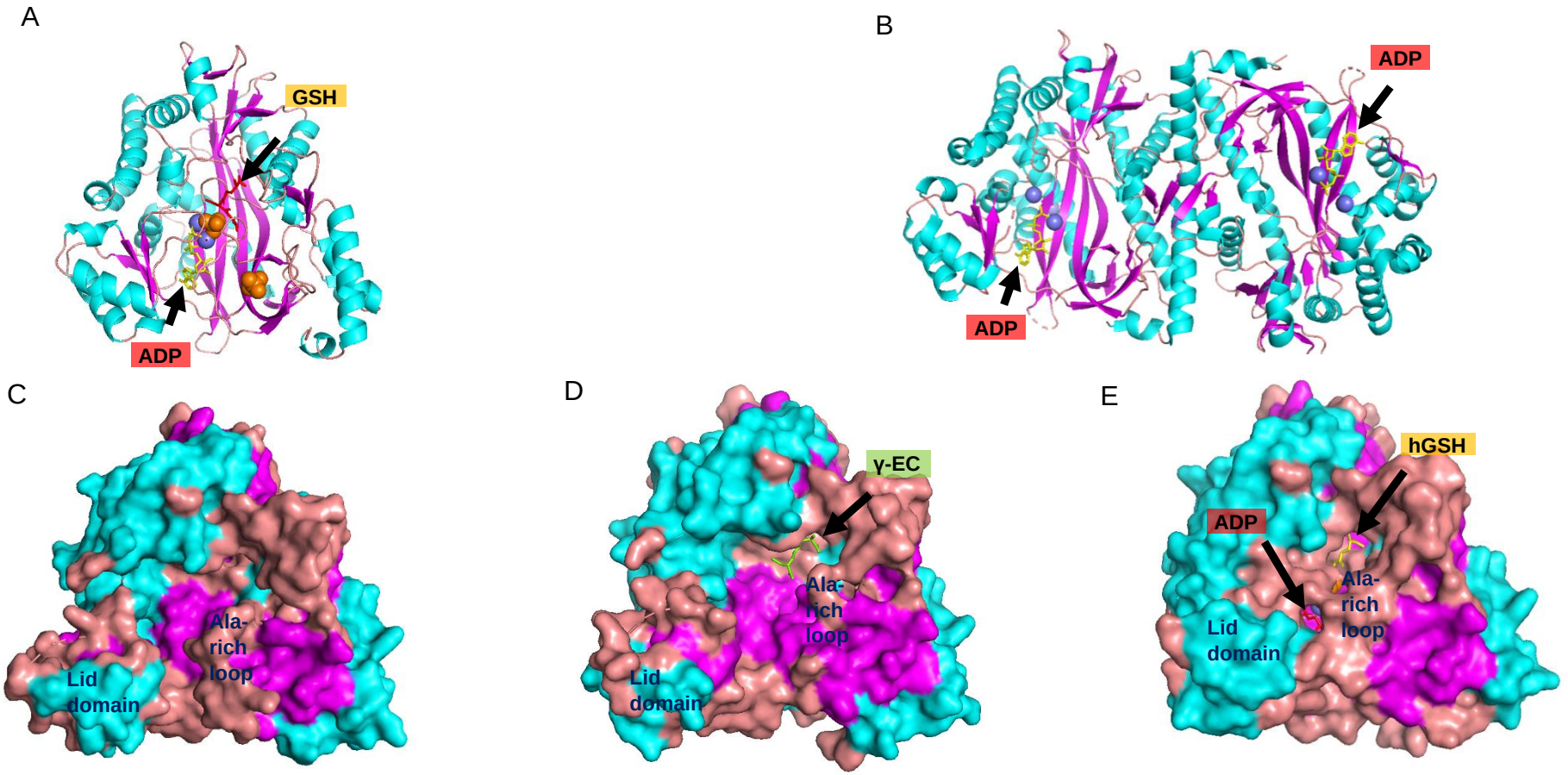


Figure 5.1: A comparison of some GS structural representatives. The structure is coloured by helix (cyan), sheet (magentas) and loop (light pink) (A) human GS (2HGS) bound with ADP (red), GSH (yellow), SO_4 (orange) and Mg^{2}+ (slate); (B) *G. pallida* GS (5OEU) in dimer bound with ADP (red) and Mg^{2}+ (slate); (C)-(E) Surface rendering of soybean hGS structure at the open form, the open form bound with γ-EC, the closed form bound with ADP and hGSH, respectively. Ala-rich loop and the lid domain enclosed the active sites when hGS bind the substrates.

5.2 Aims

1. To solve the structure of representative *R. reniformis* GS proteins.
2. To understand the active residues of *R. reniformis* GS proteins.
3. To explain the reason why non-canonical GS enzyme lacks canonical activity on the structural basis.

5.3 Materials and methods

5.3.1 Crystal production and screening

Proteins were produced as described in the previous chapter. Crystallisation screens were set up using 384 unique buffer conditions from the JCSG Core Suites (Molecular Dimensions, UK). One MRC Plate 96 well 3 Drop UV Crystallization Plate (Molecular Dimensions, UK) was exploited for each crystallisation screening experiment. Each well was filled with 30 µl of crystallisation buffer using a multi-channel pipette. Sitting drop crystallisation trials were then carried out using a NT8 robot via the RockMaker software (Forulatrix). Drops of 0.1 µl of ~10 mg/ml protein sample in 2:1, 1:1 and 1:2 drop ratios with mother buffer were mixed into the wells, respectively, meaning that it had variable effective concentration of protein and precipitant. The plate was then sealed with a ClearVue Sheet (Molecular Dimensions, UK) and stored at constant 20 °C in a RockImager 1000 (Formulatrix) that will perform automated imaging of the drops over several weeks. Both normal visible light images and Ultraviolet Two-Photon Excited Fluorescence (UV-TPEF) images were taken to identify protein crystals. UV-TPEF indicated protein crystals as proteins will absorb UV light while salt crystals will not. Crystallisation conditions and screens could be further optimised using a range of concentrations of the various precipitants, additives, or salts. If required, a 96-well additive screen HR2-138 (Hampton Research) was exploited for crystal optimisation.

5.3.2 X-ray diffraction data analysis

A number of protein crystals in good shape and size were fished from the wells using a nylon loop mounted on a cryo pin (Hampton Research, UK) and then were submerged in to 1 µl of mother buffer and appropriate cryoprotectant. Crystals were immediately immersed in liquid nitrogen for storage and transport. All X-ray data was collected at Diamond light source at Oxford, using various beamlines.

The X-ray data was integrated by DIALS (Winter *et al.* 2018). The integrated data was analysed using the CCP4I2 pipeline (Winn *et al.* 2011). A model of human GS (2HGS) monomer (Chain A) was used as a molecular replacement model as human GS shows highest sequences identity to *R. reniformis* GS among all structurally solved GS. Molecular replacement was carried out using PHASER (Mccoy *et al.* 2007). Maximum likelihood refinement was carried out on the structure using REFMAC5 (Murshudov *et al.* 2011). Initial model was auto built by BUCCANEER (Cowtan 2006) and Phenix

(Adams *et al.* 2010). Real space refinement and manual model building was then performed using sigmaA-weighted maps in Coot (Emsley *et al.* 2010). Further iterative rounds of restrained maximum likelihood refinement and real-space model building was used to build the partially disordered region and add water molecules to the model. Validation of the model was carried out using Molprobity (Chen *et al.* 2010). The figures of the structures were drawn with PyMOL (DeLano 2002).

5.4 Results

5.4.1 Crystal trails

As large amounts of *R.reniformis* GS proteins (Rre-GS1, -GS2, -GS11, -GS14, -GS20, -GS36, -GS49, -GS55) were produced, purified, concentrated and used to set up trays in 384 unique crystallography screening buffers from the JCSG Core Suites (Molecular Dimensions, UK). Although a total of 12 *R. reniformis* GS proteins representative of all three Clades has been screened for crystallisation, only GS1 from Clade 1 and GS11 from Clade 2 formed ideal crystals. A representative selection of results seen in the wells of the crystal screen is shown in Figure 5.2. In order to investigate active residues of GS1, GS1 was also co-crystallised with 2.5 mM γ-EC as a substrate. Ideal crystals were acquired and density was present in the active site where γ-EC was expected, however, that it could not unambiguously resolve the presence of the substrate.

GS1 was quick to crystallise in various crystallisation conditions, however mostly forming crystals with poor quality as shown in Figure 5.2. An initial condition containing 0.2 M potassium sodium tartrate and 20% w/v PEG 3350 were selected for optimisation based on the size and the shape of individual crystal. A 96-well optimisation screen containing different additive buffer (27 µl mother liquor plus 3 µl additive) was set up. Large crystals showed a preference in mother liquor with 0.01 M ethylenediaminetetraacetic acid disodium salt dihydrate (Figure 5.3). Crystals were fished using cryo loops and crystals dipped in a drop containing mother liquor and 25% v/v glycerol as a cryo-protectant before being flash frozen in liquid nitrogen. The samples were sent to Diamond synchrotron (Oxford) for remote data collection using beamline I04.

GS2 was slow to crystallise. Only one crystal hit was obtained for GS2, where crystals grew in a condition containing 0.2 M Lithium sulfate, 2.0 M Ammonium sulfate and 0.1 M CAPS, at pH 10.5. Crystals in good shape were fished and soaked in a drop containing mother liquor and 25% v/v glycerol as a cryo-protectant before being flash frozen in liquid nitrogen. The samples were sent to Diamond synchrotron (Oxford) for remote data collection using beamline I04-1.

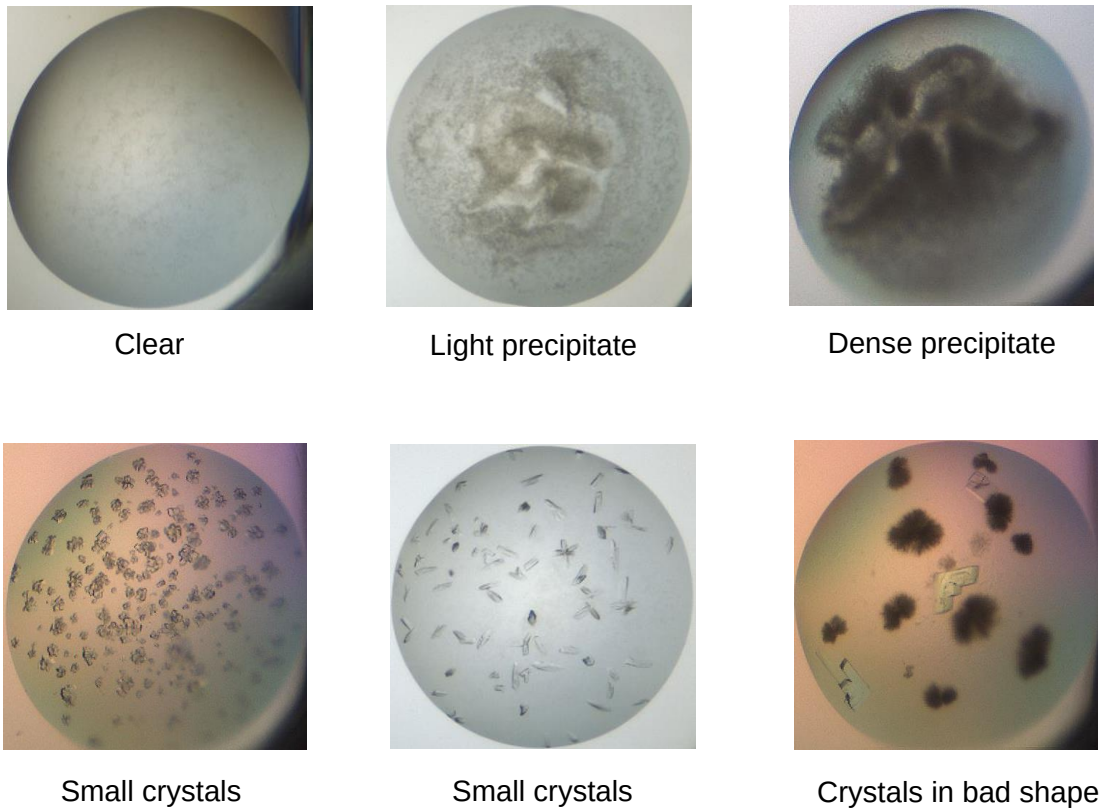


Figure 5.2: A range of images of representative results from screening conditions for growth of crystals.

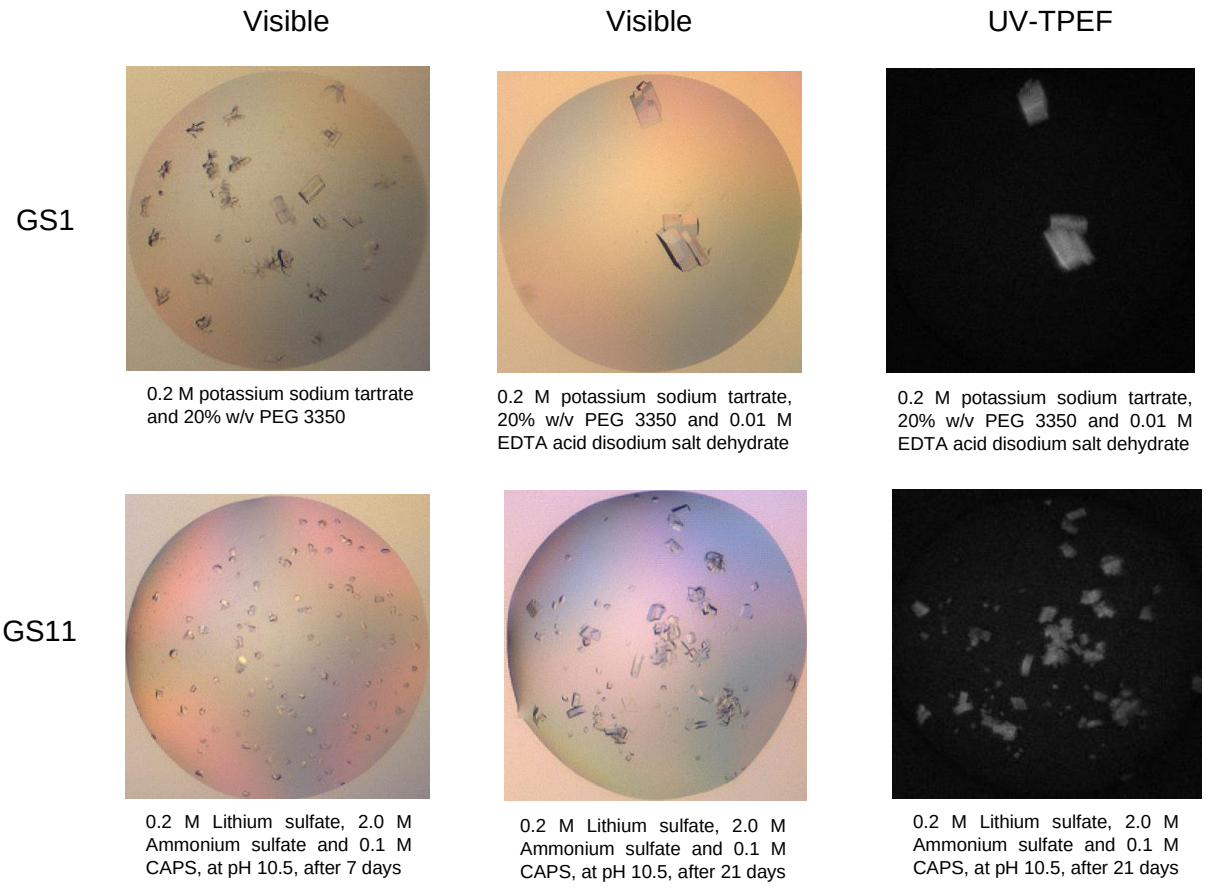


Figure 5.3: Crystal optimisation. GS1 crystals were obtained from initial condition (0.2 M potassium sodium tartrate and 20% w/v PEG 3350) and then optimised by addition of 0.01 M ethylenediaminetetraacetic acid disodium salt dehydrate. Small GS11 crystals were obtained at 0.2 M Lithium sulfate, 2.0 M Ammonium sulfate and 0.1 M CAPS, at pH 10.5 at 7 days. Ideal crystals for GS11 were obtained at the same condition at 21 days. UV images were used to confirm the protein crystals.

5.4.2 The crystal structure of GS1

5.4.2.1 Overview of GS1 structure

The structure of GS1 without any ligand (termed GS1-apo) was determined to 2.35 Å resolution. Subsequent data reduction found the space group to be C 1 2 1 with a solution probability of 86.5 %. The structure of GS1-apo was solved by molecular replacement using the structure of the human GS enzyme (PDB code: 2HGS) which shared 41.4% sequence identity with GS1, as the search model. A unique solution was found after molecular replacement, with four molecules in the asymmetric unit. The solution was then rebuilt and refined automatically by BUCCANEER (Cowtan 2006) and Phenix (Adams *et al.* 2010), and manually in Coot (Emsley *et al.* 2010). Data processing and refinement statistics are listed in Table 5.2. The refinement statistics and model geometry showed that the refinement has produced a flexible model of acceptable quality, with final R-factor and R-free as 0.21/0.27 respectively and no Ramachandran outliers observed.

In the unit cell, four molecules were present in the asymmetric unit, forming two dimers consisting of subunits A with B, and C with D (Figure 5.4A). The interface was mainly formed by a set of α-helices and an anti-parallel β-sheet (β1*). The inter-subunit contacts in the dimer were extensive and intimate, with hundreds of hydrophobic interactions between the strands of each monomer. In addition, hydrogen bonding interactions were shown between the side chain of Asp64 and the amide nitrogen of Phe448, the amide oxygen of Asp64 and the amide nitrogen of Val68, the side chains of Asp45 and Arg248, the side chains of Asp45 and Gln244, and the amide oxygen of Cys66 and the amide nitrogen of Cys66. A number of water-mediated interactions and van der Waals contacts were also discovered within the interface. Moreover, the residues involved in dimer interactions were not conserved among eukaryotic GS enzymes. Given the fact that the dimer promoted considerable stability for the molecules, these residues may play a role in GS function, which acts in an independent way with the substrate binding sites as the dimer interface was located far away from the substrate binding pocket.

These four subunits were very similar in structure when they were superimposed with SSM (Krissinel and Henrick 2004). Subunit A has five disordered sections (Lys19-Ala33, Gln134-Gly137, Glu158-Ala161, Arg377-Glu380, and Asn493-Val499). Subunit B has five (Gly22-Ala33, Glu158-Pro163, Leu402-Gly411, Leu421-Pro423, and

His491-Gly500). Subunit C has five (Ala18-Glu34, Thr157- Pro163, Glu376-Arg395, Gln401-Ala427, and His491-Gly500). Subunit D has eight (Lys19-Ala33, Tyr132-Gly137, Thr157-Gln164, Asn279-Arg286, Glu376-Leu390, Glu403-Gly407, Leu414-His429, and 492-Gly500). Compared with subunit C and D, subunit A and B were more complete in the asymmetric unit. The following discussions will be based on the subunit A. The water molecules were assigned if the distance of hydrogen bonding between relevant functional groups fell in the range of 2.5-3.2 Å.

The overall structure of GS1-apo, shown in Figure 5.4B, displayed a classic feature of eukaryotic GS members. The core structure of each monomer was a triangular fold that is around 60 Å × 60 Å in length and width. The GS1 monomer consisted of two major domains: a larger ‘core’ domain and a smaller lid domain (Figure 5.4B). The core domain was formed by a four-stranded anti-parallel β-sheet (β3, β4, β14 and β15) and two sub-domains positioned on either side of the sheet (Figure 5.4B). One of the sub-domain consisted of four parallel (β5, β6, β9 and β10) and two anti-parallel (β7 and β8) β-sheets enclosed by several α-helices. Another sub-domain was formed by a three-stranded anti-parallel β-sheet (β2, β13 and β16) surrounded with three α-helices (α1, α5 and α7).

The lid domain (residues 383 to 433) consisted of an anti-parallel β-sheet (β11 and β12), three α-helices (α15, α16 and α17) and a glycine-rich loop (residues 401 to 411). The anti-parallel β-sheet formed one lid of active sites with a further three α-helices exposed on the protein surface. Generally, the lid domain was poorly resolved in the electron density map, with only a short section of α-helices and anti-parallel β-sheet observed in subunits B, C and D. Most of residues in the glycine-rich loop were missing in the subunits B, C and D. Subunit A showed better order in this region and secondary structure within the lid domain of subunit A can be assigned with confidence. However, the positioning of some side chains was also less clear. Superimposition of all the four subunits indicated that the lid domain was extremely flexible, which was previously described in other GS (Polekhina *et al.* 1999; Gogos and Shapiro 2002; Fyfe, Alphey and Hunter 2010) and relevant ATP-grasp fold enzymes such as biotin carboxylase (Thoden *et al.* 2000). The flexibility of the lid domain explained why the electron density map of this region was poorly resolved and supported this domain functioned as a lid over the active sites and aided in the orientation of the substrates.

Table 5.2 Data collection and refinement statistics for GS1 apo and GS2 apo.

	GS1 apo	GS11 apo
Source	Diamond i04	Diamond i04-1
Wavelength (Å)	0.97950	0.91587
Resolution range (Å) *	40.78-2.35 (2.41-2.35)	59.07-1.83 (1.88-1.83)
Space group	C 1 2 1	P 21 21 2
Unit-cell parameters (Å)	*a*=154.76, *b*=100.69, *c*=154.00, α=90.00, β=120.16, γ=90.00.	*a*=111.43, *b*=118.14, *c*=39.31, α=90.00, β=90.00, γ=90.00
No. of observed reflections	539971	230144
No. of unique reflections	85157	46485
Redundancy	6.3 (6.4)	2.6 (2.2)
Completeness (%) *	100 (99.8)	97.4 (93.7)
I/σ(I) *	8.5 (1.3)	11.6 (1.1)
R_{merge} (%)*	10.5 (113.9)	5.2 (112.7)
R_{pim} (%)*	6.9 (87.5)	3.8 (90.7)
Resolution range for refinement (Å)	40.75-2.35	67.00-1.90
R factor (%)	21.1	19.5
R_{free} (%) †	26.6	23.8
$CC_{1/2}$	0.997	0.999
No. of protein atoms	1778	497
No. of water molecules	341	152
R.m.s.d bond lengths (Å)	0.0085	0.0090
R.m.s.d bond angles (˚)	1.5864	1.5766
Ramachandran analysis, the percentage of residues in the regions of plot (%) ‡		
Favoured region	96.29	97.96
Outliers	0	0
PDB code	Not deposited	Not deposited

*Values given in parentheses correspond to those in the outermost shell of the resolution range.

† R_{free} was calculated with 5% of the reflections set aside randomly.

‡ Ramachandran analysis using the program MolProbity.

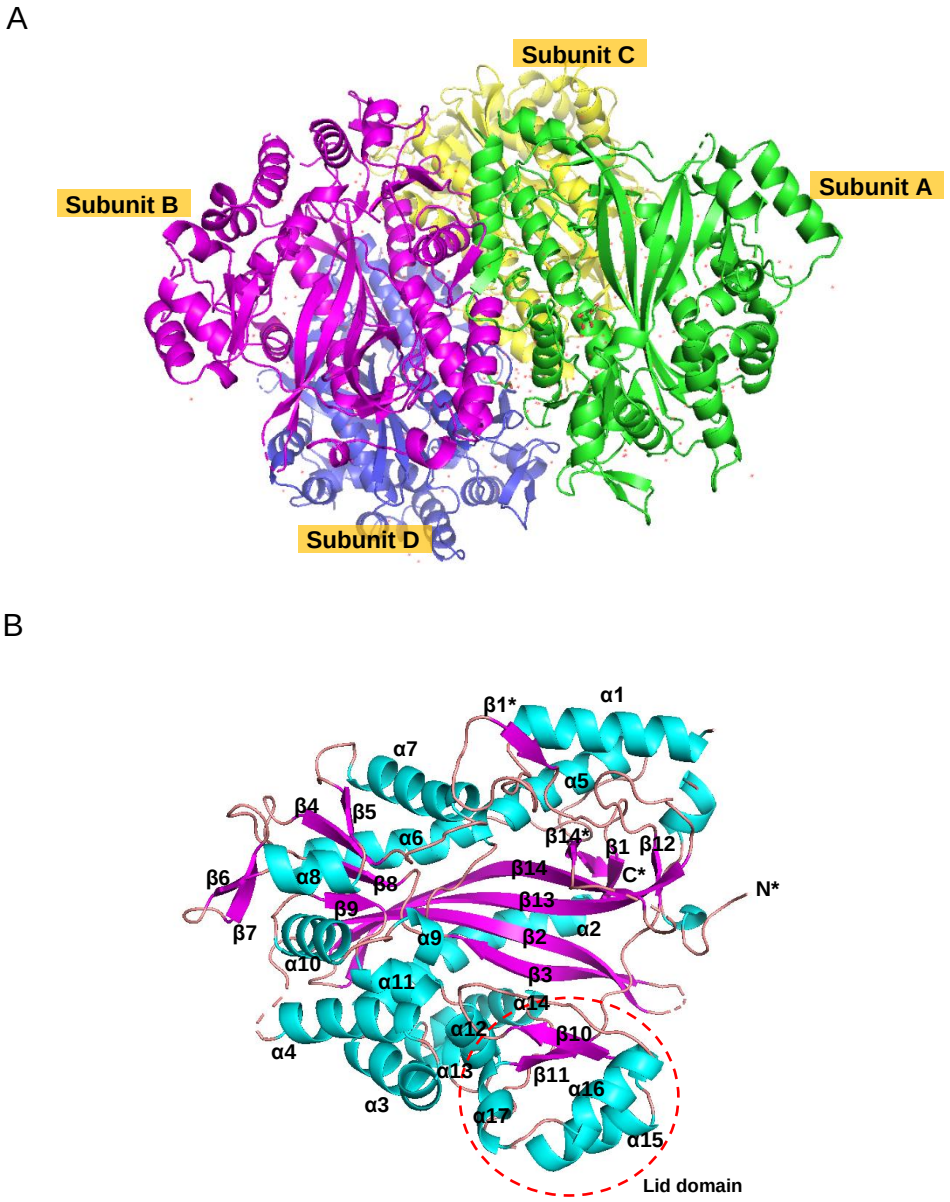


Figure 5.4: The structure of GS1. (A) Structure of GS1 dimer in the asymmetric unit. Green: subunit A. Magentas: subunit B. Blue: subunit C. Yellow: subunit D. Red spots: water molecules. (B) Overall structure of the subunit A of GS1. A ribbon representation of the monomer, indicating the location of secondary structure. Cyan: helix; Magenta: sheet; Light pink: loop. The asterisk denotes where the N- and C-terminals are located.

5.4.3 Active sites and the substrate binding

There was one central cavity on one side of the molecule that was covered by the lid domain and enclosed with four loops (residues 174 to 197, 297 to 302, 401 to 411, and 489 to 502). Previous reports suggested these loops played an essential role in interactions with the substrate (Dinescu, Anderson and Cundari 2007). Despite attempts to co-crystallise *R. reniformis* GS1 with the substrate γ-EC or ADP, these ligands were not observed. A density is present in the active site where γ-EC was expected. However, γ-EC is clearly incorrect to fit the density here (Figure 5.5).

The alignments of *R. reniformis* GS1 and potato GS (PDB code: 5OES) (Figure 5.6A) and other eukaryotic GS enzymes (not shown) showed high conservation in the putative active site regions. Consequently, superimposition of potato GS and *R. reniformis* GS1 by SSM was exploited to investigate the substrate binding sites of GS1.

As shown in Figure 5.6, based on the structural superimposition, the γ-EC and ADP molecules were predicted to be bound at one edge of the central anti-parallel β-sheet. γ-EC binding sites were positioned over the top of a loop linking to β9 to α10, with further interactions to residues from the loops that link β4 to α7 and β3 to α5. γ-EC formed extensive potential interactions with the protein, including seven hydrogen bonds and one hydrophobic bond (Figure 5.6B & C). In the cysteine moiety of γ-EC, the main-chain carbonyl oxygen of the cysteine portion likely formed hydrogen bonds with the amide nitrogen of Ser178 and the side chain of Arg148. The amide nitrogen of the cysteine portion likely formed a hydrogen bond with the amide oxygen of Ala176 (Figure 5.6B). In the γ-glutamyl moiety of predicted γ-EC, the main-chain carbonyl oxygens of γ-glutamyl moiety likely formed hydrogen bonding interactions with the side chain nitrogen of Gln247, Asn243 and Arg298. In addition, the amide nitrogen of γ-glutamyl moiety likely formed a hydrogen bond with the side-chain oxygen of Asn241. Moreover, the aromatic ring of Tyr301 probably formed a hydrophobic face against the γ-glutamyl moiety of predicted γ-EC (Figure 5.6C).

Because of the high similarity between ATP-binding pockets of Rre-GS1 and other eukaryotic GS members, ADP is predicted to interact with *R. reniformis* GS1 in a manner similar to that described in other GS and ATP-grasp proteins (Esser *et al.* 1998; St Maurice *et al.* 2007). The ADP molecule was predicted to be sandwiched with the four-stranded anti-parallel β-sheet and the lid domain based on the structural

superimposition (Figure 5.6A). All the residues that were predicted to bind the ADP were listed in the Figure 5.6D.The ADP binding pocket was largely hydrophobic with contributions from Met152, Ile170, Val397, Met432, Ile435, and the aliphatic portions of Lys399 and Lys434. Potential hydrogen bonding interactions were also found between ADP and the side-chain of Leu459, the amide oxygen of Gln433, the amide nitrogen of Ile435 and the side-chain of Lys488. In addition, the α and β phosphates likely interacted with the side chain of Leu171, Lys337, Asn408 and Arg486 by polar interactions.

Active site residues of human GS related to GSH glycyl moiety binding have been investigated (Polekhina *et al.* 1999). Accordingly, structural superimposition of Rre-GS1 and human GS was used to investigated potential residues in the glycine binding pocket. The Ala-rich loop and the Gly-rich loop of human GS were in close proximity to the glycyl portion of GSH and were shown to interact with the GSH glycyl moiety by main-chain functional groups. In the absence of glycine, the Ala-rich loop and the Gly-rich loop of *R. reniformis* GS1 exhibited their flexibility, which may be the reason why some regions of these loops were largely disordered. Although the sequence alignment suggested high conservation in these loops of *R. reniformis* GS1 and human GS, structural superimposition of these GS proteins indicated the Ala-rich loop the Gly-rich loop of *R. reniformis* GS1 extended far away from the predicted GSH binding location in the absence of glycine, leaving an open channel to the active site (data not shown).

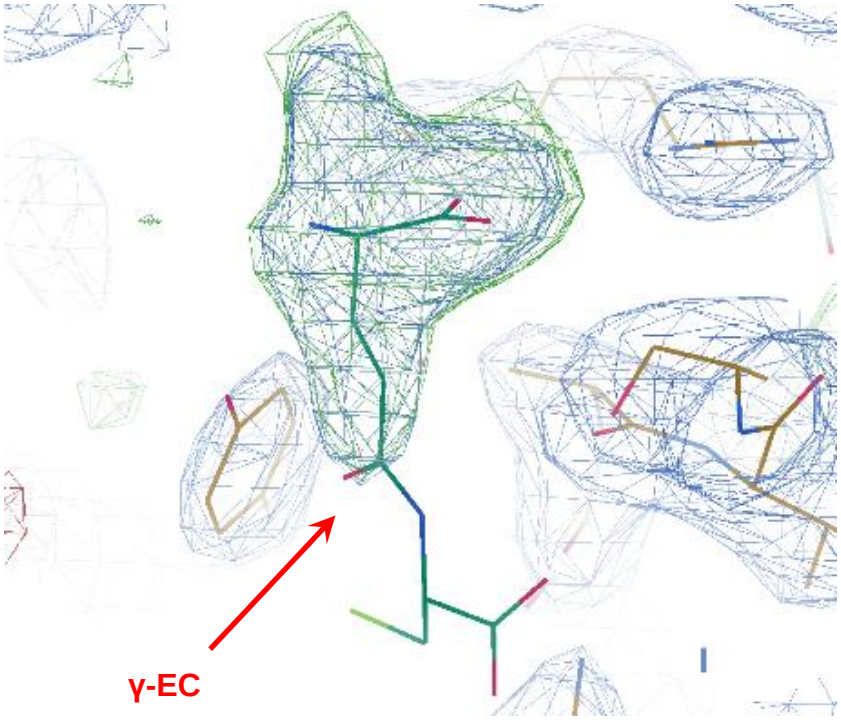


Figure 5.5: An unidentified blob of density at the active site. γ-EC is clearly not fit into the density (1.10 rmsd). Atom types are indicated by colors: red = oxygen; blue = nitrogen; yellow = sulfur; green and black = carbon.

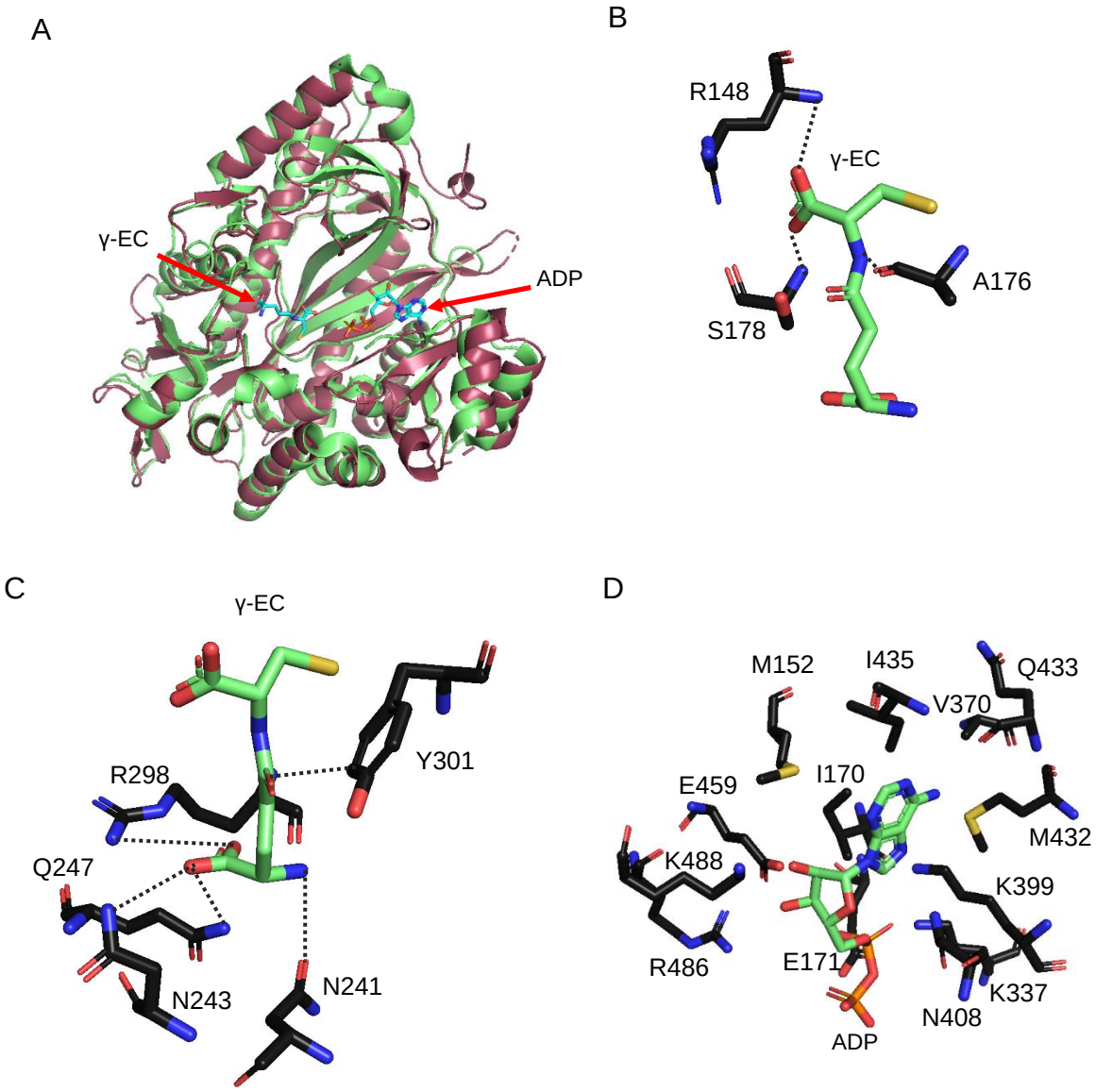


Figure 5.6: Substrate binding sites of *R. reniformis* GS1. (A) Superimposition of potato GS (green) and *R. reniformis* GS1 (raspberry). The γ-EC and ADP molecules (blue & orange) were predicted to be locate at one side of GS. (B)-(C) The potential γ-EC binding sites. Side chains of residues that form potential hydrogen bonds or hydrophobic bonds (dotted lines) with the bound ligand are shown. (B) Cysteine accommodation. (C) Glutamic acid accommodation. (D) The potential ADP binding sites. Atom types are indicated by colors: red = oxygen; blue = nitrogen; yellow = sulfur; green and black = carbon.

5.4.4 The crystal structure of GS11

5.4.3.1 Overview of GS11 structure

R. renifomis GS11 belongs to Clade 2 GS. The diffraction data for GS11 was collected to a resolution of 1.83 Å. Subsequent data reduction found the space group to be P 21 21 2. The structure of GS11 (termed GS11-apo) was solved by molecular replacement using the structure of the *G. pallida* GS-like effector in apoform (PDB code: 5OEV) which shared 39.3% sequence identity with GS11, as the search model. A unique solution was found after molecular replacement, with one monomer in the unit cell. The solution was then rebuilt automatically and manually, followed by refinements by Refmac5. Data processing and refinement statistics are listed in Table 5.2. The refinement statistics and model geometry showed that the refinement has produced a flexible model of acceptable quality, with final R-factor and R-free as 0.20/0.24 respectively and no Ramachandran outliers observed. The water molecules were assigned if the distance of hydrogen bonding between relevant functional groups fell in the range of 2.5-3.2 Å.

The overall structure of GS11-apo, shown in Figure 5.7, displayed a similar feature with GS1-apo, indicating *R. renifomis* GS11 is still a member of typical eukaryotic GS members although it lacked most of the canonical enzyme activity. The structure of *R. renifomis* GS11 had three disorder regions (Thr2-Leu29, Met423-Gly426, and Gly516-Gly522). However, it was better resolved compared with GS1-apo as it was provided higher resolution X-ray images. GS11-apo was also in a triangular shape with similar size to GS1 and was formed by two major domains: the core domain and the lid domain. For the sake of clarity, we have retained the secondary structure nomenclature used for GS1-apo, and have denoted elements unique to the GS11 structure with asterisks (Figure 5.7A and Figure 5.8). The core domain is comprised by two sub-domains placed on either side of a four-stranded anti-parallel β-sheet (β3, β4, β14 and β15). One of the sub-domain is formed by four parallel and three anti-parallel β-sheets enclosed with a set of α-helices, and the other by a two-stranded parallel β-sheet packed with four α-helices and two β-sheets. The smaller lid domain (residue391 to 455) consists of a three-stranded antiparallel β-sheet, three α-helices and a glycine-rich loop (residue421 to 428). Similar to GS1, the G-loop of GS11 was largely disordered and numerous residues in this region were not resolved. Besides the polypeptide chain, two CAPS molecules and one sulfate ion which came from the

crystallisation buffer were present in the structure (Figure 5.7A). One of CAPS molecule was assume to be bound to oxygen atom of the side chain of Leu373 by hydrogen bonds, while the other to be bound to oxygen atom of the side chain of Gln537 by hydrogen bonds and to form salt bridge with the side chain of Trp167. In addition, one sulfate ion was shown to form four hydrogen bonds with the side chain of Arg404 and Arg287 and two water molecules.

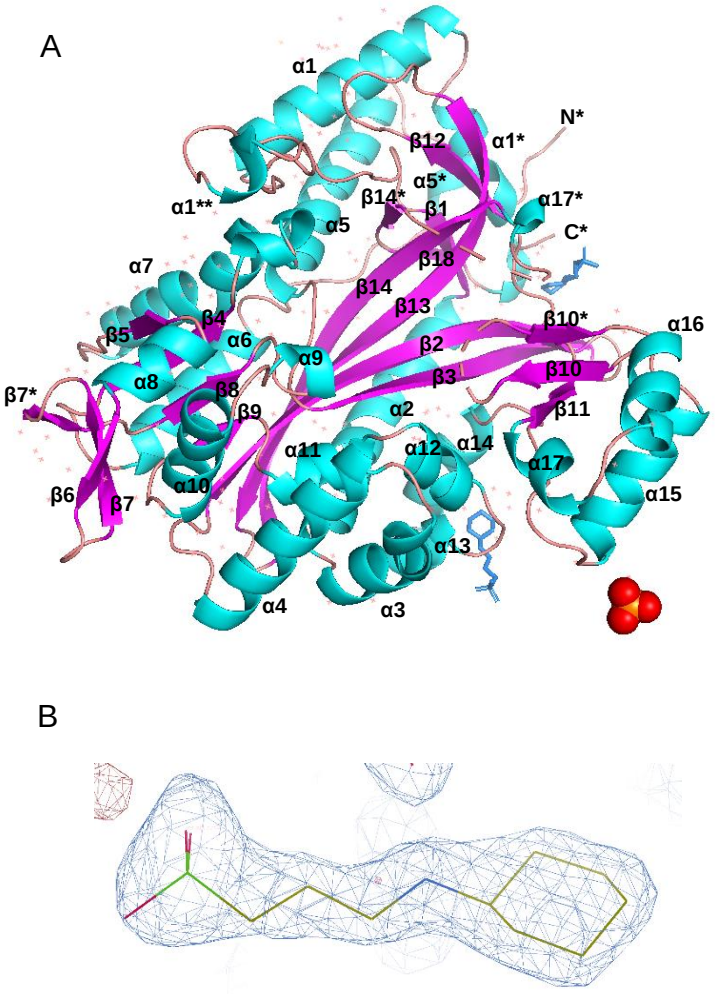


Figure 5.7: The structure of GS11. (A) Overall structure of GS11. A ribbon representation of the GS5 molecule, indicating the location of secondary structure. Cyan: helix; Magenta: sheet; Light pink: loop. The asterisk denotes where the N- and C-terminals are located. The blue molecule showed the positons of the ligand CAPS. The red and orange balls showed the position of sulfate ions. Both came from the crystallised buffer. (B) Sample electron density of CAPS molecule. The $2F_o$-F_c omit map (1.09 rmsd) for one CAPS molecule. Yellow: carbon; Green: sulphur; Red: oxygen. Light blue: nitrogen.

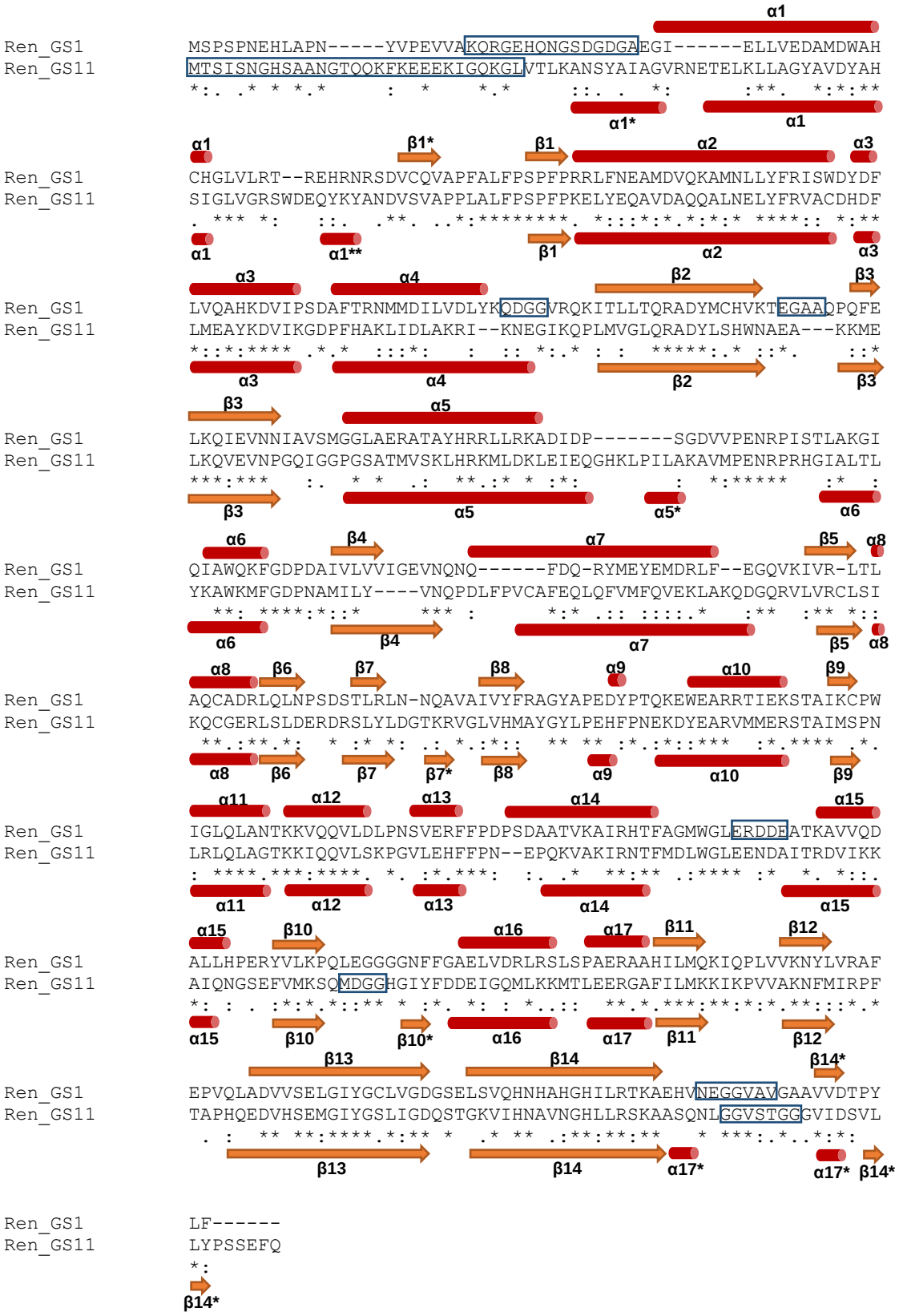


Figure 5.8: Sequence alignment of *R. reniformis* GS1 and GS11 with the sequences for *H.sapiens* GS. The assigned secondary structures of *R. reniformis* GS1 and GS11 were indicated, while those regions not resolved by well-ordered electron density (for subunit A) are enclosed in blue boxes. Strictly conserved residues were marked by star. α-helices were showed as red cans. β-sheets were showed as orange arrows.

5.4.5 Structural comparison of canonical and non-canonical GS enzymes

In order to investigate why GS11 lacked canonical enzyme activity, superimpositions of *R. reniformis* GS11 and some typical GS enzymes were performed. Human GS (*Homo sapiens*, PDB code: 2HGS) represented Animalia, yeast GS (*Saccharomyces cerevisiae*, PDB code: 1M0W) represented fungi and Potato GS (*Solanum tuberosum*, PDB code: 5OES) represented Plantae were used for the structural superimpositions. The overall structure of *R. reniformis* GS11 was highly similar with GS1 (Figure 5.9A) and other eukaryotic GS members (not shown), but also displayed some important differences. Conserved structures were only absent from one half of the acceptor di-peptide binding pocket and significant variability was shown at the glutamic acid binding pocket of γ-EC. Compared to the canonical GS, *R. reniformis* GS11 possessed an elongated loop formed by an additional five-residue insertion (Figure 5.9B), making this pocket too tight to let the substrate γ-EC access to the active sites. By contrast, all the canonical GS members shared highly conserved glutamic acid binding pocket (Figure 5.9C). In addition, similar with GS1, *R. reniformis* GS11 also had a flexible Ala-rich loop and Gly-rich loop because of the absence of glycine and ADP. Taken together, given the fact that residues in the ATP binding pocket of both canonical and non-canonical GS enzymes are highly conserved in sequence, these data supported the loss of canonical GS activity was associated with alternation of the substrate γ-EC. The novel substrate may own cysteine body and a smaller molecule than glutamate.

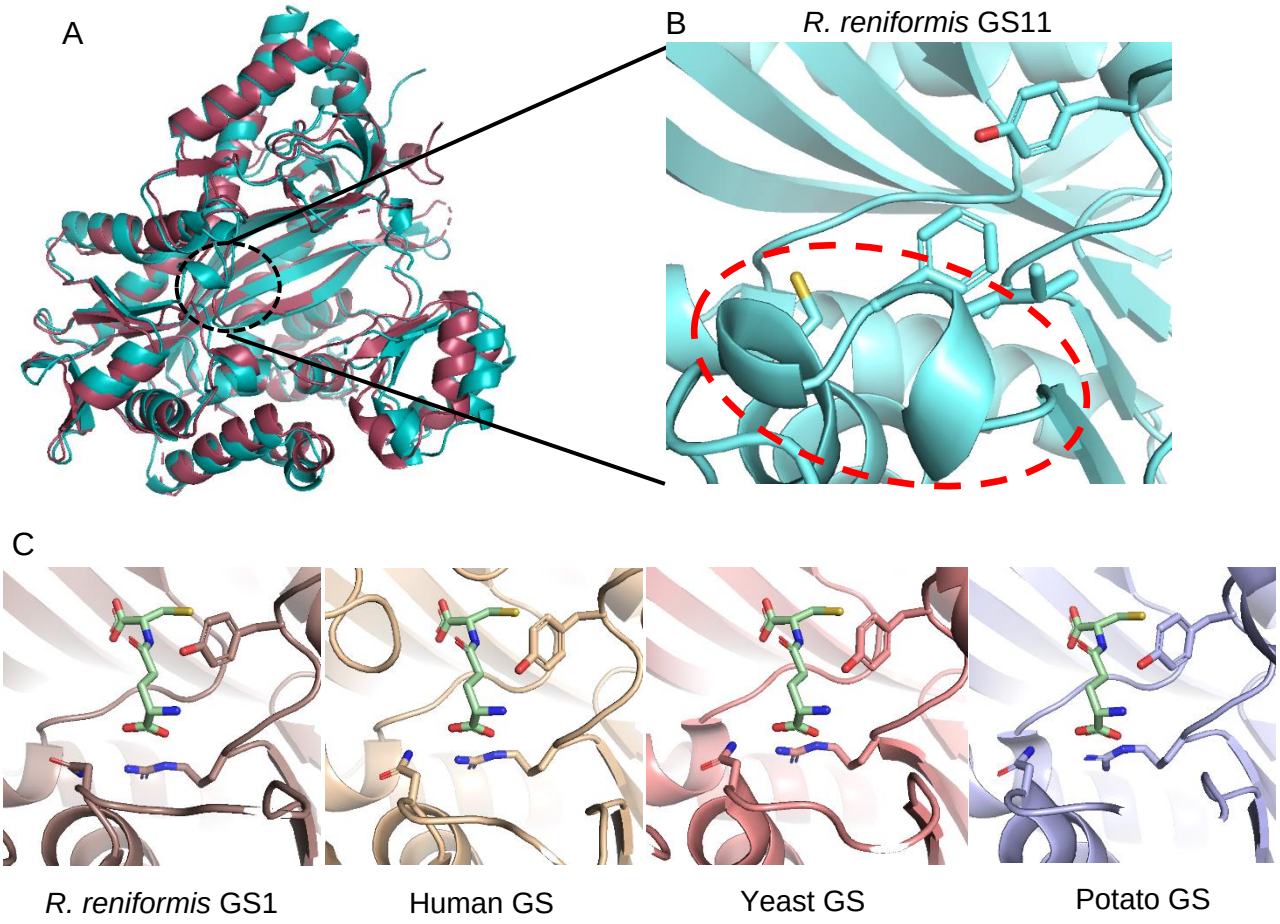


Figure 5.9: Structural comparisons of GS. (A) Superimposition of *R. reniformis* GS11 with GS1. Raspberry: GS1; Cyan: GS11. (B)-(C) Structural comparison of residues in the glutamic acid binding pocket of GS11 with the same positions of canonical GS. (B) The red dashed box shows the loop obstructing the active site in Rre-GS11. (C) The γ-EC molecule (green) was placed in the centre of the binding pocket. Canonical GS enzymes that represented different kingdoms showed high conservation in glutamic acid binding pocket, whereas the non-canonical GS displayed unusual arrangements in the same position. Atom types are indicated by colors: red = oxygen; blue = nitrogen; yellow = sulfur; green = carbon. (D) The alignment shows sequence diversification within the clades in the glutamic acid binding pocket (red dashed box) of the *R. renifomis* GS, suggesting more diversification of substrate specificity in these enzymes.

5.5 Discussion

5.5.1 The overall structures of *R. reniformis* GS enzymes

In order to investigate the reason for biochemical functional diversity across *R. reniformis* GS family, the crystal structures of GS1 (representing canonical GS) and GS11 (representing non-canonical GS) were determined with an acceptable quality. Generally, both *R. reniformis* GS members shared similar features with other eukaryotic GS like human (Polekhina *et al.* 1999). The core structure was a triangular fold that was formed by two major domains: a larger 'core' domain and a smaller lid domain (Figure 5.4B and Figure 5.6A). The core domain consisted of a four-stranded anti-parallel β-sheet enclosed in a set of α-helices and β-sheets, forming a backbone of GS enzyme. The lid domain was considered to be a flexible region, leaving an open channel for the substrate towards the central pocket. Disorder of the lid domain has been observed previously in GS (Gogos and Shapiro 2002; Fyfe, Alphey and Hunter 2010). In the structures of *R. reniformis* GS, the lid domain was also the most problematic region for model building. Some residues were unresolved and the positions of the side chain of some residues were not clear, making the Rfree values relatively high but within the acceptable range for this resolution. In addition, the lid domain was previously shown to undergo domain movements and rearrangements from an open active site form to a closed active site form (Galant *et al.* 2009; Lilley *et al.* 2018), which aided the ligands orientation in the substrate binding sites. The lid domain contained several conserved glycine residues, including a glycine-rich loop. Glycine was considered to provide flexibility necessary for the enzyme active sites to change conformation (Yan and Sun 1997). Despite no lid domain conformational change observed in the *R. reniformis* GS due to the absence of bound ligands, we have no reason to believe the lid domains of *R. reniformis* GS will not function similarly to other eukaryotic GS members.

In addition, the substrates γ-EC and ADP are predicted to locate at the one side of the GS backbone, enclosed by several flexible loops including the Ala-rich loop and the Gly-rich loop to form a central catalytic pocket. The Ala-rich loop was another disordered region in the structure and may play an important role in substrate binding.

5.5.2 The reaction mechanism of *R. reniformis* GS1

Due to no discovery of γ-EC and ADP in the active sites, the investigation of the reaction mechanism of Rre-GS1 has to rely on the structural superimposition of Rre-GS1 with other eukaryotic GS. The reaction mechanism of a canonical GS has been previously described: the C-terminal carboxylate of γ-glutamylcysteine was phosphorylated by the γ-phosphate portion of ATP to form an acylphosphate intermediate and release the ADP, followed by nucleophilic attack of glycine on the acylphosphate intermediate to form a tetrahedral carbon intermediate that dissociated to produce the glutathione and caused the release of inorganic phosphate (Hara *et al.* 1996). The crystal structure of *R. reniformis* GS1, which represented nematode canonical GS enzymes, fully supported this proposed reaction mechanism. Based on the structural superimposition of Rre-GS1 with other eukaryotic canonical GS, the active residues of *R. reniformis* GS1 involved in the substrate binding and ATP binding were highly conserved in sequence and position with other eukaryotic GS members, indicating these residues also likely play an essential role in canonical GS catalytic activity.

In addition, a lot of potential polar interactions were shown between the G-loop and the γ-phosphate moiety of predicted ATP, suggesting the G-loop played a significant role in stabilising the pentavalent phosphate intermediate during the phosphorylation step of the catalytic cycle. Also, potential hydrogen bonding interactions were found between the main-chain carbonyl oxygen of the cysteine moiety of predicted γ-EC and the amide nitrogen of Ser178 and the side chain of Arg148 (Figure 5.5B), indicating these two residues may have a role in stabilising the tetrahedral carbon intermediate. In the γ-glutamyl moiety of predicted γ-EC, many potential polar interactions were also identified between γ-EC and Gln247, Asn241, Asn243, Arg298 and Tyr301 (Figure 5.5C), supporting these residues stabilised the substrate γ-EC backbone during the catalytic activity.

Both *R. reniformis* GS members belonged to ATP-grasp enzymes as the active residues of ADP binding were highly conserved in sequence and position. Most of the residues were predicted to be involved in ADP binding (Val397, Lys399, Asn408, Met432, Gln433, Lys434 and Ile435) located at the lid domain (Figure 5.5D) and the lid domain formed a wall for the predicted ATP binding pocket. As introduced in other eukaryotic GS, the lid domain undergoes conformational changes when the ligands

bind. It is possible that this domain of Rre-GS1 moved during the catalytic cycle: the lid domain was located far away from the core of the molecule when no substrates or cofactors entered. The lid would move in to cover the active sites in the presence of substrates. The Ala-rich loop and the Gly-rich loop would also shroud over the active site cleft, forming a substrate pocket together with the lid. These two loops, the Ala-rich loop and the Gly-rich loop, were shown to have many potential interactions with the predicted substrates (Figure 5.5B & C), suggesting the loops may also move when the substrates enter and the products exit. Such lid domain movements were also revealed in other ATP-grasp enzymes such as biotin carboxylase (Thoden *et al.* 2000), DNA topoisomerase (Wei *et al.* 2005) and pyruvate carboxylase (St Maurice *et al.* 2007), which indicates domain motions may facilitate the transfer of ATP between active sites.

Of course, the reaction mechanism of Rre-GS1 is largely predicted according to the structural superimpositions of Rre-GS1 and other eukaryotic GS as no real γ-EC and ATP identified in the active sites. Given that Rre-GS1 is able to consume the canonical GS substrates: γ-EC and ATP, and also exhibited high structural conservation at all the active positons with other eukaryotic GS members, we believe Rre-GS1 exploits similar reaction mechanism to those of other eukaryotic GS during the catalytic cycle. Further investigation of Rre-GS1 reaction mechanism will use more trials and test more crystals to find the exact location of the substrates in the Rre-GS1, and how they interact with the enzyme.

5.5.3 Structural comparison of canonical and non-canonical GS enzymes

Non-canonical GS representative *R. reniformis* GS11 was superposed with several canonical GS members: Human GS (*Homo sapiens*, PDB code: 2HGS) represented Animalia, yeast GS (*Saccharomyces cerevisiae*, PDB code: 1M0W) represented fungi Potato GS (*Solanum tuberosum*, PDB code: 5OES) represented Plantae and *R. reniformis* GS1 represented nematode canonical GS enzymes. Despite the lack of significant sequence identity (~40%) between these eukaryotic GS members, they shared similar overall structures (Figure 5.8A), but also displayed some significant differences in the substrate binding pocket. Compared with the canonical GS enzymes, *R. reniformis* GS11 had an extra loop formed by five amino acids in the glutamic acid binding pocket (Figure 5.8B &C), which resulted in a narrower pocket that presumably prevent the substrate γ-EC accessing the active sites. By contrast, all the investigated

GS members shared highly conserved cysteine binding pocket. Taken together, these results supported the hypothesis that the non-canonical GS enzymes may exploit an alternative substrate rather that γ-EC, perhaps a cysteine containing compound, and probably still dependent on ATP.

Actually, the GS enzymes which used alternative substrates have been previously described. For example, soybean produces homoglutathione which the terminal glycine is replaced by β-alanine (Matamoros *et al.* 1999). The structure of hGS showed that two amino acid differences in an active site loop provided additional space to accommodate the longer β-alanine moiety of homoglutathione in comparison to the glycinyl group of glutathione (Galant *et al.* 2009). Similarly, this structural variation in the size of the substrate binding pocket may also provide a hint on the shape of the novel substrate. Given the fact that γ-EC is formed by two amino acids: glutamate and cysteine, the novel substrate may be consisted by cysteine and a smaller molecule than glutamate.

The phylogeny of *R. reniformis* GS family suggested GS1 was the progenitor sequence while Clade 2 & 3 GS represented the first and the second expansions, respectively (Figure 3.5). Accordingly, the non-canonical GS enzymes were hypothesised to evolve from the canonical GS enzymes, with a replacement of the substrate. This evolvement may cause a gain of novel function involved in nematode parasitism.

5.6 Summary

1. Crystal structures of two *R. reniformis* GS enzymes which represented canonical and non-canonical GS respectively were solved, with acceptable qualities.
2. Active residues involved in substrate and ADP binding were identified by structural superimposition with other eukaryotic GS members.
3. Structural differences between canonical and non-canonical GS were revealed, suggesting non-canonical GS may exploit alternative substrate rather than γ-EC.

Chapter 6

Functional analysis of *R. reniformis* glutathione synthetases

6 Functional analysis of *R. reniformis* glutathione synthetases

6.1 Introduction

6.1.1 Roles of glutathione in plant-pathogen interactions

Glutathione has been shown to play multiple crucial roles in plant development and responses to abiotic and biotic stress. In recent years, numerous studies have investigated the involvement of glutathione in plant-pathogen interactions.

Early in the 1980s, many reports revealed that the treatment of cultured plant cells with exogenous glutathione could induce the accumulation of plant defence-related proteins (Wingate, Lawton and Lamb 1988). Moreover, treatment with pathogen-derived elicitors was demonstrated to induce glutathione accumulation in plant tissues (Edwards, Blount and Dixon 1991). The Arabidopsis *PAD2* gene was shown to encode GCL which is involved in the first step of glutathione biosynthesis (Parisy *et al.* 2007). The Arabidopsis *pad2-1* mutant had a significantly reduced glutathione level and showed enhanced susceptibility to a broad range of plant pathogens, such as virulent bacterial strains of *Pseudomonas syringae* (Parisy et al. 2007), the oomycete pathogen *Phytophthora porri* (Roetschi *et al.* 2001), and the pathogenic fungus *Botrytis cinerea* (Ferrari *et al.* 2003). In addition, reduced expression levels of pathogenesis-related protein 1, oxidative stress-related genes and salicylic acid were shown in the *pad2-1* mutant (Dubreuil-Maurizi *et al.* 2011). Similarly, a clear link between glutathione metabolism and plant defence mechanisms was shown in other Arabidopsis glutathione-deficient mutants, *cad2-1* and *rax1-1* (Ball *et al.* 2004). Taken together, these studies highlighted the importance of glutathione in disease resistance of plants. Interestingly though, one report described an increase in homoglutathione and in relevant gene expression in root-knot nematode-induced root galls of *Medicago truncatula*. In addition, pharmacological depletion of glutathione content impaired nematode egg mass formation and modified the sex ratio of *M. incognita*, suggesting that glutathione has a key role in the regulation of giant cell metabolism and promotes the success of root-knot nematode infection (Baldacci-Cresp *et al.* 2012).

However, the details of how glutathione affects plant defence are still unclear. As plants lack mobile immunity cells, the basal resistance of each cell and effective signal transduction from infected cells are important for plant immunity. A change of redox status of the host plant is considered as a key response to attempted pathogen

invasion (Shetty *et al.* 2008). A massive burst of reactive oxygen species (ROS), hydrogen peroxide, reactive nitrogen species and nitric oxide were detected during pathogen infection and were shown to activate the downstream defence mechanisms (Matika and Loake 2014). As introduced in Chapter 1, under the environmental and cellular conditions that cause oxidative stress, particularly those that produce ROS, glutathione is a key moderator of cellular redox potential in many physiological processes, protecting cells from the negative oxidative environment (Galant *et al.* 2011). In this process, ascorbate and glutathione are positioned between oxidants, such as ROS, and cellular reductants, such as NADP/NADPH, to form a gradient of redox potential, which buffers oxidative changes resulting from ROS. Glutathione is oxidised to the disulfide form (GSSG) and recycled to glutathione (GSH) by NADPH-dependent glutathione reductase (Matika and Loake 2014; Noctor *et al.* 2012). Moreover, a wide range of glutathione conjugates can be formed while interactions with the nitric oxide system by formation of S-nitrosoglutathione, broaden the scope of glutathione as a reservoir of signalling potential in plant immunity (Lindermayr, Saalbach and Durner 2005). Thus, glutathione may play an important role in plant defence mechanisms by regulating redox status of a potential host.

6.1.2 Roles of glutathione in plant-beneficial microbe interactions

Glutathione also plays an essential role in interactions between plants and beneficial microbes. The best example can be described in nitrogen-fixing symbiosis. Legumes interact symbiotically with Rhizobiaceae to form root nodules, the nitrogen-fixing organs. An increased glutathione content was demonstrated in soybean root nodules, which was positively correlated with nitrogen fixation efficiency (Dalton *et al.* 1986). At the same time, glutathione-depleted plants, by both pharmacological and genetic approaches, showed lower nitrogen fixation efficiency and smaller nodules (Dalton, Langeberg and Treneman 1993). A correlation between glutathione level and nitrogen fixation efficiency has also been reported during the early stage of nitrogen-fixing symbiosis (Frendo *et al.* 2005), in mature nodules (El Msehli *et al.* 2011) and during the natural and stress-induced senescence of root nodules of other legumes (Matamoros *et al.* 1999). However, the role of glutathione in other interactions between plants and beneficial microbes is still not well defined. One study provides some evidence that glutathione level affected the growth of plant-associated fungus,

suggesting glutathione may be involved in mycorrhizal symbiosis (Ruíz-Sánchez *et al.* 2011).

6.1.3 Regulation of glutathione biosynthesis

Glutathione biosynthesis consists of two conserved chemical steps: GCL catalyses formation of γ-EC from cysteine and glutamate and GS catalyses the addition of glycine to γ-EC to produce glutathione. Considering all the vital functions of glutathione, complete knockout lines for either GCL or GS have lethal phenotypes and both GCL and GS are therefore considered as essential genes in most organisms. For example, homozygous knockout of mouse GCL or GS gene led to embryonic lethality, but heterozygous mice survived with no distinct phenotype (Dalton *et al.* 2000; Winkler *et al.* 2011). Similarly, glutathione depletion in homozygous *Arabidopsis* knockouts lacking GCL (*GSH1*) caused embryo lethality (Cairns *et al.* 2006), while insertion mutant lines of GS (*GSH2*) showed a bleached seedling-lethal phenotype after germination (Pasternak *et al.* 2008). This difference in phenotype may be due to partial substitution of glutathione functions by γ-EC which significantly accumulates in *gsh2* mutants (Pasternak *et al.* 2008).

Many factors play a role in the synthesis of plant GSH, but the most important is γ-EC activity (Noctor *et al.* 2012). Multiple experiments indicated that GCL is the rate-limiting enzyme in glutathione biosynthesis. Taking Arabidopsis as an example, over-expression and knockout of the Arabidopsis GCL gene *GSH1* resulted in significant increase (200%) and decrease (to 3%) of glutathione level, respectively (Xiang *et al.* 2001), while over-expression of the Arabidopsis GS gene *GSH2* barely affected glutathione content (Parisy *et al.* 2007). Also, the addition of immediate precursors of glutathione biosynthesis such as cysteine, glutamate, or glycine to Arabidopsis suspension culture cells did not improve glutathione biosynthesis (Meyer and Fricker 2002). Another factor that may affect plant GSH content is feedback inhibition of GCL by GSH. Alleviation of feedback inhibition is likely to be an essential mechanism driving accelerated rates of GSH synthesis under conditions in which GSH is consumed (Noctor *et al.* 2002). Therefore, under most conditions, GCL probably works at considerably less than maximal activity due to the feedback inhibition of GCL by GSH.

While much is known about GCL regulation, little attention has been paid to GS. Although GS is generally thought not to play a major role in the regulation of glutathione biosynthesis, and is not the rate-limiting step, there is accumulating evidence that GS is involved in manipulating overall glutathione synthetic capacity in certain tissues and under stressful conditions. Transgenic GS over-expressing Indian mustard plants accumulated higher concentrations of glutathione during exposure to heavy-metals like cadmium (Zhu *et al.* 1999). In addition, it is conceivable that when GCL is induced tremendously, the step catalysed by GS may become limiting. In rat hepatocytes, an increase in both GCL and GS expression further enhanced glutathione production above that observed with increased GCL expression alone (Huang *et al.* 2000).

As introduced previously, host GSH biosynthesis has a positive relationship with *M. incognita* and *H. schachtii* parasitism (Baldacci-Cresp *et al.* 2012; Lilley *et al.* 2018). Accordingly, plant parasitic nematodes may manipulate host GSH metabolism to promote their parasitism. In fact, several enzymes like glutathione peroxidase and glutathione S-transferase that play a key role in GSH metabolism have been identified as potential ‘effectors’ in many plant parasitic nematodes like *M. incognita* (Bellafiore *et al.* 2008) and *G. rostochiensis* (Jones *et al.* 2004). Interestingly, ‘effector-like’ GS were only identified from syncytium-forming nematodes such as *R. reniformis*, *G. pallida* (Lilley *et al.* 2018) and *H. glycines* (Masonbrink et al. 2019). In addition, unlike a typical plant GS, these ‘effector-like’ GS enzymes lacked canonical GS catalytic activity. As a result, the detailed roles of these secreted nematode GS during plant-nematode interactions are still unclear. In this final part of the work, a number of experimental approaches were undertaken to start to understand the role of these novel effectors in the host plant.

6.2 Aims

1. To reveal if the nematode GS can substitute for the function of plant GS *in planta*.
2. To understand the roles of *R. reniformis* Clade 1 & 2 GS in nematodes.
3. To investigate the functions of *R. reniformis* Clade 3 GS in nematode parasitism.

6.3 Materials and methods

6.3.1 Expression of *R. reniformis* GS proteins in Arabidopsis

6.3.1.1 Constructs for ectopic expression

The vector used for GS ectopic expression in Arabidopsis was based on the binary vector pBI121 (Chen *et al.* 2003) with some minor modifications (the GUS gene has been removed and a FLAG tag (amino acids sequence: DYKDDDDK) and a Kpn I site introduced). A map of the T-DNA region of the new vector is shown in the Figure 6.1. The transgene of interest is fused to the FLAG tag at its N-terminus and is expressed under control of the CaMV35S promoter. A typical cloning procedure for insertion of the genes of interest into the FLAG-tag pBI121 vector involved the following:

Each selected GS protein coding sequence with its stop codon but without the start codon and signal peptide (where present) was amplified from plasmid template by PCR using Phusion proof-reading enzyme (New England Biolabs, UK) and gene specific primers with the addition of appropriate restriction enzyme sites (Table 6.1). Both FLAG-tag pBI121 vector and the purified amplified GS gene fragment were digested with the relevant restriction enzymes at 37 °C for 3 hours. The quality of linearised vector and digested gene fragment were then examined by agarose gel electrophoresis. Once a clear single band was obtained in each case, the target bands were isolated from the gel and purified. Genes of interest and vectors with the same cohesive ends were ligated with T4 DNA ligase, followed by transformation into *E. coli* DH5α competent cells and downstream selection of transformants using kanamycin. After confirmation of successful GS sequence insertion by colony PCR, the positive single colonies were cultured in 5 ml LB medium and mini-prepped. Subsequent sequencing was performed using primers 35S1 and pBI seq R (Table 6.1). The correct plasmids were stored at -20 °C for further use.

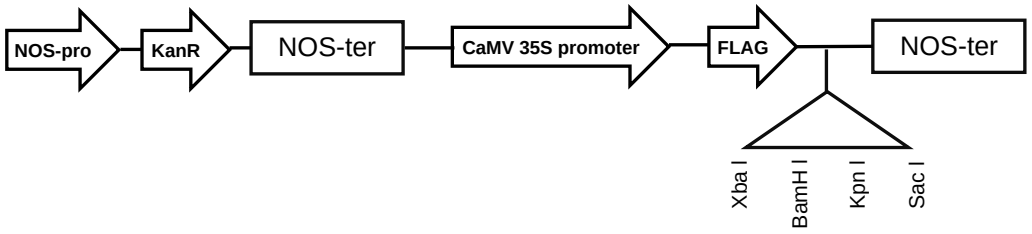


Figure 6.1. A map of the T-DNA region of the FLAG-tag pBI121 vector. The T-DNA region contains the nopaline synthase promoter (NOS-pro), aminoglycoside phosphotransferase gene to confer resistance to kanamycin (KanR), two nopaline synthase terminators (NOS-ter), the cauliflower mosaic virus 35S (CaMV 35S) promoter, a FLAG tag and a .multiple cloning site.

Table 6.1. Primers used in GS ectopic expression in Arabidopsis. The restriction enzyme sites and additional 5' bases to allow efficient digestion are in lower case.

Primer name	Seq 5'-3'	TM (°C)	Enzyme site
pBI-GS1-F	tgctctagaTCCCCATCACCGAACGAA	64	Xba I
pBI-GS1-R	acaggtaccCTAGTGATTTACAGCAACTCCTC		Kpn I
pBI-GS2-F	tgctctagaGTGGTGACACTCCCTCCCAA	67	Xba I
pBI-GS2-R	acagagctcTCATTCTTGGTGAAATTGGCTGG		Sac I
pBI-GS5-F	tgctctagaTCGATCACTGTGCTGAACAG	66	Xba I
pBI-GS5-R	acagagctcTCACTGGTGGAATTCGCGAG		Sac I
pBI-GS11-F	tgctctagaACATCGATCAGCAACGGACA	65	Xba I
pBI-GS11-R	acaggatccTCACTGAAACTCGCTAGACG		BamHI
pBI-GS14-F	tgctctagaGCCCATATTCCGGAAGGTAA	65	Xba I
pBI-GS14-R	acagagctcCTACACCAGGAAAGGCGAGT		Sac I
pBI-GS18-F	cgcggatccGAGGATGAGACACAAAAATCT	60	BamHI
pBI-GS18-R	acaggtaccCTAGTACAAGTACGGTGTGTC		Kpn I
pBI-GS20-F	tgctctagagaAGCTGATGCCGAAATAACT	61	Xba I
pBI-GS20-R	acagagctcCTAGTACAAGTACGGAGTGTC		Sac I
pBI-GS23-F	tgctctagaGGGCCTGTCGATGAAAATG	61	Xba I
pBI-GS23-R	acagagctcCTAATACAGGTATGCACTATCG		Sac I
pBI-GS36-F	tgctctagaGCTCCCACAAATTTAGCAG	59	Xba I
pBI-GS36-R	acaggtaccCTAGTAGAGATAGGGATTGTAG		Kpn I
pBI-GS44-F	cgcggatccATGAAATTGGTGCAAACCAA	61	BamHI
pBI-GS44-R	acaggtaccTCAGAACAGAAGGGGTGAAT		Kpn I
pBI-GS49-F	tgctctagaGTGCCAACCCACAAGGGG	68	Xba I
pBI-GS49-R	acagagctcCTAGACCACCAGGTATGGCG		Sac I
pBI-GS55-F	tgctctagaACTGAAGATGCTTCTACTGA	62	Xba I
pBI-GS55-R	acaggtaccCTACACAAGCAATGGTGAAT		Kpn I
pBI-GS57-F	tgctctagaACTAAAGATGCTTCTACTGATC	60	Xba I
pBI-GS57-R	acaggtaccCTATACGAGCAATGGCGA		Kpn I
pBI-GS61-F	tgctctagaCGCATTCTGATTGCGGACA	67	Xba I
pBI-GS61-R	acagagctcTCAGAACAGGTAGGGCGAGT		Sac I
pBI-GS67-F	tgctctagaCAACAAGACATCGAAGTGCA	64	Xba I
pBI-GS67-R	acagagctcTCAAATCAGCAACGGCGAAT		Sac I
pBI-GS72-F	tgctctagaACCCCGAGGGGAAATGAT	63	Xba I
pBI-GS72-R	acaggtaccCTAATAAAGGATGGGCGAATCAA		Kpn I
pBI-AtGS-F	cgcggatccGAATCACAGAAACCCATTTTCG	63	BamHI
pBI-AtGS-R	acaggtaccTCAAATCAGATATATGCTGTCCAAGA		Kpn I
35S1	GATGTGATATCTCCACTGACG	N/A	N/A
pBI Seq R	AACGACGGCCAGTGAATTC		

6.3.1.2 Preparation and transformation of competent *Agrobacterium tumefaciens* cells

Agrobacterium tumefaciens strain GV3101 was used for plant transformation. A 5 ml culture of Agrobacterium inoculated from a single colony in LB medium containing 50 µg/ml rifampicin and 25 µg/ml gentamycin was grown at 28 °C overnight, with shaking at 200 rpm. A 2 ml aliquot of the overnight culture was transferred to 50 ml LB medium with rifampicin and gentamycin of the same concentrations in a 250 ml flask and continued to grow at 28 °C until the OD_{600} reached 0.5. The cells were chilled on ice and pelleted at 4000 *g* at 4 °C for 5 minutes. The culture supernatant was discarded and the cells gently re-suspended in 5 ml of ice cold 20 mM $CaCl_2$. The tubes were centrifuged as before at 4000 *g* at 4 °C for 5 minutes, followed by removing the supernatant and gently re-suspending the pellet in 1.0 ml of ice cold 20 mM $CaCl_2$. The bacteria were divided into 200 µl aliquots and flash-frozen in liquid nitrogen. The tubes were stored at -70 to -80 °C until required.

A 200 µl aliquot of frozen cells was placed on ice. 1 µl of plasmid DNA was added to the cells as they started to thaw. The tube was transferred to 37 °C water bath for 5 min to allow for complete thawing. 1 ml of LB medium was added and then the cells transferred to a 50 ml polypropylene tube and incubated at 28 °C with shaking at 200 rpm for around 4 hours. The cells were plated onto selective LB agar plates containing 50 µg/ml rifampicin and 50 µg/ml kanamycin. The plates were sealed with Parafilm and incubated at 28 °C for 48 hours.

Individual colonies were re-streaked onto fresh selection plates and grown again at 28 °C for 24 hrs. To confirm the presence of the introduced plasmid, PCR was carried out directly on the Agrobacterium cells by re-suspending a small amount of bacterial growth in 100 µl of sterile ELGA water in a 0.5 ml tube. The tube was incubated at 99 °C for 10 min to lyse the cells. The tube was cooled on ice, centrifuged at 12000 *g* for 2 min, and then 2 µl of cell lysate was used in a PCR reaction with primers specific for the GS sequence in the introduced plasmid.

6.3.1.3 Floral dip transformation of *Arabidopsis thaliana*

Five wild-type Col-0 plants and one heterozygous GS mutant *gsh2 A. thaliana* (SAIL 301_C06; here designated *gsh2-T1*) (Pasternak *et al.* 2008) plant were used for each ectopic expression construct. Plants were grown in the glasshouse at around 20 °C with a 16 hour day length until a number of inflorescences each with several unopened

flower buds could be seen. A 20 ml LB medium with 50 µg/ml kanamycin was incubated with a single colony of Agrobacterium containing the relevant vector. The culture was incubated at 28 °C overnight with shaking at 200 rpm. The entire 20 ml culture was used to inoculate 200 ml LB medium containing 50 µg/ml kanamycin with 200 rpm shaking until the OD_{600} reached 0.5-0.8 (usually 4-5 hours). The bacterial cells were pelleted by centrifugation at 2455 *g* for 10 min at room temperature and then re-suspended in 200 ml 5% sucrose containing 100 µl Silwet (LEHLE seeds, USA). *A. thaliana* inflorescences were then submerged within the solution for 2 min with a little agitation. Plants were placed under a propagator lid for 24 h, prior to the lid being removed and plant growth continued, under standard conditions.

6.3.1.4 Selection and growth of transformed plants

Arabidopsis thaliana seeds were surface sterilised in 20% household bleach (Domestos, UK) in a 50 ml polypropylene tube for 20 min with slow rotation during this time, followed by 5 washes in the TC flow hood with sterile distilled water. After the final wash, the seeds were kept in the water at 4 °C overnight before plating out.

For selection of transformed seeds, the seeds were plated on ½ MS 10 agar plates with 50 µg/ml kanamycin. Successfully transformed seeds were able to germinate and subsequently progress to the two true leaves stage on ½ MS 10 agar plates with 50 µg/ml kanamycin. Successfully transformed plants were transferred to compost for seed collection. Meanwhile, a single leaf was taken for PCR to confirm the presence of the transgene with the relevant gene specific primers.

6.3.1.5 Genotyping of *Arabidopsis* mutant plants

A single *Arabidopsis* leaf was removed from each plant and was ground using a micro-pestle with 500 µl DNA extraction buffer (0.1 M Tris (pH 8.0), 0.5 M NaCl, 50 mM EDTA), vortexed and incubated at 65 °C for 15 min. 500 µl phenol : chloroform : isoamylalcohol (25:24:1) was added into each sample, mixed and then centrifuged at 12000 rpm for 10 min. The top phase was transferred into a separate tube containing 300 µl isopropanol, mixed and then centrifuged at 12000 rpm for 10 min. The supernatant was discarded and the DNA pellet was washed with 70% ethanol. The pellet was collected by centrifugation at 12000 rpm for 2 min, followed by removal of all the supernatant. The pellet was allowed to air dry and then re-suspended in sterile deionised water. The DNA was stored at -20 °C for downstream experiments.

PCR followed by gel electrophoresis was carried out to identify any transgenic Arabidopsis plants that were homozygous for the Arabidopsis *ghs2* mutant allele using the primers described before (Pasternak et al. 2008). Genotyping of T-DNA insertion lines was done with the genomic primers P1 (5'-TTC CAC TTG TTT GCA GGT CAT TGC-3') and P2 (5'-AAT AAA CCA CTG CGA CTG CTT GGC-3') for amplification of the wild-type allele and the primers P2 and P3 (5'-TAG CAT CTG AAT TTC ATA ACC AAT CTC GAT ACA C-3') for amplification of the mutant allele.

6.3.1.6 Identification of homozygous transgenic lines

For all wild type background T1 lines containing the desired transgene, three highest expressing lines per construct were identified by semi-quantitative RT-PCR with the relevant gene specific primers. Around twenty T2 seeds from each T1 transgenic high expressing line were grown in tissue culture with 50 µg/ml kanamycin as described in section 6.3.1.4. Six green surviving T2 seedlings of each line were transferred to compost to produce T3 seeds. Around 30 T3 seeds per line were selected on ½ MS 10 agar plates with 50 µg/ml kanamycin. Those that were 100% kanamycin resistant indicated that the mother plant was homozygous for the T-DNA insertion. Semi-quantitative RT-PCR was performed to confirm the expression of the transgene in these lines. These homozygous T3 seeds from each line were utilised for the downstream experiments.

6.3.2 Nematode infection assay

For each GS gene, three homozygous transgenic lines were used for infection assays with *H. schachtii*. Wild type Col-0 plants were used as the control. Three homozygous *A. thaliana* seedlings were sown on a 10 x10 cm square plate (Sterilin, UK) of ½ MS 10 agar in a row across the top of the plate and the plates were then placed vertically in a growth chamber to encourage the downward growth of the roots. Growth conditions were 20 °C with a cycle of 16 h light and 8 h dark. At least 15 plants were sown per line.

Seedlings were infected with nematodes three weeks after sowing. Freshly hatched *H. schachtii* J2s were re-suspended in sterile water to a concentration of 1 nematode/µl after surface sterilisation as described in the General methods section. Five infection points on the root tips were selected per seedling and approximately 20 sterilised *H. schachtii* J2s were pipetted on the root surface per infection point.

6.3.3 RNA interference (RNAi) soaking assay

RNAi soaking assays were performed as previously described (Urwin, Lilley and Atkinson 2002; Roderick, Urwin and Atkinson 2018). Approximately 350 bp fragments of the coding region of *Rre-gs1* representing Clade 1, *Rre-gs11* representing Clade 2, *Rre-gs14* and *Rre-gs65* representing Clade 3 were cloned between the Xbal and Hind III sites of the L4440 vector (Timmons and Fire 1998) that contains opposing T7 promoters for *in vitro* transcription of double stranded RNA (dsRNA) complementary to *Rre-gs1*, *Rre-gs11*, *Rre-gs14* and *Rre-gs65*. A green fluorescent protein (GFP) sequence (Chalfie *et al.* 1994) was exploited as a control of a non-nematode gene. The DNA templates for complementary single stranded RNA (ssRNA) were produced by linearisation of each clone with either Xbal or Hind III. A minimum of 1 µg digested DNA was needed for each *in vitro* transcription reaction. The synthesis of complementary ssRNAs and subsequent production of dsRNA were carried out using the Megascript T7 RNAi kit (Invitrogen, UK), according to the manufacturer's instructions. The concentration of dsRNA was then measured by a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies).

The following primers were used for preparation of RNAi constructs in the L4440 vector (restriction site underlined):

GS1-RNAi-F: GCTCTAGAGTTATGCTCCCGAAGATTATCC

GS1-RNAi-R: CCCAAGCTTCAGTCTGTCCACCAGTTCC

GS11-RNAi-F: GCTCTAGAGAGGCGAAGAAAATGGAATTGAAAC

GS11-RNAi-R: CCCAAGCTTACAAGCATCGGACCAACAC

GS14-RNAi-F: GCTCTAGACATCATTGTGGTGGAGATGC

GS14-RNAi-R: CCCAAGCTTCGATTTTGTCCGCATCTTTT

GS65-RNAi-F: GCTCTAGATCATTTTGACGAACCGTTGA

GS65-RNAi-R: CCCAAGCTTAAGCCCAGAGACCAGCATAA

GFP-RNAi-F: GCTCTAGAGCACTATTGCGGACTTGAAACA

GFP-RNAi-R: CCCAAGCTTCCATATTACGCGCTCCAGTT

For each RNAi assay, a total of 3000 freshly hatched J2s of *R. reniformis* were soaked in a solution of 100 µg/ml dsRNA and 100 mM octopamine in M9 buffer (3 g/L KH_2PO_4, 6 g/L Na_2HPO_4, 0.5 g/L NaCl and 1 g/L NH_4Cl) at 25 °C on a rotator. Control nematodes were soaked in solutions without dsRNA or with dsRNA targeted against GFP. At 6 hours after dsRNA treatment, approximately 500 nematodes were removed for qRT-PCR analysis to assess the reduction of target gene transcript. The survival rate of around 100 J2 nematodes was monitored under a microscope at 24 hours after dsRNA treatment. The remaining nematodes were used for the measurement of total glutathione content. Each treatment was repeated four times.

6.3.4 Determination of total glutathione content

The J2 nematodes were collected, pelleted in a microcentrifuge tube, and ground with a pellet pestle in PBS buffer (8 g/L NaCl, 0.2 g/L KCl, 1.44 g/L Na_2HPO_4 and 0.24 g/L KH_2PO_4), followed by three freeze-thaw cycles using liquid nitrogen to lyse the cells. Cell lysate was then centrifuged for 10 min at 12000 rpm at 4 °C. The supernatant was collected and the protein concentration was measured using Quick Start™ Bradford dye reagent (Bio-rad, UK). Total glutathione content of each test sample was assayed using a SensoLyte Glutathione Assay Kit (AnaSpec, Inc. Canada).

6.3.5 Quantitative Reverse-Transcriptase PCR (qRT-PCR)

Total RNA extraction from dsRNA treated J2s was performed as described in section 2.4 and the residual genomic DNA was then removed. cDNA was synthesised from 500 ng DNase-treated RNA using the iScript™ Select cDNA Synthesis Kit with oligo dT and random primers (Bio-Rad, UK).

Each reaction mixture for qRT-PCR analysis in a 96-well plate contained 10 µl SsoAdvanced™ Universal SYBR Green Supermix (Biorad, UK), 0.5 µM forward and 0.5 µM reverse primers (detailed in Table 6.2), ~50 ng cDNA template and ddH_2O to make a final volume of 20 µl. The plate was sealed with optical quality sealing film and centrifuged briefly to eliminate air bubbles, followed by PCR reaction running using a CFX Connect Real-Time PCR detection System (Biorad, UK). A two-step amplification profile was used for all reactions: initial denaturation of 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 sec and 60 °C for 10 sec.

All primer sets were tested for their amplification efficiency prior to running the experimental samples. This was performed by generating a standard cure with five

10-fold dilution of standard cDNA. Each sample was amplified with both target gene-specific primers and *R. reniformis* 18S ribosomal RNA primers for normalisation of expression levels as previously described (Ganji, Jenkins and Wubben 2014). Each reaction was run in three technical repeats and negative controls for each primer pair contained no cDNA template.

Fluorescence signals were collected at each 60 °C stage. Amplification plots, dissociation curves and threshold fluorescence were viewed in CFX Manager™ Software (Biorad, UK). Gene expression change was calculated using the ($2^{-\Delta\Delta CT}$) method (Wagner 2013): C_T value is the threshold cycle determined by CFX Manager™ Software when threshold fluorescence was reached. ΔC_T is C_T (target gene) - is C_T (normalisation control). $\Delta\Delta C_T$ is ΔC_T (sample 1) - ΔC_T (sample 2). Fold change (sample 1 vs. sample 2) equals $2^{-\Delta\Delta CT}$ value.

Table 6.2. Primers used for the quantification of *R. reniformis* GS gene expression following RNAi, alongside the annealing temperatures and product sizes.

Target	Primer set	Ta (°C)	Product size	Sequences (5'-3')
Rre-GS1	qRr-GS1-F qRr-GS1-R	64	158	TGGAAGTGATGGAGATGGAG GAAATGGCGATGGGAATAGG
Rre-GS11	qRr-GS11-F qRr-GS11-R	63	140	ACTTCTTCCCGAACGAACC TTTCATCACAAACTCCGATCC
Rre-GS14	qRr-GS14-F qRr-GS14-R	63	82	GCAAAACCTACATTGTCAAACC ATGAAAAAGCCGTAAGCCC
Rre-GS65	qRr-GS65-F qRr-GS65-F	64	94	CGAAAAGGATGAACCCGAG CCGATGGTGTAGTGAGGTAAG
18S ribosomal RNA	qRr-18s-F qRr-18s-R	68		TCGCCACACTAACAAACCGT GCAACAACTGCTCAACAACGCA

6.4 Results

6.4.1 *R. reniformis* GS failed to complement Arabidopsis glutathione deficiency mutant *gsh2*

Homozygous *gsh2* mutants bleach after germination and can be distinguished from green, phenotypically wild-type seedlings, but heterozygous *gsh2* mutants survive with no distinct phenotype (Pasternak *et al.* 2008). So Arabidopsis *gsh2* mutants were exploited to examine whether *R. reniformis* GS can complement a lack of Arabidopsis GS. As homozygous mutants are not viable, each pool of seeds will contain both wild type plants and heterozygous mutant plants. Therefore, heterozygous *gsh2* mutants were first identified by PCR genotyping to select plants for transformation. This *gsh2* mutant line contains two T-DNA insertions head to head. Figure 6.2A shows the physical map of the *GSH2* gene (At5g27380) and the insertion site for one T-DNA insertion allele, *gsh2-T1*. The primer at the left border of the T-DNA insertion (P3) therefore amplifies a DNA fragment in combination with either of the gene-specific primers P1 or P2 when a mutant locus is present (Figure 6.2B left). In the wild type locus of a heterozygous mutant, no fragment is amplified with P3 in combination with either P1 or P2. The P1 and P2 primer pair together will amplify a DNA fragment from the wild type locus but not from the mutant locus. The wild type Col-0 was used as the control for genotype determination of *gsh2* plants. Because of no *gsh2-T1* insertion in Col-0, the DNA band was absent using P3 and P2 amplification (Figure 6.2B right). Accordingly, any homozygous *gsh2* progeny following transformation with a GS construct can be identified using PCR genotyping: a fragment will be amplified with the P2/P3 primer pair but not with the P1/P2 primer pair.

To ensure the validity of the experimental strategy, and to act as a positive control, *gsh2-T1* heterozygous plants were transformed with a FLAG-tagged pBI121 construct expressing the wild type Arabidopsis *GSH2* cDNA lacking the start codon and the N-terminal plastid transit peptide that would usually direct the enzyme to the plastids. PCR genotyping of 24 T1 plants was then performed as described. No DNA fragments were amplified from genomic DNA of lines 5, 11, 13, 17 and 18 using P1 and P2 primers while clear DNA bands were present using P3 and P2 (Figure 6.2C), indicating that these lines were homozygous *gsh2* mutants. Considering that homozygous *gsh2* mutants cannot survive after germination, this indicated that the transgenic *GSH2* expression from the CaMV35S promoter was sufficient to rescue the homozygous

gsh2 mutants. These results were consistent with the previous report (Pasternak *et al.* 2008) and validated the vectors being used. As expected, both heterozygous *gsh2* (lines 1, 2, 6, 7, 9, 10, 12, 14, 15, 16, 19, 20, 21, 24) and wild type backgrounds (lines 3, 4, 8, 22, 23) were also identified.

A total of 16 *R. reniformis* GS-like genes, with representatives from each of the three clades, were cloned into the binary vector FLAG-tagged pBI121. The relevant transgenic GS Arabidopsis plants in a *gsh2* heterozygous background were then produced as described. Genotyping PCR was carried out on DNA from at least 24 individual *Arabidopsis* T1 plants for each construct to identify any that were homozygous mutants for the *gsh2* T-DNA insertion allele. None of the *R. reniformis* GS were demonstrated to rescue the homozygous *gsh2* mutants, including the canonical *R. reniformis* GS: Rre-GS1 (Figure 6.2D). All the 24 segregating T1 plants were identified to contain either two wild type alleles (wild type) or one wild type and one mutant (heterozygous), suggesting that homozygous *gsh2* mutants still failed to survive even when *R. reniformis gs1* was expressed.

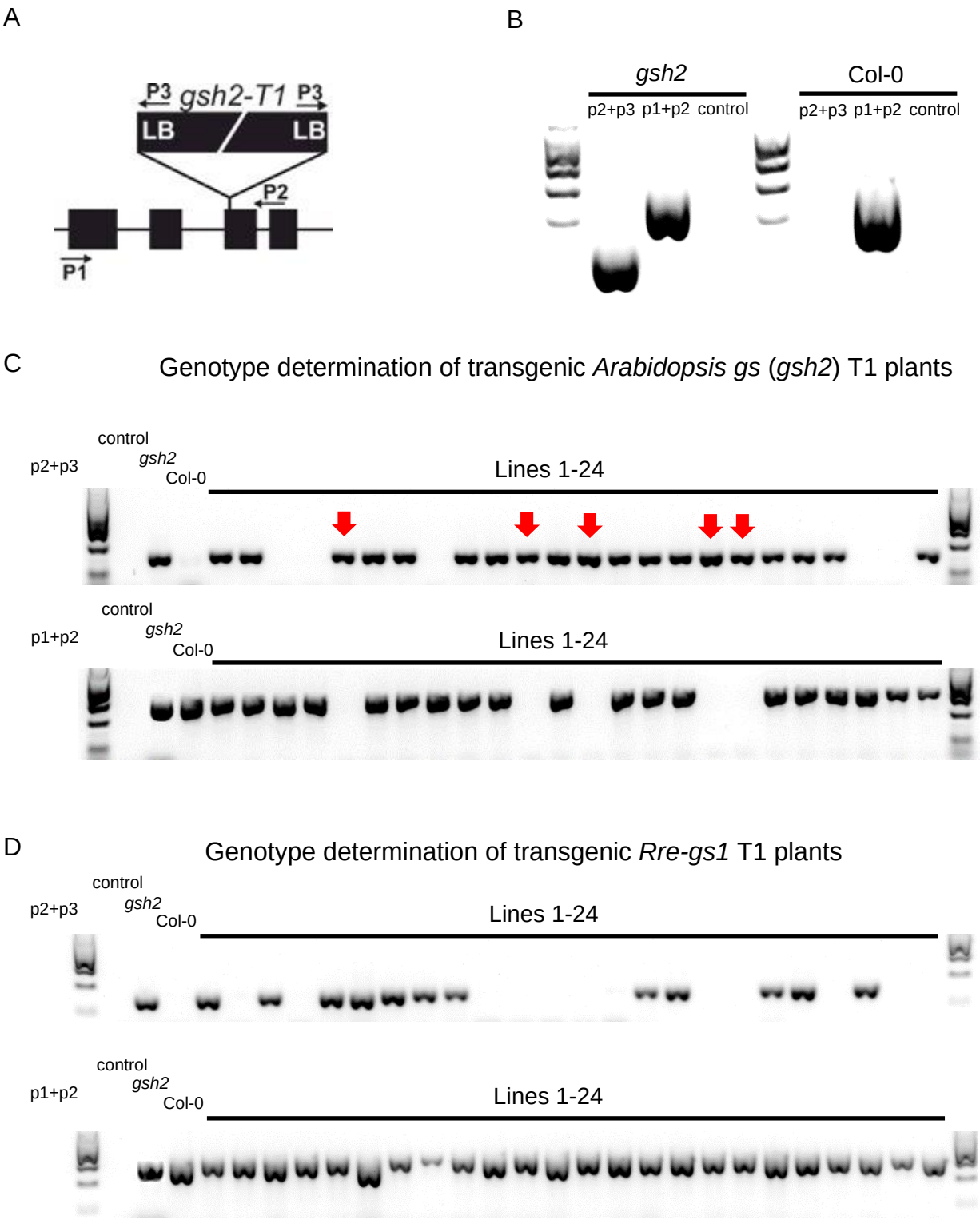


Figure 6.2. **Example of identification of heterozygous/homozygous mutants for the *gsh2* T-DNA insertion allele**. **(A)** Physical map of the *GSH2* gene (At5g27380) and insertion sites for the T-DNA insertion alleles, *gsh2-T1*. Exons are represented as boxes and introns as lines. **(B)** Genotype determination of heterozygous *gsh2* plants. A DNA fragment was amplified by the gene-specific primer P2 in combination with both a T-DNA region-specific primer P3 and another gene-specific primer P1 in a heterozygous mutant. Only the P1/P2 primer combination amplified a fragment from wild type Col-0 DNA. **(C)** Genotype determination of transgenic *GSH2* T1 plants. The gel images show that lines 5, 11, 13, 17 and 18 were homozygous *gsh2* mutants (red arrows). The survival of the homozygous *gsh2* mutants was due to the complementation of the *gsh2* mutant. **(D)** Genotype determination of transgenic *Rre-gs1* T1 plants. No homozygous *gsh2.*

6.4.2 RNAi of *R. reniformis* GS genes

In order to investigate the essential role of the different GS-like genes for the survival of *R. reniformis*, RNAi assay of four *R. reniformis* GS genes to represent the three clades and a non-nematode gene GFP as control was performed. *Rre-gs1* and *Rre-gs11* were used to represent Clade 1 and Clade 2 GS genes, respectively, as both were highly expressed in the J2 stage. For Clade 3 GS genes, *Rre-gs14* and *Rre-gs65* were selected because they displayed different expression profiles between the non-parasitic stage and the parasitic stage: *Rre-gs14* was more highly expressed in the parasitic female whereas *Rre-gs65* was more highly expressed in the J2 stage.

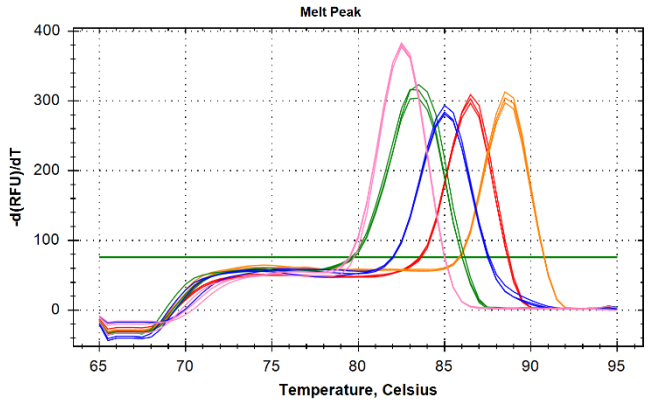
First, the expression of each target *R. reniformis* GS gene was evaluated by qRT-PCR following soaking in dsRNA. Dissociation curve was used to determine the primer specificity. The specific melting temperature of a product is related to its size and C/G content. Accordingly, one single peak in the dissociation curve indicates a pure product and therefore primer specificity (Figure 6.3A). All primer sets were tested for their amplification efficiency prior to running the experimental samples by generating a standard curve (Figure 6.3B).

Treatment of J2 *R. reniformis* with a dsRNA solution targeting *Rre-gs1*, *Rre-gs11*, *Rre-gs14* and *Rre-gs65* significantly reduced the transcript of these genes by around 75%-80% (Figure 6.4A). A control dsRNA treatment that targeted a *gfp* sequence had no significant effect on the expression of the genes. Total GSH content was measured for RNAi-treated worms at 24 hours after dsRNA treatment. In the absence of RNAi, *R. reniformis* J2 contained 22.5 ± 4.8 nmol/mg glutathione (Figure 6.4B) which is within the range found in hepatic cells of 20-30 nmol/mg protein (Brigelius *et al.* 1983). A significant decrease (~60%) of glutathione content was demonstrated in RNAi-treated worms targeting *Rre-gs1*. By contrast, there was no obvious difference in glutathione content in RNAi-treated worms targeting *Rre-gs11*, *Rre-gs14* and *Rre-gs65* genes as compared with the control worms (Figure 6.4B). Taken together, this supports the hypothesis that only Clade 1 GS functions as a typical GS while both Clade 2 and Clade 3 GS do not produce glutathione.

The survival rate of the RNAi-treated worms was then monitored at 24 hours after dsRNA treatment. In the absence of RNAi, nearly all the nematodes were still alive after 24 hours. There were no apparent major differences in survival rate and obvious

phenotypic alterations between Clade 2 and Clade 3 GS RNAi-treated and no RNAi-treated nematodes from observation, which was consistent with the expectation due to two possible reasons: 1. Clade 2 and Clade 3 GS were not predicted to play an essential role in nematode development. 2. There are a large number of Clade 2 and Clade 3 GS genes. Knocking down a single Clade 2 or Clade 3 GS gene cannot contribute to an obvious effect. Interestingly, there was also no difference in survival rate and obvious phenotype of RNAi-treated nematodes targeting *Rre-gs1* as compared to control worms despite *Rre-gs1* being considered as an essential housekeeping gene. Given the fact that RNAi-treated nematodes targeting *Rre-gs1* showed a significantly reduced level in both *Rre-gs1* transcript and total glutathione content, these results indicated that *Rre-gs1* can maintain its ‘housekeeping’ function even in very low transcript level. In addition, nematodes can survive in a low glutathione level under laboratory conditions.

A



B

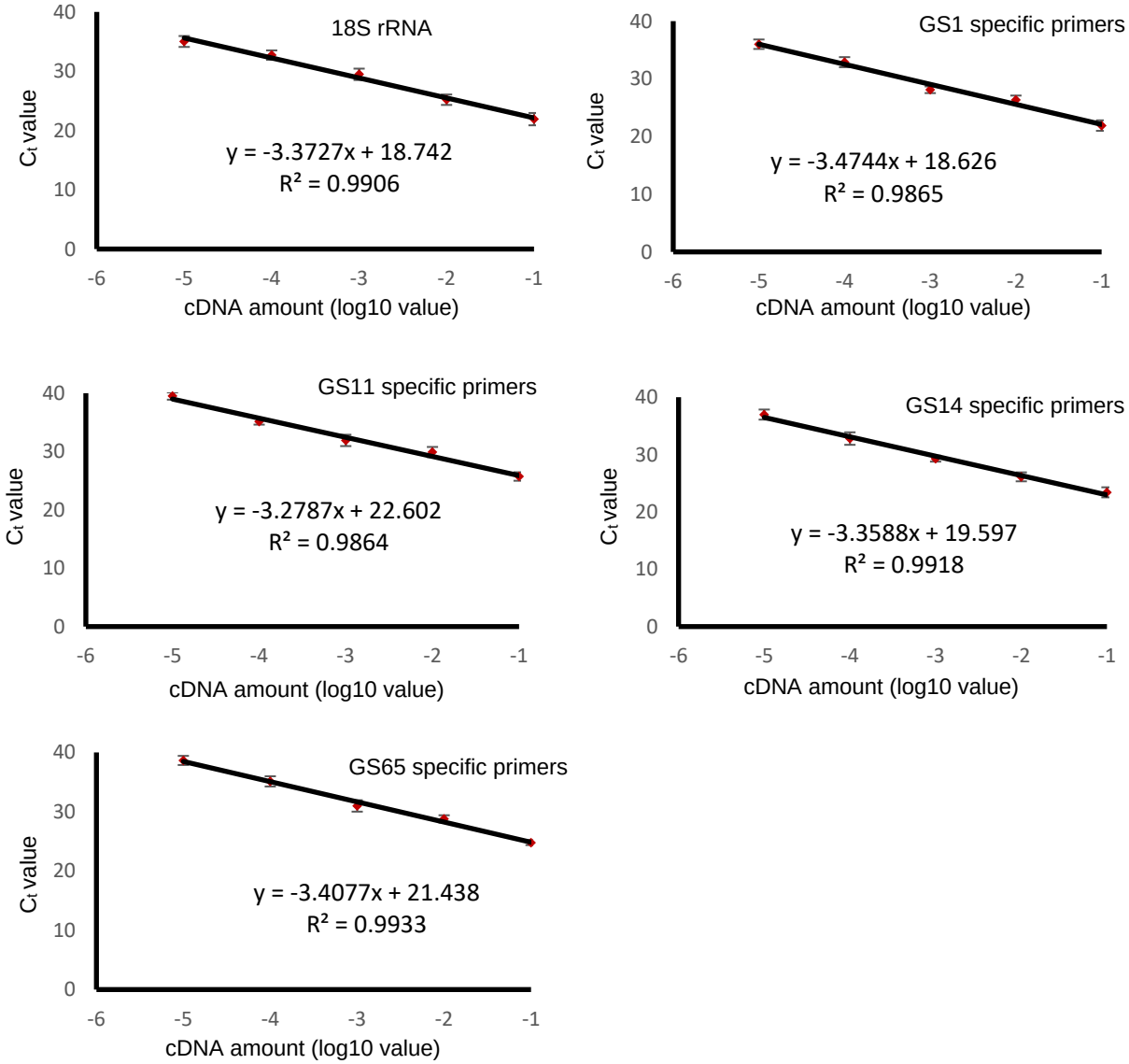


Figure 6.3. Test of qRT-PCR primers. (A) Dissociation curves for all the primer pairs used to amplify transcripts in *R. reniformis*. One single peak in the dissociation curve indicated a pure product amplified by corresponding primers. (B) Standard curves was constructed for each primer pairs used in qRT-PCR assay by exploiting 10-fold dilution series of standard cDNA.

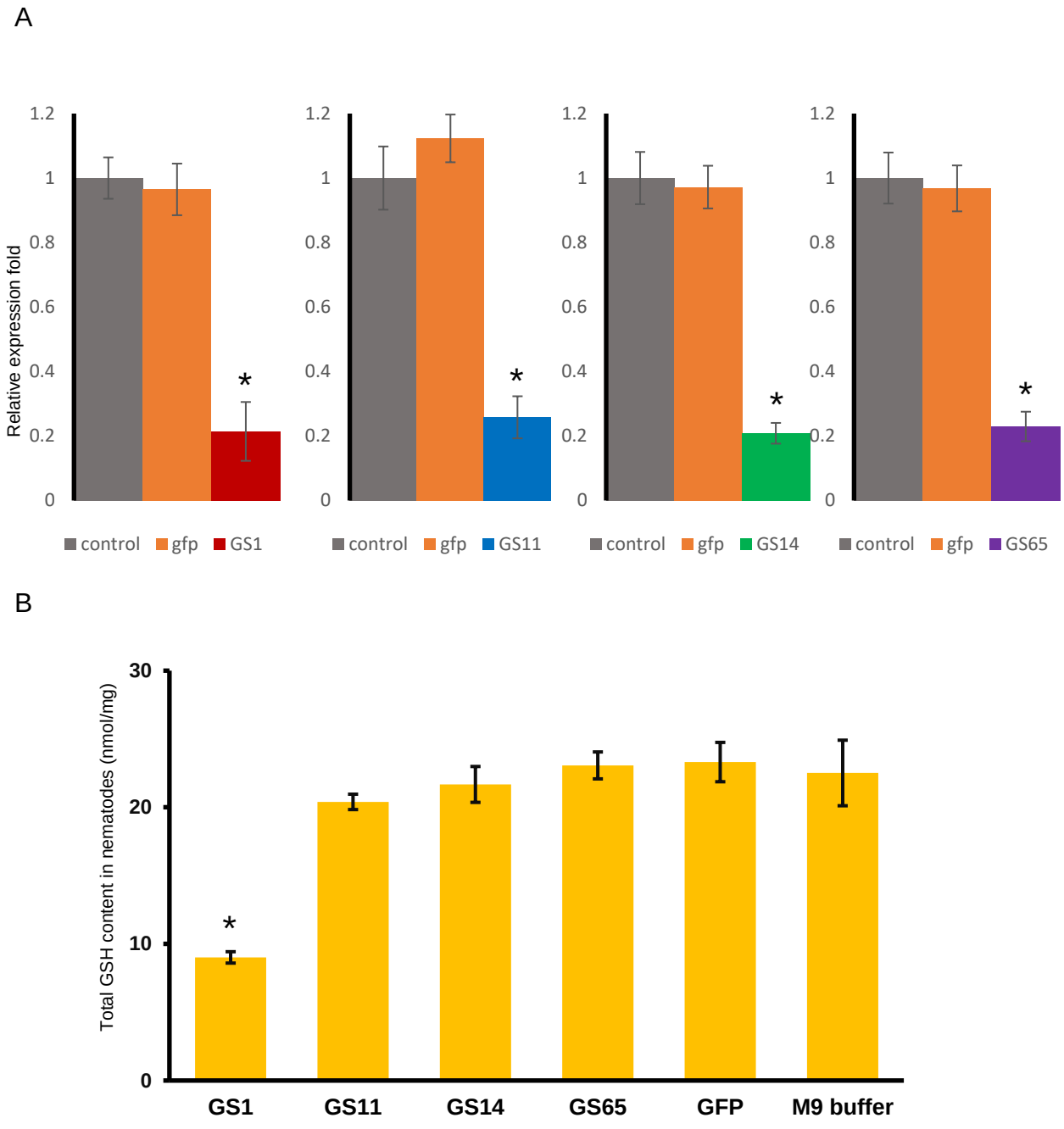


Figure 6.4. **Total GSH content following RNAi of *R. reniformis* GS genes.** (A) qRT-PCR expression analysis of relevant GS genes in J2 RNAi nematodes. Treatment of J2 R. reniformis with a dsRNA solution targeting Rre-gs1, Rre-gs11, Rre-gs14 and Rre-gs65 significantly reduced the transcript of these genes. Data are reported as means ± standard error. * indicates a statistically significant difference (One-way ANOVA, P<0.05). (B) Total GSH content measurement of J2 *R. reniformis* after RNAi treatment. The dsRNA molecules targeted *R. reniformis gs1*, *gs2*, *gs14* and *gs65* with control dsRNA against GFP and incubation in M9 buffer only. Values are means ± standard error (n=4 pools of J2 *R. reniformis*). * indicates a statistically significant difference from the controls (One-way ANOVA, P<0.05).

6.4.3 Functional analysis of *R. reniformis* Clade 3 GS

Considering that *R. reniformis* Clade 3 GS enzymes were hypothesised to act as 'effectors', it is necessary to investigate their direct effects on nematode parasitism within the plant-nematode interaction. This was achieved using ectopic over-expression of *R. reniformis gs* genes in wild type Arabidopsis. The Arabidopsis-*H. schachtii* pathosystem was then exploited for nematode infection assay as it is easier to acquire transgenic Arabidopsis plants than the common *R. reniformis* hosts like cotton. Furthermore, Arabidopsis is not an ideal host for *R. reniformis* and *H. schachtii* was reported to encode a number of putatively secreted Clade 3 GS-like effectors by transcriptomic analysis (Lilley *et al.* 2018), so making it a relevant test system.

R. reniformis GS14 and GS23 were selected to represent Clade 3 GS enzymes. Independent homozygous *Rre-gs14* or *Rre-gs23*-expressing Arabidopsis lines (lines 8-1, 8-2, 11-1 for *Rre-gs14* and lines 1-6, 2-6, 8-5 for *Rre-gs23*) and wild-type (Col-0) as the control were infected with *H. schachtii* J2s. Figure 6.5A shows an example for confirmation of *R. reniformis* GS expression in homozygous T3 plants by reverse transcription PCR.

The number of both males and females of *H. schachtii* were counted per root system two weeks after infection for both the transgenic and wild-type lines. Figure 6.5B & C show images of *R. reniformis* males and females in transgenic plant roots. A clear effect of transgenic expression on nematode susceptibility was observed. All these transgenic lines were significantly more susceptible to *H. schachtii* than the wild-type control (approximately 30% increased), as evidenced by the statistically significant higher total number of females (Figure 6.5D). In addition, higher female: male ratios were shown on the GS transgenic Arabidopsis plants (Figure 6.5D). Taken together, given the fact that GS transgenic Arabidopsis plants showed no obvious phenotypic difference from wild-type, including root growth, it suggested a key role of Clade 3 GS effectors in nematode parasitism.

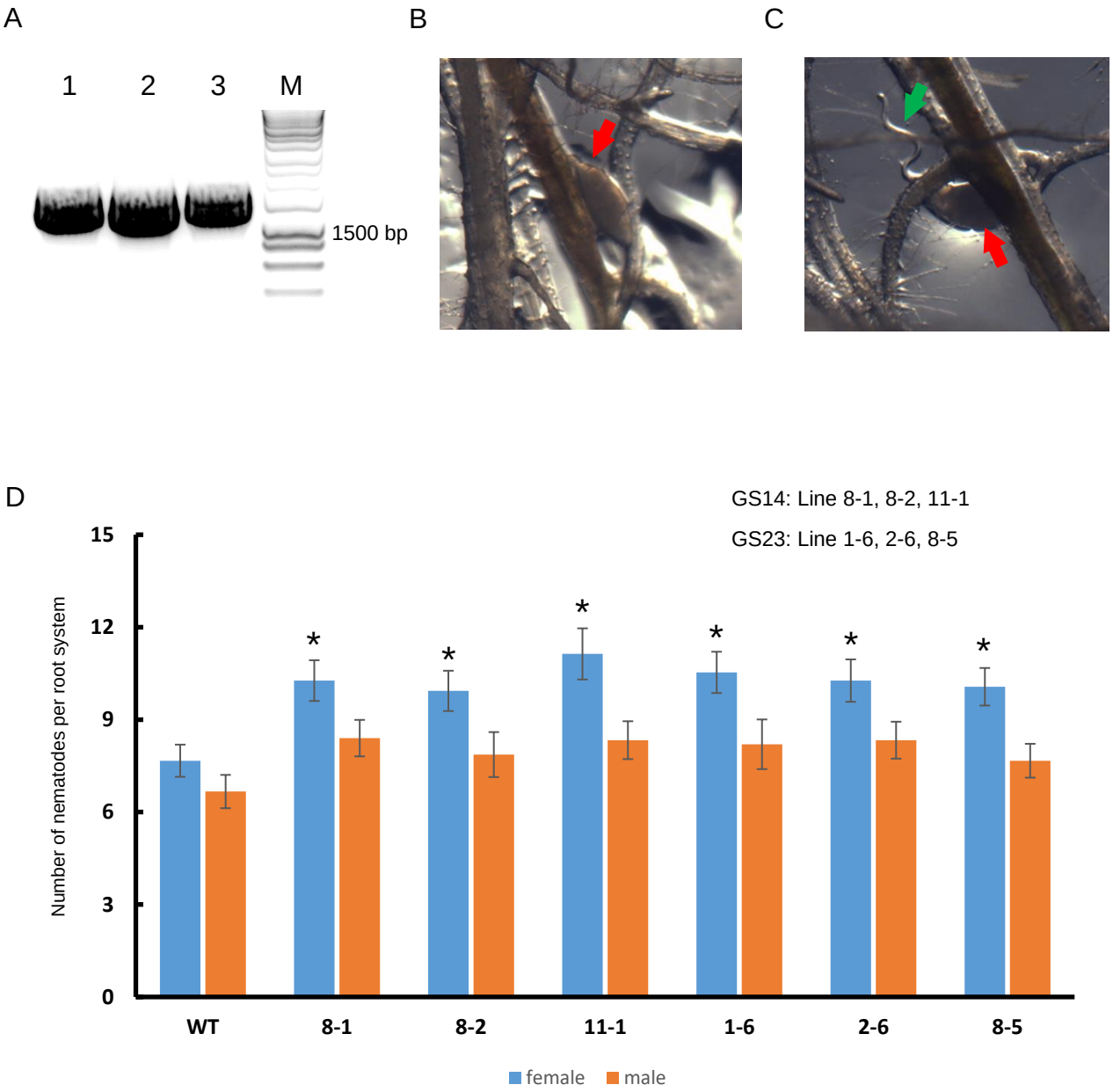


Figure 6.5. **Functional analysis of *R. reniformis* Clade 3 GS.** (A) Confirmation of *R. reniformis gs14* expression by RT-PCR in homozygous T3 plants. (B)-(C) Images to show females (red arrow) and males (green arrow) on transgenic Arabidopsis roots. (D) Transgenic Arabidopsis plants expressing *Rre-gs14* and *Rre-gs23* showed enhanced susceptibility to *H. schachtii*. 15 plants per line were used in this assay. The numbers of adult males and females of *H. schachtii* per root system were determined. Values are expressed as a mean ± Standard error. Mean values significantly different from the wild type are denoted by asterisks, as determined by One-way ANOVA test (n=15, P<0.05).

6.5 Discussion

6.5.1 Functional analysis of *R. reniformis* GS genes

The expanded *R. reniformis* GS gene family has diversified in function between the three major clades as demonstrated in this work by phylogenetic, biochemical, and structural evidence. However, the function of these diversified GS genes is still unclear. A total of 16 *R. reniformis* GS-like genes, representatives of all the three clades, were used to complement the Arabidopsis GS mutant *gsh2*. None of the investigated *R. reniformis* GS genes were able to complement the homozygous Arabidopsis GS mutant (Figure 6.1), indicating that the *R. reniformis* GS cannot substitute plant GS *in planta*. Given the fact that Arabidopsis GS is a canonical GS enzyme, it is not surprising that *R. reniformis* Clade 2 and 3 GS genes could not rescue the Arabidopsis GS mutant as both Clade 2 and 3 GS enzymes have negligible glutathione synthetic activity. However, it was surprising that the canonical GS enzyme, *R. reniformis* GS1, also failed to rescue the Arabidopsis GS mutant. The earlier biochemical analysis indicated that although Rre-GS1 displayed typical GS activity, which was similar to that reported for *C. elegans* GSS1, the GSH production rate was seven-fold lower than that of Arabidopsis GS. Nevertheless, it seems unlikely that this lower activity is wholly responsible for the failure of *Rre-gs1* to rescue the Arabidopsis mutant. For example, various GSH-deficient mutants of *Arabidopsis* have been identified, which are mapped to GSH1. These include *cadmium-sensitive2* (*cad2*) (Howden *et al.* 1995), *regulator of ascorbate peroxidase2 1* (*rax1*) (Ball *et al.* 2004), *phytoalexin-deficient2* (*pad2*) (Parisy *et al.* 2007) and *root meristemless1* (*rml1*) (Cheng, Seeley and Sung 1995) mutants. Despite much lower GSH content in these lines: *cad2*, approximately 30% of wild type GSH content; *rax1*, 50%; *pad2*, 20%, they are otherwise phenotypically wild type under normal growth conditions. However, *rml1* which has only 2% of wild type GSH content failed to develop a primary root after germination (Cheng, Seeley and Sung 1995). These *GSH1* mutants do not accumulate γ-EC, which has been shown to cause perturbation of ER morphology in some *gsh2* mutants (Au *et al.* 2012), so that may be a contributing factor in the lack of mutant rescue.

Another possible factor is due to the subcellular localisation of the GS enzymes. Although Arabidopsis GS is localised both to chloroplasts and cytosol in plant cells (Noctor *et al.* 2002), only 8% of *GSH2* transcripts were shown to encode the entire chloroplast target peptide (Wachter *et al.* 2005), suggesting the majority of GSH2

protein is present in the cytosol rather than the plastids. In addition, assays of GS activity in the cytosol and chloroplast fractions of *Arabidopsis* leaf tissue showed cytosolic GS contributed ~69% of total activity while chloroplast GS contributed ~31% of the total activity (Galant *et al.* 2011). Complementation of *gsh2* mutant with both cytosol-specific wild type GSH2 and *E. coli* GS fused with GFP fully rescued the mutant, and fluorescence analysis indicated the fusion protein was exclusively located in the cytosol (Pasternak *et al.* 2008). Taken together, these results suggested the compartmentation of GS affects its function. None of the *R. reniformis* GS constructs for plant transformation contained any obvious sub-cellular targeting signals and the chloroplast transit peptide was not included in the Arabidopsis *GSH2* control construct. Therefore all transgenic GS proteins were expected to be cytosolic, however it is possible that cryptic signals in *R. reniformis* GS1 could have caused mis-targeting. This, in combination with the lower activity of Rre-GS1, could then lead to a lack of complementation. Given the failure of *Rre-gs1* to rescue the Arabidopsis mutant, it is difficult to then draw any conclusions about the similar results for the non-canonical *R. reniformis* GS.

RNAi was then used to investigate the knock-down effect of *R. reniformis* GS genes in nematodes. As expected, a significantly reduced total glutathione content was showed in RNAi-treated *R. reniformis* J2 targeting *Rre-gs1* while no obvious change of glutathione content was observed in RNAi-treated *R. reniformis* J2 targeting both Clade 2 and 3 GS genes (Figure 6.3). However, there was no obvious difference in survival rate and obvious phenotypic alterations between RNAi-treated and no RNAi-treated nematodes despite *Rre-gs1* being predicted as an essential 'housekeeping' GS gene. Given the fact that there was still 40% glutathione content left in the *Rre-gs1* RNAi-treated nematodes (Figure 6.3), the nematodes were shown to survive at a low glutathione level under laboratory conditions. This is perhaps not surprising as in fact, surviving with incomplete depletion of GSH to a certain extent has been described previously as described earlier for Arabidopsis *gsh1* mutants (Parisy *et al.* 2007). Another possible explanation is that nematodes may be able to compensate the loss of glutathione synthesis by exploiting other functional thiols such as γ-EC. Decreased GS level usually leads to a hyperaccumulation of γ-EC because of the slow consumption of γ-EC and the alleviation of feedback inhibition of GCL by GSH (Grant, MacIver and Dawes 1997). γ-EC was reported to play a substitute role to GSH in many

organisms provided it is present in sufficient quantities (Newton and Javor 1985; Grant, MacIver and Dawes 1997). Accordingly, nematodes may utilise γ-EC to functionally replace the role of GSH.

For *R. reniformis* Clade 2 GS genes, no visible RNAi effect may be due to the large number of genes. Knock down of a single Clade 2 GS gene could be compensated by the activity of other Clade 2 GS genes. For *R. reniformis* Clade 3 GS genes, no obvious RNAi effect may be because Clade 3 GS genes do not play an essential role in nematode survival at the pre-parasitic J2 stage that was tested.

In order to further investigate the roles of Clade 3 GS genes during plant parasitism, nematode infection assays were performed with GS transgenic Arabidopsis using the Arabidopsis-*H. schachtii* pathosystem. The Arabidopsis-*H. schachtii* pathosystem was exploited as it is easier to acquire homozygous transgenic plants for Arabidopsis than the typical *R. reniformis* hosts like cotton (Sijmons *et al.* 1991). In addition, a large number of GS-like genes were also identified from *H. schachtii* transcriptomes and displayed similar topology in the phylogeny (Lilley *et al.* 2018), indicating that *H. schachtii* may also utilise GS genes as ‘effectors’ to promote parasitism. *R. reniformis* GS14 and GS23 were selected to represent Clade 3 effector GS in the nematode infection assay as these two genes were both highly expressed at the parasitic female stage and were shown to be expressed in the gland cell by *in situ* hybridisation. Both homozygous *Rre-gs14* and *Rre-gs23*-expressing Arabidopsis lines showed increased susceptibility to *H. schachtii* infection compared to the wild-type Col-0 plants (Figure 6.5D). Interestingly, higher female: male ratios were demonstrated on the GS transgenic plants than the controls. Taken together, given that the proportion of female nematodes in the adult population increases when juveniles are exposed to favourable conditions (Lilley, Atkinson and Urwin 2005), *Rre-gs14* and *Rre-gs23* may play an essential role in successful nematode parasitism.

6.6 Summary

1. None of the *R. reniformis* GS were able to complement the Arabidopsis GS *in planta*.
2. RNAi-treatment of J2 nematodes with dsRNA targeting Clade 1 GS caused a significant reduction in glutathione level while RNAi-treatment targeting Clade 2 and Clade 3 GS caused no difference in glutathione level.
3. An increased nematode parasitic success and higher female: male ratios were shown in transgenic Arabidopsis expressing *R. reniformis* Clade 3 GS using the Arabidopsis-*H. schachtii* pathosystem, confirming that Clade 3 GS may play a role in nematode parasitism.

Chapter 7

General discussion

7 General discussion

7.1 Gene birth and evolution

A large GS gene family with up to ~260 members was identified from the *R. reniformis* genome and life-stage specific transcriptomes. As introduced above, most eukaryotic organisms investigated to date have only one gene coding for glutathione synthetase. These canonical eukaryotic GS enzymes share similar primary sequence and structural features and also have the same catalytic ability: they catalyse the addition of glycine to γ-EC to form glutathione. By contrast, most of the *R. reniformis* GS enzymes are likely to be non-canonical GS enzymes. The functional diversity within the *R. reniformis* GS family was revealed by phylogenetic, biochemical, structural and functional evidence. Whilst all those GS analysed in this work shared similar characteristics for their Clade, in terms of spatial expression and enzyme activity, it must be noted that these represent only a proportion of the likely diversity with the large *R. reniformis* GS gene family. This was addressed in part by selecting representative genes from across the phylogeny, however it is possible that some members may have different attributes. Nevertheless, taken together, the abnormal expansion of the GS gene family in *R. reniformis* represents the adaptation of this economically important plant pathogen by generation of novel genes that we hypothesise have gained novel functions.

Several well-characterised mechanisms can be responsible for the emergence of new genes within a plant pathogen species, such as horizontal gene transfer, gene duplication and divergence, gene fusion, gene fission, *de novo* gene birth and retroposition (Long *et al.* 2003; Van Oss and Carvunis 2019). Given the fact that plant pathogens possess large numbers of effector genes, which generally share little sequence homology, even for closely related species, *de novo* gene birth likely plays an important role in creating effector diversity (Plissonneau *et al.* 2017; Frantzeskakis *et al.* 2019). *De novo* gene birth is the process by which ancestrally non-genic and non-coding DNA is transformed into a functional sequence with an open reading frame and a *cis*-regulatory element, to produce a new gene (Carvunis *et al.* 2012). Such novel genes are often shorter than established genes to evolve more rapidly (Carvunis *et al.* 2012). *De novo* gene evolution was once considered to be rare. However, there are now several reports to indicate *de novo* gene birth is an essential source of gene functional diversity. A typical example of *de novo* gene birth is an effector gene of the

barley powdery mildew fungal pathogen *Blumeria graminis* f.sp. *hordei* (*Bgh*) (Nottensteiner *et al.* 2018). A *Bgh* virulence factor termed ROPIP1 that is encoded on the active non-long terminal repeat retroelement Eg-R1 of *Bgh* was demonstrated to act as an 'effector' during *Bgh*-barley interactions, suggesting a possible *de novo* effector birth from the retroelement-derived transcripts. Another good example is the large and diverse genes family of C-TERMINALLY ENCODED PEPTIDE (CEP) plant hormone mimics (RrCEP) from *R. reniformis*. With the exception of the CEP domain, RrCEPs share no sequence similarity with any other CEPs from plants or animals, suggesting that RrCEPs may evolve *de novo* and arise independently of other plant and animal CEPs (Eves-Van Den Akker *et al.* 2016b).

For the *R. reniformis* GS family, the phylogeny suggests that the large family is divided into three major clades and both Clade 2 and 3 GS originated from Clade 1 that is considered to contain the only canonical GS enzyme in the *R. reniformis* GS family. Given the fact that both Clade 2 and 3 GS still share the ATP-grasp domain and similar substrate binding domain with Clade 1 GS, these non-canonical GS genes are highly unlikely to be produced via *de novo* gene birth events.

In addition to *de novo* gene birth, many genes are acquired by horizontal gene transfer events. Horizontal gene transfer is the transmission of genes between different organisms other than vertical inheritance from an ancestor to an offspring (Keeling and Palmer 2008). One of the best studied examples in a plant pathogen is the effector gene *ToxA* that was first identified to be transferred from the wheat fungal pathogen *Stagonospora nodorum* to *Pyrenophora tritici-repentis* (Friesen *et al.* 2006), and subsequently found in other cereal fungal pathogens (Ma *et al.* 2010; McDonald *et al.* 2018). The acquisition of *ToxA* significantly enhanced virulence of *P. tritici-repentis* on wheat. In all species investigated to carry *ToxA*, this gene was located in a chromosomal region that was rich in repetitive transposable elements and underwent large rearrangements. Of particular relevance to this work, horizontal gene transfer has made major contributions to the effector complements of plant parasitic nematodes (Danchin *et al.* 2010; Paganini *et al.* 2012). For example, a series of plant cell wall-degrading enzymes, which are not usually found in animals and are similar to bacterial homologues, were indicated to play essential roles in successful nematode parasitism (Danchin *et al.* 2010). Cell wall-degrading enzymes are diverse and abundant in *M. incognita* with more than 60 genes covering six different protein

families for the degradation of cell wall oligo- and polysaccharides (Abad *et al.* 2008). These ‘effector’ genes originated from different bacteria by multiple independent horizontal gene transfer events, followed by gene duplications.

However, horizontal gene transfer is unlikely to have played a role in the evolutionary history of the *R. reniformis* GS family. Eukaryotic and prokaryotic forms of GS share extremely low sequence identity despite similar enzymatic activity. The ancient eukaryotic and prokaryotic GS genes are hypothesised to arise from different progenitors and have evolved independently, with the ATP-grasp domain being somehow acquired by both eukaryotic and prokaryotic members during convergent evolution (Copley and Dhillon 2002). The non-canonical *R. reniformis* GS are genetically closer to the Clade 1 *R. reniformis* GS gene than to prokaryotic GS genes or those from other species, discounting horizontal gene transfer as a mechanism for their acquisition and evolution. Similarly, *R. reniformis* GS are more closely related to GS from other animals than from plants - ruling out the possibility of horizontal gene transfer from the host.

Hybridisation is another major route for introduction of foreign genes into a pathogen’s gene pool as it generates mosaic sequences from those that are optimally adapted to the new host and environment in the parental species (Stukenbrock 2016). The best-understood example of pathogen hybridisation was demonstrated in the powdery mildew strains (*Blumeria graminis*) of triticale, an artificial hybrid of wheat and rye (Menardo *et al.* 2016). Mirroring the hybridisation between the hosts, *B. graminis* f. sp. *triticale*, which grows on triticale and wheat, is a hybrid between wheat powdery mildew (*B. graminis* f. sp. *tritici*) and mildew specialised on rye (*B. graminis* f. sp. *secalis*). Generally, hybridisation leads to rapid genomic changes, including chromosomal rearrangements and genome expansion, which contributes to beneficial new phenotypes (Baack and Rieseberg 2007).

Hybridisation has also been described to increase the gene content in the genome of plant parasitic nematodes, especially root-knot nematodes. *Meloidogyne* species except for the automictic, diploid *M. hapla* contain divergent genomic copies of many loci, likely due to multiple hybridisation events (Lunt 2008; Szitenberg *et al.* 2017). These peculiar hybrid genome structures are believed to provide root-knot nematodes with a potential for adaptation and may explain their paradoxical success in the

absence of sex (Blanc-Mathieu *et al.* 2017). Interestingly, only Clade 1 and 2 GS genes were identified in *M. incognita*, suggesting that that particular hybridisation event may not play a major role in 'effector' GS expansion. In support of this, many of the cyst nematode species that also have an expansion of 'effector' GS are not hybrids. Nevertheless, *R. reniformis* appears to have more Clade 3 GS genes than the cyst nematode species analysed to date. It is uncertain if hybridisation has contributed in any way to Clade 2 and 3 GS expansion because of little information about genomic analysis of *R. reniformis*.

Gene duplication is a very common phenomenon in all eukaryotic and prokaryotic organisms (Kaessmann 2010). Gene duplication is often linked to evolutionary innovations via one of three basic scenarios: 1) extra gene copies can increase protein levels; 2) ancestral genes can be split over different paralogs and evolve independently after gene duplication events. 3) one of the copies can develop a novel function (Voordeckers *et al.* 2012). Many pathogen genes, especially effector genes, are likely candidates for 'young genes' which arose from an 'ancient gene' by multiple gene duplication events (Fouche, Plissonneau and Croll 2018). Effector gene sequences tend to be altered at significantly higher rates than more conserved genes (Hartmann and Croll 2017). A large family of glucosidase genes has been identified in some yeast species that metabolise a broad spectrum of natural disaccharides found in plants and fruits (Kurtzman and Robnett 2003), and are believed to have undergone several gene duplication events (Voordeckers *et al.* 2012). The ancestral enzyme from which all the others originated via repeated gene duplications was identified. This very first enzyme was active against both maltose-like and isomaltose-like substrates. Interestingly, gene duplications spawned daughter genes in which mutations near the active sites optimised either maltase or isomaltase activity. Taken together, these results indicated that all the three basic scenarios for gene duplications cannot be taken into consideration separately (Voordeckers *et al.* 2012).

Adaptive gene gains by gene duplication are also well-described in pathogens of plants. In smut fungi *Microbotryum* species, a large number of effector genes were mainly driven by tandem gene duplications within gene clusters (Schirawski *et al.* 2010; Dutheil *et al.* 2016). The effector-like genes usually evolved from a pool of young and largely non-functional genes in the transposable element-rich region of the genome. The duplicates rapidly accumulated mutations after gene duplication events, followed

by gaining of a signal peptide for secretion (Poppe *et al.* 2015). Similarly, large gene family expansions and evolution of new genes were also shown in the genomes of plant parasitic nematodes, which is very important in successful plant-nematode interactions (Kikuchi, Eves-van den Akker and Jones 2017). New functions can be adapted from endogenous genes through a process of duplication followed by diversification (Mei *et al.* 2015). For example, the SPRY domain is wide-spread among eukaryotes and thought to be involved in mediating protein-protein interactions (Woo *et al.* 2006). Interestingly, a large expansion of SPRY domain-containing proteins were identified in some plant parasitic nematodes such as *G. rostochiensis* and *G. pallida*, which are hypothesised to result from gene duplications and recombinations (Cotton *et al.* 2014; Diaz-Granados *et al.* 2016). Phylogenetic analysis further suggested that the conserved SPRY core is probably the most ancient part of the SPRY domain architecture (Diaz-Granados *et al.* 2016). Some members of the expanded family of SPRY domain-containing proteins in *Globodera* species carry a N-terminal signal peptide for secretion, localise to the gland cells, and they are therefore considered as 'effectors' (termed SPRYSEC effectors) (Mei *et al.* 2015). For comparison, far fewer SPRY domain-containing proteins were encoded in the genome of the free-living nematode *C. elegans*, and plant parasitic nematodes *B. xylophilus* and *M. incognita*, none of which harbour a signal peptide for secretion, suggesting that SPRYSEC effectors may be an adaptation specific to cyst nematodes (Mei *et al.* 2015).

For the *R. reniformis* GS family, the conservation of the ATP-binding domain suggests the expanded *gs* sequences were most likely derived from the ancestral *gs* gene by a series of gene duplication events. Furthermore, the overall sequence identities between the *R. reniformis* canonical GS and all non-canonical GS are very low, suggesting that the duplicated genes may have undergone a large number of mutations to create the non-canonical GS. In addition, Clade 3, containing the so-called 'effector' GS is clearly larger than Clade 2, which may reflect that effector gene sequences change at significantly higher rates due to the strong evolutionary pressure (Hartmann and Croll 2017). Also, the fact that average sequence identities between Clade 2 *gs* genes are much higher than those within Clade 3 supports the argument that 'effector' GS underwent more mutation events during evolution.

Taken together, an overview of the evolutionary history of the *R. reniformis* GS family is summarised here. The canonical *gs* gene is considered as an ancestral gene that

shares similar structural and functional characteristics with the *gs* genes from other species. The non-canonical GS genes, Clade 2 and 3 GS, represent parallel evolutions from the canonical GS via multiple gene duplication events, followed by independent diversifications to create novel and as yet undetermined functions. Meanwhile, both the Clade 2 and 3 GS maintain basic GS structural features such as binding sites for ATP and a likely small peptide substrate. However, these non-canonical *gs* genes also obtained novel unique domains during evolution. For example, the Clade 2 *gs* genes acquired a short and somewhat variable extension at the C-terminal, which has as yet unknown function. The Clade 3 *gs* genes gained a signal peptide for secretion at the N-terminal and also experienced more mutations than Clade 2 *gs*, presumably as a result of the evolutionary arms race between host and pathogen. Given the fact that the Clade 3 *gs* expansions were only identified in syncytia-forming nematodes, 'effector' GS may play an essential role in syncytia formation. Furthermore, the evolution of novel gene functions often also involves the recruitment of new transcriptional regulation patterns (Kikuchi, Eves-van den Akker and Jones 2017). In the case of nematode effectors, the new genes should be expressed in the effector-producing tissue, the pharyngeal gland cell(s), at the parasitic stage of the life-cycle. Therefore, the birth of these 'effector' genes may be linked to the translocation of the regulatory element and motif in the promoter regions of the associated genes (Fouche, Plissonneau and Croll 2018). The dorsal gland box (DOG box) was recently identified as a putative promoter element for dorsal gland effectors of cyst nematodes (Eves-van den Akker *et al.* 2014; Eves-van den Akker and Birch 2016). An enrichment of DOG boxes was demonstrated in the promoters of some GS-like genes from *Globodera* spp. although there was no direct correlation between the number of DOG box motifs per promoter and temporal expression (Lilley *et al.* 2018). In the genome of *R. reniformis*, a variant of the DOG box was also identified (Showmaker *et al.* 2019). Accordingly, one of the future plans of *R. reniformis* GS project will be the investigation of the enrichment in Clade 3 GS.

7.2 Neofunctionalisation: from endogenous genes to effectors

As introduced above, three distinct trajectories are suggested to be responsible for the evolution of functions after gene duplication events. In one, the duplicated gene can develop a novel function (Voordeckers *et al.* 2012). As early as the 1930s, a report described how copies of existing genes may contribute to novel genes with new

functions, highlighting for the first time the potential importance of neofunctionalisation from refashioned copies of old genes (Haldane 1933). Until now, a lot of molecular, genetic and genomic studies have confirmed the hypothesis that subtle genetic modifications of pre-existing ancestral genes may have significantly contributed to the evolution of lineage- or species-specific phenotypic traits (Kaessmann 2010). In *R. reniformis*, a large number of non-canonical *gs* genes were identified. Considering that these enzymes possess extremely low typical GS activities and an alternative substrate rather than γ-EC may be applied, neofunctionalisation of an endogenous ancestral *gs* gene is likely to create a re-purposed gene.

Several evolutionary models have been built to describe the neofunctionalisation of an old gene (Innan and Kondrashov 2010). The most classical model is Ohno's neofunctionalisation (Hahn 2009). This model suggests a single gene copy is enough to fulfil the required gene function. Hence, the extra copies are redundant and are subject to genetic drift in the population. The original copy will maintain its function and the novel copy will be either pseudogenised or lost from negative selection by the accumulation of neutral loss-of-function mutations. In addition, this model also suggests the dying copy can acquire a novel gene function that will be maintained by selection pressure (Hahn 2009). For the *R. reniformis* GS family, the functional diversity between Clade 2 and Clade 3 GS such as differing temporal and spatial expression may be involved in Ohno's model-driven evolution because Clade 3 *gs* genes are only found in syncytia-forming cyst and reniform nematodes. The Ohno's model may explain the absence of Clade 3 GS in the non-syncytia-forming nematodes as the specification of nematode lifestyle may provide the selective advantage responsible for the remarkable conservation of Clade 3 GS in the syncytia-forming nematodes. However, Ohno's model does not explain the large number of non-canonical *R. reniformis* GS members because it assumes the duplicated gene copies are not necessary to maintain the same functions.

One recent model is called 'escape from adaptive conflict' (EAC) (Des Marais and Rausher 2008). The EAC model indicates that the ancestral single copy gene is selected to perform a novel function in addition to its primary function, leading to further constraints for optimisation of each function. This model resolves the adaptive conflict between the old and an emerging new function within a single gene by allowing each daughter gene to specialise to perform either the ancestral or the novel function after

gene duplication (Des Marais and Rausher 2008). A clear experimental evidence for EAC-driven evolution is the type-III antifreeze protein gene that has been duplicated from an old sialic acid synthetase (SAS) gene in an Antarctic zoarcid fish (Deng *et al.* 2010). The SAS has both sialic acid synthetase and rudimentary ice-binding activity. Interestingly, in the new duplicate, the N-terminal SAS domain was removed and substituted with a nascent signal peptide, eliminating the biochemical conflict between SAS and ice-binding activities and allowing rapid evolution to become a secreted protein capable of non-colligative freezing-point depression. Considering the non-canonical GS possess extremely low enzyme activity but can still produce glutathione, the overall evolution of non-canonical GS genes supports the EAC evolutionary model. The EAC evolutionary model suggests in addition to the canonical GS activity, the ancestral canonical *gs* gene has potential to perform an extra function which remains unknown yet. During the evolution of *R. reniformis* GS family, the duplicated genes lost the original main function but maintained the extra functions and may have acquired some novel functions due to selection pressures.

Neofunctionalisation of endogenous genes has been described in plant pathogens to produce ‘effector’ genes. For example, peptidases are key endogenous regulators of many physiological processes such as embryogenesis and peptidases are known to affect spore formation and germination in fungal pathogens (Yuan and Cole 1989). By transcriptomics, comparative genomics and evolutionary analyses, numerous secreted peptidases were identified in many fungal wheat pathogens such as *Zymoseptoria tritici* (Krishnan *et al.* 2018). These secreted peptidases act as effectors that suppress apoplastic immunity by breaking down plant-derived pathogenesis-related proteins during the biotrophic phase (Doehlemann and Hemetsberger 2013). Some of these ‘effector’ peptidases were proposed to arise from a single ancestral gene, constantly evolving to acquire new functions, which is consistent with the EAC evolutionary model.

Another interesting example for neofunctionalisation of endogenous genes in plant pathogens is the GALA type III effectors from the plant pathogenic bacterium *Ralstonia solanacearum* (Remigi *et al.* 2011). The GALA family consists of six to nine members and is highly conserved within *R. solanacearum* species. These effectors were demonstrated to target the host proteins for ubiquitination, leading either to their degradation or to modification of their activity by ubiquitination (Angot *et al.* 2006).

GALAs were already present in the ancestral strain and have subsequently evolved within the *R. solanacearum* genome. Evolutionary analysis indicated a distinct pattern of selection and functional diversification that was confirmed by experimental evidence. Differential *GALAs* were required for the pathogenicity of *R.solanacearum* on different plant hosts, which is tightly linked to the difference of selection pressures between *GALAs* (Remigi *et al.* 2011). A similar situation has been described for the Tin2 effector in the fungal pathogen *Ustilago maydis* (Tanaka *et al.* 2019). Tin2 has acquired a specialised function, which is probably connected to the distinct pathogenic lifestyle of this fungus. Taken together, the evolution of the *GALA* gene family and Tin2 effector are more likely to support the Ohno's neofunctionalisation model that gene neofunctionalisation was significantly driven by selection pressures such as host specialisation. Given that *R. reniformis* has also a wide range of hosts (Robinson *et al.* 1997), the large expansion of 'effector' GS may be required for the pathogenicity on different hosts. However, this hypothesis still remains controversial because in cyst nematodes which have a narrow host range, 'effector' GS expansion was also demonstrated despite fewer members than *R. reniformis* 'effector' GS (Lilley *et al.* 2018).

An extension of the concept of gene neofunctionalisation is the catalytically inactive secreted enzymes of some fungal plant pathogens. Chitin is a polymer of N-acetylglucosamine and a structural component of the cell wall in fungi. Chitin fragments can be recognised by plant hosts and hence elicit related immunity responses in many species of plants (Kaku *et al.* 2006). Interestingly, catalytically inactive chitinases were shown to function as effectors in two cacao fungal pathogens. These effector genes encode chitinases with mutations that abolish the enzymatic activity. Despite the lack of chitinolytic activity, these inactive chitinases still retain the ability to bind chitin, preventing plant immunity by sequestering free chitin fragments (Fiorin et al. 2018). Other examples of inactive enzymes as effectors include enzymatically inactive proteases in *Phytophthora* that function as plant glucanase inhibitors (Damasceno *et al.* 2008), a truncated inactive xyloglucanase secreted by *P. sojae* as a decoy to protect its enzymatically active paralogue from the plant's defence protein (Ma et al. 2017), and a large family of inactive RNase-like effectors in cereal powdery mildews interferes with degradation of host ribosomal RNA to suppress plant immunity (Pennington *et al.* 2019). The catalytically inactive *R.reniformis* GS enzymes also support this strategy for novel effector birth, suggesting

that neofunctionalisation of inactive enzymes may constitute a widespread strategy for the evolution of effectors in plant pathogens.

7.3 Possible functions of non-canonical *R. reniformis* GS

The crystal structure of the non-canonical *R. reniformis* Clade 2 GS suggested that it accepted an alternative substrate rather than the canonical γ-EC. In addition, the glycine could also be replaced by another small molecule as the binding pocket is very flexible. As a result, the novel substrate should share a similar structure with γ-EC but be smaller at the glutamic acid portion of the di-peptide and the new product should have a carbon backbone resembling that of glutathione. Figure 7.1A shows the structures of the three canonical substrates of GS.

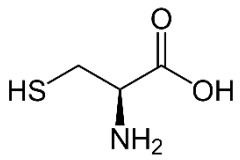
Moreover, as introduced previously, a low glutathione level *in planta* can hamper nematode parasitism (Baldacci-Cresp *et al.* 2012) and a series of low molecular weight thiols were found in the syncytia induced by *H. schachtii* (Lilley *et al.* 2018), and it is GCL rather than GS that is the rate-limiting enzyme in glutathione biosynthesis (Noctor *et al.* 2012). Therefore, the plant parasitic nematodes may utilise alternative thiols to substitute glutathione to benefit their infection. In fact, many small molecule thiols are biologically relevant to glutathione in plants (Pivato, Fabrega-Prats and Masi 2014) and in prokaryotes (Fahey 2013) because of the intrinsic reactivity of the nucleophilic sulfhydryl group. Recent mass spectroscopy analysis indicated the presence of about 300 sulfur metabolites in *Arabidopsis*. However, most of them remained unidentified and many of these could be small molecule thiols (Glaser *et al.* 2014), making it extremely challenging to predict possible plant substrates of the non-canonical *R. reniformis* Clade 3 GS. According to the structural criteria of the possible substrate, several functional small molecule thiols are predicted as putative candidates for the novel substrate or final thiol product (Figure 7.1B) despite their function in host-pathogen interactions remaining unknown. A future aim would be to explore the exact substrates and products of the non-canonical GS enzymes.

The ‘effector’ GS have been demonstrated to be highly expressed in the early parasitic stage of corresponding nematodes (highest at 7 dpi for *G. pallida* GS (Cotton *et al.* 2014); sedentary female stage for *R. reniformis* GS (Lilley *et al.* 2018)). Additional transcriptomic data for *R. reniformis* analysed for this thesis provided more precise expression information and in the study of *G. pallida* GS, the transcriptomic data

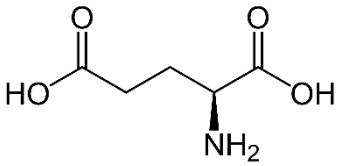
covered egg, pre-infective J2, 7 dpi, 14 dpi, 21 dpi, 28 dpi, 35 dpi and adult male. Interestingly, a very recent life-stage specific transcriptome analysis of *H. schachtii* (Sebastian Eves-van den Akker, personal communication) with more focus on the very early stages of infection showed that most 'effector' *gs* genes of *H. schachtii* were highest expressed at 48 hours post infection, suggesting that 'effector' GS may play significant roles at this earliest stage of nematode parasitism. As introduced in Chapter 1, at the early stage of syncytia-forming nematode infection, the nematodes select a suitable place and then become sedentary to establish the feeding site syncytium. Accordingly, the 'effector' GS may be involved in the syncytium establishment and formation. Of course, more detailed expression profiles of *gs* genes from 48 hpi to 7 dpi are needed to support this hypothesis as expression may actually peak between these time points. Nevertheless, the new data suggest an earlier role in parasitism than previously assumed.

Plant parasitic nematodes are believed to manipulate host redox homeostasis to facilitate successful parasitism (Siddique *et al.* 2014; Siddique and Grundler 2018). Also, syncytia were shown to be a pool of novel thiols with unknown origin (Lilley *et al.* 2018) and glutathione has been considered as a positive regulator of both cyst nematode and root-knot nematode (Baldacci-Cresp *et al.* 2012; Lilley *et al.* 2018). Taken together, the future work may focus on whether these 'effector' GS play any role in host redox status regulations and how?

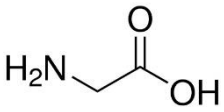
A



Cysteine

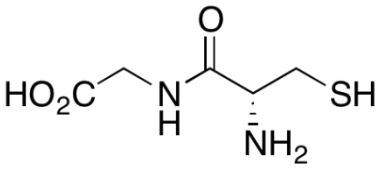


Glutamate

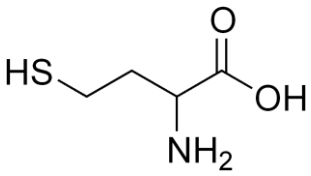


Glycine

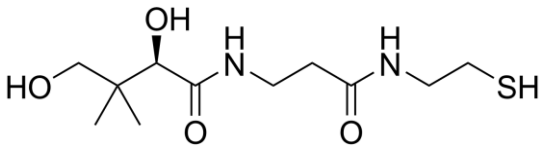
B



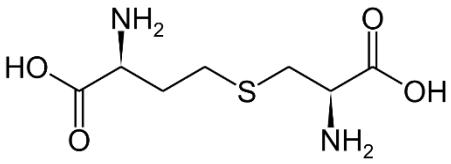
Cysteinylglycine



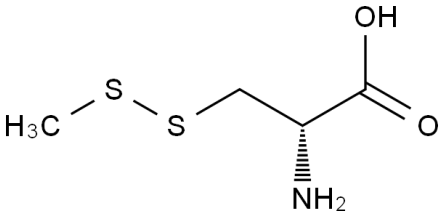
Homocysteine



Pantetheine



Cystathionine



Thiocysteine

Figure 7.1: Known small molecule thiols of plants. (A) The structures of canonical substrates of GS. (B) Putative candidates for the novel substrate or final thiol product of *R. reniformis* non-canonical GS.

7.3 Future plan

The future work will focus on two scientific questions:

1) what are the novel substrates for the non-canonical GS enzymes.

The novel substrates are hypothesised to be sulfydryl-containing compounds. In addition, Clade 3 GS are predicted to be secreted into host and play a role in plant-nematode interactions. Therefore, direct thiol-related compounds examination of *R. reniformis*-infected root tissues by mass spectrometry or high-performance liquid chromatography may contribute to the discovery of the novel product of the non-canonical *R. reniformis* GS enzymes. Furthermore, homozygous GS transgenic Arabidopsis could also be exploited to examine possible substrates for non-canonical GS enzymes. In addition, the crystal structure of GS1 with ligands needs to be solved, which can help to better understand the reaction mechanism of canonical GS enzyme.

2) why *R. reniformis*/plant parasitic nematodes have such a large number of GS?

The nematode GS phylogeny indicated that GS family is divided into three major clade and is hypothesised to witness at least two expansion during gene evolutions. As discussed in the Chapter 3, the Clade 1 contains only one sequences from each nematode species in the phylogeny except *M. incognita* and *S. ratti* due to their polyploid genome while Clade 2 shows an expansion of genes from plant parasitic nematodes belonging to the order Tylenchida. Interestingly, Clade 3 GS were present in syncytium-forming reniform and cyst nematodes, indicating GS effector may have a role in syncytium formation. Therefore, *in planta* RNAi will be exploited to examine whether syncytium formation will be influenced by a decrease of Clade 3 GS transcripts. Furthermore, GS effectors were predicted to manipulate host redox status as discussed in the Chapter 6. ROS burst level can be examined in nematode-infected *in planta* RNAi plants or homozygous GS transgenic Arabidopsis.

8 References

ABAD, P., J. GOUZY, J. M. AURY, P. CASTAGNONE-SERENO, E. G. DANCHIN, E. DELEURY, L. PERFUS-BARBEOCH, V. ANTHOUARD, F. ARTIGUENAVE, V. C. BLOK, M. C. CAILLAUD, P. M. COUTINHO, C. DASILVA, F. DE LUCA, F. DEAU, M. ESQUIBET, T. FLUTRE, J. V. GOLDSTONE, N. HAMAMOUCH, T. HEWEZI, O. JAILLON, C. JUBIN, P. LEONETTI, M. MAGLIANO, T. R. MAIER, G. V. MARKOV, P. MCVEIGH, G. PESOLE, J. POULAIN, M. ROBINSON-RECHAVI, E. SALLET, B. SEGURENS, D. STEINBACH, T. TYTGAT, E. UGARTE, C. VAN GHELDER, P. VERONICO, T. J. BAUM, M. BLAXTER, T. BLEVE-ZACHEO, E. L. DAVIS, J. J. EWBANK, B. FAVERY, E. GRENIER, B. HENRISSAT, J. T. JONES, V. LAUDET, A. G. MAULE, H. QUESNEVILLE, M. N. ROSSO, T. SCHIEX, G. SMANT, J. WEISSENBACH and P. WINCKER. 2008. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nat Biotechnol,* **26**(8), pp.909-15.

ADAMS, P. D., P. V. AFONINE, G. BUNKOCZI, V. B. CHEN, I. W. DAVIS, N. ECHOLS, J. J. HEADD, L. W. HUNG, G. J. KAPRAL, R. W. GROSSE-KUNSTLEVE, A. J. MCCOY, N. W. MORIARTY, R. OEFFNER, R. J. READ, D. C. RICHARDSON, J. S. RICHARDSON, T. C. TERWILLIGER and P. H. ZWART. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography,* **66**, pp.213-221.

AFGAN, E., D. BAKER, B. BATUT, M. VAN DEN BEEK, D. BOUVIER, M. CECH, J. CHILTON, D. CLEMENTS, N. CORAOR, B. A. GRUNING, A. GUERLER, J. HILLMAN-JACKSON, S. HILTEMANN, V. JALILI, H. RASCHE, N. SORANZO, J. GOECKS, J. TAYLOR, A. NEKRUTENKO and D. BLANKENBERG. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research,* **46**(W1), pp.W537-W544.

AGUDELO, P., R. T. ROBBINS, J. M. STEWART and J. B. MURPHY. 2004. Glycoproteins in the gelatinous matrix of Rotylenchulus reniformis. *Nematropica,* **34**(2), pp.229-234.

ALFANO, J. R. and A. COLLMER. 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu Rev Phytopathol,* **42**, pp.385-414.

ALI, M. A., A. ABBAS, F. AZEEM, N. JAVED and H. BOHLMANN. 2015. Plant-nematode Interactions: From Genomics to Metabolomics. *International Journal of Agriculture and Biology,* **17**(6), pp.1071-1082.

ANGOT, A., N. PEETERS, E. LECHNER, F. VAILLEAU, C. BAUD, L. GENTZBITTEL, E. SARTOREL, P. GENSCHIK, C. BOUCHER and S. GENIN. 2006. Ralstonia solanacearum requires F-box-like domain-containing type III effectors to promote disease on several host plants. *Proceedings of the National Academy of Sciences of the United States of America,* **103**(39), pp.14620-14625.

AU, K. K. C., J. PEREZ-GOMEZ, H. NETO, C. MULLER, A. J. MEYER, M. D. FRICKER and I. MOORE. 2012. A perturbation in glutathione biosynthesis disrupts endoplasmic reticulum morphology and secretory membrane traffic in Arabidopsis thaliana. *Plant Journal,* **71**(6), pp.881-894.

AYALA, A. and C. T. J. J. O. A. O. T. U. O. P. R. RAMÍREZ. 1964. Host-range, distribution, and bibliography of the reniform nematode, Rotylenchulus reniformis, with special reference to Puerto Rico. **48**(2), pp.140-161.

BAACK, E. J. and L. H. RIESEBERG. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development,* **17**(6), pp.513-518.

BALDACCI-CRESP, F., C. CHANG, M. MAUCOURT, C. DEBORDE, J. HOPKINS, P. LECOMTE, S. BERNILLON, R. BROUQUISSE, A. MOING, P. ABAD, D. HEROUART, A. PUPPO, B. FAVERY and P. FRENDO. 2012. (Homo)glutathione Deficiency Impairs Root-knot Nematode Development in Medicago truncatula. *PLoS Pathog,* **8**(1).

BALL, L., G. P. ACCOTTO, U. BECHTOLD, G. CREISSEN, D. FUNCK, A. JIMENEZ, B. KULAR, N. LEYLAND, J. MEJIA-CARRANZA, H. REYNOLDS, S. KARPINSKI and P. M. MULLINEAUX. 2004. Evidence for a direct link between glutathione biosynthesis and stress defense gene expression in Arabidopsis. *Plant Cell,* **16**(9), pp.2448-62.

BASYONI, M. M. and E. M. RIZK. 2016. Nematodes ultrastructure: complex systems and processes. *J Parasit Dis,* **40**(4), pp.1130-1140.

BELLAFIORE, S., Z. SHEN, M. N. ROSSO, P. ABAD, P. SHIH and S. P. BRIGGS. 2008. Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential. *PLoS Pathog,* **4**(10), pe1000192.

BERNARD, G. C., M. EGNIN and C. BONSI. 2017. The Impact of Plant-Parasitic Nematodes on Agriculture and Methods of Control. *Nematology-Concepts, Diagnosis and Control.* Intechopen.

BIRD, A. F. 1961. The ultrastructure and histochemistry of a nematode-induced giant cell. *J Biophys Biochem Cytol,* **11**, pp.701-15.

BIRD, L. E. 2011. High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods,* **55**(1), pp.29-37.

BLANC-MATHIEU, R., L. PERFUS-BARBEOCH, J. M. AURY, M. DA ROCHA, J. GOUZY, E. SALLET, C. MARTIN-JIMENEZ, M. BAILLY-BECHET, P. CASTAGNONE-SERENO, J. F. FLOT, D. K. KOZLOWSKI, J. CAZARETH, A. COULOUX, C. DA SILVA, J. GUY, Y. J. KIM-JO, C. RANCUREL, T. SCHIEX, P. ABAD, P. WINCKER and E. G. J. DANCHIN. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *Plos Genetics,* **13**(6).

BLAXTER, M. L., P. DE LEY, J. R. GAREY, L. X. LIU, P. SCHELDEMAN, A. VIERSTRAETE, J. R. VANFLETEREN, L. Y. MACKEY, M. DORRIS, L. M. FRISSE, J. T. VIDA and W. K. THOMAS. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature,* **392**(6671), pp.71-75.

BOBAY, B. G., P. DIGENNARO, E. SCHOLL, N. IMIN, M. A. DJORDJEVIC and D. M. BIRD. 2013. Solution NMR studies of the plant peptide hormone CEP inform function. *Febs Letters,* **587**(24), pp.3979-3985.

BOHLMANN, H. and M. SOBCZAK. 2014. The plant cell wall in the feeding sites of cyst nematodes. *Front Plant Sci,* **5**, p89.

BOLGER, A. M., M. LOHSE and B. USADEL. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30**(15), pp.2114-2120.

BRIGELIUS, R., C. MUCKEL, T. P. AKERBOOM and H. J. B. P. SIES. 1983. Identification and quantitation of glutathione in hepatic protein mixed disulfides and its relationship to glutathione disulfide. **32**(17), pp.2529-2534.

BROWN, T. R., M. L. DRUMMOND, S. BARELIER, A. S. CRUTCHFIELD, A. DINESCU, K. D. SLAVENS, T. R. CUNDARI and M. E. ANDERSON. 2011. Aspartate 458 of human glutathione synthetase is important for cooperativity

and active site structure. *Biochemical and Biophysical Research Communications,* **411**(3), pp.536-542.

BURKE, M., E. H. SCHOLL, D. M. BIRD, J. E. SCHAFF, S. D. COLMAN, R. CROWELL, S. DIENER, O. GORDON, S. GRAHAM, X. G. WANG, E. WINDHAM, G. M. WRIGHT and C. H. OPPERMAN. 2015. The plant parasite Pratylenchus coffeae carries a minimal nematode genome. *Nematology,* **17**, pp.621-637.

BUZIE, A. and C. ENJUAKWEI. 2007. Characterization of the enzymes involved in the glutathione synthesis and the redox enzyme glutathione reductase in Caenorhabditis elegans (Maupas, 1900).

CAIRNS, N. G., M. PASTERNAK, A. WACHTER, C. S. COBBETT and A. J. MEYER. 2006. Maturation of Arabidopsis seeds is dependent on glutathione biosynthesis within the embryo. *Plant Physiology,* **141**(2), pp.446-455.

CAO, W. and E. M. DE LA CRUZ. 2013. Quantitative full time course analysis of nonlinear enzyme cycling kinetics. *Sci Rep,* **3**, p2658.

CARVUNIS, A. R., T. ROLLAND, I. WAPINSKI, M. A. CALDERWOOD, M. A. YILDIRIM, N. SIMONIS, B. CHARLOTEAUX, C. A. HIDALGO, J. BARBETTE, B. SANTHANAM, G. A. BRAR, J. S. WEISSMAN, A. REGEV, N. THIERRY-MIEG, M. E. CUSICK and M. VIDAL. 2012. Proto-genes and de novo gene birth. *Nature,* **487**(7407), pp.370-374.

CHALFIE, M., Y. TU, G. EUSKIRCHEN, W. W. WARD and D. C. J. S. PRASHER. 1994. Green fluorescent protein as a marker for gene expression. **263**(5148), pp.802-805.

CHEN, J. S., B. R. LIN, Q. L. HUANG, L. L. HU, K. ZHUO and J. L. LIAO. 2017. A novel Meloidogyne graminicola effector, MgGPP, is secreted into host cells and undergoes glycosylation in concert with proteolysis to suppress plant defenses and promote parasitism. *Plos Pathogens,* **13**(4).

CHEN, P. Y., C. K. WANG, S. C. SOONG and K. Y. TO. 2003. Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants. *Molecular Breeding,* **11**(4), pp.287-293.

CHEN, V. B., W. B. ARENDALL, J. J. HEADD, D. A. KEEDY, R. M. IMMORMINO, G. J. KAPRAL, L. W. MURRAY, J. S. RICHARDSON and D. C. RICHARDSON. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Structural Biology,* **66**, pp.12-21.

CHENG, J.-C., K. A. SEELEY and Z. R. J. P. P. SUNG. 1995. RML1 and RML2, Arabidopsis genes required for cell proliferation at the root tip. **107**(2), pp.365-376.

CHEVREUX, B. 2007. MIRA: an automated genome and EST assembler.

CHITWOOD, D. J. and R. N. J. R.-K. N. PERRY. 2009. Reproduction, physiology and biochemistry. pp.182-200.

CONSORTIUM, C. E. S. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science,* **282**(5396), pp.2012-8.

COPLEY, S. D. and J. K. DHILLON. 2002. Lateral gene transfer and parallel evolution in the history of glutathione biosynthesis genes. *Genome Biology,* **3**(5).

COTTON, J. A., C. J. LILLEY, L. M. JONES, T. KIKUCHI, A. J. REID, P. THORPE, I. J. TSAI, H. BEASLEY, V. BLOK, P. J. COCK, S. EVES-VAN DEN AKKER, N. HOLROYD, M. HUNT, S. MANTELIN, H. NAGHRA, A. PAIN, J. E. PALOMARES-RIUS, M. ZAROWIECKI, M. BERRIMAN, J. T. JONES and P. E. URWIN. 2014. The genome and life-stage specific transcriptomes of Globodera

pallida elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol,* **15**(3), pR43.

COWTAN, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D-Biological Crystallography,* **62**, pp.1002-1011.

DALTON, D. A., L. LANGEBERG and N. C. TRENEMAN. 1993. Correlations between the Ascorbate-Glutathione Pathway and Effectiveness in Legume Root-Nodules. *Physiologia Plantarum,* **87**(3), pp.365-370.

DALTON, D. A., S. A. RUSSELL, F. J. HANUS, G. A. PASCOE and H. J. EVANS. 1986. Enzymatic-Reactions of Ascorbate and Glutathione That Prevent Peroxide Damage in Soybean Root-Nodules. *Proceedings of the National Academy of Sciences of the United States of America,* **83**(11), pp.3811-3815.

DALTON, T. P., M. Z. DIETER, Y. YANG, H. G. SHERTZER, D. W. J. B. NEBERT and B. R. COMMUNICATIONS. 2000. Knockout of the mouse glutamate cysteine ligase catalytic subunit (Gclc) gene: embryonic lethal when homozygous, and proposed model for moderate glutathione deficiency when heterozygous. **279**(2), pp.324-329.

DAMASCENO, C. M., J. G. BISHOP, D. R. RIPOLL, J. WIN, S. KAMOUN and J. K. J. M. P.-M. I. ROSE. 2008. Structure of the glucanase inhibitor protein (GIP) family from Phytophthora species suggests coevolution with plant endo-β-1, 3-glucanases. **21**(6), pp.820-830.

DANCHIN, E. G. J., M. N. ROSSO, P. VIEIRA, J. DE ALMEIDA-ENGLER, P. M. COUTINHO, B. HENRISSAT and P. ABAD. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences of the United States of America,* **107**(41), pp.17651-17656.

DASGUPTA, D., D. RASKI and S. SHER. 2011. A Revision of the Genus Rotylenchulus Liiiford and Oliveira, 1940 (Nematoda: Tylenchiclae).

DAVIS, E. L., R. S. HUSSEY and T. J. BAUM. 2004. Getting to the roots of parasitism by nematodes. *Trends Parasitol,* **20**(3), pp.134-41.

DAVIS, E. L., R. S. HUSSEY, T. J. BAUM, J. BAKKER and A. SCHOTS. 2000. Nematode parasitism genes. *Annual Review of Phytopathology,* **38**, pp.365-396.

DAVIS, E. L., R. S. HUSSEY, M. G. MITCHUM and T. J. J. C. O. I. P. B. BAUM. 2008. Parasitism proteins in nematode–plant interactions. **11**(4), pp.360-366.

DE BOER, J.M., Yan, Y., Smant, G., Davis, E.L. and Baum, T.J., 1998. In-situ hybridization to messenger RNA in Heterodera glycines. Journal of nematology, 30(3), p.309.

DECRAEMER, W. and D. J. HUNT. 2006. Structure and Classification. *Plant Nematology*, pp.3-32.

DEL ANGEL, V. D., E. HJERDE, L. STERCK, S. CAPELLA-GUTIERREZ, C. NOTREDAME, O. V. PETTERSSON, J. AMSELEM, L. BOURI, S. BOCS and C. J. F. KLOPP. 2018. Ten steps to get started in Genome Assembly and Annotation. **7**.

DELANO, W. L. J. H. W. P. O. 2002. The PyMOL molecular graphics system.

DENG, C., C. H. C. CHENG, H. YE, X. M. HE and L. B. CHEN. 2010. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proceedings of the National Academy of Sciences of the United States of America,* **107**(50), pp.21593-21598.

DES MARAIS, D. L. and M. D. RAUSHER. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature,* **454**(7205), pp.762-U85.

DIAZ-GRANADOS, A., A. J. PETRESCU, A. GOVERSE and G. SMANT. 2016. SPRYSEC Effectors: A Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity. *Frontiers in Plant Science,* **7**.

DINESCU, A., M. E. ANDERSON and T. R. CUNDARI. 2007. Catalytic loop motion in human glutathione synthetase: A molecular modeling approach. *Biochemical and Biophysical Research Communications,* **353**(2), pp.450-456.

DINESCU, A., T. R. CUNDARI, V. S. BHANSALI, J. L. LUO and M. E. ANDERSON. 2004. Function of conserved residues of human glutathione synthetase: implications for the ATP-grasp enzymes. *J Biol Chem,* **279**(21), pp.22412-21.

DOEHLEMANN, G. and C. HEMETSBERGER. 2013. Apoplastic immunity and its suppression by filamentous plant pathogens. *New Phytologist,* **198**(4), pp.1001-1016.

DROPKIN, V. J. E. B. 1972. Pathology of meloidogyne—Galling, giant cell formation, effects on host physiology. **2**(6), pp.23-32.

DUBREUIL-MAURIZI, C., J. VITECEK, L. MARTY, L. BRANCIARD, P. FRETTINGER, D. WENDEHENNE, A. J. MEYER, F. MAUCH and B. J. P. P. POINSSOT. 2011. Glutathione deficiency of the Arabidopsis mutant pad2-1 affects oxidative stress-related events, defense gene expression, and the hypersensitive response. **157**(4), pp.2000-2012.

DUBREUIL, G., E. DELEURY, M. MAGLIANO, M. JAOUANNET, P. ABAD and M. N. ROSSO. 2011. Peroxiredoxins from the plant parasitic root-knot nematode, Meloidogyne incognita, are required for successful development within the host. *International Journal for Parasitology,* **41**(3-4), pp.385-396.

DUTHEIL, J. Y., G. MANNHAUPT, G. SCHWEIZER, C. MK SIEBER, M. MÜNSTERKÖTTER, U. GÜLDENER, J. SCHIRAWSKI, R. J. G. B. KAHMANN and EVOLUTION. 2016. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. **8**(3), pp.681-704.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* **32**(5), pp.1792-1797.

EDWARDS, R., J. W. BLOUNT and R. A. J. P. DIXON. 1991. Glutathione and elicitation of the phytoalexin response in legume cell cultures. **184**(3), pp.403-409.

EISENBACK, J. D. and H. H. J. M. O. A. N. TRIANTAPHYLLOU. 1991. Root-knot nematodes: Meloidogyne species and races. **1**, pp.191-274.

EL-GEBALI, S., J. MISTRY, A. BATEMAN, S. R. EDDY, A. LUCIANI, S. C. POTTER, M. QURESHI, L. J. RICHARDSON, G. A. SALAZAR, A. SMART, E. L. L. SONNHAMMER, L. HIRSH, L. PALADIN, D. PIOVESAN, S. C. E. TOSATTO and R. D. FINN. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res,* **47**(D1), pp.D427-D432.

EL MSEHLI, S., A. LAMBERT, F. BALDACCI-CRESP, J. HOPKINS, E. BONCOMPAGNI, S. A. SMITI, D. HEROUART and P. FRENDO. 2011. Crucial role of (homo)glutathione in nitrogen fixation in Medicago truncatula nodules. *New Phytologist,* **192**(2), pp.496-506.

ELZAHABY, H. M., G. GULLNER and Z. KIRALY. 1995. Effects of Powdery Mildew Infection of Barley on the Ascorbate-Glutathione Cycle and Other Antioxidants in Different Host-Pathogen Interactions. *Phytopathology,* **85**(10), pp.1225-1230.

EMSLEY, P., B. LOHKAMP, W. G. SCOTT and K. COWTAN. 2010. Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography,* **66**, pp.486-501.

ESCOBAR, C., M. BARCALA, J. CABRERA and C. FENOLL. 2015. Overview of root-knot nematodes and giant cells. *Advances in botanical research.* Elsevier, pp.1-32.

ESPADA, M., S. E. V. DEN AKKER, T. MAIER, P. VIJAYAPALANI, T. BAUM, M. MOTA and J. T. JONES. 2018. STATAWAARS: a promoter motif associated with spatial expression in the major effector-producing tissues of the plantparasitic nematode Bursaphelenchus xylophilus (vol 19, 553, 2018). *Bmc Genomics,* **19**.

ESSER, L., C. R. WANG, M. HOSAKA, C. S. SMAGULA, T. C. SUDHOF and J. DEISENHOFER. 1998. Synapsin I is structurally similar to ATP-utilizing enzymes. *Embo Journal,* **17**(4), pp.977-984.

EVES-VAN DEN AKKER, S. and P. R. J. M. P. BIRCH. 2016. Opening the effector protein toolbox for plant–parasitic cyst nematode interactions. **9**(11), pp.1451-1453.

EVES-VAN DEN AKKER, S., D. R. LAETSCH, P. THORPE, C. J. LILLEY, E. G. J. DANCHIN, M. DA ROCHA, C. RANCUREL, N. E. HOLROYD, J. A. COTTON, A. SZITENBERG, E. GRENIER, J. MONTARRY, B. MIMEE, M. O. DUCEPPE, I. BOYES, J. M. C. MARVIN, L. M. JONES, H. B. YUSUP, J. LAFOND-LAPALME, M. ESQUIBET, M. SABEH, M. ROTT, H. OVERMARS, A. FINKERS-TOMCZAK, G. SMANT, G. KOUTSOVOULOS, V. BLOK, S. MANTELIN, P. J. A. COCK, W. PHILLIPS, B. HENRISSAT, P. E. URWIN, M. BLAXTER and J. T. JONES. 2016a. The genome of the yellow potato cyst nematode, Globodera rostochiensis, reveals insights into the basis of parasitism and virulence. *Genome Biology,* **17**.

EVES-VAN DEN AKKER, S., C. J. LILLEY, E. G. J. DANCHIN, C. RANCUREL, P. J. A. COCK, P. E. URWIN and J. T. JONES. 2014. The Transcriptome of Nacobbus aberrans Reveals Insights into the Evolution of Sedentary Endoparasitism in Plant-Parasitic Nematodes. *Genome Biology and Evolution,* **6**(9), pp.2181-2194.

EVES-VAN DEN AKKER, S., C. J. LILLEY, H. B. YUSUP, J. T. JONES and P. E. URWIN. 2016b. Functional C-TERMINALLY ENCODED PEPTIDE (CEP) plant hormone domains evolved de novo in the plant parasite Rotylenchulus reniformis. *Mol Plant Pathol,* **17**(8), pp.1265-75.

FAHEY, R. C. 2013. Glutathione analogs in prokaryotes. *Biochimica Et Biophysica Acta-General Subjects,* **1830**(5), pp.3182-3198.

FAWAZ, M. V., M. E. TOPPER and S. M. FIRESTINE. 2011. The ATP-grasp enzymes. *Bioorganic Chemistry,* **39**(4-6), pp.185-191.

FERRARI, S., J. M. PLOTNIKOVA, G. DE LORENZO and F. M. J. T. P. J. AUSUBEL. 2003. Arabidopsis local resistance to Botrytis cinerea involves salicylic acid and camalexin and requires EDS4 and PAD2, but not SID2, EDS5 or PAD4. **35**(2), pp.193-205.

FIORIN, G. L., A. SANCHEZ-VALLET, D. P. D. THOMAZELLA, P. F. V. DO PRADO, L. C. DO NASCIMENTO, A. V. D. FIGUEIRA, B. P. H. J. THOMMA, G. A. G. PEREIRA and P. J. P. L. TEIXEIRA. 2018. Suppression of Plant Immunity by Fungal Chitinase-like Effectors. *Current Biology,* **28**(18), pp.3023-+.

FLETCHER, L. C., U. BRAND, M. P. RUNNING, R. SIMON and E. M. MEYEROWITZ. 1999. Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science,* **283**(5409), pp.1911-1914.

FODOR, J., G. GULLNER, A. L. ADAM, B. BARNA, T. KOMIVES and Z. KIRALY. 1997. Local and systemic responses of antioxidants to tobacco mosaic virus infection and to salicylic acid in tobacco - Role in systemic acquired resistance. *Plant Physiol,* **114**(4), pp.1443-1451.

FOSU-NYARKO, J., P. NICOL, F. NAZ, R. GILL and M. G. JONES. 2016. Analysis of the Transcriptome of the Infective Stage of the Beet Cyst Nematode, H. schachtii. *PLoS One,* **11**(1), pe0147511.

FOUCHE, S., C. PLISSONNEAU and D. CROLL. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology,* **46**, pp.34-42.

FRANTZESKAKIS, L., A. DI PIETRO, M. REP, J. SCHIRAWSKI, C. H. WU and R. PANSTRUGA. 2019. Rapid evolution in plant-microbe interactions - a molecular genomics perspective. *New Phytologist*.

FRENDO, P., J. HARRISON, C. NORMAN, M. J. H. JIMENEZ, G. VAN DE SYPE, A. GILABERT and A. PUPPO. 2005. Glutathione and homoglutathione play a critical role in the nodulation process of Medicago truncatula. *Molecular Plant-Microbe Interactions,* **18**(3), pp.254-259.

FRENDO, P., M. J. H. JIMENEZ, C. MATHIEU, L. DURET, D. GALLESI, G. VAN DE SYPE, D. HEROUART and A. PUPPO. 2001. A Medicago truncatula homoglutathione synthetase is derived from glutathione synthetase by gene duplication. *Plant Physiology,* **126**(4), pp.1706-1715.

FRIESEN, T. L., E. H. STUKENBROCK, Z. H. LIU, S. MEINHARDT, H. LING, J. D. FARIS, J. B. RASMUSSEN, P. S. SOLOMON, B. A. MCDONALD and R. P. OLIVER. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics,* **38**(8), pp.953-956.

FYFE, P. K., M. S. ALPHEY and W. N. HUNTER. 2010. Structure of Trypanosoma brucei glutathione synthetase: Domain and loop alterations in the catalytic cycle of a highly conserved enzyme. *Molecular and Biochemical Parasitology,* **170**(2), pp.93-99.

GALANT, A., K. A. ARKUS, C. ZUBIETA, R. E. CAHOON and J. M. JEZ. 2009. Structural basis for evolution of product diversity in soybean glutathione biosynthesis. *Plant Cell,* **21**(11), pp.3450-8.

GALANT, A., M. L. PREUSS, J. C. CAMERON and J. M. JEZ. 2011. Plant glutathione biosynthesis: diversity in biochemical regulation and reaction products. *Front Plant Sci,* **2**, p45.

GALI, R. R. and P. G. BOARD. 1995. Sequencing and Expression of a Cdna for Human Glutathione Synthetase. *Biochemical Journal,* **310**, pp.353-358.

GALPERIN, M. Y. and E. V. KOONIN. 1997. A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci,* **6**(12), pp.2639-43.

Ganji, S., Sanders, W.S., Stokes, J.V., Showmaker, K., Bartlett, B., Wang, H., et al. 2013. The genome of reniform nematode, *Rotylenchulus reniformis* 2013. In Proceedings of the 10th Conference of the Midsouth Computational Biology and Bioinformatics Society, April 5–6, 2013, Columbia, Mo. 102 pp.

GANJI, S., J. N. JENKINS and M. J. J. G. WUBBEN. 2014. Molecular characterization of the reniform nematode C-type lectin gene family reveals a likely role in mitigating environmental stresses during plant parasitism. **537**(2), pp.269-278.

GANJI, S., M. J. WUBBEN and J. N. JENKINS. 2013. Two Simple Methods for the Collection of Individual Life Stages of Reniform Nematode, Rotylenchulus reniformis. *Journal of Nematology,* **45**(2), pp.87-91.

GAO, B., R. ALLEN, E. L. DAVIS, T. J. BAUM and R. S. HUSSEY. 2004. Molecular characterisation and developmental expression of a cellulose-binding protein gene in the soybean cyst nematode Heterodera glycines. *Int J Parasitol,* **34**(12), pp.1377-83.

GAO, B., R. ALLEN, T. MAIER, E. L. DAVIS, T. J. BAUM and R. S. HUSSEY. 2003. The parasitome of the phytonematode Heterodera glycines. *Mol Plant Microbe Interact,* **16**(8), pp.720-6.

GAUR, H. and R. J. A. Z. R. PERRY. 1991. The biology and control of the plant parasitic nematode Rotylenchulus reniformis. **4**, pp.177-212.

GELADOPOULOS, T. P., T. G. SOTIROUDIS and A. E. EVANGELOPOULOS. 1991. A Malachite Green Colorimetric Assay for Protein Phosphatase-Activity. *Analytical Biochemistry,* **192**(1), pp.112-116.

GHEYSEN, G. and M. G. MITCHUM. 2011. How nematodes manipulate plant development pathways for infection. *Current Opinion in Plant Biology,* **14**(4), pp.415-421.

GHEYSEN, G. and M. G. MITCHUM. 2019. Phytoparasitic Nematode Control of Plant Hormone Pathways. *Plant Physiology,* **179**(4), pp.1212-1226.

GIACOVAZZO, C., H. L. MONACO, G. ARTIOLI, D. VITERBO, G. FERRARIS and C. GIACOVAZZO. 2002. *Fundamentals of crystallography.* Oxford University Press Oxford.

GLASER, K., B. KANAWATI, T. KUBO, P. SCHMITT-KOPPLIN and E. GRILL. 2014. Exploring the Arabidopsis sulfur metabolome. *Plant Journal,* **77**(1), pp.31-45.

GOGOS, A. and L. SHAPIRO. 2002. Large conformational changes in the catalytic cycle of glutathione synthase. *Structure,* **10**(12), pp.1669-76.

GOLINOWSKI, W., F. GRUNDLER and M. J. P. SOBCZAK. 1996. Changes in the structure ofArabidopsis thaliana during female development of the plant-parasitic nematodeHeterodera schachtii. **194**(1-2), pp.103-116.

GOVERSE, A. and G. SMANT. 2014. The activation and suppression of plant innate immunity by parasitic nematodes. *Annu Rev Phytopathol,* **52**, pp.243-65.

GRABHERR, M. G., B. J. HAAS, M. YASSOUR, J. Z. LEVIN, D. A. THOMPSON, I. AMIT, X. ADICONIS, L. FAN, R. RAYCHOWDHURY and Q. J. N. B. ZENG. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **29**(7), p644.

GRANT, C. M., F. H. MACIVER and I. W. J. M. B. O. T. C. DAWES. 1997. Glutathione synthetase is dispensable for growth under both normal and oxidative stress conditions in the yeast Saccharomyces cerevisiae due to an accumulation of the dipeptide gamma-glutamylcysteine. **8**(9), pp.1699-1707.

GUO, X., J. WANG, M. GARDNER, H. FUKUDA, Y. KONDO, J. P. ETCHELLS, X. WANG and M. G. MITCHUM. 2017. Identification of cyst nematode B-type CLE peptides and modulation of the vascular stem cell pathway for feeding cell formation. *PLoS Pathog,* **13**(2), pe1006142.

GUSHIMA, H., T. MIYA, K. MURATA and A. KIMURA. 1983. Purification and characterization of glutathione synthetase from Escherichia coli B. *J Appl Biochem,* **5**(3), pp.210-8.

GUSHIMA, H., S. YASUDA, E. SOEDA, M. YOKOTA, M. KONDO and A. J. N. A. R. KIMURA. 1984. Complete nucleotide sequence of the E. coli glutathione synthetase gsh-II. **12**(24), pp.9299-9307.

HABASH, S. S., Z. S. RADAKOVIC, R. VANKOVA, S. SIDDIQUE, P. DOBREV, C. GLEASON, F. M. W. GRUNDLER and A. ELASHRY. 2017. Heterodera schachtii Tyrosinase-like protein - a novel nematode effector modulating plant hormone homeostasis. *Scientific Reports,* **7**.

HAEGEMAN, A., L. BAUTERS, T. KYNDT, M. M. RAHMAN and G. GHEYSEN. 2013. Identification of candidate effector genes in the transcriptome of the rice root knot nematode Meloidogyne graminicola. *Mol Plant Pathol,* **14**(4), pp.379-90.

HAEGEMAN, A., J. JACOB, B. VANHOLME, T. KYNDT, M. MITREVA and G. GHEYSEN. 2009. Expressed sequence tags of the peanut pod nematode Ditylenchus africanus: the first transcriptome analysis of an Anguinid nematode. *Mol Biochem Parasitol,* **167**(1), pp.32-40.

HAHN, M. W. J. J. O. H. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. **100**(5), pp.605-617.

HALDANE, J. B. S. 1933. The part played by recurrent mutation in evolution. *American Naturalist,* **67**, pp.5-19.

HARA, T., H. KATO, Y. KATSUBE and J. I. J. B. ODA. 1996. A pseudo-Michaelis quaternary complex in the reverse reaction of a ligase: structure of Escherichia coli B glutathione synthetase complexed with ADP, glutathione, and sulfate at 2.0 Å resolution. **35**(37), pp.11967-11974.

HARA, T., T. TANAKA, H. KATO, T. NISHIOKA, J. I. J. P. E. ODA, DESIGN and SELECTION. 1995. Site-directed mutagenesis of glutathione synthetase from Escherichia coli B: mapping of the γ-L-glutamyl-L-cysteine-binding site. **8**(7), pp.711-716.

HARMS, K., P. VON BALLMOOS, C. BRUNOLD, R. HOFGEN and H. HESSE. 2000. Expression of a bacterial serine acetyltransferase in transgenic potato plants leads to increased levels of cysteine and glutathione. *Plant Journal,* **22**(4), pp.335-343.

HARTMANN, F. E. and D. CROLL. 2017. Distinct Trajectories of Massive Recent Gene Gains and Losses in Populations of a Microbial Eukaryotic Pathogen. *Molecular Biology and Evolution,* **34**(11), pp.2808-2822.

HENKLE-DUHRSEN, K. and A. KAMPKOTTER. 2001. Antioxidant enzyme families in parasitic nematodes. *Mol Biochem Parasitol,* **114**(2), pp.129-42.

HERRERA, K., R. E. CAHOON, S. KUMARAN and J. JEZ. 2007. Reaction mechanism of glutathione synthetase from Arabidopsis thaliana: site-directed mutagenesis of active site residues. *J Biol Chem,* **282**(23), pp.17157-65.

HEWEZI, T., P. HOWE, T. R. MAIER, R. S. HUSSEY, M. G. MITCHUM, E. L. DAVIS and T. J. BAUM. 2008. Cellulose binding protein from the parasitic nematode Heterodera schachtii interacts with Arabidopsis pectin methylesterase: cooperative cell wall modification during parasitism. *Plant Cell,* **20**(11), pp.3080-93.

HEWEZI, T., P. J. HOWE, T. R. MAIER, R. S. HUSSEY, M. G. MITCHUM, E. L. DAVIS and T. J. BAUM. 2010. Arabidopsis Spermidine Synthase Is Targeted by an Effector Protein of the Cyst Nematode Heterodera schachtii. *Plant Physiology,* **152**(2), pp.968-984.

HEWEZI, T., P. S. JUVALE, S. PIYA, T. R. MAIER, A. RAMBANI, J. H. RICE, M. G. MITCHUM, E. L. DAVIS, R. S. HUSSEY and T. J. BAUM. 2015. The cyst nematode effector protein 10A07 targets and recruits host posttranslational

machinery to mediate its nuclear trafficking and to promote parasitism in Arabidopsis. *Plant Cell,* **27**(3), pp.891-907.

HIBI, T., H. KATO, T. NISHIOKA, J. ODA, H. YAMAGUCHI, Y. KATSUBE, K. TANIZAWA and T. J. B. FUKUI. 1993. Use of adenosine (5') polyphospho (5') pyridoxals to study the substrate-binding region of glutathione synthetase from Escherichia coli B. **32**(6), pp.1548-1554.

HOFMANN, J., A. EL ASHRY, S. ANWAR, A. ERBAN, J. KOPKA and F. GRUNDLER. 2010. Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism. *Plant Journal,* **62**(6), pp.1058-1071.

HOGENHOUT, S. A., R. A. VAN DER HOORN, R. TERAUCHI and S. KAMOUN. 2009. Emerging concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact,* **22**(2), pp.115-22.

HOGG, N., R. J. SINGH and B. KALYANARAMAN. 1996. The role of glutathione in the transport and catabolism of nitric oxide. *FEBS Lett,* **382**(3), pp.223-8.

HOLBEIN, J., F. M. W. GRUNDLER and S. SIDDIQUE. 2016. Plant basal resistance to nematodes: an update. *Journal of Experimental Botany,* **67**(7), pp.2049-2061.

HOLTERMAN, M., G. KARSSEN, S. VAN DEN ELSEN, H. VAN MEGEN, J. BAKKER and J. HELDER. 2009. Small Subunit rDNA-Based Phylogeny of the Tylenchida Sheds Light on Relationships Among Some High-Impact Plant-Parasitic Nematodes and the Evolution of Plant Feeding. *Phytopathology,* **99**(3), pp.227-235.

HOWDEN, R., C. R. ANDERSEN, P. B. GOLDSBROUGH and C. S. J. P. P. COBBETT. 1995. A cadmium-sensitive, glutathione-deficient mutant of Arabidopsis thaliana. **107**(4), pp.1067-1073.

HU, L. J., X. Q. WU, H. Y. LI, Q. ZHAO, Y. C. WANG and J. R. YE. 2019. An Effector, BxSapB1, Induces Cell Death and Contributes to Virulence in the Pine Wood Nematode Bursaphelenchus xylophilus. *Mol Plant Microbe Interact*, pMPMI10180275R.

HU, Y. and T. HEWEZI. 2018. Nematode-secreted peptides and host factor mimicry. *J Exp Bot,* **69**(12), pp.2866-2868.

HUANG, C. S., W. X. HE, A. MEISTER and M. E. ANDERSON. 1995. Amino-Acid-Sequence of Rat-Kidney Glutathione Synthetase. *Proceedings of the National Academy of Sciences of the United States of America,* **92**(4), pp.1232-1236.

HUANG, Z.-Z., H. YANG, C. CHEN, Z. ZENG, S. C. J. B. E. B. A.-G. S. LU and EXPRESSION. 2000. Inducers of γ-glutamylcysteine synthetase and their effects on glutathione synthetase expression. **1493**(1-2), pp.48-55.

HUSSEY, R. S. 1989. Disease-Inducing Secretions of Plant-Parasitic Nematodes. *Annual Review of Phytopathology,* **27**, pp.123-141.

HUSSEY, R. S., E. L. DAVIS and T. J. J. B. J. O. P. P. BAUM. 2002. Secrets in secretions: genes that control nematode parasitism of plants. **14**(3), pp.183-194.

INGLE, B. L. 2015. *The Mechanisms of Human Glutathione Synthetase and Related Non-Enyzmatic Catalysis*. thesis, University of North Texas.

INGLE, B. L., B. SHRESTHA, M. C. DE JESUS, H. M. CONRAD-WEBB, M. E. ANDERSON, T. R. J. C. CUNDARI and S. B. JOURNAL. 2019. Genetic Mutations in the S-loop of Human Glutathione Synthetase: Links Between Substrate Binding, Active Site Structure and Allostery. **17**, pp.31-38.

INNAN, H. and F. J. N. R. G. KONDRASHOV. 2010. The evolution of gene duplications: classifying and distinguishing between models. **11**(2), p97.

JAOUANNET, M., M. MAGLIANO, M. J. ARGUEL, M. GOURGUES, E. EVANGELISTI, P. ABAD and M. N. ROSSO. 2013. The Root-Knot Nematode Calreticulin Mi-CRT Is a Key Effector in Plant Defense Suppression. *Molecular Plant-Microbe Interactions,* **26**(1), pp.97-105.

JAOUANNET, M. and M. N. ROSSO. 2013. Effectors of root sedentary nematodes target diverse plant cell compartments to manipulate plant functions and promote infection. *Plant Signal Behav,* **8**(9).

JEZ, J. M. and R. E. CAHOON. 2004. Kinetic mechanism of glutathione synthetase from Arabidopsis thaliana. *Journal of Biological Chemistry,* **279**(41), pp.42726-42731.

JIMÉNEZ-QUESADA, M. J., J. Á. TRAVERSO and J. D. D. J. F. I. P. S. ALCHÉ. 2016. NADPH oxidase-dependent superoxide production in plant reproductive tissues. **7**, p359.

JONES, J. T., A. HAEGEMAN, E. G. DANCHIN, H. S. GAUR, J. HELDER, M. G. JONES, T. KIKUCHI, R. MANZANILLA-LOPEZ, J. E. PALOMARES-RIUS, W. M. WESEMAEL and R. N. PERRY. 2013. Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol Plant Pathol,* **14**(9), pp.946-61.

JONES, J. T., A. KUMAR, L. A. PYLYPENKO, A. THIRUGNANASAMBANDAM, L. CASTELLI, S. CHAPMAN, P. J. COCK, E. GRENIER, C. J. LILLEY, M. S. PHILLIPS and V. C. BLOK. 2009. Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode Globodera pallida. *Mol Plant Pathol,* **10**(6), pp.815-28.

JONES, J. T., B. REAVY, G. SMANT and A. E. PRIOR. 2004. Glutathione peroxidases of the potato cyst nematode Globodera Rostochiensis. *Gene,* **324**, pp.47-54.

JONES, L. M., A. K. KOEHLER, M. TRNKA, J. BALEK, A. J. CHALLINOR, H. J. ATKINSON and P. E. URWIN. 2017. Climate change is predicted to alter the current pest status of Globodera pallida and G. rostochiensis in the United Kingdom. *Glob Chang Biol,* **23**(11), pp.4497-4507.

JONES, M. and V. J. P. P. P. DROPKIN. 1975. Cellular alterations induced in soybean roots by three endoparasitic nematodes. **5**(2), pp.119-124.

JONES, M. G. K. 1981. Host-Cell Responses to Endo-Parasitic Nematode Attack - Structure and Function of Giant-Cells and Syncytia. *Annals of Applied Biology,* **97**(3), pp.353-+.

JUNG, H., C. WINEFIELD, A. BOMBARELY, P. PRENTIS and P. WATERHOUSE. 2019. Tools and Strategies for Long-Read Sequencing and De Novo Assembly of Plant Genomes. *Trends in Plant Science,* **24**(8), pp.700-724.

KAESSMANN, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research,* **20**(10), pp.1313-1326.

KAKU, H., Y. NISHIZAWA, N. ISHII-MINAMI, C. AKIMOTO-TOMIYAMA, N. DOHMAE, K. TAKIO, E. MINAMI and N. SHIBUYA. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences of the United States of America,* **103**(29), pp.11086-11091.

KEELING, P. J. and J. D. PALMER. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics,* **9**(8), pp.605-618.

KENDREW, J. C., G. BODO, H. M. DINTZIS, R. G. PARRISH, H. WYCKOFF and D. C. PHILLIPS. 1958. 3-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature,* **181**(4610), pp.662-666.

KIKUCHI, T., J. A. COTTON, J. J. DALZELL, K. HASEGAWA, N. KANZAKI, P. MCVEIGH, T. TAKANASHI, I. J. TSAI, S. A. ASSEFA, P. J. A. COCK, T. D.

OTTO, M. HUNT, A. J. REID, A. SANCHEZ-FLORES, K. TSUCHIHARA, T. YOKOI, M. C. LARSSON, J. MIWA, A. G. MAULE, N. SAHASHI, J. T. JONES and M. BERRIMAN. 2011. Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus. *Plos Pathogens,* **7**(9).

KIKUCHI, T., S. EVES-VAN DEN AKKER and J. T. JONES. 2017. Genome Evolution of Plant-Parasitic Nematodes. *Annu Rev Phytopathol,* **55**, pp.333-354.

KIM, J., R. YANG, C. CHANG, Y. PARK and M. L. TUCKER. 2018. The root-knot nematode Meloidogyne incognita produces a functional mimic of the Arabidopsis INFLORESCENCE DEFICIENT IN ABSCISSION signaling peptide. *J Exp Bot,* **69**(12), pp.3009-3021.

KLAPHECK, S. 1988. Homoglutathione - Isolation, Quantification and Occurrence in Legumes. *Physiologia Plantarum,* **74**(4), pp.727-732.

KLAPHECK, S., B. CHROST, J. STARKE and H. ZIMMERMANN. 1992. Gamma-Glutamylcysteinylserine - a New Homolog of Glutathione in Plants of the Family Poaceae. *Botanica Acta,* **105**(3), pp.174-179.

KOENNING, S. R., T. L. KIRKPATRICK, J. L. STARR, J. A. WRATHER, N. R. WALKER and J. D. MUELLER. 2004. Plant-parasitic nematodes attacking cotton in the United States - Old and emerging production challenges. *Plant Disease,* **88**(2), pp.100-113.

KRISHNAN, P., X. MA, B. A. MCDONALD and P. C. BRUNNER. 2018. Widespread signatures of selection for secreted peptidases in a fungal plant pathogen. *Bmc Evolutionary Biology,* **18**.

KRISSINEL, E. and K. HENRICK. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr,* **60**(Pt 12 Pt 1), pp.2256-68.

KROGH, A., M. BROWN, I. S. MIAN, K. SJOLANDER and D. HAUSSLER. 1994. Hidden Markov-Models in Computational Biology - Applications to Protein Modeling. *Journal of Molecular Biology,* **235**(5), pp.1501-1531.

KUMAR, M., N. P. GANTASALA, T. ROYCHOWDHURY, P. K. THAKUR, P. BANAKAR, R. N. SHUKLA, M. G. JONES and U. RAO. 2014. De novo transcriptome sequencing and analysis of the cereal cyst nematode, Heterodera avenae. *PLoS One,* **9**(5), pe96311.

KURTZMAN, C. P. and C. J. ROBNETT. 2003. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *Fems Yeast Research,* **3**(4), pp.417-432.

KYNDT, T., D. FERNANDEZ and G. GHEYSEN. 2014. Plant-parasitic nematode infections in rice: molecular and cellular insights. *Annu Rev Phytopathol,* **52**, pp.135-53.

KYNDT, T., P. VIEIRA, G. GHEYSEN and J. DE ALMEIDA-ENGLER. 2013. Nematode feeding sites: unique organs in plant roots. *Planta,* **238**(5), pp.807-18.

LAWRENCE, K., M. OLSEN, T. FASKE, R. HUTMACHER, J. MULLER, J. MARIO, R. KEMERAIT, C. OVERSTREET, G. SCIUMBATO and G. LAWRENCE. 2015. Cotton disease loss estimate committee report, 2014. *In: Proceedings of the 2015 Beltwide Cotton Conferences, San Antonio, TX. Cordova: National Cotton Council*, pp.188-190.

LEACH, M., P. AGUDELO and A. J. P. D. LAWTON-RAUH. 2012. Genetic variability of Rotylenchulus reniformis. **96**(1), pp.30-36.

LEE, C., D. CHRONIS, C. KENNING, B. PERET, T. HEWEZI, E. L. DAVIS, T. J. BAUM, R. HUSSEY, M. BENNETT and M. G. MITCHUM. 2011. The novel cyst

nematode effector protein 19C07 interacts with the Arabidopsis auxin influx transporter LAX3 to control feeding site development. *Plant Physiol,* **155**(2), pp.866-80.

LEE, D., O. REDFERN and C. J. N. R. M. C. B. ORENGO. 2007. Predicting protein function from sequence and structure. **8**(12), p995.

LEELARASAMEE, N., L. ZHANG and C. GLEASON. 2018. The root-knot nematode effector MiPFN3 disrupts plant actin filaments and promotes parasitism. *PLoS Pathog,* **14**(3), pe1006947.

LI, S. M., C. M. ARMSTRONG, N. BERTIN, H. GE, S. MILSTEIN, M. BOXEM, P. O. VIDALAIN, J. D. J. HAN, A. CHESNEAU, T. HAO, D. S. GOLDBERG, N. LI, M. MARTINEZ, J. F. RUAL, P. LAMESCH, L. XU, M. TEWARI, S. L. WONG, L. V. ZHANG, G. F. BERRIZ, L. JACOTOT, P. VAGLIO, J. REBOUL, T. HIROZANE-KISHIKAWA, Q. R. LI, H. W. GABEL, A. ELEWA, B. BAUMGARTNER, D. J. ROSE, H. Y. YU, S. BOSAK, R. SEQUERRA, A. FRASER, S. E. MANGO, W. M. SAXTON, S. STROME, S. VAN DEN HEUVEL, F. PIANO, J. VANDENHAUTE, C. SARDET, M. GERSTEIN, L. DOUCETTE-STAMM, K. C. GUNSALUS, J. W. HARPER, M. E. CUSICK, F. P. ROTH, D. E. HILL and M. VIDAL. 2004. A map of the interactome network of the metazoan C-elegans. *Science,* **303**(5657), pp.540-543.

LI, W., A. COWLEY, M. ULUDAG, T. GUR, H. MCWILLIAM, S. SQUIZZATO, Y. M. PARK, N. BUSO and R. LOPEZ. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res,* **43**(W1), pp.W580-4.

LILLEY, C. J., H. J. ATKINSON and P. E. URWIN. 2005. Molecular aspects of cyst nematodes. *Molecular Plant Pathology,* **6**(6), pp.577-588.

LILLEY, C. J., A. MAQBOOL, D. WU, H. B. YUSUP, L. M. JONES, P. R. J. BIRCH, M. J. BANFIELD, P. E. URWIN and S. EVES-VAN DEN AKKER. 2018. Effector gene birth in plant parasitic nematodes: Neofunctionalization of a housekeeping glutathione synthetase gene. *PLoS Genet,* **14**(4), pe1007310.

LIN, B., K. ZHUO, S. CHEN, L. HU, L. SUN, X. WANG, L. H. ZHANG and J. LIAO. 2016. A novel nematode effector suppresses plant immunity by activating host reactive oxygen species-scavenging system. *New Phytol,* **209**(3), pp.1159-73.

LINDERMAYR, C., G. SAALBACH and J. DURNER. 2005. Proteomic identification of S-nitrosylated proteins in Arabidopsis. *Plant Physiology,* **137**(3), pp.921-930.

LIU, J., H. PENG, J. K. CUI, W. K. HUANG, L. G. KONG, J. L. CLARKE, H. JIAN, G. L. WANG and D. L. PENG. 2016. Molecular Characterization of A Novel Effector Expansin-like Protein from Heterodera avenae that Induces Cell Death in Nicotiana benthamiana. *Scientific Reports,* **6**.

LOBSTEIN, J., C. A. EMRICH, C. JEANS, M. FAULKNER, P. RIGGS and M. BERKMEN. 2012. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories,* **11**.

LOHAR, D. P. and D. M. BIRD. 2003. Lotus japonicus: a new model to study root-parasitic nematodes. *Plant Cell Physiol,* **44**(11), pp.1176-84.

LONG, M., E. BETRAN, K. THORNTON and W. WANG. 2003. The origin of new genes: Glimpses from the young and old. *Nature Reviews Genetics,* **4**(11), pp.865-875.

LOZANO-TORRES, J. L., R. H. P. WILBERS, P. GAWRONSKI, J. C. BOSHOVEN, A. FINKERS-TOMCZAK, J. H. G. CORDEWENER, A. H. P. AMERICA, H. A. OVERMARS, J. W. VAN 'T KLOOSTER, L. BARANOWSKI, M. SOBCZAK, M. ILYAS, R. A. L. VAN DER HOORN, A. SCHOTS, P. J. G. M. DE WIT, J.

BAKKER, A. GOVERSE and G. SMANT. 2012. 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,* **109**(25), pp.10119-10124.

LOZANO-TORRES, J. L., R. H. P. WILBERS, S. WARMERDAM, A. FINKERS-TOMCZAK, A. DIAZ-GRANADOS, C. C. VAN SCHAIK, J. HELDER, J. BAKKER, A. GOVERSE, A. SCHOTS and G. SMANT. 2014. Apoplastic Venom Allergen-like Proteins of Cyst Nematodes Modulate the Activation of Basal Plant Innate Immunity by Cell Surface Receptors. *Plos Pathogens,* **10**(12).

LU, S. W., S. CHEN, J. WANG, H. YU, D. CHRONIS, M. G. MITCHUM and X. WANG. 2009. Structural and functional diversity of CLAVATA3/ESR (CLE)-like genes from the potato cyst nematode Globodera rostochiensis. *Mol Plant Microbe Interact,* **22**(9), pp.1128-42.

LUERSEN, K., D. STEGEHAKE, J. DANIEL, M. DRESCHER, I. AJONINA, C. AJONINA, P. HERTEL, C. WOLTERSDORF and E. LIEBAU. 2013. The Glutathione Reductase GSR-1 Determines Stress Tolerance and Longevity in Caenorhabditis elegans. *Plos One,* **8**(4).

LUNT, D. H. J. B. E. B. 2008. Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. **8**(1), p194.

MA, L. J., H. C. VAN DER DOES, K. A. BORKOVICH, J. J. COLEMAN, M. J. DABOUSSI, A. DI PIETRO, M. DUFRESNE, M. FREITAG, M. GRABHERR, B. HENRISSAT, P. M. HOUTERMAN, S. KANG, W. B. SHIM, C. WOLOSHUK, X. H. XIE, J. R. XU, J. ANTONIW, S. E. BAKER, B. H. BLUHM, A. BREAKSPEAR, D. W. BROWN, R. A. E. BUTCHKO, S. CHAPMAN, R. COULSON, P. M. COUTINHO, E. G. J. DANCHIN, A. DIENER, L. R. GALE, D. M. GARDINER, S. GOFF, K. E. HAMMOND-KOSACK, K. HILBURN, A. HUA-VAN, W. JONKERS, K. KAZAN, C. D. KODIRA, M. KOEHRSEN, L. KUMAR, Y. H. LEE, L. D. LI, J. M. MANNERS, D. MIRANDA-SAAVEDRA, M. MUKHERJEE, G. PARK, J. PARK, S. Y. PARK, R. H. PROCTOR, A. REGEV, M. C. RUIZ-ROLDAN, D. SAIN, S. SAKTHIKUMAR, S. SYKES, D. C. SCHWARTZ, B. G. TURGEON, I. WAPINSKI, O. YODER, S. YOUNG, Q. D. ZENG, S. G. ZHOU, J. GALAGAN, C. A. CUOMO, H. C. KISTLER and M. REP. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature,* **464**(7287), pp.367-373.

MA, Z. C., L. ZHU, T. Q. SONG, Y. WANG, Q. ZHANG, Y. Q. XIA, M. QIU, Y. C. LIN, H. Y. LI, L. KONG, Y. F. FANG, W. W. YE, Y. WANG, S. M. DONG, X. B. ZHENG, B. M. TYLER and Y. C. WANG. 2017. A paralogous decoy protects Phytophthora sojae apoplastic effector PsXEG1 from a host inhibitor. *Science,* **355**(6326), pp.710-714.

MAIER, T. R., T. HEWEZI, J. PENG and T. J. BAUM. 2013. Isolation of whole esophageal gland cells from plant-parasitic nematodes for transcriptome analyses and effector identification. *Mol Plant Microbe Interact,* **26**(1), pp.31-5.

MANTELIN, S., S. BELLAFIORE and T. KYNDT. 2017. Meloidogyne graminicola: a major threat to rice agriculture. *Mol Plant Pathol,* **18**(1), pp.3-15.

MASONBRINK, R., T. R. MAIER, U. MUPPIRALA, A. S. SEETHARAM, E. LORD, P. S. JUVALE, J. SCHMUTZ, N. T. JOHNSON, D. KORKIN, M. G. MITCHUM, B. MIMEE, S. E. DEN AKKER, M. HUDSON, A. J. SEVERIN and T. J. BAUM. 2019. The genome of the soybean cyst nematode (Heterodera glycines) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics,* **20**(1), p119.

MATAMOROS, M. A., J. F. MORAN, I. ITURBE-ORMAETXE, M. C. RUBIO and M. BECANA. 1999. Glutathione and homoglutathione synthesis in legume root nodules. *Plant Physiology,* **121**(3), pp.879-888.

MATIKA, D. E. F. and G. J. LOAKE. 2014. Redox Regulation in Plant Immune Function. *Antioxidants & Redox Signaling,* **21**(9), pp.1373-1388.

MATSUDA, K., K. MIZUGUCHI, T. NISHIOKA, H. KATO, N. GO and J. ODA. 1996. Crystal structure of glutathione synthetase at optimal pH: Domain architecture and structural similarity with other proteins. *Protein Engineering,* **9**(12), pp.1083-1092.

MAY, M. J. and C. J. LEAVER. 1994. Arabidopsis-Thaliana Gamma-Glutamylcysteine Synthetase Is Structurally Unrelated to Mammalian, Yeast, and Escherichia-Coli Homologs. *Proceedings of the National Academy of Sciences of the United States of America,* **91**(21), pp.10059-10063.

MCCARTER, J. P., M. D. MITREVA, J. MARTIN, M. DANTE, T. WYLIE, U. RAO, D. PAPE, Y. BOWERS, B. THEISING, C. V. MURPHY, A. P. KLOEK, B. J. CHIAPELLI, S. W. CLIFTON, D. M. BIRD and R. H. WATERSTON. 2003. Analysis and functional classification of transcripts from the nematode Meloidogyne incognita. *Genome Biol,* **4**(4), pR26.

MCCOY, A. J., R. W. GROSSE-KUNSTLEVE, P. D. ADAMS, M. D. WINN, L. C. STORONI and R. J. READ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography,* **40**, pp.658-674.

MCDONALD, M. C., D. AHREN, S. SIMPFENDORFER, A. MILGATE and P. S. SOLOMON. 2018. The discovery of the virulence gene ToxA in the wheat and barley pathogen Bipolaris sorokiniana. *Molecular Plant Pathology,* **19**(2), pp.432-439.

MEI, Y., P. THORPE, A. GUZHA, A. HAEGEMAN, V. C. BLOK, K. MACKENZIE, G. GHEYSEN, J. T. JONES and S. J. N. MANTELIN. 2015. Only a small subset of the SPRY domain gene family in Globodera pallida is likely to encode effectors, two of which suppress host defences induced by the potato resistance gene Gpa2. **17**(4), pp.409-424.

MEI, Y. Y., K. M. WRIGHT, A. HAEGEMAN, L. BAUTERS, A. DIAZ-GRANADOS, A. GOVERSE, G. GHEYSEN, J. T. JONES and S. MANTELIN. 2018. The Globodera pallida SPRYSEC Effector GpSPRY-414-2 That Suppresses Plant Defenses Targets a Regulatory Component of the Dynamic Microtubule Network. *Frontiers in Plant Science,* **9**.

MEIERJOHANN, S., R. D. WALTER and S. MULLER. 2002. Glutathione synthetase from Plasmodium falciparum. *Biochem J,* **363**(Pt 3), pp.833-8.

MEISTER, A. 1983. Transport and metabolism of glutathione and gamma-glutamyl amino acids. *Biochem Soc Trans,* **11**(6), pp.793-4.

MEISTER, A. 1995. Glutathione Biosynthesis and Its Inhibition. *Biothiols, Pt B,* **252**, pp.26-30.

MENARDO, F., C. R. PRAZ, S. WYDER, R. BEN-DAVID, S. BOURRAS, H. MATSUMAE, K. E. MCNALLY, F. PARLANGE, A. RIBA, S. ROFFLER, L. K. SCHAEFER, K. K. SHIMIZU, L. VALENTI, H. ZBINDEN, T. WICKER and B. KELLER. 2016. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nature Genetics,* **48**(2), pp.201-205.

MEUWLY, P., P. THIBAULT, A. L. SCHWAN and W. E. RAUSER. 1995. Three families of thiol peptides are induced by cadmium in maize. *Plant J,* **7**(5), p857.

MEYER, A. J. and M. D. J. P. P. FRICKER. 2002. Control of demand-driven biosynthesis of glutathione in green Arabidopsis suspension culture cells. **130**(4), pp.1927-1937.

MITCHUM, M. G., R. S. HUSSEY, T. J. BAUM, X. WANG, A. A. ELLING, M. WUBBEN and E. L. DAVIS. 2013. Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol,* **199**(4), pp.879-94.

MITCHUM, M. G., X. WANG, J. WANG and E. L. DAVIS. 2012. Role of nematode peptides and other small molecules in plant parasitism. *Annu Rev Phytopathol,* **50**, pp.175-95.

MOENS, M., R. N. PERRY and J. L. J. R.-K. N. STARR. 2009. Meloidogyne species–a diverse group of novel and important plant parasites. **1**, p483.

MOOZ, E. D. and A. J. B. MEISTER. 1967. Tripeptide (glutathione) synthetase. Purification, properties, and mechanism of action. **6**(6), pp.1722-1734.

MURSHUDOV, G. N., P. SKUBAK, A. A. LEBEDEV, N. S. PANNU, R. A. STEINER, R. A. NICHOLLS, M. D. WINN, F. LONG and A. A. VAGIN. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Biological Crystallography,* **67**, pp.355-367.

NEAL, D. C. 1954. The Reniform Nematode and Its Relationship to the Incidence of Fusarium Wilt of Cotton at Baton-Rouge, Louisiana. *Phytopathology,* **44**(8), pp.447-450.

NEWTON, G. L. and B. J. J. O. B. JAVOR. 1985. gamma-Glutamylcysteine and thiosulfate are the major low-molecular-weight thiols in halobacteria. **161**(1), pp.438-441.

NGUYEN, C. N., L. PERFUS-BARBEOCH, M. QUENTIN, J. L. ZHAO, M. MAGLIANO, N. MARTEU, M. DA ROCHA, N. NOTTET, P. ABAD and B. FAVERY. 2018. A root-knot nematode small glycine and cysteine-rich secreted effector, MiSGCR1, is involved in plant parasitism. *New Phytologist,* **217**(2), pp.687-699.

NICOL, J. M., S. J. TURNER, D. L. COYNE, L. DEN NIJS, S. HOCKLAND and Z. T. MAAFI. 2011. Current Nematode Threats to World Agriculture. *Genomics and Molecular Genetics of Plant-Nematode Interactions*, pp.21-43.

NJALSSON, R., K. CARLSSON, B. OLIN, B. CARLSSON, L. WHITBREAD, G. POLEKHINA, M. W. PARKER, S. NORGREN, B. MANNERVIK, P. G. BOARD and A. LARSSON. 2000. Kinetic properties of missense mutations in patients with glutathione synthetase deficiency. *Biochem J,* **349**(Pt 1), pp.275-9.

NJALSSON, R., S. NORGREN, A. LARSSON, C. S. HUANG, M. E. ANDERSON and J. L. LUO. 2001. Cooperative binding of gamma-glutamyl substrate to human glutathione synthetase. *Biochemical and Biophysical Research Communications,* **289**(1), pp.80-84.

NOCTOR, G., L. GOMEZ, H. VANACKER and C. H. FOYER. 2002. Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signalling. *Journal of Experimental Botany,* **53**(372), pp.1283-1304.

NOCTOR, G., A. MHAMDI, S. CHAOUCH, Y. HAN, J. NEUKERMANS, B. MARQUEZ-GARCIA, G. QUEVAL and C. H. FOYER. 2012. Glutathione in plants: an integrated overview. *Plant Cell Environ,* **35**(2), pp.454-84.

NOCTOR, G., M. STROHM, L. JOUANIN, K. J. KUNERT, C. H. FOYER and H. RENNENBERG. 1996. Synthesis of glutathione in leaves of transgenic poplar overexpressing gamma-glutamylcysteine synthetase. *Plant Physiology,* **112**(3), pp.1071-1078.

NOON, J. B., M. S. QI, D. N. SILL, U. MUPPIRALA, S. EVES-VAN DEN AKKER, T. R. MAIER, D. DOBBS, M. G. MITCHUM, T. HEWEZI and T. J. BAUM. 2016. A Plasmodium-like virulence effector of the soybean cyst nematode suppresses plant innate immunity. *New Phytologist,* **212**(2), pp.444-460.

NOTTENSTEINER, M., B. ZECHMANN, C. MCCOLLUM and R. HUCKELHOVEN. 2018. A barley powdery mildew fungus non-autonomous retrotransposon encodes a peptide that supports penetration success on barley. *Journal of Experimental Botany,* **69**(15), pp.3745-3758.

NYAKU, S. T., V. R. SRIPATHI, R. V. KANTETY, S. B. CSEKE, R. BUYYARAPU, R. MC EWAN, Y. Q. GU, K. LAWRENCE, Z. SENWO, P. SRIPATHI, P. GEORGE and G. C. SHARMA. 2014. Characterization of the reniform nematode genome by shotgun sequencing. *Genome,* **57**(4), pp.209-21.

NYAKU, S. T., V. R. SRIPATHI, G. WILEY, F. Z. NAJAR, L. J. CSEKE, G. C. SHARMA, B. A. ROE, S. B. CSEKE, E. MOSS and R. V. J. A. J. O. P. S. KANTETY. 2013. The expressed parasitism genes in the reniform nematode (Rotylenchulus reniformis). **4**(04), p780.

OHTAKE, Y., K. WATANABE, H. TEZUKA, T. OGATA, S. YABUUCHI, K. MURATA and A. KIMURA. 1989. Expression of the Glutathione Synthetase Gene of Escherichia-Coli-B in Saccharomyces-Cerevisiae. *Journal of Fermentation and Bioengineering,* **68**(6), pp.390-394.

ONKENDI, E. M., G. M. KARIUKI, M. MARAIS and L. N. MOLELEKI. 2014. The threat of root-knot nematodes (Meloidogyne spp.) in Africa: a review. *Plant Pathology,* **63**(4), pp.727-737.

OPPENHEIMER, L., V. P. WELLNER, O. W. GRIFFITH and A. J. J. O. B. C. MEISTER. 1979. Glutathione synthetase. Purification from rat kidney and mapping of the substrate binding sites. **254**(12), pp.5184-5190.

OPPERMAN, C. H., D. M. BIRD, V. M. WILLIAMSON, D. S. ROKHSAR, M. BURKE, J. COHN, J. CROMER, S. DIENER, J. GAJAN, S. GRAHAM, T. D. HOUFEK, Q. LIU, T. MITROS, J. SCHAFF, R. SCHAFFER, E. SCHOLL, B. R. SOSINSKI, V. P. THOMAS and E. WINDHAM. 2008. Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. *Proc Natl Acad Sci U S A,* **105**(39), pp.14802-7.

PAGANINI, J., A. CAMPAN-FOURNIER, M. DA ROCHA, P. GOURET, P. PONTAROTTI, E. WAJNBERG, P. ABAD and E. G. J. DANCHIN. 2012. Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes. *Plos One,* **7**(11).

PARISY, V., B. POINSSOT, L. OWSIANOWSKI, A. BUCHALA, J. GLAZEBROOK and F. MAUCH. 2007. Identification of PAD2 as a gamma-glutamylcysteine synthetase highlights the importance of glutathione in disease resistance of Arabidopsis. *Plant Journal,* **49**(1), pp.159-172.

PASTERNAK, M., B. LIM, M. WIRTZ, R. HELL, C. S. COBBETT and A. J. MEYER. 2008. Restricting glutathione biosynthesis to the cytosol is sufficient for normal plant development. *Plant Journal,* **53**(6), pp.999-1012.

PATRO, R., G. DUGGAL, M. I. LOVE, R. A. IRIZARRY and C. J. N. M. KINGSFORD. 2017. Salmon provides fast and bias-aware quantification of transcript expression. **14**(4), p417.

PENG, H., B. L. GAO, L. A. KONG, Q. YU, W. K. HUANG, X. F. HE, H. B. LONG and D. L. PENG. 2013. Exploring the host parasitism of the migratory plant-parasitic nematode Ditylenchus destuctor by expressed sequence tags analysis. *PLoS One,* **8**(7), pe69579.

PENNINGTON, H. G., R. JONES, S. KWON, G. BONCIANI, H. THIERON, T. CHANDLER, P. LUONG, S. N. MORGAN, M. PRZYDACZ and T. J. P. P. BOZKURT. 2019. The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. **15**(3), pe1007620.

PETERSEN, T. N., S. BRUNAK, G. VON HEIJNE and H. NIELSEN. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* **8**(10), pp.785-786.

PIVATO, M., M. FABREGA-PRATS and A. MASI. 2014. Low-molecular-weight thiols in plants: Functional and analytical implications. *Archives of Biochemistry and Biophysics,* **560**, pp.83-99.

PLISSONNEAU, C., J. BENEVENUTO, N. MOHD-ASSAAD, S. FOUCHE, F. E. HARTMANN and D. CROLL. 2017. Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. *Frontiers in Plant Science,* **8**.

POGORELKO, G., P. S. JUVALE, W. B. RUTTER, T. HEWEZI, R. HUSSEY, E. L. DAVIS, M. G. MITCHUM and T. J. BAUM. 2016. A cyst nematode effector binds to diverse plant proteins, increases nematode susceptibility and affects root morphology. *Mol Plant Pathol,* **17**(6), pp.832-44.

POGORELKO, G. V., P. S. JUVALE, W. B. RUTTER, M. HUTTEN, T. R. MAIER, T. HEWEZI, J. PAULUS, R. A. L. VAN DER HOORN, F. M. W. GRUNDLER, S. SIDDIQUE, V. LIONETTI, O. A. ZABOTINA and T. J. BAUM. 2019. Retargeting of a Plant Defense Protease by a Cyst Nematode Effector. *Plant J*.

POLEKHINA, G., P. G. BOARD, R. R. GALI, J. ROSSJOHN and M. W. PARKER. 1999. Molecular basis of glutathione synthetase deficiency and a rare gene permutation event. *EMBO J,* **18**(12), pp.3204-13.

POPPE, S., L. DORSHEIMER, P. HAPPEL and E. H. STUKENBROCK. 2015. Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen Zymoseptoria tritici (Mycosphaerella graminicola). *Plos Pathogens,* **11**(7).

QUEVAL, G., E. ISSAKIDIS-BOURGUET, F. A. HOEBERICHTS, M. VANDORPE, B. GAKIERE, H. VANACKER, M. MIGINIAC-MASLOW, F. VAN BREUSEGEM and G. NOCTOR. 2007. Conditional oxidative stress responses in the Arabidopsis photorespiratory mutant cat2 demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H2O2-induced cell death. *Plant Journal,* **52**(4), pp.640-657.

QUEVAL, G., D. THOMINET, H. VANACKER, M. MIGINIAC-MASLOW, B. GAKIERE and G. NOCTOR. 2009. H2O2-Activated Up-Regulation of Glutathione in Arabidopsis Involves Induction of Genes Encoding Enzymes Involved in Cysteine Synthesis in the Chloroplast. *Molecular Plant,* **2**(2), pp.344-356.

RADEWALD, J. and G. TAKESHITA. 1964. DESICCATION STUDIES ON 5 SPECIES OF PLANT-PARASITIC NEMATODES IN HAWAII. *In: Phytopathology*: AMER PHYTOPATHOLOGICAL SOC 3340 PILOT KNOB ROAD, ST PAUL, MN 55121, pp.903-&.

RADZICKA, A. and R. J. B. WOLFENDEN. 1988. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. **27**(5), pp.1664-1670.

REBOIS, R. V. 1980. Ultrastructure of a Feeding Peg and Tube Associated with Rotylenchulus-Reniformis in Cotton. *Nematologica,* **26**(4), pp.396-405.

REHMAN, S., V. K. GUPTA and A. K. GOYAL. 2016. Identification and functional analysis of secreted effectors from phytoparasitic nematodes. *BMC Microbiol,* **16**, p48.

REMIGI, P., M. ANISIMOVA, A. GUIDOT, S. GENIN and N. J. N. P. PEETERS. 2011. Functional diversification of the GALA type III effector family contributes to Ralstonia solanacearum adaptation on different plant hosts. **192**(4), pp.976-987.

REPLOGLE, A., J. WANG, A. BLECKMANN, R. S. HUSSEY, T. J. BAUM, S. SAWA, E. L. DAVIS, X. WANG, R. SIMON and M. G. MITCHUM. 2011. Nematode CLE signaling in Arabidopsis requires CLAVATA2 and CORYNE. *Plant J,* **65**(3), pp.430-40.

REPLOGLE, A., J. WANG, V. PAOLILLO, J. SMEDA, A. KINOSHITA, A. DURBAK, F. E. TAX, X. WANG, S. SAWA and M. G. MITCHUM. 2013. Synergistic interaction of CLAVATA1, CLAVATA2, and RECEPTOR-LIKE PROTEIN KINASE 2 in cyst nematode parasitism of Arabidopsis. *Mol Plant Microbe Interact,* **26**(1), pp.87-96.

RHODES, G. 2010. *Crystallography made crystal clear: a guide for users of macromolecular models.* Elsevier.

RINALDI, M. A., J. LIU, T. A. ENDERS, B. BARTEL and L. C. STRADER. 2012. A gain-of-function mutation in IAA16 confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. *Plant Molecular Biology,* **79**(4-5), pp.359-373.

ROBINSON, A. F. 2007. Reniform in U.S. cotton: when, where, why, and some remedies. *Annu Rev Phytopathol,* **45**, pp.263-88.

ROBINSON, A. F., R. N. INSERRA, E. P. CASWELL-CHEN, N. VOVLAS and A. TROCCOLI. 1997. Rotylenchulus species: Identification, distribution, host ranges, and crop plant resistance. *Nematropica,* **27**(2), pp.127-180.

RODERICK, H., P. E. URWIN and H. J. ATKINSON. 2018. Rational design of biosafe crop resistance to a range of nematodes using RNA interference. *Plant Biotechnology Journal,* **16**(2), pp.520-529.

ROETSCHI, A., A. SI-AMMOUR, L. BELBAHRI, F. MAUCH and B. MAUCH-MANI. 2001. Characterization of an Arabidopsis-Phytophthora pathosystem: resistance requires a functional PAD2 gene and is independent of salicylic acid, ethylene and jasmonic acid signalling. *Plant Journal,* **28**(3), pp.293-305.

ROUHIER, N., S. D. LEMAIRE and J. P. JACQUOT. 2008. The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxins and glutathionylation. *Annu Rev Plant Biol,* **59**, pp.143-66.

ROZE, E., B. HANSE, M. MITREVA, B. VANHOLME, J. BAKKER and G. SMANT. 2008. Mining the secretome of the root-knot nematode Meloidogyne chitwoodi for candidate parasitism genes. *Molecular Plant Pathology,* **9**(1), pp.1-10.

RUÍZ-SÁNCHEZ, M., E. ARMADA, Y. MUÑOZ, I. E. G. DE SALAMONE, R. AROCA, J. M. RUÍZ-LOZANO and R. J. J. O. P. P. AZCÓN. 2011. Azospirillum and arbuscular mycorrhizal colonization enhance rice growth and physiological traits under well-watered and drought conditions. **168**(10), pp.1031-1037.

SAWA, S., A. KINOSHITA, I. NAKANOMYO and H. FUKUDA. 2006. CLV3/ESR-related (CLE) peptides as intercellular signaling molecules in plants. *Chemical Record,* **6**(6), pp.303-310.

SCHIRAWSKI, J., G. MANNHAUPT, K. MUNCH, T. BREFORT, K. SCHIPPER, G. DOEHLEMANN, M. DI STASIO, N. ROSSEL, A. MENDOZA-MENDOZA, D. PESTER, O. MULLER, B. WINTERBERG, E. MEYER, H. GHAREEB, T. WOLLENBERG, M. MUNSTERKOTTER, P. WONG, M. WALTER, E.

STUKENBROCK, U. GULDENER and R. KAHMANN. 2010. Pathogenicity Determinants in Smut Fungi Revealed by Genome Comparison. *Science,* **330**(6010), pp.1546-1548.

SCHLAEPPI, K., N. BODENHAUSEN, A. BUCHALA, F. MAUCH and P. REYMOND. 2008. The glutathione-deficient mutant pad2-1 accumulates lower amounts of glucosinolates and is more susceptible to the insect herbivore Spodoptera littoralis. *Plant Journal,* **55**(5), pp.774-786.

SCHMUTZ, J., S. B. CANNON, J. SCHLUETER, J. X. MA, T. MITROS, W. NELSON, D. L. HYTEN, Q. J. SONG, J. J. THELEN, J. L. CHENG, D. XU, U. HELLSTEN, G. D. MAY, Y. YU, T. SAKURAI, T. UMEZAWA, M. K. BHATTACHARYYA, D. SANDHU, B. VALLIYODAN, E. LINDQUIST, M. PETO, D. GRANT, S. Q. SHU, D. GOODSTEIN, K. BARRY, M. FUTRELL-GRIGGS, B. ABERNATHY, J. C. DU, Z. X. TIAN, L. C. ZHU, N. GILL, T. JOSHI, M. LIBAULT, A. SETHURAMAN, X. C. ZHANG, K. SHINOZAKI, H. T. NGUYEN, R. A. WING, P. CREGAN, J. SPECHT, J. GRIMWOOD, D. ROKHSAR, G. STACEY, R. C. SHOEMAKER and S. A. JACKSON. 2010. Genome sequence of the palaeopolyploid soybean. *Nature,* **463**(7278), pp.178-183.

SEINHORST, J. 1961. Plant-Nematode Inter-Relationships. *Annual Review of Microbiology,* **15**, pp.177-&.

SHETTY, N. P., H. J. L. JØRGENSEN, J. D. JENSEN, D. B. COLLINGE and H. S. J. E. J. O. P. P. SHETTY. 2008. Roles of reactive oxygen species in interactions between plants and pathogens. **121**(3), pp.267-280.

SHI, Y. G. 2014. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell,* **159**(5), pp.995-1014.

SHOEMAKER, S. C. and N. J. B. ANDO. 2018. X-rays in the cryo-electron microscopy era: Structural biology's dynamic future. **57**(3), pp.277-285.

SHOWMAKER, K. C., W. S. SANDERS, S. DEN AKKER EVES-VAN, B. E. MARTIN, J. STOKES, C. HSU, B. BARTLETT, D. PETERSON and M. J. J. O. N. WUBBEN. 2019. A genomic resource for the sedentary semi-endoparasitic reniform nematode, Rotylenchulus reniformis Linford & Oliveira. **51**, pp.1-2.

SHOWMAKER, K. C., W. S. SANDERS, S. EVES-VAN DEN AKKER, B. E. MARTIN, F. E. CALLAHAN, D. G. PETERSON and M. J. WUBBEN. 2018. Five life stage-specific transcriptome assemblies for the reniform nematode, Rotylenchulus reniformis Linford & Oliveira. *Journal of Nematology,* **50**(4), pp.529-530.

SIDDIQUE, S. and F. M. J. C. O. I. M. GRUNDLER. 2018. Parasitic nematodes manipulate plant development to establish feeding sites. **46**, pp.102-108.

SIDDIQUE, S., C. MATERA, Z. S. RADAKOVIC, M. S. HASAN, P. GUTBROD, E. ROZANSKA, M. SOBCZAK, M. A. TORRES and F. M. W. GRUNDLER. 2014. Parasitic Worms Stimulate Host NADPH Oxidases to Produce Reactive Oxygen Species That Limit Plant Cell Death and Promote Infection. *Science Signaling,* **7**(320).

SIDDIQUE, S., Z. S. RADAKOVIC, C. M. DE LA TORRE, D. CHRONIS, O. NOVAK, E. RAMIREDDY, J. HOLBEIN, C. MATERA, M. HUTTEN, P. GUTBROD, M. S. ANJAM, E. ROZANSKA, S. HABASH, A. ELASHRY, M. SOBCZAK, T. KAKIMOTO, M. STRNAD, T. SCHMULLING, M. G. MITCHUM and F. M. W. GRUNDLER. 2015. A parasitic nematode releases cytokinin that controls cell division and orchestrates feeding site formation in host plants. *Proceedings of the National Academy of Sciences of the United States of America,* **112**(41), pp.12669-12674.

SIJMONS, P. C., F. M. GRUNDLER, N. VON MENDE, P. R. BURROWS and U. J. T. P. J. WYSS. 1991. Arabidopsis thaliana as a new model host for plant-parasitic nematodes. **1**(2), pp.245-254.

SIVAKUMAR, C. and A. J. I. J. O. N. SESHADRI. 1971. Life history of the reniform nematode Rotylenchulus reniformis Linford and Oliveira, 1940. **1**, pp.7-20.

SKIPSEY, M., B. G. DAVIS and R. EDWARDS. 2005. Diversification in substrate usage by glutathione synthetases from soya bean (Glycine max), wheat (Triticum aestivum) and maize (Zea mays). *Biochemical Journal,* **391**, pp.567-574.

SLAVENS, K. D., T. R. BROWN, K. A. BARAKAT, T. R. CUNDARI and M. E. ANDERSON. 2011. Valine 44 and valine 45 of human glutathione synthetase are key for subunit stability and negative cooperativity. *Biochemical and Biophysical Research Communications,* **410**(3), pp.597-601.

SMANT, G. and J. JONES. 2011. Suppression of Plant Defences by Nematodes. *Genomics and Molecular Genetics of Plant-Nematode Interactions*, pp.273-286.

SOBCZAK, M. and W. GOLINOWSKI. 2011. Cyst nematodes and syncytia. *Genomics and molecular genetics of plant-nematode interactions.* Springer, pp.61-82.

SOMMER, R. J. and A. STREIT. 2011. Comparative Genetics and Genomics of Nematodes: Genome Structure, Development, and Lifestyle. *Annual Review of Genetics, Vol 45,* **45**, pp.1-20.

SOMSSICH M, JE B I, SIMON R, Jackson D. 2016. CLAVATA-WUSCHEL signaling in the shoot meristem. Development, **143**(18): 3238-3248.

ST MAURICE, M., L. REINHARDT, K. H. SURINYA, P. V. ATTWOOD, J. C. WALLACE, W. W. CLELAND and I. RAYMENT. 2007. Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science,* **317**(5841), pp.1076-1079.

STARR, J. L., S. R. KOENNING, T. L. KIRKPATRICK, A. F. ROBINSON, P. A. ROBERTS and R. L. NICHOLS. 2007. The future of nematode management in cotton. *J Nematol,* **39**(4), pp.283-94.

STONE, A. R. 1986. Taxonomy and phylogeny of cyst nematodes. *Cyst nematodes.* Springer, pp.1-21.

STROHM, M., L. JOUANIN, K. J. KUNERT, C. PRUVOST, A. POLLE, C. H. FOYER and H. J. T. P. J. RENNENBERG. 1995. Regulation of glutathione synthesis in leaves of transgenic poplar (Populus tremula x P. alba) overexpressing glutathione synthetase. **7**(1), pp.141-145.

STUKENBROCK, E. H. 2016. Hybridization speeds up the emergence and evolution of a new pathogen species. *Nature Genetics,* **48**(2), pp.113-115.

SZITENBERG, A., L. SALAZAR-JARAMILLO, V. C. BLOK, D. R. LAETSCH, S. JOSEPH, V. M. WILLIAMSON, M. L. BLAXTER and D. H. LUNT. 2017. Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biol Evol,* **9**(10), pp.2844-2861.

TANAKA, S., G. SCHWEIZER, N. RÖSSEL, F. FUKADA, M. THINES and R. J. N. M. KAHMANN. 2019. Neofunctionalization of the secreted Tin2 effector in the fungal pathogen Ustilago maydis. **4**(2), p251.

TANAKA, T., H. KATO, T. NISHIOKA and J. J. B. ODA. 1992. Mutational and proteolytic studies on a flexible loop in glutathione synthetase from Escherichia coli B: the loop and arginine 233 are critical for the catalytic reaction. **31**(8), pp.2259-2265.

TANG, H. 2007. Genome assembly, rearrangement, and repeats. *Chemical Reviews,* **107**(8), pp.3391-3406.

TAYLOR, A. and J. J. N. C. S. U. G. SASSER. 1978. Biology, identification and control of root-knot nematodes. **111**.

TCHATCHOUA, J. and R. A. SIKORA. 1983. Alterations in Susceptibility of Wilt Resistant Cotton Varieties to Verticillium-Dahliae Induced by Rotylenchulus-Reniformis. *Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection,* **90**(3), pp.232-237.

THODEN, J. B., C. Z. BLANCHARD, H. M. HOLDEN and G. L. WALDROP. 2000. Movement of the biotin carboxylase B-domain as a result of ATP binding. *Journal of Biological Chemistry,* **275**(21), pp.16183-16190.

THORPE, P., S. MANTELIN, P. J. COCK, V. C. BLOK, M. C. COKE, S. EVES-VAN DEN AKKER, E. GUZEEVA, C. J. LILLEY, G. SMANT, A. J. REID, K. M. WRIGHT, P. E. URWIN and J. T. JONES. 2014. Genomic characterisation of the effector complement of the potato cyst nematode Globodera pallida. *BMC Genomics,* **15**, p923.

TIMMONS, L. and A. FIRE. 1998. Specific interference by ingested dsRNA. *Nature,* **395**(6705), pp.854-854.

TORUNO, T. Y., I. STERGIOPOULOS and G. COAKER. 2016. Plant-Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. *Annu Rev Phytopathol,* **54**, pp.419-41.

TRIANTAPHYLLOU, A. J. A. R. O. P. 1973. Environmental sex differentiation of nematodes in relation to pest management. **11**(1), pp.441-462.

TURNER, S. and J. J. P. N. ROWE. 2006. Cyst nematodes. pp.91-122.

URWIN, P. E., C. J. LILLEY and H. J. ATKINSON. 2002. Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference. *Molecular Plant-Microbe Interactions,* **15**(8), pp.747-752.

VADHERA, I., B. N. SHUKLA and J. BHATT. 1995. Interaction between Reniform Nematode (Rotylenchulus-Reniformis) and Fusarium-Solani Causing Root-Rot of French Bean (Phaseolus-Vulgaris). *Indian Journal of Agricultural Sciences,* **65**(10), pp.774-777.

VAN DER HOEVEN, R., K. C. MCCALLUM, M. R. CRUZ and D. A. GARSIN. 2011. Ce-Duox1/BLI-3 Generated Reactive Oxygen Species Trigger Protective SKN-1 Activity via p38 MAPK Signaling during Infection in C. elegans. *Plos Pathogens,* **7**(12).

VAN DOORN, W. G., E. P. BEERS, J. L. DANGL, V. E. FRANKLIN-TONG, P. GALLOIS, I. HARA-NISHIMURA, A. M. JONES, M. KAWAI-YAMADA, E. LAM, J. MUNDY, L. A. J. MUR, M. PETERSEN, A. SMERTENKO, M. TALIANSKY, F. VAN BREUSEGEM, T. WOLPERT, E. WOLTERING, B. ZHIVOTOVSKY and P. V. BOZHKOV. 2011. Morphological classification of plant cell deaths. *Cell Death and Differentiation,* **18**(8), pp.1241-1246.

VAN OSS, S. B. and A. R. CARVUNIS. 2019. De novo gene birth. *Plos Genetics,* **15**(5).

VERMA, A., C. LEE, S. MORRISS, F. ODU, C. KENNING, N. RIZZO, W. G. SPOLLEN, M. LIN, A. G. MCRAE, S. A. GIVAN, T. HEWEZI, R. HUSSEY, E. L. DAVIS, T. J. BAUM and M. G. MITCHUM. 2018. The novel cyst nematode effector protein 30D08 targets host nuclear functions to alter gene expression in feeding sites. *New Phytol,* **219**(2), pp.697-713.

VIEIRA, P. and C. J. C. O. I. P. B. GLEASON. 2019. Plant-parasitic nematode effectors—insights into their diversity and new tools for their identification. **50**, pp.37-43.

VIJAYAPALANI, P., T. HEWEZI, F. PONTVIANNE and T. J. BAUM. 2018. An Effector from the Cyst Nematode Heterodera schachtii Derepresses Host rRNA Genes by Altering Histone Acetylation. *Plant Cell,* **30**(11), pp.2795-2812.

VOORDECKERS, K., C. A. BROWN, K. VANNESTE, E. VAN DER ZANDE, A. VOET, S. MAERE and K. J. VERSTREPEN. 2012. Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *Plos Biology,* **10**(12).

WACHTER, A., S. WOLF, H. STEININGER, J. BOGS and T. J. T. P. J. RAUSCH. 2005. Differential targeting of GSH1 and GSH2 is achieved by multiple transcription initiation: implications for the compartmentation of glutathione biosynthesis in the Brassicaceae. **41**(1), pp.15-30.

WAGNER, E. M. 2013. Monitoring gene expression: quantitative real-time RT-PCR. *Lipoproteins and Cardiovascular Disease.* Springer, pp.19-45.

WANG, C. L. and D. J. OLIVER. 1996. Cloning of the cDNA and genomic clones for glutathione synthetase from Arabidopsis thaliana and complementation of a gsh2 mutant in fission yeast. *Plant Mol Biol,* **31**(6), pp.1093-104.

WANG, J., C. LEE, A. REPLOGLE, S. JOSHI, D. KORKIN, R. HUSSEY, T. J. BAUM, E. L. DAVIS, X. WANG and M. G. MITCHUM. 2010. Dual roles for the variable domain in protein trafficking and host-specific recognition of Heterodera glycines CLE effector proteins. *New Phytol,* **187**(4), pp.1003-17.

WANG, J., A. REPLOGLE, R. HUSSEY, T. BAUM, X. WANG, E. L. DAVIS and M. G. MITCHUM. 2011. Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode Heterodera schachtii. *Mol Plant Pathol,* **12**(2), pp.177-86.

WANG, X., M. G. MITCHUM, B. GAO, C. LI, H. DIAB, T. J. BAUM, R. S. HUSSEY and E. L. DAVIS. 2005. A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of Arabidopsis thaliana. *Mol Plant Pathol,* **6**(2), pp.187-91.

WANG, X., B. XUE, J. DAI, X. QIN, L. LIU, Y. CHI, J. T. JONES and H. LI. 2018. A novel Meloidogyne incognita chorismate mutase effector suppresses plant immunity by manipulating the salicylic acid pathway and functions mainly during the early stages of nematode parasitism. *Plant Pathology,* **67**(6), pp.1436-1448.

WATANABE, K., Y. YAMANO, K. MURATA and A. KIMURA. 1986. Glutathione Production by Escherichia-Coli-Cells with Hybrid Plasmid Containing Tandemly Polymerized Genes for Glutathione Synthetase. *Applied Microbiology and Biotechnology,* **24**(5), pp.375-378.

WATERHOUSE, A. M., J. B. PROCTER, D. M. A. MARTIN, M. CLAMP and G. J. BARTON. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* **25**(9), pp.1189-1191.

WEI, H., A. J. RUTHENBURG, S. K. BECHIS and G. L. VERDINE. 2005. Nucleotide-dependent domain movement in the ATPase domain of a human type IIA DNA topoisomerase. *Journal of Biological Chemistry,* **280**(44), pp.37041-37047.

WESTPHAL, A., A. F. ROBINSON, A. W. SCOTT and J. B. SANTINI. 2004. Depth distribution of Rotylenchulus reniformis under crops of different host status and after fumigation. *Nematology,* **6**, pp.97-107.

WHISSON, S. C., P. C. BOEVINK, L. MOLELEKI, A. O. AVROVA, J. G. MORALES, E. M. GILROY, M. R. ARMSTRONG, S. GROUFFAUD, P. VAN WEST, S. CHAPMAN, I. HEIN, I. K. TOTH, L. PRITCHARD and P. R. J. BIRCH. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature,* **450**(7166), pp.115-+.

WILBERS, R. H. P., R. SCHNEITER, M. H. M. HOLTERMAN, C. DRUREY, G. SMANT, O. A. ASOJO, R. M. MAIZELS and J. L. LOZANO-TORRES. 2018. Secreted venom allergen-like proteins of helminths: Conserved modulators of host responses in animals and plants. *Plos Pathogens,* **14**(10).

WINGATE, V. P., M. A. LAWTON and C. J. J. P. P. LAMB. 1988. Glutathione causes a massive and selective induction of plant defense genes. **87**(1), pp.206-210.

WINKLER, A., R. NJÅLSSON, K. CARLSSON, A. ELGADI, B. ROZELL, L. ABRAHAM, N. ERCAL, Z.-Z. SHI, M. W. LIEBERMAN, A. J. B. LARSSON and B. R. COMMUNICATIONS. 2011. Glutathione is essential for early embryogenesis–analysis of a glutathione synthetase knockout mouse. **412**(1), pp.121-126.

WINN, M. D., C. C. BALLARD, K. D. COWTAN, E. J. DODSON, P. EMSLEY, P. R. EVANS, R. M. KEEGAN, E. B. KRISSINEL, A. G. W. LESLIE, A. MCCOY, S. J. MCNICHOLAS, G. N. MURSHUDOV, N. S. PANNU, E. A. POTTERTON, H. R. POWELL, R. J. READ, A. VAGIN and K. S. WILSON. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography,* **67**, pp.235-242.

WINTER, G., D. G. WATERMAN, J. M. PARKHURST, A. S. BREWSTER, R. J. GILDEA, M. GERSTEL, L. FUENTES-MONTERO, M. VOLLMAR, T. MICHELS-CLARK, I. D. YOUNG, N. K. SAUTER and G. EVANS. 2018. DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D-Structural Biology,* **74**, pp.85-97.

WOO, J. S., J. H. IMM, C. K. MIN, K. J. KIM, S. S. CHA and B. H. J. T. E. J. OH. 2006. Structural and functional insights into the B30. 2/SPRY domain. **25**(6), pp.1353-1363.

WUBBEN, M. J., F. E. CALLAHAN and B. S. SCHEFFLER. 2010. Transcript analysis of parasitic females of the sedentary semi-endoparasitic nematode Rotylenchulus reniformis. *Mol Biochem Parasitol,* **172**(1), pp.31-40.

WUBBEN, M. J., S. GANJI and F. E. J. J. O. N. CALLAHAN. 2010. Identification and molecular characterization of a β-1, 4-endoglucanase gene (Rr-eng-1) from Rotylenchulus reniformis. **42**(4), p342.

WUBBEN, M. J., L. GAVILANO, T. J. BAUM and E. L. DAVIS. 2015. Sequence and Spatiotemporal Expression Analysis of CLE-Motif Containing Genes from the Reniform Nematode (*Rotylenchulus reniformis* Linford & Oliveira). *J Nematol,* **47**(2), pp.159-65.

WYSS, U. 1997. Root parasitic nematodes: an overview. *Cellular and molecular aspects of plant-nematode interactions.* Springer, pp.5-22.

XIANG, C., B. L. WERNER, E. M. CHRISTENSEN and D. J. OLIVER. 2001. The biological functions of glutathione revisited in arabidopsis transgenic plants with altered glutathione levels. *Plant Physiol,* **126**(2), pp.564-74.

YAMAGUCHI, H., H. KATO, Y. HATA, T. NISHIOKA, A. KIMURA, J. ODA and Y. KATSUBE. 1993. 3-Dimensional Structure of the Glutathione Synthetase from Escherichia-Coli-B at 2.O-Angstrom Resolution. *Journal of Molecular Biology,* **229**(4), pp.1083-1100.

YAMAZAKI, S., K. OCHIAI and T. J. P. D. MATOH. 2019. Rice plants have three homologs of glutathione synthetase genes, one of which, OsGS2, codes for hydroxymethyl-glutathione synthetase. **3**(2), pe00119.

YAN, B. X. and Y. Q. SUN. 1997. Glycine residues provide flexibility for enzyme active sites. *Journal of Biological Chemistry,* **272**(6), pp.3190-3194.

YANG, S., Y. DAI, Y. CHEN, J. YANG, D. YANG, Q. LIU and H. JIAN. 2019. A Novel G16B09-Like Effector From Heterodera avenae Suppresses Plant Defenses and Promotes Parasitism. *Front Plant Sci,* **10**, p66.

YUAN, L. and G. T. COLE. 1989. Characterization of a Proteinase-Inhibitor Isolated from the Fungal Pathogen Coccidioides-Immitis. *Biochemical Journal,* **257**(3), pp.729-736.

ZHENG, J. S., D. H. PENG, L. CHEN, H. L. LIU, F. CHEN, M. C. XU, S. Y. JU, L. F. RUAN and M. SUN. 2016. The Ditylenchus destructor genome provides new insights into the evolution of plant parasitic nematodes. *Proceedings of the Royal Society B-Biological Sciences,* **283**(1835).

ZHU, Y. L., E. A. H. PILON-SMITS, L. JOUANIN and N. TERRY. 1999. Overexpression of glutathione synthetase in Indian mustard enhances cadmium accumulation and tolerance. *Plant Physiology,* **119**(1), pp.73-79.

Supplementary figure 1

>Rre_GS1
MSPSPNEHLAPNYVPEVVAKQRGEHQNGSDGDGAEGIELLVEDAMDWAHCHGLVLRTREH
RNRSDVCQVAPFALFPSPFPRRLFNEAMDVQKAMNLLYFRISWDYDFLVQAHKDVIPSDA
FTRNMMDILVDLYKQDGGVRQKITLLTQRADYMCHVKTEGAAQPQFELKQIEVNNIAVSM
GGLAERATAYHRRLLRKADIDPSGDVVPENRPISTLAKGIQIAWQKFGDPDAIVLVVIGE
VNQNQFDQRYMEYEMDRLFEGQVKIVRLTLAQCADRLQLNPSDSTLRLNNQAVAIVYFRA
GYAPEDYPTQKEWEARRTIEKSTAIKCPWIGLQLANTKKVQQVLDLPNSVERFFPDPSDA
ATVKAIRHTFAGMWGLERDDEATKAVVQDALLHPERYVLKPQLEGGGGNFFGAELVDRLR
SLSPAERAAHILMQKIQPLVVKNYLVRAFEPVQLADVVSELGIYGCLVGDGSELSVQHNH
AHGHILRTKAEHVNEGGVAVGAAVVDTPYLF
>Rre_GS2
MVVTLPPNNGSATIVTNLNAQELKEQLQQQQNGGEGTNGTHQNGTNGTHAATNGKTHQNG
TNGTQTNGTNGSVHDVNAYVMDAVHDEQQLQALHEYALDYAHSIGLCARMVDYKFQSDIA
ATPPLALLPSPFPRELFYQAMDVQNILNELYFRISRDHAFLLDAYKDVIKGDPYIKRCVE
IADQIQREGIHQSISFCVQRADYMSHWNTQQQCMQLKQVEVNIGQIGGPGCATQARKYHC
KMLDKVDILRGESVEIGTEFPENQPRHKMAQSLYQGWKLFGDPNAVLLFVNQPDLFPLCH
FEQLQFTTFQVEKLAKKDGHRVQVIRMTLKQCAERCHLDESDWSLYADGKRVALCHMAYG
YIPEHYPTEAEWQIRIAMERSTCIMSPNIRLQLTGTKKIQQVLSKPGVLEHFFPEEPERI
AQLKQTFTGLWGLEDNDETTTAVIEDAIRRPRDYVLKAQMGAGKGNYFDDEMVHKLRTMS
LEERGAYILQQKIWPVIVQNYMMRPFKAPYPEHVVSELGIYGAIIGDSSTGKVLHNSAEG
YLCRSKPANVNQGGVCEGAGVVDSLLLFPASQFHQE
>Rre_GS3
MVVTLPPNSGSATIVTNLNAQELKEQLQQQQQNGGEGTNGTHQNGTNGTHAATNGKTHQN
GTNGTQTNGINGSVHDVNAYVMDAVQEEQQLQALHEYALDYAHSIGLCARMVDHKFQSDI
AATPPLALLPSPFPRELFYQAMDVQNILNELYFRISRDHAFLLDAYKDVIKGDPYIKRCV
EIADQIQREGIHQPISFCVQRADYMSHWNTQQQCMQLKQVEVNIGQIGGPGCATQARKYH
RKMLDKVDILRGASVEIGTEFPENQPRHKMAQSLYQGWKLFGDPNAVLLFVNQPDLFPLC
HFEQLQFTTFQVEKLAKKDGHRVQVIRMTLKQCAERCHLDESDWSLYADGKRVALCHMAY
GYIPEHYPTEAEWQIRIAMERSTCIMSPNIRLQLTGTKKIQQVLSKPGVLERFFPEEPER
IAQLKQTFTGLWGLEDNDDTTTAVIEDAIRRPRDYVLKAQMGAGKGNYFDDEMVHKLRTM
SLEERGAYILQQKIWPPENGTDAKNYMMRPFKAPYSEHVVSELGIYGAIIGDSSTGKVLH
NSAEGYLCRSKPANVNQGGVCEGAGVVDSLLLFPASQFHQE
>Rre_GS4
MSITVLNSKLLDNTKGEEKGHDKQKEETVTVTTIMRHHVADGTFSVDVSKSVEQLSRKCD
EQNVKDYVLELIRDQRELGELADYATSYAHSIGLVSRTLDRSFSSEPAVLVPVALLPSPF
PKELYDQAVEVQKALSELYFRVSCDHQFLMDSLKDVIKTDPFVARMVDMSQRVHAEGVQQ
PLTVALQRADYLAHWDPDQKTMELKQVEMNIGPIGGPGCASQVSKLHAKLLDKLEAIHGE
TSPVMDHAQLPDQNVRKNMARTLYKAWKLFGDPNAILLYISNSVADPLCHFEQLQFVQFE
VEKLAKREGQLVEVIQMTLSEAARRLTLDESRDFSLFVDGKKRVALAHITEGNMPEEYPT
EREWEARLIMERSNAILSPTIRLELSSSKKIQQVLANPGVLERFLPDDPQSCAALRLTFA
GLWGLEHDDAQTRDVIKEAIRNPQNFVLKSQIEAGKGNFFDEALAQKLGEMGLEERGAFI
LQQKIKPVAVKNFMLRPFKPLELDDVVGELGIFGSLIGDQCSRTVMWNTVDGHELKTRSA
SVNQAGVCAGYGVVDSAVLFPAREFHQ
>Rre_GS5
MSSAFKMNGHHNTPASNGVHIEHCQKNNSNESAKKMVKLLNGIGLKNRSDAKQYVPDMVR
DPAELQMLTKYAIDYAHCIGLEAPVPSDPPEANFSNVLAMCPPITLFPSPFPRELYEQAV
DVQQSMNELYFRIASDHDFLMDAFKDVVKGDPFMARFVHIAKQIHEEGVRQPLAIALQRA
DYLAHWDPQNGQVALKQVEVNIGPVGGPGFGSGVSKLHRKMLAKLTIEKGGTPVVLAKAD
APENRASQNLSCALYQAWKLFGDPSAILLFLDSPNITHFEQLQFIQFGVEKLGMQDGKLV
SVMSMTMVEAAKRLSLDEAGDFSLYLDGDKRIALVHIADGNMPDEYPSEREWQARIMMER
STAILSPNIRFQLSSTKKIQQVLAKPGMLDRFFPKEPQRVAQIRNTFAGLWGLDEDDEPT
RTVIQNAIRCPQDFVLKSQLEAGLGNFFDAQMAEMLRTMNGEQRSAYILQQRIKPLVVKN
FLLRRPSEPAELESVVGELGIYGTLIGDQRSGRILHNSVDGHTIRSKPSDLNQGGIGSGG
GSVDSALLFSATDLMNNDDREEGQEMVMINGK
>Rre_GS6
MPINKPVLAKPGVLDRFFPEEPQRVAQIRNTFAGLWGLDEDDEPTRVVIQEAIRCPQDFV
LKSQLEAGLGNFFDAQMAEMLRTMNGEQRSAYILQQRIKPLVVKPSDLNQGGIGSGGGSV
DSALLFTATELMNNNDQEEGQEMMVWLGAARTSSTKLYEQAVDAQQALNELYFRVACDHD
FLMEAYKDVIKGDPFHAKLIDLAKRINEEGIKQPIMLCLQRADYLSHWNDQTQQMELKQL
PILSKAVIPENHAGQNFAIALYQAWKMFGDPNAIILYVSQPDLFPVLALEQLQFVMFQVE

KLAKQDGQLVLVRCLTLKQCGERLSLDEHDRSLYLDGTKRVGLVHMAYGYLPQHFASEKD
FETRVMMERSTAILSLNLRLQLVNAKKIQQVLSKPGVLERFFPNDPQKVAKIRNTFMDLW
GLEEDDAITRDVIKKAIQNGSDFVMKSQMDGGHGIYFDDDICQMLKKMTLEERGAFILMK
KIKPLVAKNVIIRPFEAPKEEELNSEMGICGSLIGNQSTGQILRNNVDGHLLRSKPISQN
AGGICFGGGVFDSLLLFPSSEFQ
>Rre_GS7
MTSFPTHNVIDGNSQLMDQNKLESAQAMKAKDYTKVVIRNDEELHLLAEYAVNYAHTIGL
VGRSKDEQYKYANDVSVAPPITLLPSPFPKELYEQAVDAQQALNELYFRVACDNDFLMEA
YKDVIKGDPFHAKLIDLAKRINEEGIKQPIMLCLQRADYLSHWNDQTQQMELKQYPRNIG
LVGLSKAVIPENHAGQNFAIALYQAWKMFGDPNAIILYVSQPDLFPVLALEQLQFVMFQV
EKLAKQDGQLVLVRCLTLKQCGERLSLDEHDRSLYLDGTKRVGLVHMAYGYLPQHFASEK
DFETRVMMERSTAILSPNLRLQLVNAKKIQQVLSKPGVLERFFLNDPQKVAKIRNTFMDL
WGLEEDDVITRDVIKKAIQNGSDFVMKSQMDGGHGIYFDDEIGQMLKKMTLEERGAFILM
KKIKPLVTKNVIIRPFEALKEEEMNSEMGICGSLIGNQSTGQILHNKVDGHLLRSKPSSQ
KAGGICFGGGVFDSLLLFPSSELQ
>Rre_GS8
MTSLLIRNPIDDIGKQKLKSDQENLDLAHVVKAKDYAKVVIRNEEELRVLVEYAVDYAHS
IGLVGRSWDEQYKYANDVSVAPPMTLLPSPFPKELYEQAVDAQQALNELYFRIACDHDFL
MDAYKDVIKGDPFHAKLINLAKRINDEGIKQPIMLCLQRADYLSHWNDQAQQMELKQVEV
NVGQIGGPGSTTGMTKLHRQMLDKIETLHGQKLPIVAKAMMPENRVRHSVSKALYQAWKM
FGDPNAMILYVNQPDLFPVCAFEQLQFVMFQVEKLAKQDGQRVLVRCLSFKQCGERLSLD
ERDRSLYLDGTKRVGLVHMAYGYLPEHFSNEKDYEARVMMERSTAILSPNLHLQLAGTKK
IQQLLSKPCVLEYFFPNDPQKVAKIRNTFMDLWGLEEDDTITQDVIQNAMQNGMDFVMKS
QMDGGHGIYFDDEIGQMLKKMTLEERGAFILMKKIKPVMAKNVFIRPFEVPKEVEVNSEM
GIYGTLIGNQSTGQIFHNNVDGHLLRSKPVSQNMGGICSGGGVFDSLLLFPASEFQ
>Rre_GS9
MTSLLIRNPIDGIGQQKLKKDQDKADLEQVIKAKDYAKVAIRNEEELHLLAEYAVDYAHS
IGLVGRSWDEQYKYSNDVSVAPPMTLLPSPFPKELYEQAVDAQQALNELYFRIACDHDFL
MDAYKDVIKGDPFHAKLIDLAKRINDEGIKQPIMLCLQRADYLSHWNDQAQQMELKQVEV
NVGQIGGPGSTTGMIKLHRQMLDKVESLHGQKLPILSKAVMPENRVRHSVSKALYQAWKM
FGDPNAMILYVNQPDLFPVCAFEQLQFVMFQVEKLAKQDGQRVLVRCLSFKQCGERLSLD
ERDRSLYLDGTKRVGLVHMAYGYLPEHFPNEKDYEARVMMERSTAILSPSLHLQLAGTKK
IQQLLSKPDVLERFFPNDPQKVAKIRNTFMDLWGLEEDDAITRDVIQNAIQNGMDFVMKS
QMDGGHGIYFDDQIGQMLKKMTMEERGAFILMKKIKPVVAKNVFIRPFEAPKEEKVNSEM
GIYGTLIGNQSNGQILLNNVDGHLLRSKPISQNMGGICCGGGVFDSLLLFPSSEFQ
>Rre_GS10
MTSLLIRNPIDGIGQQKLKKDQDKADLEQVIKAKDYAKVAIRNEEELHLLAEYAVDYAHS
IGLVGRSWDEQYKYANDVSVAPPMTLLPSPFPKELYEQAVDAQQALNELYFRIACDHDFL
MDAYKDVIKGDPFHAKLIDLAKRIKNEGIKQPLMVGLQRADYLSHWNAEAKKMELKQVEV
NPGQIGGPGSATMVSKLHRKMLDKLEIEQGHKLPILAKAVMPENRPRHGIALTLYKAWKM
FGDPNAMILYVNQPDLFPVCAFEQLQFVMFQVEKLAKQDGQRVLVRCLSIKQCGERLSLD
ERDRSLYLDGTKRVGLVHMAYGYLPEHFPNEKDYEARVMMERSTAIMSPNLRLQLAGTKK
IQQVLSKPGVLEHFFPNEPQKVAKIRNTFMDLWGLEENDAITRDVIKKAIQNGSEFVMKS
QMDGGHGIYFDDEIGQMLKKMTLEERGAFILMKKIKPVVAKNFMIRPFTAPHQEDVHSEM
GIYGSLIGDQSTGKVIHNAVNGHLLRSKAASQNLGGVSTGGGVIDSVLLYPSSEFQ
>Rre_GS11
MTSISNGHSAANGTQQKFKEEEKIGQKGLVTLKANSYAIAGVRNETELKLLAGYAVDYAH
SIGLVGRSWDEQYKYANDVSVAPPLALFPSPFPKELYEQAVDAQQALNELYFRVACDHDF
LMEAYKDVIKGDPFHAKLIDLAKRIKNEGIKQPLMVGLQRADYLSHWNAEAKKMELKQVE
VNPGQIGGPGSATMVSKLHRKMLDKLEIEQGHKLPILAKAVMPENRPRHGIALTLYKAWK
MFGDPNAMILYVNQPDLFPVCAFEQLQFVMFQVEKLAKQDGQRVLVRCLSIKQCGERLSL
DERDRSLYLDGTKRVGLVHMAYGYLPEHFPNEKDYEARVMMERSTAIMSPNLRLQLAGTK
KIQQVLSKPGVLEHFFPNEPQKVAKIRNTFMDLWGLEENDAITRDVIKKAIQNGSEFVMK
SQMDGGHGIYFDDEIGQMLKKMTLEERGAFILMKKIKPVVAKNFMIRPFTAPHQEDVHSE
MGIYGSLIGDQSTGKVIHNAVNGHLLRSKAASQNLGGVSTGGGVIDSVLLYPSSEFQ
>Rre_GS12
MTSISNGHSAANGTQQKFKEEEKIGQKGLVTLKANSYAIAGVRNETELKLLAGYAVDYAH
SIGLVGRSWDEQYKYANDVSVAPPLALFPSPFPKELYEQAVDAQQALNELYFRVACDHDF
LMEAYKDVIKGDPFHAKLIDLAKRIKNEGIKQPLMVGLQRADYLSHWNAEAKKMELKQVE
VNPGQIGGPGSATMVSKLHRKMLDKLEIEQGHKLPILAKAVMPENRPRHGIALTLYKAWK
MFGDPNAMILYVNQPDLFPVCAFEQLQFVMFQVEKLAKQDGQRVLVRCLSIKQCGERLSL

DERDRSLYLDGTKRVGLVHMAYGYLPEHFPNEKDYEARVMMERSTAIMSPNLRLQLAGTK
KIQQVLSKPGVLEHFFPNEPQKVAKIRNTFMDLWGLEENDAITRDVIKKAIQNGSEFVMK
SQMDGGHGIYFDDEIGQMLKKMTLEERGAFILMKKIKPVMAKNVFIRPFEVPKEVEVNSE
MGIYGTLIGNQSTGQIFHNNVDGHLLRSKPVSQNMGGICSGGGVFDSLLLFPASEFQ
>Rre_GS13
MTSISNGHSAANGTQQKFKEEEKIGQKGLATLKANSYAIAGVRNETELKLLAEYAVDYAH
SIGLVGRSWDEQYKYANDVSVAPPLALFPSPFPKELYEQAVDAQQALNELYFRIACDHDF
LMEAYKDVIKGDPFHAKLIDLAKRIKNEGIKQPLMVGLQRADYLSHWNADAKKMELKQVE
VNPGQIGGPGSATMVSKLHRKMLDKLEIEQGQKLPILAKSMMPENRPRHGIALTLYKAWK
MFGDPNAMILYVNQPDLFPVCAFEQLQFVMFQVEKLAKQEGQRVLVRCLSFKQCGERLSL
DERDRSLYLDGTKRIGLVHMAYGYLPEHFPNEKDYEARVMMERSTAIMSPNLRLQLAGTK
KIQQVLSKPGVLEHFFPNEPQKVAKVRNTFMDLWGLEEDDAITRDVIKKAIQNGSDFVMK
SQMDGGHGIYFDDEIGQMLKKMTLEERGAFILMKKIEPVVAKNFMIRPFTAPHQEEVHSE
MGIYGSLIGDQSTGKVIHNAVNGHLLRSKAASQNLGGVSTGGGVIDSVLLYPSSEFQ
>Rre_GS14
MNINGAKLFEFLLIFAIFCTILTGTPAAHIPEGNAIIADYSAYVQEGTEQFRDLVENSKC
WAQRRGLILRTEEHVDKSDVAEIAPYTLFPTNIPRDVYRKVMAVHQAMMLLYYRVSQNFK
FVKAVTGRSDLHMEHWEEHLKQEAKQPIVMFFARSDYVLHETRLANGEMHYELKLIDMSS
GNIAMAGLSQKASKLHRRILSEIGKEVPKGALPVNMPAATLAQGLIYAWKLFNDPEAIII
VVEMPNQANKLHYDQRQVDWEVEQMTGGEIKIVYISYEQCAEKCQLDPEDNSLSLDGQKV
AVVYQRTILTPGSRSETYWDIYQKIHSSSAIKCPTLGTTVASSKKIQQALMRPGMLERFF
PDTKDADKIATIRATFPRQWYLKNIRDEKTGAAVKDAILHPDNYVLKHVHDGEEQEYHGA
EVAGKIIALKPDKHLAEYTLMERIRPMSGKTYIVKPWKQPELLDTVTELGAYGFFIADKS
KNVVRKSITSGHLMRARWAKSPADNDAPGDTSAWDSPFLV
>Rre_GS15
MRINGAKLFEFLLIFAIFGTILKGIPAAHNPEDNGLIADYSAYVQEGTEQFRDLVENSKC
WAQRRGLILRTEEHVDKSDVAEIAPYTLFPTNVPWDIYKKVMAVHQAMMLLYYRVSQNFK
FVKAVTGRSDLHMEHWEQHLKQEAKQPIVMFFARSDYVLHETKLANGETHYELKLIDMSS
GNIAMAGLSQKASKLHRRILSEIGKEVPEGALPVNMPAATLARGLIYAWKLFDDPEAIII
VVEMPNQANKLHYDQRQVDWEVEQMTGGEIKIVYISYEQCAEKCQLDPEDNSLSLDGQKV
AVVYQRTILTPGSRSETYWDIYQKIHSSSAIKCPTLGTTVASSKKIQQALMRPGMLERFF
PDQKDADKIATIRATFPRQWYLKNIRDEKTGAAIKDAILHPDNYVLKHVHDCEEQEYHGA
EVANKIIALKPGKHLAEYTLMERIRPMGGKTYIVKPWKQPELLDTVTELGAYGFFIADKS
KNVVRKSITSGHLMRARWVKSPADNDVPGDTSAWDSPFLV
>Rre_GS16
MKNFFFLTLFLLFAIVINVRGDDEAPKTIEDNEELDLDVLVLDAKDFAQNNGLTWPSRDY
PWKTGITKIAPLTLFPSPVPRKLFRQAYDVHQAMQLLYFRIAQDHKFILDSLHDILKSGE
ESYMTDMLKIHEETLREGVKQPITFFFSRADYMFHDKDQKGEGQNATGVMERTARLHRRM
LTKAHLEVSDEVTPINTPATVLGQGLAYAWEYFNDPNAAMLIIRSLSTRHVELEVERILA
SKGKKLKVFYLSSKECAKQVELDPNDFTLWVKGHKIAVVYLRDGYSSNAIRPPEDILEAF
RKIHRSTAIKCPTVIAEIVSSKKFQQVLAQPKVLEHFFPDDAEDVAAIRQTFARMWALDK
EDEETKNVIQDAKDHPGHYVLKSMGEGGGNNHFDEDIPKKLNEFTPAELSAHILMERLYP
LKFQNYMIKAFEKVQLGEMVTELGIYGYLMADSRDSSVLHNYAAGHLMRSKWSHQKEGGI
SHGVGVCDTPYLY
>Rre_GS17
MKNFFFLTLFLLFAIVINVRGDDEAPKTIAKDFAQNNGLTWPSRDYPWKTGITKIAPLTL
FPSPVPRKLFRQAYDVHQAMQLLYFRIAQDHKFILDSLHDILKSGEESYMTDMLKIHEET
LREGVKQPITFFFSRADYMFHDKDQKGEGQNCMERTARLHRRMLTKAHLEVSDEVTPINT
PATVLGQGLAYAWEYFNDPNAAMLIIRSLDTRHVELEVERILASKGKKLKVFYLSSKECA
KHVELDPNDFTLWVKGHKIAVVYLRDGYSSNAIRPPEDILEAFRKIHRSTAIKCPTVIAE
IVSSKKFQQVLAQPKVLEHFFPDDAEDVAAIRQTFARMWALDKEDEETKNVIQDAKDHPG
HYVLKSMGEGGGNNHFDEDIPKKLNEFTPAELSAHILMERLYPLKFNNYMIKAFEKVQLG
EMVTELGIYGYLMADSRDSSVLHNYAAGHLMRSKWSHQKEGGISHGVGVCDTPYLY
>Rre_GS18
MKNFFYLRLFLLFTIVLSVRGEDETQKSIEDNEELDLDVLVLDAKDFAQNNGLTWPSRDY
PWKTGITKIAPMTLFPSPVPRKLFRQAYDVHQAMQLLYFRIAQDHEFILKSLHDILKSGE
ESYMTDMLKIHEETLREGVKQPITFFFSRADYMFHDKEQKGEGQNSTGVMERTARLHRRM
LTKAHMEVSDDVTPINTPATVFGQGLAYAWEYFNDPNAAMLIIRAQDTRHVELEVERILA
SKGKKLKIFYLSSQECAKYVELDPNDFSLWIKGHKIAVVYLRDGYSSNAIRPPKDILEAF
RKIHRSTAIKCPTVIAEIVSSKKFQQVLAQPNVLEHFFPDDTEAVSAIRETFARMWALDK
EDEETKNVIQDAKDHPDRYVLKSMGEGGGNNFFDEDIPKKLNEFTPAELSAHILMERLYP

LKFENYMMKAFEKVELGEMVTELGIYGYLMADSRDSSVLHNYAAGHLMRSKLSHQKEGGI
NHGVGVCDTPYLY
>Rre_GS19
MKNFFYLRLFSLFVIVINVKGEADVQKTTEKEELDLDVLVLDAKDFAQNNGLTWPSRDYP
QKTGITKIAPLTLFPSPVPKKLFKQAYDVHQAMQLLYFRIAQDHKFILDSLHDILKSGEE
SYMTDMLKIHEETLREGIKQPITFFFSRADYMFNDKEQKGEGQNCELKQVIATGIMERTA
RMHRRVLTKAHLEVSDEVTPINTPATVLGQGLVYAWEHFNDPKAAMLIIRGQETRHVELE
VERILASKGKKLKIIYLSSQEAADSVRLDPNDFSLWVKNHKIAIVYLRDGFSSRALRPPK
DILEAFRKIHRSTAIKCPTIIAEIVSSKKIQQVLAQPNVLEHLFPDDAEAVSAIRETFAR
MWALDKEDEETKNVIQDAKDHPDRYVLKSMGEGGGNNYFEDDIPKKLNEFTPAELSAHIL
MERLYPLKFDVNYMMKAFEKVELGEMVTELGIYGYLMADSRDSSVLHNYAAGHLMRTKWS
HQKEGGISHGVGVCDTPYLY
>Rre_GS20
MKYFFLLTLFSLSIFVINVRGEADAEITTEEEELDLDVLVLDAKDFAQNNGLTWPSRDYP
IGTGITKIAPLTLFPSPVPKKLFKQAYDVHQAMQLLYFRIAQDHKFILDSLHDILKSGEE
SYMTDMLKIHEETLREGVKQPITFFFSRADYMFNDKEQKGEGQNATGIMERTARMHRRVL
TKAHMEVSDEVTPINTPATVLGQGLVYAWEHFNDPKAAMLIVRGQETRHVELEVERILAS
KGKKLKIIYLSSQEAAFSVRLDPNDFTLWVKDHKIAVVYLRDGFSSRALRPPEDILEAFR
KIHRSTAIKCPTVIAEIVSSKKIQQVLAQPNILEHFFPDDAEAVSAIRETFARMWALDKE
DEETKNVIQDAKDHPDRYVLKSMGEGGGNNYFEDDIPKKLNEFTPAELSAHILMERLYPL
KFDNYMMKAFEKVELGEMVTELGIYGYLMADSRDSRVLHNYAAGHLMRTKWSHQKEGGIS
HGIGVCDTPYLY
>Rre_GS21
MKNFFYLRLFLLFTIVLSVRGEDETQKSIEDNEELDLDVLVLDAKDFAQNNGLTWPSRDY
PQKTGITKIAPLTLFPSPVPKKLFKQAYDVHQAMQLLYFRIAQDHKFILDSLHDILKSGE
ESYMTDMLKIHEETLREGIKQPITFFFSRADYMFNDKEQKGEGQNCELKQVEINLGNVGA
TGIMERTARMHRRVLTKAHLEVSDEVTPINTPATVLGQGLVYAWEHFNDPKAAMLIIRGQ
DTRHVELEVERILASKGKKLKIIYLSSQEAADYVRLDPNDFSLWVKNHKIAIVYLRDGFS
SRALRPPKDILEAFRKIHRSTAIKCPTIIAEIVSSKKIQQVLAQPNVLEHLFPDDAEAVS
AIRETFARMWALDKEDEETKNVIQDAKDHPDRYVLKSMGEGGGNNYFEDDIPKKLNEFTP
AELSAHILMERLYPLKFNNYMIKAFEKVQLGEMVTELGIYGYLMADSRDSSVLHNYAAGH
LMRTKWSHQKEGGISHGVGVCDTPYLY
>Rre_GS22
MFLQIFYFVLFNILYTLGTAGEVSEPYWAHFLKDGGGTFSVQELADQAREQAAKIKMHKF
LRGGAKGQNDANKPMAPFTLFPSKFPRKMDLKFLENALEEVAKSDEKIRNMLSILQDIYK
EGGYRQPVAVQSQRSDYMLNVPQEDEKELELKQIEVNVGGVGGAPMGRRMTMVHRHFLRS
LGLEAGLEVVRDNHPFTTLAAGLVLGWHKFGDPEAILLMVHGPSLVKTNWDLEEEIGRLS
SGKLKYVAMPLKDCDERLKLDPEDFTLRLDDGRKVAIVYHRFPTRNLEEEWRARRMIERS
TAVKCPTMGMELMGTKKMQQVLAKPGVLEQFFSESDDAHYVEAIRQSFAGLWGMDNHQDK
ETMERIQDAIDHPERYVLKPMKEGGGNNIFGQEIAEKLQEFVPEERAAHILMQLLRPVVS
QNYFIRPNKEPELINVNSEFSTFGCLVGNVNDGTVYHNHGHGYLMRTKWEHSAEGGILVG
HGAYDTPYLC
>Rre_GS23
MAKSINFFLACFFIIFSTAFCGPVDENGKNGLINKEDDGDKLDDEHSLNYWSGVVKNKHE
LQVLVEQAREYADSIKHYEWTYASQHYGKPDFKPIAPMTLFPSKFPRQLFDLAQAVQPGM
NLLYFRASLDFEFLINNFSEMAKSDLYIGNLLKILKETKEEGYRQQIAVQLQRSDYMTHL
EKAGDDQEIQLKQIEVNVGGGGGPPMAKRGTKVHRKTLTQLGLDASVEVLRDNQPYTTYA
EALYNGWRSFGDEDAIMVILAGLLRKGEHGGKKTQWDLEEQLARLSGGRLKYIAMSIEQA
NERLYLDPKDFSLRVKKDDRKVAIVFHRYPMDPNDPAEWNARAFIERSTAVKAPTIGMEL
LGTKKGQQLLAMPGVAEKFLTSPDEAHYVDSIRQTFAGLWGLDQKGADIEAVKQDAIAHP
EKYVMKPMKEGGGHNFFDQTLVDNLENFSPKELAAHILMQKLQPMAVPNYFVRPCEEPQF
VMAASELGVFGGLVGNVQDGTVLYNHAHGYLVRSKSVKSNEGGVLSGNGAYDSAYLY
>Rre_GS24
MAKSINFFLAFFFIIFSTAVCGPVDENGKNGLINKEDDGDKLDDEHPLNYWSGVVKNKHE
LQVLVEQAREYADSIKHYEWTYASQHYGKPDFKPIAPMTLFPSKFPRQLFDLAQAVQPGM
NLLYFRASLDFEFLINNFSEMAKSDLYIGNLLKILKETKEEGYRQQIAVQLQRSDYMTHL
EKAGDDQEIQLKQIEVNVGGGGGPPMAKRGTKVHRKTLTQLGLDASVEVLRDNQPYTTYA
EALYNGWRSFGDEDAIMVILPGLLRKGEHGGKKTQWDLEEQLARLSGGRLKYIAMSIEQA
NERLYLDPKDFSLRVKKDDRKVALVFHRYPMDPNDPAEWNARTFIERSTAVKAPTIGMEL
LGTKKGQQLLAMPGVAEKFLTSPDEAHYVDSIRQTFAGLWGLDQKGAEVEAVKQDAIEHP
EKYVMKPMKEGGGHNFFDQTLVDNLKNFSPKELAAHILMQKILPMAVPNYFVRPCEEPQF

VMAASELGVFGGLVGNVQDGTVLYNHAHGYLVRSKGVKSNEGGVLSGNGAYDSAYLY
>Rre_GS25
MYISIFISILCLIISSYGAPPTIDRNHGHHQEHQHPLIYWSHLVNGKEALQNLANYERQC
AADVGHIEYLKTFDYTDKGPDFKPIAPMMLFPSKYPRRIFDQAEALQPAMNLLYFRASRD
YEFLQQAFAETAKGDHYIQTLLDILGEMQEHGYRHQPLAFQLQRADYYMCYGKNAAKDDV
SKYELKQIEVNIGPAGGPPNAIRQTKVHRKVLTKLGLDASLELLRDNHPYTTYVEPLILA
WHQFGDPDAIMLIAVGMGVKGKTVGKKTQWDLEEEIERLSSGKLKYRLTTYRQCASRCRL
AEDNSLMFKPRDKDGNDLPEVKVAVVMHRYKTNDDEEEWKARRLMERSTAIKGPSIGMEL
IGTKKGQQALAKPGAVERFFGPGEEHYAEAVRQTFAGLWGLDNEEAETKRVIEDAIAHPE
NYVMKPMKEGGGNNFFGQTVADKLQQLDQNELAAHILMERLQPVSVPNYFVRTADEQPAF
GMVVPELSTFGGLLGNIQDGTVLHNHGHGVLTRTKPEDSDEGGIFSGIGAYDTPYLY
>Rre_GS26
MKISLFLFVFLFINDLNNYDAAPVEIEDAHATNVLAFWTHLVKDEEAVRELAAFESKCAE
EVDHLEILEKDDLKDSPGYVPVAPMTLFPSAFPRHLFDLAEAVQPAMNLLYFRASRDYDF
LVKTFKEVAKQDRFISKSMDIIEQIQQEGGHRQPIAVQVERADYMVHVAENKEGKEEFQL
KQIEVNVANGGGSNNAIRQTKVHRKVLAKLGLDSSIEVLRDNHPYMSYAEPAYLGWLKFG
DPDAIMVITVKGRKDPTKATDLQKKYGKRAGHFHTDLHADFDMLSGGKMQIEYLTMEECD
DRLTLDPEDFSLRLDDGRKVGIVLYRWAGSSDKAWSARLKIERSTSVKSPTVAMNLLGSK
KGQQALAAKGVVEQYFPDPDEAHYVEAIRQTFAGLWSLEKDDERTKQLIKDAMNHPEQYV
LKPNKDGGGHNFFDQDLVEKLQSLSTPAERSAFILMERLKPMTVPNYFIRSARAPLKLDH
VVPELGIYGGLVGDLSNGNVLHNHGHGYMIRTKKADSAEGGIWEFAGVYDSPYLY
>Rre_GS27
MSIFLFVILFIFGISSYEAAPLEKEEPHSNSVLAYWAHLVKDEQAVRELVDFESKCAEEL
DHLEILEKDDLKDSPGYVPVAPMTLFPSAFPRHLFDLAEAVQPAMNLLYFRASRDYDFLV
KTFKEVAKQDKFIGRCMEIVEKIHQEGGHRQPIGVQVERADYMVHVEKNEDGTEEFQLKQ
IEVNVANGGGSNNAIRQTKVHRKVLAKLGLDSSIEVLRDNHPYMSYAEPAYLGWLKFGDP
EAIMVITFKGRKDYTATDVQKKYGKRAHHFHTDLHADLDMLSGGKMQIEYLTMEECDDRL
TLDPEDFTLRLDDGRKVGIVLYRWAGSSDKAWSARLKIERSTAVKSPTVAMNLLGAKRGQ
QALAEDGVVERYFPDPAEAHYVEAIRNTFAGLWSLENDDEQTQKLIQDAMAHPEKYVLKP
NKDGGGHNFFDQDLVEKLQSLSTPSERSAFILMQRLNPMTVPNYFIRSARAPLKLDHVVP
ELGIYGGLVGDLSNGNILFNHQHGYMIRTKKEDSPEGGIWEFAGVYDSPYLY
>Rre_GS28
MSFSLQVMSFLEPNSNSVLAYWAHLVKDEQAVRELVDFESKCAEELDHLEILEKDDLKDS
PGYVPVAPMTLFPSAFPRHLFDLAEAVQPAMNLLYFRASRDYDFLVKTFKEVAKQDKFIG
RCMEIVEKIHQEGGHRQPIGVQVERADYMVHVEKNEDGTEEFQLKQIEVNVANGGGSNNA
IRQTKIHRKVLAKLGLDSSIEVLRDNHPYMSYAEPAYLGWLKFGDPEAIMVITFKGRKDY
TATDVQKKYGKRAHHFHTDLHADLDMLSGGKMQIEYLTMEECDDRLTLDPEDFSLRLDDG
RKVGIVLYRWAGASEKAWSARLKIERSTAVKSPTVAMNLLGAKRSQQALAEEGVVERYFP
DPAEAHYVEAIRNTFAGLWSLENDDEQTQKLIQDAMAHPERYVLKPNKDGGGHNFFDQRL
VEKLQTLSTPAERSAFILMQRLNPMTVPNYFIRSAREPLKLDHVVPELGIYGGLVGDLSN
GNVLFNHQHGYMIRTKKEDSPEGGIWEFAGVYDSPYLY
>Rre_GS29
MAKIFSSKLFLFIFVLANIQVVELQDFDGLIDYVPNVVQEQKDRGLFDELVNFGRNYANS
IHMLKRTREHKHLPKDLEPLAPFTLLPSKFPRKLFDQAMAVQPAMNLLYFRGSLDHKFLA
TTLKEVAKTDEMVRTLVDIMQEVDREGGYRRQPISVNIQRADYMLNVIEVNVGAPGGGPM
APRMTKVHRKMLAKVGMNSSLEVLRDNRPYKTLAEAMYLAWLQFGDPKAIMVMVMGRGYL
ETKAFVLPWGLDLYKEIERASSGQLNYVRMSVELCEKRLKLDDDMNLWLDDGRRVGLVYY
RAIPTLTDEFGWKARRMMERSNAIKAPTIGLELMGMKKVQQVLAGPNVVEQFFSNPEEAH
YVEAIRAVFTGLWSLDNDDEDTNKVIKDAIAHPDGYVLKPMKEGGGNNFFGKDVAEKLQQ
FTPEERRAHILMQRITPPTFQNYLIKPFNEPKLDEMVTEFSTFGSLVGNATSGTVLYNRG
DGYLMRSKWKNSNEGGIIYGTGAYDSPYLY
>Rre_GS30
MVSIDDKQDYVPALVTIKGHNLNDLVEYQRSLANEIKMHKRTNEHKHDPNLEPLAPFTVL
PSRFPRKQFIQAMDVQQALNLLYYRASLDRDFLRKNLEEVAKSDGYFQKLIDIMDEVEQE
GGYRKQPVSVNVQRSDYMLNVVKDEENKDNYELKQIEVNVGAPGGCPMAQRMTMVHRNTL
AKLGMDASPAMLPENRPYDTIVQSLYVAWRMFDDPNAVVLMVMARGFKDGVKRLYWAFDI
ERELTRISSGEIPIERMELTTCDERLKLDDDMTLRMDDGRKVAVVYYRSIPYRFSAKPGA
EEMCWSARRTIERSTAIKAPTIGLELVGTKKMQQVLAIPSTLEHFFPKPEEAHYVEAIRE
VFTGLWDLENDDEDAIAHPDGYVLKPMKEGGGNNFFDEKVAEKLQQFIPEERRAHILMQR
ITPPTFQNILVKEKEPLKFGEMVTEFSTFGSLVGNEIDGTVLYNLGDGYLMRSKWKDSNE
GGILYGAGAYDSPYLF

>Rre_GS31
MLCLAQFSLLLFDLAIISSQWIPFQPTEDDECPTKMHTAEISLWHRFNPPSCFADPSLDT
MRNYVPHLAQLQDKRHLQQLVHEAHRHAYDLCMHMDLLEEEGKWEFIPVAPQTLLPSRFP
RSQFDQAKALQPALNLLYFRASWDYEFLYGALSKVMASDAIVRNMMAIMLDVLKEGGYRR
QPISLNWIRTDYMLHMDPQNSEKLQQDSGYTLKQIEVNAAAAGGAPRAKRMTKVHRRTLA
QLGMDASLEALPDNSAETMLIEALYRAWLQFGDPNAVVLSITHGRGKISPILIKDVDREI
ERISGGKMKFVRLSLEQCDDRLTLDPDNFTLHLDDGRSVAIAYFRVFPFPVNVPNSEKAW
RARRAIERSTAIKAPSIGFHLMNLKKVQQVLAKPGMVERFFWRHDEKHYIDAIRQTFAGL
WGLDTDEDQEIETVIRDAIAHPDRYVLKPMKEGGGHNFFGKDIVDKLKAFTPEEKATHIL
QQRLRPPVFQNFMIAPLGRLDFGSMVTELSMFGTLLGNGHFGTVQYNHAGGYLARSKWEH
SPEGGISIGEGFYDSPYLY
>Rre_GS32
MVSMLFISPLFFFVVTLAAPPTSRRQPEYTPEDPFPNKEKRIIQYLKEYLRPVPGSADSN
LCEIRNYVPYLAQLQDPKRLHELVVSMHEHAYSIGMHMWNKKEDDLRGYFFPVAPMTLLP
SKFPRNKFDQAKALQPAVNLLYFRASLDFEFLRDALGEVAKSDQLVQNMLDIMRDVIEEG
GYKRQPISANWLRADYMLSITDQDNQYELKQIEVNAGAAACAPMVHRMTKVHRKALTQLG
LDASLETLRDNEPFRMAIEAFYLAWLKFGDPNAVLLAIGHKSPVRYIPSLLMKDIDDGIK
HISGGKMKVVFMGLEECDDRLTLDPKDFTLQMDDGRLVAIANYRPFPFPVKVPTSEKAWR
ARKMIERSTAIKAPSIGFHLMNLKKTQQVLSKPGKVEQFFSSEEGHYAAAIRQTFAGLWG
LDEDQDEETKLAIEDAIAYPEKYVLKPLKEGGGNNFFGQVIAEKLRTMEPKERSSHVLQQ
RLQPPVFQNYLLEPLVGFEFGKMVSELSIYGTIVGNANDRTVLYNKGEGYMVRSKWEHKD
EGGILIGEGFYDSPYLY
>Rre_GS33
MGPPKQIREDNSLLRQPAWNQLLSHVTSSTDISSERDYVPYIVQPLPLDRLFSLVRSTHQ
YADEIQLHQMTVKNWAMEPPPITLFPSKFPRKQFDQAKTLQPALNLLYFRASSNYEFLEE
SLGEVAESDKMVRGMLDIIRDVIREDGYKRQPISVNWMRADYMLNTKDKDKENDSMEDEY
ELKQIEVNAAAAGGSAQAHRMTRVHRRVLSQLGLDASMEVLRDNNPFKMQIEATHLAWLK
FGNPEAVLVLITSRYKDMHGMKHLPDFELGIKRISGGQMQIVSMSLEEIDERLTLDPDFS
LRLDDGRLVAVVSCRLSPSSPRNSKKQNNEPILRAMRKIEFSTAIKAPTIGFHLMNLKKM
QQVLAKPAVVEKFFPDLEEAHYIEAIRQSFAGLWGMDADDEGTKAAVQDAFAHPERYVLK
PMKESGGNNYFGINIPRKLRAMTPAERKTYILQQRVRPPEFHNFAILSNYDRPLFGKMVS
ELSIYGALVGDVHDKVVLYNKGEGYMVRTKWAHSDEGGILIGQGFYDAPYLY
>Rre_GS34
MLKIQFLNFLAKLLNSGLLASEHELDFLVRPLIFKFYAKFVLGIYLPLRDYVPRLAQLQD
PDQVRKFAKFARQYADEIKLHRETARNWAKEVPPITLLPSKFPQKYFDQAKTVQPAMNLL
YFRASIDYEFLSEALGEVAKTDHLVKEALDIMSEVIREGGYKRQPMSVNWMRADYMLHME
DENGDGGDDEQYELKQIEVNAGAAGGATQSHRMTKVHRKMLTELGLDSSLAVLPDNTPYN
MLVEALYQAWLRFGNPNAVMLEVFYRYTYHLVLPDLEQGIERLSGGKMQLVQIPLEVLEQ
MITLDPVDFSLRLDDWRLVALTLCRLSPILRFLLNPAWSSDSFRTQKKRIMPLQFGKPLP
DCGPWTERMKAAIQDAMAHPEGYVLKPMLESGGNNFFRDGIVQKLQQMQPHERATYILQQ
RIHPPEFHNYAIIPNYDHPEFGKMTSELSIYGVLVGNADKDNKKTVLYNKGEGHMVRSKW
AHSDEGGILIGQGFYDSPYLY
>Rre_GS35
MIFMNILAFLIAFVFTVPSYSAPTNLADNAVCTLRDYVPHFAYLKDVKKLQELEQFARKY
ADEIEFHHERTKNWVGEPIAPFTLLPSKFPRRSFDMAMELQPALNLLYWRASFDYEFLSD
AFGEVAKTDAAVRQILEIIRDAVNEGAYQRQGISMNWMRADYMLHMEDDAQAEGNNDQFE
LKQIEVNAGAAGGIPMCRRMTVIHRHILEHLGMDASPEVLRPNTPFDMMVEALLLAWNKF
GDPDAVVLAFTRTRSKMPGAESKIQALFDLHTDGKLKMEYMPYADCDDELILDPKDFSLR
MKDGRKVAIADYRNAPFPPVIKTTEKLMRCLRKIEFSTAIKHPTLGFHLMNLKKAQQYLA
KPGMVERFFTKPEEAHYAKAIRQTFAGLWGLDSEDDETKAIIEDAMEHPDRYVLKPMKEG
GGNNFFGEDIVKKLNELTPEQRGTFILQQRINPPTYQNYIVEPFATERQDPNPRFGTMIA
ELSIWAMFVGDGHKDTVLYNHGHGYMVRTKWADDNEGGILNGRGFYDSPYLY
>Rre_GS36
MQILPFLIAFSFIVPSYSAPTNLADNAVCTLRDYVPHFAYLKDEKRLQELEQFARKYADE
IEFHHERTKNWVGEPIAPFTLLPSKFPRRSFDMAMELQPSLNLLYWRASFDYEFLSEVFG
EVAKTDAAVRQMLEIIRDAVNEGAYQRQGISMNWMRADYMLHMEDDAQAEGNNDQFELKQ
IEVNAGAAGGIPMCRRMTVIHRHILEHLGMDASPEVLRPNTPLDMMVEALLLAWHKFGDP
DAVVLAFTRTRSKMPGAESTIQALFDLHTDGKLKMEYVPYADCDDELVLDPKDFSLRMKD
GRKVAIADYRNAPFPPVIKTTDKLIRCLRKVEFSTAIKHPTLGFHLMNLKKAQQYLAKPG
MVERFFTKPEEAHYAKAIRQTFAGLWGLDSEDEETKAIIEDAMEHPDRYVLKPMKEGGGN
NFYGEDVAKKLKELTPEQRGTFILQQRINPPTYQNYIVEPFATERQDPNPRLGTMIAELS

IWAMFVGDGHKGTVLYNHGHGYMVRAKWADDNEGGILNGRGFYYNPYLY
>Rre_GS37
MIFMQILPFLIAFSFIVPSYSAPTNLADNAVCTLRDYVPHFAYLKDEKRLQELEQFARKY
ADEIEFHHERTKNWVGEPIAPFTLLPSKFPRRSFDMAMELQPSLNLLYWRASFDYEFLSE
VFGEVAKTDAAVRQMLEIIRDAVNEGAYQRQGISMNWMRADYMLHMEDDAQAEGNNDQFE
LKQIEVNAGAAGGIPMCRRMTVIHRHILEHLGMDASPEVLRPNTPLDMMVEALLLAWHKF
GDPDAVVLAFTRTRSKMPGAESTIQALFDLHTDGKLKMEYVPYADCDDELVLDPKDFSLR
MKDGRKVAIADYRNAPFPPVIKTTDKLIRCLRKVEFSTAIKHPTLGFHLMNLKKAQQQYL
AKPGMVERFFTKPEEAHYAKAIRQTFAGLWGLDSEDEETKAIIEDAMEHPDRYVLKPMKE
GGGNNFYGEDVAKKLKELTPEQRGTFILQQRINPPTYQNYIVEPFATERQDPNPRLGTMI
AELSIWAMFVGDGHKGTVLYNHGHGYMVRAKWADDNEGGILNGRGFYDNPYLY
>Rre_GS38
MISMQILPLFFLFAVINLSYTALAKTGSDLICAMRDYVPQLAQLKDEKQLEEIEKFGRKY
ANSIGFHHETKRHWKWTSEPISPFPLVPFRYPRKSFDMAMQLQPAMNLLYFRASFDYDFL
NMALGEMAKSDDFVKTILDIMHEMKQENGYQRQPISMHWQRADYMLHVNEQENECDNNIE
LKQIEVNCSAASGMPMSQHVTKIHRQVLKHMGLDASPEALRENNPFPMIAEGFLLAWLKF
ADPNAVVLSIMYRSVESSKMSKMEVETRQLFERFTGGKMQFVYLGIEECDEKLTLDPEDF
SLRLEDGRKVAIADYRILVRPKNKPVSEKVKRVLRKIERSTAIKHPTVGSYLMDLKKVQQ
VLAKPGMVERFFTKPEEAHYAKAIRQTFAGLWGLDSEDEETKAIIEDAKANPDRYVLKPC
KEGGGNNFFGEDIVKKLEEFTPEERTSHILMQRVKPPVVQNYLIEPYIGERPKPKFGEMV
IELGTYGMLVGNSQDGTILYNHGQGYMARSKFSHSDEGGIFEGAGFYDSPYLF
>Rre_GS39
MIFLQILPFFFLCTVINLNYTALAETDNDLILAMRDYVPQLAQLQDEKQLEEIEKFGRKY
ADSIGFHHETKRHWKWTSEPISPFPLVPFRYPRKSFDMAMQLQPAMNLLYFRASFDYDFL
NMALGEMAKSDDFVKTILDIMHEMKQENGYQRQPISMHWQRADYMLHVNEQENECDNSIE
LKQIEVNCSAASGMPMSQHVTKIHRQILTHMGLDASPEALRENNPFPMIAEGFLLAWLKF
ADPNAVVLSIMYRSVESSKMSKMEVGTRQLFERFTGGKMQFVYLGIEECDEKLTLDPEDF
SLRLEDGRKVAIADYRILVRPKNKPVSEKVKRVLRKIERSTAIKHPTVGSYLMDLKKVQQ
VLAKPGMVERFFTKPEEAHYAKAIRQTFAGLWGLDSEDEETKAIIEDAKANPDRYVLKPC
KEGGGNNFFGKDIVKKLEEFTPEERTSHILMQRVKPPVVKNYLVEPYIGERPKPKFGEMV
IELGTYGMLVGNSQDGTILYNHGQGYMARSKFSHSDEGGIFEGAGFYDSPYLF
>Rre_GS40
MILTKILEFFFALTFIGCSFSASAHTDNELICAMRDYVPKLAQIQDEKQLEEIERFARKY
ADSIGFHHETKRHWKWTSEPITPFPLVPNRFPRKSFDMAMHLQPAMNLLYFRASFDYDFL
NMALGEMAKSDDFVKTILDIMHEMKQENGYQRQPISMHWQRADYMLHINDAKNADNKLEL
KQIEVNTSAASGIPMSQHMTKVHRQVLKHMGLDASPEVLRVNNPLPMIVEAFLLAWLKFA
DPNAVVLSIMYKPTESAKMSKTEAESRQLFSRLTGGKMQFVYMGIEECDEKLTLDPEDFS
LRLEDGRKVAIADYRILVRPKTKPLSEKVKRVLCKIERSTAIKHPSVGSYLMDLKKVQQV
LAKPDMVERFFTKPEEAHYAKAIRQTFAGLWGLDSEDEETKAIIEDAKANPDRYVLKPCK
EGGGNNFFGEDIVKKLEEFTPKERTSHILMQRVKPPVVKNYLVEPYIGARPKPKFGEMVA
ELGIFGLLVGNNQQDGTILFNHGQGYAVRTKFSHSDEGGIFEGAGFYDSPYLF
>Rre_GS41
MSLIQILEFFFALVFIDCSFSTSVLNDNELICAMRDYVPQLAQLQDEKQLEEIEQFGRKY
ADSIGFHHETKRHWKWTSEPITPFPLMPYRFPRKSFDMAMQLQPAVNLLYFRAAFDYEFL
NMALGEMARSDDFVKTILDIMFEMKQENGYKRQPISMHWQRADYMLHINDAKDTDNNLEL
KQIEVNTSAASGIPMSQHMTKVHRQILTHIGLDASPEVLRVNNPLPMILEAFLIAWLKFA
DPNAVVLSIMYKSVENSKMSKNEAETRQLFERFTGGKMQFVYLGIEECDENLTLDPEDFS
LRLEDGRKVAIADYRILVRPKTKPLSEKKHPSVGSYLMDLKKVQQVLAKPGMVERFFTKP
EEAHYAKAIRQTFAGLWGLDSEDEETKAIIEDAKANPDRYVLKPCKEGGGNNFFGQDIVK
KLEEFTPEERTSHILMQRVKPPVVKNYLVEPYIGARPKPKFGEMVAELGIFGLLVGNNQD
GTILYNHGQGYAVRSKFSHSNEGGIFEGAGFYDTPYLF
>Rre_GS42
MYLLFSISSLLLTFALIHSGNAETSESESPLFIPNMMMNMIKDEEHLKELKNDAIDWAHY
VGLKFRVQDHKDKSDLATIVPISLFPSPFPRKVYDQAVAVQEAMALLYFRASWNFEFMTN
ALAEVEKSDEVVQKMMGIYREAHVAGIKQPIAVLPIRNDYMMHINKSSKNASQFQLKQVE
VNIGYMGGGSRGPGATKVHRRITSLIGFDQTRVPENHALETVCKGIYYAWKKLADPKAVL
VMLINPISYGHFEPRAYEYELERVSKFEMPILLLSMEQANERLTLEDDSTLRLDDGRRVG
IVFSRATALASSRKLPDFIWEVKRKVELSTAIKIPTLGEELASTKKVQQVLAVPGMVEKF
FTKPEEKPMVEAIRRTFAGLWGLQNNNEEAKQIIAHAIAHPEKYVLKPQKEGGGRNLWNK
EMVEKLENASESELAEYILQQRIESAITENYAIRPKDEEPEKNMLITEKGNVITELSTFG
CLVGNGVDGTVFQNEGSKEGTEGYFMRTKWSHVTEGGVLKGEGVYDSPYLI

>Rre_GS43
MNTNFLIEISIIFSFFVVLLICAQNEVQNTKNEVEDDNLGELVKDAMDWAQQVGMIWRPK
ESKNKSDVAMVTPVALLPTPFPRKMFWQAYELQKYMNLLYHRAASDYKFLVDNLTEPAKS
DPTIGIWLAALKEMHEEGVINRQPITMVYQRSDYMAHVKPLCDQCVAEGRREFELKQIEV
NTGAPAGLICERVTHFHERMVVTAGITEDPSEVLPDNPVTSTIAQGLFVAWQKFGNPDAI
LLMLNTPGGNQHQFEKRYIQYELERLSKGKMKVVLMSLGHCHTHLRLADDYSLMLGDLVV
GVVFFRASLLISVEGLHGINKVLEARKMIERSTAIKCPSIEHILTGTKKFQQIMAKPGNV
EHFFPKPEEAHIVRGIRATFAGLWGLDDESEKTKAVIKDAIAHPDNYVLKPNREGGGHNL
WGQDIADKLSNFVPEERDQFILMERLTPLKTENYFLRNLLPPEKGEVTSELSVFGVLVAN
ARDGTVIHNGRGGHLYRTKWANVNEGGITPGTGVYDSPLLFD
>Rre_GS44
MMKLVQTKIVKLFLEKRCLYVFTLFLLCTSSMGELSGENEHNNNRDEDVQVLVDDVMDFS
HFVGSQLRSKDLPNKSDLYAVAPVTLFPSLVPRKPFEVANDIQETMNLLYFRISCDFDFL
ANTLSGADKYVKTLIDMAREAIEEGHQVPVKLLLERADYMFHINRDLDENDYNRYELKQI
EVNIGPVGGTVNDNVYLLHRRVLSKAGISDKNLAPSRGIDTLAWGFYTAWQQFGDPNAII
ILVHQEQSTMATPEQRQLQYEIERISGNKIKIVRISTQACGEKLQLREDEDFSLYLDHMK
VAIVYFRMSFMVIKDKESLDIRRMLERSSAIKVPSLAEEMSGSKKVQQVMAEPGMLEKFL
PEPEHAKHIEAIRGTFAGLWGLEGEDEQTKAIIQSAIDHPDRYVLKPNREGGGHNIWGED
IATKLKEFTPEQLAGHILMQKLDPIVVKNYFVRPKIDVEFGHVISEISIFGALVGDQKTG
KVAYNKGGQGYMMRTKWSDVNEGGILKGTGVYDSPLLF
>Rre_GS45
MAKLSFFVAFFFIYISNVMITVVNAEVSKAETGNIVYAANSEFQALFEEAVDYAHQMGVQ
FRIKDHRDKSDVLGVAPVTLFPSAVPRKLFEQANALQEALNLLYFRVSCDQKFLMDQLVS
GMDPTIKILADMAWEAQAEGHQLPIKLFLERADYMFHSTGNDPKVGPTHELKQIEVNIGS
VGGLVNDRVTQLHRRMVSKAGIDRTLLPQNNAINTFAQAFYVAWQKFNNPDAIMLLVHEE
QSTMATPEQRHLQYEIERISNEKMKVIRITVQACAERLSLADDFSLLLDDLVVAIVYFRI
SFMGPNSAAKMQARQIIERSTAIKAPSLPQELASAKKIQQVLAQPGMLEKFLHEPEHAKN
IEAIRNTFAGLWGLEGEDEKTKAVIQDAIDHPERYVLKPNREGGGHNIWDDEISTSLQKF
TPEELAGHILMQKLNPLIVQNFFLRPLMPLEFGNVINELSIFGALVGDEKDGKIHYNKGH
GHMVRTKWAHLNEAGVIKGTGANDTPLLF
>Rre_GS46
MAAEQQANCPSWEIRHDLKALEDEARDWAHYVGAVMRTKKSRDRSDVMQFAPMALFPTPV
PRKLFQQAVDVHNAMTLLYYRASSDYQFMSKALHEVGQTDETTRTMLDIMHQVQAEGHQS
PIKLLFERCDYMFHVNKDEKDEAKKHQLKQIEVNIGPMGGNLTERVTEFHRRVLSKAGIA
TPPEVLPINKPTNTMAQGLYLAWQKFGDPDALLVMVYYKPSTLVISEHRMVEYELERISE
GKVKIVRLSTEDCARMTLNDDFSLRLDGQRVGLVYFRVTFLVQKFSATWETRRMIERSTA
IKVPTIAQELASTKKMQQIMAQPGMVEQFFKDPKDKDKITAIRQTFAGLWGLDRDDEETT
AAIQDAILHPDRYVLKPNREGGGYNLWGVDIAKKLRNLKAEERADYILMQKLEPMVVKNF
FLRPGMDLEFGPVITELSIFGALVANEKDGTLYQNICQGHLMRTKWADCNEGGILKMTGV
FDSPLLIG
>Rre_GS47
MSTIFRVIFAAVFLQVFLLCSASTDEEENGPKAKNADGEDLRVLQDDAMDWAHKVGAIMR
RKESRNRSDTMQVAPMALFPTLVPRKLFNHGNAIQETMNLLYWRVSSDYEFMSKSLTELA
KTDVTTGRMLDIMHQVHKEGNQPEISMFLERADYMFHVNKAAMDEDSQYELKQIEVNIGP
IGGTLSQQVTEFHRRMLTNAGMPTSDDVLPKNDSTGTLAQGLFLAWQKFGNPDAVFVFVY
HRISNLVVSEHRTIQYELERLSEKMTGRKMKILRLSWDDCERLELADDFSLRLENHVIGM
MYFRFTFLCDVYTEKSLERRRMIERSNAIKSPSIAQELASTKKIQQIMARSGMLEQFLPE
AEHAESIKAIRQTFAGLWGLEGEDQKTEEVIQDAIAHPDRYVLKPNREGGQNNIWGQEIA
EKLRNFTHADRSEHILMQRLEPMVVQNYFLRPEMDLELFDVITELSIFGALVGNKKTGEV
LHNKGHGHLMRTKFAEINEGGVLKGTGVFDSPYLF
>Rre_GS48
MSTIIRVIFVAVYLELFLLCSASTDKEGIFSRSDTMQVAPMALFPTLVPRKLFNHGNAIQ
ETMNLLYWRVSSDYEFMSKSLAELAKTDVTTGRMLNIMHQVHKEGNQPEISMFLERADYM
FHVNKAAEDEDSKYELKQIEVNIGPIGGTLSQQVTEFHRRMLTNAGMPTSEDVLPRNDST
GTLAQGLFLAWQKFGNPDAVFVFVYHRISNLVVSEHRTIQYELERLSEKMTGRKMKILRL
SWDDCERLELADDFSLRLENYVIGMMYFRFTFLSDVYTEKSLERRRMIERSNAINLLRSR
RNWPAPRKSNNIMARPEMLEQFLPEAEHAESIKAIRQTFAGLWGLEGEDEKTEEVIQDAI
ANPDRYVLKPNREGGQNNIWGQEIAEKLRSFTKADRSEHILMQRLEPMVVQNYFLRPEMD
LELFDVITELSIFGALVGNKKTGEVLHNKGHGHLMRTKFAEINEGGVLKGTGVFDSPYLF

>Rre_GS49

MANNNFLSTFGFIFVLISIGHAVPTHKGDFDAPLENGHAVVCNGYVVNDGCPLMASSSSS
VEDEDVGHTAACNSDVADSSCCLSPSSTTTTTSEDVENLVENNAEWMEVFVEDAKDWAYR
VGMIVKPKEHLGSSDTSQFAPFALFPTKIPRELFHQAVGVQEALTLLYFRASCDYEFLVK
HLTEASQTDEVLRKLMDIHEEVHREGVIRQPISMVFQRADYMFHTKNDTEEEYELKQIEV
NSGAVAGLLIQRVTEVHRRVLTKAQLPTTTEFLPENNPIGNVALGLYTAWKAFGNPNAIM
VMVVSKIGSPNHYEQRLVEYELERISSGQMEVVRLRHTECAQRLHLAKDFTLMLGNRVVS
VVYFRVSHLATKDYDGDTFEGRRMIDRSTAIKCPSIGMLLANTKKVQQALTMPGMLEHFF
PLPEETKMIVAIRSTFADMWGLDKSDKKTRRVIKDAIAHPDRYVLKPNKEGGGNNFWGPD
IADKLITMPHAELAKHILMQKLKPMVSQNYFIRPFLAPEFGPVVSELGVFGALVGNQVTG
RVLYNRSQGHAMRTKWERVHEGGISCGSAVYDSPYLVV
>Rre_GS50
MASTNGTILMLLFGICFYCCSLLSFVFGETETIPNGDIAEITKSEEPEDIQVLVDDAKDY
AHYQGLIFRENRESSDIARFSPFTLFPTPVPKKQFHQAVAVEDAMTLLYFRISRDYKFLV
KILGEMAKTDKIIGALMDILHEVQEEGIRQPLTMVVQRADYFFHEKPEGGNNGDKYELKQ
IEVNSGAIAGTFVQHLTDLHRRTLKWASMDASEARLPVNNAISTIADGLYQAWLAYKNPD
AILLMLTSPRGSPTRFDQRYIEYELDRISEGKLVTVIMNIKESGRLTLGNDFELLLDHRE
VAVVYFRVSFLQEFEYLEPYLPARKLIEHSKAIKCPSIGMSLASTKKIQQVLAQPQMVEL
FFPGSENKPTVDAIRATFAGLWALEEDNRDIIQDAIDHPEGYVLKPNRECGGHNIWGQEI
TEKLKTMKPEERKDYILMQRLHPMATKNYFVHPGSEPKLVSVVVELGVFGCLVGNLEDGT
TSYNRGHGHLMRTKWATAQEGGVIEGSGALDSPYLF
>Rre_GS51
MKHSVKFDYVFIIFAIFLLNHGYAEEKNIEEGSKAEVPADRKVSEDLAVLLDDAKDYAHH
VGMKMRTKEMVNRSDVAEFAPFALYPTPFPRKIFKQAMDVQEAMMELYFRVSNDFDFLVE
HLTEVSKSDETIRTLIEIMREAREEDGPKQSISMIWARSDYMCHENKNANEDTPKYELKQ
IEMNIGAAGGYIGERTTQVHRRVLEKAAMDVSNEVIPDNSSSMTTAQGLYEAWKLFGNPD
AVVAFYMPRNGIKTHFVDRAVEYVLRELSKKKVKTIFVDVVECRNNRLTVDPVDKSLRFD
GQLIGVFFYRVTLLLPLDTMPNGVMDVRRTVERSTAIKCPTVGMILASTKKIQQVLAMPN
MVERFFPLPEDADTVEAIRKTFTGMWALDKDDDMTKAVIKDAMEHPERYVLKPNREGGGH
NFWNEQIPEKLKNFSKSELAEHILMQRIQPRIAQNYFVRPHEAVEFGDVVTELSPFGVLV
GNMDKNKVIFTRSHGHFARTKWSNVTEGGIRIGSAVYDSPYLFDTD
>Rre_GS52
MKTTVIFTNHNIHNQYVLFDEDNEVENEEDLQVPEQNGQKSVQTNKISENSIPSSSHQNV
HNLYVLFDEDNIVVPNEPAEQMEELESEEDLQVLVQDAQDWAHNIGMKFRLKDKKDCSDG
AEVPPFTLFPTPIPLHEEMIKTETDDVIKDMIAMMEDMYEEGIKQEYAMIWQRADYMTHG
KPTVDVNNGAAGGFMADHVTEVHRRLLSKAGLDASPEVLPRNETLDTVAEGIILAWELFG
DPNAVVVIMESSTGILTQFVDLEIGEHVYKLSNGRVKTVYMNYLQCSKRMTITEDRSLIL
DGKVKVAVVYYRVSLLRARSQNPPVAWEMRRMIERSTAIKCPTIGMLLASTKKVQQVLAM
PGMVEKFFPDEEDAGKVVAIRETFADLWGLENDDAETKAAIKDAIEHPERYVMKANKDGG
GFTYWNEQISEKLQTFTKAQLAEYILMQKFQPPKSKNYVIRALQEPEYGEVVNEISTFGY
AIGNFVEKKVMYNKAGGYLMRTKWAHVNEGGILIGTGVYDTPYLTV
>Rre_GS53
MRTPESKNRSGIAAVAPVTLFPTQVPREIYEQATAVQEAMTLLYFRVSRDYKFLVKVLGE
VAKTDKTINFLMNILEQVQREGIKQPYNMIWQRADYMIHEEDGDGTNIGKHYQLKQIEVN
NGAAAGYIAEHVTEFHQRMLSKAGIAASLPENPVFNTVSNAIYSAWKQFDDPNAIIIVMV
HKNGALTHFVDHEILYHLEELGKGELEIVYMNVVDAYEKLTMADDFSLRFGDKKVGVVFY
RVTFLRPWDFFPEKAWELRRTIERSTAIKCPTVGTLLASTKKVQQVLAQPNMVEHFFPDP
KDEDKVKSIRATFAGLWGLENDDAKTKHIIKDAIEHPDRYVLKPNREGGGNNIWGEEIAE
KLQSFTADQRAEHILMQRLNPKLSKNFFIRPMEEPEYGDVSVELGTFGYIVGNSNDGTVL
HSGANGYFLRTKWADTNECGINVGSGVYDTPLLI
>Rre_GS54
MRTPGSKNRSDIAAVAPVTLFPTQVPREIYEQATAVQEAMTLLYFRVSRDYKFLKKILGE
VAKTDRTINFLMDVLEQVQREGVKQHYNMIWQRADYMIHEEDDDDTDKGKHYQLKQIEVN
NGAAAGYIAEHVTEFHQRMLSKAGIAASLPENPVFDTVSNAIYSAWKQFDDPNAIIIVMV
HKNGALTHFVDHEILYHLEKLGKDELEIVYMNVVEAYEKRLTLADDSSLRFGDKKVGVVF
YRVTFLRPWDFFPEQAWEVRRTIERSTAIKCPTVGTLLASTKKVQQVLAQPNVVEHFFPD
PKDADKVKAIRATFAGLWGLEEDDAKTKHIIQDAIAHPERYVLKPNREGGGNNIWGEEIA
EQLQSFTPQQRAEHILMQRLQPKLSKNFFIRPMEEPEYGDVAVELGTFGYIVGNSSDGTV
IHSGAKGYFLRTKWADTNECGINVGSGVYDTPLLI
>Rre_GS55
MLMFLSKSCLLFSFLIIATDIALGTEDASTDQLLDTKDLKVLVEEANDRAHHIGMIMRQR
DARDRSDVAEVAPYTIFPSPFPRVLFQEAINVQEAMTLLYFRVAMDHEFLKEQLKEVSET

DETIKRLISIMDDVREEIGEKGIHQPLALMWQRADYMTHKHIDDEGQLHHELKQIEVNNG
AAGGFIGYYATELHRRMLSKAGIDTDYLPENNPLNTLGKGLYEAWLKFGNPNAIVAVVEP
RGGSLTYFADHKIAQELDRISGGKIKTEFVHWQSCLERMTLADDFSLMLDDKVVAVIFYR
VTFLSPIEKIPPESWAVRRLIERSTAIKSPTIGMLLASTKKIQQVLAMPNMVERFFQDPK
DSDKVKAIRATFTGLWGLEHGDEKTKAVIADAKAHPENYVLKPNREAGGYNIWGNDIVDK
LNAFTPVERAQHILMQKLNPIVTKNFFVRPLKEPEYGDVVTEFSPFGVILGNVQDGNVLY
QNAHGHFMRTKWVWANEGGIMKGTGVYDSPLLV
>Rre_GS56
MMLMFLSKSCLLFSFLVIAIDIALGTKDSSTDQLLDTNDLKVLVEEANDWAHNIGMIMRQ
RDARDRDVAEVAPYTIFPSPFPRALFQEATDVQEAMTLLYFRVAMDHKFLKEQLKEVSEI
DETIKRLISIMDDVREEIGEKGIHQPLALMWQRADYMTHKHIDDEGKLHHELKQIEVNNG
AAGGFIGYYAMELHRRMLSKAGIDTDNLPENNPLNTLRKGLYEAWLKFGNPNAIVAVVEP
RGGSLTYFTDHKIAQELDRISGGKIKTEFVHWQSCLERMRLADDFSLMLDDKVVAVIFYR
VTFLSPIEKIPPESWAVRRLIERSTAIKSPTIGMLLASTKKIQQVILAMPNMVERFFQDP
KDADKVKAIRATFTGLWGLEHGDEKTKTVIADAKAHPENYVLKPNREAGGYNIWGHDIVD
KLNAFTPVERAQYILMQKLNPIVTKNFFVRPLKEPEYGEVVTEFSPFGVLLGNVQDGTVL
YQNAHGHFMRTKWAWANEGGIMKGTGVYDSPLLV
>Rre_GS57
MMLIFLSKSCLLFSFLVIAIDIALGTKDASTDQLLDTNDLKVLVEEANDWAHNIGMIMRQ
RDARDRSDVAEVAPYTIFPSPFPRELFKEATEVQEAMLLLYFRVAMDHEFLKEQLKEVSE
TDETIKRLISIMDDVREEIGEKGIHQSLALMWQRADYMTHKHIDDEGQLHHELKQIEVSN
GAAGGFIGYYATELHRRMLSKAGIDTDYLPENNPLNTLGKGLYEAWLKFGNPNAIVAVVE
PRGGSLTYFADHKIAKELNRISDGKIKTEFVHWQSCLERMRLADDFSLMLDDKVVAVIFY
RVTFLSPIEKIPPESWAVRRLIERSTAIKSPTIGMLLASTKKIQQVLAMPNMVERFFQDP
KDADKVKAIRATFTGLWGLEHDDDKTQAVIADAKVHPENYVLKPNREAGGYNIWGNDIVD
KLNAFTPVERAQHILMQKLNPIVTKNFFVRPLKEPEYGEVVTEFSPFGVLLGNVQDGTVL
YQNAHGHFMRTKWAWANEGGIMKGTGVYDSPLLV
>Rre_GS58
MLNLLANNFLVIIVVAMAFPIIVAANPISKEVQNTRTKQDQETKNSDVNEEEIKLLWLDA
LDYAHNIGLIIRTNDQPHRSNLSGITPVTLFPTTVPRKMFEKANGVREAMALLYFRVARN
YDFLNEVLGEAAKSDYSTRELLSIVKNVQEEGVHQPIALDYMFNEIIDADTKEKDYELKQ
IEVNNGPVGGLVVEHATKLHRRMLELSNMDAGEDVLPENRAYETIAEGLYRAWKAFDDEE
AIVIMIVGRIKNPFQYEQRQLEYMLEEMSGGKLKIWRLNLFQCDEKLKLGEDFSLMLDDK
KVGLVYFRLNLLIPERLQTTEGLRVRRLIERSTAIKSPPMSLELATTKKVQQYLAKPGML
ERFFTPAEADMVTAIRSTFAGLWGLDQNDEKTEAIIQDAIDHPENYVLKPNREGGGHNFW
GEEIAEKLKTFTPTDRVEHILQQRLHPPVTQNYLMKQLAEPKLENVVTELSTYCALLGNF
EDGTVLYNKGYGHLMRTKIESVTEGGIMEGSGYYDTPYLID
>Rre_GS59
MSKFQPIFVTFIVVLLRCYGDSTPDSDKDAINVSSIEDEVDLKILADDAIDFAQNNGLII
RTNDHPTESDISAFAAFTLFPTQFPRKQFHQAYDVQEAMSLLYFRISRDYDFLVKIASEI
TKNDYAVEKMLEIVQKIHEEAKLGKINQPISLVLQRSDYMCHMNPKAQGKEDQYQLKQIE
VNNGPIGAILVERVRKLHQRMLAKANMDGGSMLPENRPFNTIAEGIYLAWQQFKNPNAIV
VTIIGSKRNRFRFEQAQLEYELERISGNKIKNIVYMNMNEAHESLRLAKDNSLMLGDRVV
GVVYFRRGFLIKPHPLADQQFVTRLLIERSTAIKSPTVALELASMKKIQQVLAKPNMVEQ
FFPDPKDADKVAVIRATFANLWGLEKEDEETEAVIQDAIAHPENYVLKPNREGGGHNYWG
HEISEKLSAFSMTDRKEHILMERLRPFVAQNYPIRAGGDVRLENIVTEFSTYGYLVGNIQ
DGEVLYNKGHGHLMRTKIESVTEGGILEGSGFYDSPYLID
>Rre_GS60
MSKFLPIFVAIIVVTLICYCDSTTDSAKDAINVSSIEDEVDLKILADDAIDFAQNNGLII
RTNDHPTESDISAFAAFTLFPTQFPRKQFHQAYDVQEAMSLLYFRISRDYDFLIKIASEI
TKNDYAVEKMLEIVQTIHEEAKLGKINQPISLVLQRSDYMCHMNPKAQGKEDQYQLKQIE
VNNGPIGAILVERVRKLHQRMLAKANMDGGSMLPENRPFNTIAEGIYLAWQQFKNPNAIV
VTIIGSKRNRFRFEQAQLEYELERISENKIKNIVYMNMNEAHESLRLAEDNSLMLGDRVV
GVVYFRRGFLIKPHPLADQQFVTRLLIERSTAIKSPTVALELASMKKIQQVLAKPNMVEH
FFPNPEDADKMAAIRATFAKLWGLEKEDAETETVIQDAIAHPENYVLKPNREGGGHNYWG
HEISEKLRVFSMTDRKEHILMERLRPFVAQNYPIRAGGDVRLENIVTEFSTYGYLVGNIE
DGEVLYNKGHGHLMRTKIEGVTEGGILEGSGFYDSPYLID
>Rre_GS61
MRILIADTMDWANKVDSSIFNFSNEFFPKVGFVMRTKEHPDRSDYTQVGPFTLYPSPVPR
RLFFEAQAVQEALSLLYFRVASNHEFLMRALGKVAKGDETIRRLLTIVEEVHRDGVVRQP
ISLMPIRADYMIHVNDAEGNQENDCNEASDDRKYELKQIEVNIGAVGGFLAECVTPIHQR

VLAKSGRSAKFIKKRLPPNNIYHTMADGIFLAWQKFDNPKAIVLMVVPQLGGPVNRELDY
RMEELSDGKMVVVEQTLPECIEKLKLGKDNELLLEGLVVAVVNYRLARTIHPKFITDEKL
RLWLMVERSTAIVMKSPSITLDLASTKKVQQVLAQPGEVERFFADPEDADKVAAIRATYA
GLWSLDESNEKSKQIMQDAIEHPERYVLKPNRDGGGYNYWDEEIPQMLNKMTPDERTGFI
LMQRLNPFVTQNYLMKLLTEPQVEDVVTELSIFGFLLGNQVDGTVDANRVGGHMMRTKPE
HVREGGTSRGNGFYDSPYLF
>Rre_GS62
MFMHVIFFTLTIISHNKSAYADFGKDDNIDVKPLAEDAVDMAQNIGMKMRTKEHLDKSDV
AKFVPFTLFPTPVPRELYEQALKVQEAFAELYFRIASDLDFLTKTMAEVAKSDMIIRILL
SLVKKAHKEGIRQPVGLMQIRSDYMIHVNGNQQNELKQIEVNIGSIGGFGVDKTSTIHRR
MVSKAAGMDVSEEVLPTNKVVDTLAEGLFKAWQHFGDPKAVVLMVAGRRDPLHFDEATME
YKLEQLSGGQIRCFRLNLIDCHENLKLGDDFSLSLGGYTVGVVYYLVLRTGNERFITQEV
VDVWRMIEASTAIKSPTIAMDLASTKKIQQVLAQPGVVEQFFPDPKDADKVKAIRQVQAG
LWALDRDDENTRKVIEDAIEHPDRYVVKANRDGGGNNLWDQEMAQKLREWTPLERSRFIL
MERLRPLVVQNYVVVPSQEPRLESVVTELGIHGALLGDERSGKVLHNRAGGHLMRTKPAG
SVEGGISEGSGFFDSPLLY
>Rre_GS63
MFMHVIFFTLTIISHNKSAYADFGKSDVAKFVPFTLFPTPVPRELYEQALKVQEAFAELY
FRIASDLDFLTKTMAEVAKSDMIIRILLSLVKKAHKEGIRQPVGLMQIRSDYMIHVNGNQ
QNELKQIEVNIGSIGGFGVDKTSTIHRRMVSKAAGMDVSEEVLPTNKVVDTLAEGLFKAW
QHFGDPKAVVLMVAGRRDPLHFDEATMEYKLEQLSGGQIRCFRLNLIDCHENLKLGDDFS
LRLGDHTVAVVYYLILRTGNERFLTQEVVEVWRMIEASTAIKSPTIAMDLASTKKIQQVL
AQPGVVEQFFPDPKDADKVRAIRQVQAGLWALDRDDENTRKVIKDAIEHPDRYVVKANRD
GGGNNLWDQEMAQKLREWTPLERSRFILMERLRPFVVQNYVVVPSQEPRLESVVTELGIH
GALLGDERSGKVLHNRAGGHLMRTKPAGSVEGGISEGSGFFDSPLLY
>Rre_GS64
MVILCYFLFALISINGQAEETTGIDAENKKDKPIADGLDLPSLAADAIDWAQNAGIKMRV
KGKLTSSDVASFVPLTLFPTPYPRQLFESAYNVEEAMMTLYFRVASDYEFLSNSLTELAE
QDETVRKLLGIYQQSQQGTIQQPNGLMLMRSDYFCHLNEKDEPELKQIEVNIGTIGGYNI
EQLPKLHERMLAKAGMPASVDRLPVNNMQDTSAEGLYQAWLKFGNPEAALLIVINAGGDP
FHSDEPLIQYKLEELSNGRMKVFMLTIMECYRRLKLADDFSLHLEDYVVGLVWYRVTRAQ
SPRMLTEEKLDIWLKIEKSTAIKSPTLGMELASMKKIQQILAQPGTVEQFFPDPKDAEKV
AAIRRTYAGLWGLEGEDAESLIEDAIEHPDRYVLKPNREGGGYNFWDDKMVEKLRKLDPK
ERGQFILMQRLRPMVHPNYVLRPNADVQCQLENVVAELGMMGYLLGDASAKTVHGTRTGG
HLIRTKMSESREGGLTIGAGSYDSPFLY
>Rre_GS65
MELLFLFALTFFCGQAKEMEGVGAENDKPTDELDLRMLAEDAIDWAQNVGIKMRIKEQNN
SDVAAFAPFTLFPTPVPRQLFESAHEVQEAMMCLYFRVASDYEFLSNSLAELAKADEAVE
KLLWLFQQSHQDTVQQPNGLMFMRSDYFCHLNEKDEPELKQVEANIGAIGGFGVGHLTSL
HHRMLTKAGLPVSADRLPANKVHDTMAEGLYQAWLKFGNPKAVLLIVARFGGYDPLHFDE
PLILYKLEELSEGQIKVFMLSNLECYQRLKLADDFSLHLDDYVVGVVWYRATRAVHPKLL
TQQKLDIWLKMEKSTAIKCPTIGMELAGMKKIQQILAQPGTVERFFPDTEDGRNKVAAIR
QTYAGLWALEGEAAENIVKDAIEHPDRYVLKPNREGGGYNFWDNELVEKLRTMSSTQRGQ
FILMQRLRPMVHQNYVLKPKADGARLENVVTELGIIGFLLGDASAKVVHGTRVGGHLMRT
KLEQSREGGIVIGTGFYDTPYLV
>Rre_GS66
MSIQKQIEPKFLETVDMKVLVEDAQDWSHAVGLIQRTKDHAKISDVAEVPPFALFPSPLP
RKIFYQAVAVQETLAELYFKIASDYKFLVDTYREVRKADKTIAILFLDILDDVRKKGTHQ
PIGLMMMRADYMCHAEEYESAEHYELKQIEVNIGAVGGSTCEAATLLHRHVLSKAGLQMV
VLPENRTTDTLSEGLYQAWKAFGNDNAIIVTIIGKKGNKTHFEMRATEYKLEELSDGKIR
SVYLNLTEANEKLKLADDFTLLLDDKDIVAVVNHRLARLIYSWFIDEEKIAVWTKIEQST
AIKSPSLSMELSGTKKMQQLLAEDGIVERFFPGPEDAQKVRAIRQFQARMWSLENDDEET
NAIIKDAILHPERYVLKPNRDGGGNNFWEDDIRNKLQKMEPIERNQFILMQRLRPLITKN
FLKCPMEEVRYEDGVVTELSIFGTLLGNQENSKILHNLVGGHMMRTKPRHINEGGVQNGA
GFLDSPLLF
>Rre_GS67
MAILLNICLVCCCYYCIFGETSGQQDIEVQVLVEDALDYGHYVGLIHRANDHLKSSDLSE
VSAMALFPSPFPRQVFEDANNVQEALAELYFRVANDYEFLMNAYREVRKVDKTVDKLLSL
LDDIRKKGIHQPIGLMMMRADYMANMNEQNSESPYEIKQIEVNIGAVGGATCEKATLVHR
RVLAKAGITSVVLPDNNATDTLAMGMYQAWKAFGNENAIIVTIIGKLGQKTQYEMRKAEY
KATQLSGGKIRTVCMNLTEANEKLTLDADFNLRLDDLIVAVVNYRLARNIHEKFLTDEKM

EVWTKMEVSTAIKSPPLNYEIACTKKMQQVLAEKDVVEKFFPEPKDAKKVAAIRKFQARM
WSLDHNDEKTQAVIQDAIEHPDRYVLKPNKDGGGNNLWEEEMKIKLEALKPEERSQYILM
QRIRPFVGKNYLKRPLEQARYEDQVVTELSIFGALLGNQENGKILHNKGGGHMMRSKPKH
VNEGGLEMGAGFYDSPLLI
>Rre_GS68
MAILLNICIACCCFYCVFGETSGQQEIDVQVLVEDALDYGHYVGLIHRAKDHLKSSDLSE
VSAMALFPSPFPRQVFEDANNVQEALAELYFRVANDYEFLMNAYREVRKVDKTVDKLMNL
LEDIRKKGIHQPIGLMMMRADYMANMNEQNSESPYEIKQIEVNIGAVGGATCEKATLVHR
RVLAKAGMTSVVLPDNNATDTLAMGMYQAWKAFNNEQAIIVTIIGKLGQKTQYEMRKAEY
KATELSGGKIRTVCMNLTEANEKLTLDDNFNLRLDDQIVAVVNYRLARNIPEKFLTDEKI
DVWTKMEVSTAIKSPPLNYEIACTKKMQQVLAEKDVVEKFFLEPKDAKKVAAIRKFQARM
WSLDHNDEKTQAVIQDAIEHPDRYVLKPNKDGGGNNLWEEEMKIKLETLKPEERSQYILM
QRIRPFVGKNFLKRPLEQARYEDQVVTELSIFGALLGNQENGKILHNKGGGHMMRSKPKH
VNEGGLEMGAGFYDSPLLI
>Rre_GS69
MTKINIILLFFFLWHTCFAKIVTENGEVTNDGSATDVLVLMDQARDWAQHLGMHDRSKNH
PQKSDVAELPPFALLPTAIPRKMYEQAYAVQEPMAELYFKVASDYKFLADCFRDVRKVDQ
TMSKLMDLLDELHQEEMRQPITLLVSRSDYMFHTCEDCDESQRHQLKQIEMNVGPVGGSL
SQKTTLLHQRIFRKAGIDPKFLPDNKPNDTLAEGLYEAWKSFGDPNAIFLSVLSKRNISH
YEMRDIEYRLEELSDDKIRIIHLSPVEAYEQLRLADDFKLMLGDNVVAVVHFKHARVLDP
KFLHPKQISLWRMVERSTAIKCPTIGMDLASSKKVQQVLAQDGVLERFFPDAKDAHKIAA
IRQIQAGLWPLDHIDAKTEALIRDAISHPEDYVLKPNREAGGNNFWDEALKHLLERMEPD
QLGAFVLMQRLRPLVTKNFLVKPMEDGARFEEVVTELGVFGALLGSQQTGQVLYNRVGGH
LMRTKPKDVKEGGVDHGAGFFDSPILY
>Rre_GS70
MLQINVLLLFFFLCHVCFAKTITENGEVAVANEGSATDVLVLVDQARDWAQRVGMHDRSK
NHPQTSDAAELPPFALFPTAIPRKMYEQAYAVQEALAELYFKVASDYKFLAECFRDVRKV
DQTMSKLMDLLDELHQEELRQPITLLVTRAIEVNVGGIGGPLSQKTTLLHQRILRKAGID
PKLLPDNKPNDTLAEGLYQAWKSFGDPNAIFLSVVSKKRNIPHYEMRDIEYRLEELSDDK
IRIVHLSPDEAYEQLRLTDDFKLMLDDNVVAVMHYKHARAIHPKFLTEERIGLWRMIERS
TAIKCPTIGMDLADSKKVQQVLAQDGVLEQFFPDAKDAHKIAAIRQTQAGLWPLDHIDAK
TEALIKDAISHPERYVLKPNREGGGNNFWDEALKAQLEQKQMEPDQLGAFVLMQRLRPLV
TKNFLVKSMEDGARFEEVVTELSVFGALLGNQQTGKVLYNRVGGHLMRTKPKDVKEGGVD
HGAGFFDSPILY
>Rre_GS71
MPQINLLLTILFLYDACMMTSGDGTPRGNDGSTPANVEVLVEQARSWAQHVGLHYRYKKF
LLRSYFAVLPPFTLFPTAIPRKMFEEAYAVQEALAELYFKVASDYKFLSECFKDVRKADE
TISKLMGLLDELQQEEIRQPIALQLMRSDYMFHTCEDCEVSQPTQLKQVEMNIGVGGSIS
EKATLLHQRIMAKAGIDPKLLPENKPNDTWAEGFHKAWQQFGDPNAIFLIIINQKYDLAH
YEMRDLDYRLEELSDHKIRIVHLSPNEAYEQLRLANDHKLMLDDNVVGVVHFSTARLINP
KFLTEKRIDLWRMVERSNAIKNPTIGMDLADSKKVQQALAEDGIVEKFFPDPKDAHKVEA
IRKIQAGMWPLDRVDEKTVVIIKDAISHPGRYVLKPNREGGGHNYWDDELKSLLEKTKGE
ELGKFILMQRLRPLVTKNLLVQPMEADARLEEVVTELGVFGSLLGNLKTGKVLYNRGGGH
LMRTKPKDIKEGGVFHGTGFFDSPILY
>Rre_GS72
MPKINLLLTFLVLYIACLMTSGDGTPRGNDGSTPANVEVLVEQARSWAQHVGLHYRYKEH
LATSDFAVLPPFTLFPTAIPRKMFEEAYAVQEALAELYFKVASDYKFLTECFKEVRKADE
TISKLMGLLDELQQEEIRQPIALQLMRSDYMFHTCEDCEVSQPTQLKQVEMNMGVGGSIS
EKATLLHQRIMTKAGIDPKLLPENKPNDTWAEGFHKAWKQFGDPNAIFLIIINQKYDIAH
YEMRDLDYRLEELSGHKIRIIHLSPNEAYEQLRLADDHKLMLDDNVVGVVHFSTARLINP
KFLTEKRIDLWRMVERSNAIKNPTIGMDLADSKKVQQALAEDGIVEKFFPDPKDAHKVEA
IRKIQAGMWSLDRVDEKTDAIIKDAISHPGRYVLKPNREGGGHNYWDDELKSLLEKTKGE
ELGKFILMQRLRPLVTKNLLVRPLEEDARLEEVVTELGVFGSLLGNLKTGKVLYNRGGGH
LMRTKQKDIKEGGVFHGTGFFDSPILY
>Rre_GS73
MPKINLLLTFLFLYIACLMTSGDGTPRGNDGSTPANVEVLVEQARSWAQHVGLHYRYKEH
LATSDFAVLPPFTLFPTAIPRKMFEEAYAVQEALAELYFKVASDYKFLSECFKEVRKADE
TISKLMGLLDELQQEEIRQPITLQLMRSDYMFHTCEDCEESQPTQLKQVEMNMGVGGSIS
EKATLLHQRIMTKAGIDPKLLPENKPNDTWAEGFHKAWQQFGDPNAIFLIIINQKYDIAH
YEMRDLDYRLEELSDHKIRIVHLSPNEAYEQLRLADDHKLMLDDNVVGVVHFSTARLINP
KFLTEKRIDLWRMVERSNAIKNPTIGMDLADSKKVQQALAEDGIVEKFFPDPKDAHKVEA

IRKIQAGMWPLDRVDEKTDAIIKDAISHPGRYVLKPNREGGGHNYWDDELKSLLEKTKGE
ELGKFILMQRLRPLVTKNLLVQPLEEDARLEEVVTELGVFGSLLGNLKTGKVLYNRGGGH
LMRTKQKDIKEGGVFHGTGFFDSPILY

Supplementary figure 2

>NP_496011.1 Glutathione synthetase [Caenorhabditis elegans]
MAQKDDRILLLNAPRLPLEDDKLNELTADLHDWAHANGLVMRLSTDKLSSEVCQTTPLTLLPSPFPKNVF
EEAVHIQNLFASLYHFIAYEFDFLIDIHKNVVKTDDFTRNMVEILKKVKAQGLKQPVTLAIQRSDYMCHK
DQYSAEYGLKQIEINNIASSMGAHALRLTEWHIRVLKALNISDDVIQRAIPENKPIPMIAEALFKAWSHF
SNPAAVVLVVVENVNQNQIDQRHVEYELEKLGVPMTCIIRRNLTQCYEQLSLNDRSDLMIDGRQVAIVYF
RAGYSPDHYPSTKEWEARERMELSTAIKTPWIGLQVANTKKTQQVLSEDGVLERFIGKPREARDIRASFA
GMWALENTDEVTMKVVAGAQKHPEAFVLKPQTEGGAALHTGDEMVQMLRELPEEERGAFILMEKLKPMII
ENYLVLAKKPITFAKAVSELGVYGYAFGRKDAPELKTAGHLLRTKPESTAMGGVAAGHAVVDTPFLYEFI

>XP_024508894.1 Glutathione synthetase [Strongyloides ratti]
MNLYEEFSTIKNNNPKLFNYIIDEAKDFAVVHGNILRLSNSKDSSDIVQHAPITLLPSPFPKNLYIQALK
VQPIMNKLYFKISLDYEFLKSSLKCVIDTDDFTKKLFNIYDIVMKESLKQPITLLLQRSDYFCHVNTSNP
NKIVYELKQIEVNNIAAGMASLSQITTQMHKHILRNFSNIYNEDNHPSNRGNDTVGKGLAEAWKMYGNTN
AYVLQLVEDTNKNQLDHRHIQYATDYFTSYKCKTIRIPLSECRNRLKLGKNYELIMDDIYEIGVVYFRTG
YSPENYIDEEDWNSRLLIEKSCAIKSPWIGLQLANTKRIQQVLIENGVVEKFLNDIEEIKDVKKTFAMMW
SLNDNMDEIKKKVINNCKDYVLKEQLEGGAGNYFDDDIIEKMNDVEKLKSCILMERLNPMKSKNYIMVSG
KEYEESEVVSELGIMGTYLGNITTNQEYFNYSGGYLLRTKKSNETKGGVTIGASSIDSIYLI

>XP_024500287.1 Glutathione synthetase [Strongyloides ratti]
MEKIISIEKKSLIEYAKDFALLNGLSMRTREHPDSSDTIEHAPFTLFPSYFDETSFYYVKNLQKHIQMLY
YKVSQDIPFLIETHKDIVNQDNLIKGLCNVLIKSSIDPSPQQFNLILQRSDYMPHVNSNNKIEIKQVEVN
NIAVSMGGLGNAIENLHRNILNIFFKDVILNQEDIYLPEKSNPAKLCAEGLVTALKYYHIIDYNDNPKRG
MLLYGNSLPLNCASILTVTEDVSRNIFDQRHIEAEIQYLTNYNVKNFRIPLSQLNSRLTLDENKKLFLDK
KYEIGLVYYRTGYSIDQYNDETNDWDTRLLIEKSAAIKCPSIGLQLANTKKMQQVLANKCVLKKYVYDDE
VVEKIFKSFAKLWPLGGDSDEEKLVIKDAKSNPDKYVLKPQTEGGGGNFFGKDIPNLLNSLSKTELKCYI
LMEKLHPTPTENFLIRPNQKVEKSQVVSELGIYGWLIADKKNIFPNRNSTYYSYMMRTKESTTNEGGICV
GAACLDSIIFKNFEKDFYTNFI

>XP_001892534.1 glutathione synthetase family protein [Brugia malayi]
MDEMPIRNNDDSVLKYYPKLELIDGKLIKQLVEDTVDWAHAHGMVMRTAMTTDRSDICQTAPCTLFPSPF
PYNLFQEAMDIQRTFSLLYFRISWDFDFLIKSHAEVVKTDDFTRHFVEILNAVXTSNFCQKKTLLIQRND
YMCHEDSYGNRSLKQIEVNNIAASMGXLAERATCVHKRTLETVQLPSKIIEKAIXDNHPTVTIARGIYEA
WYDFGVPEAIVLFVVEDANRNQIDQRHVEYCIDELSNRNARCLRITLTDGAKRLKLNESNHHLVLDNILR
VAVVYFRAGYSPSNYPTEMEWTARRIIELSDAVKCPWIGLQLANTKKVQQVLSENGVLEKYITDDKMCAR
IRQTFAGMWGLENDDEKTQRIIQDAIAHPEKFVLKPQLEGGGGNYYGKEVAEKLKTMNRDEMAAYIIMER
ITPMVVKNYVIRPQEEPLLMDVVGELGVYAYLYGSAAVDNIIVENIMKNHVSGHIIRSKDKSVDKGGVAI
GAAVIDSPYLF

>XP_003142623.1 glutathione synthetase [Loa loa]
MDETSVRSENDSVLEYYPKLELIDRKLIKQLVEDTVDWAHAHGMVMKTETAADRGDICQIAPCTLFPSPF
PYSLFQEAMDIQQAFSLLYFRISWDFDFLIKSHAEVVKTDDFTKHFVEILNAVRTSDFCQRKTLMIQRND
YMCHEDNYGNRSLKQIEVNNIAASMGSLAERATCVHRRTLETLQLPNKIIEKAILNNHPTVTVAKGICEA
WYDYGVPEAIILFVVEDANQNQIDQRHVEYCIDELSSRSIRCLRITLTDGAKRLKVNETNHYLVLDNMLR
VAVVYFRAGYSPNNYPTEAEWTARRIIELSDAVKCPWIGLQLANTKKIQQVLSENGVLERYITDGRMSTR
IRKTFAGMWGLENDDDRTQKIIQDAVTHPEKFVLKPQLEGGGGNYYGKEVAEKLKTMSRDEMAAYIIMER
ITPMVVKNYVIRPRQEPVLMDVVGELGIYAYLYGSPAVDYIPAENVITNYVSGHIIRSKDKNVDKGGVAI
GAAVIDSPYLF

>XP_002630028.1 Hypothetical protein CBG13395 [Caenorhabditis briggsae]
MAQKDGRVLLLNAPRLPLPEDKVKELAGDLHDYAHAHGLVMRLSTDKMSSEVCQTTPLTLLPSPFPKNVF
DEAVRIQNLYASLYHYIAYDFDFLIDIHKNVVKTDEFTRKMIEILKKVKEQGLKQPVTLAIQRSDYMCHK
DQSSAEYDLKQIEINNIASSMGAHAERLTQWHIRVLKALDVPDDVIRRAIPENKPIPMIAEALFKAWSHF
NNPDAWVLVVVENVNQNQIDQRHVEYELEKLGVPMTRIIRRTLTQCYEQLSLNESSELLIDGKPVAIVYF
RAGYSPDHYQSDKEWEARERMELSTALKTPWIGLQVANTKKTQQVLSEDGVLERFIGKPREARDIRRSFA
GMWALENSDEVTLKVVQGAQKHPEAFVLKPQTEGGAALHTGDEMVQMLKELPEEERGAYILMEKLRPMII
ENYLVFARKPVTFAKAVSELGIYGYAFGAKDAPELKTAGHLLRSKPESTAMGGVAAGYAVVDTPFLYEFI

>XP_003117464.1 hypothetical protein CRE_02056 [Caenorhabditis remanei]
MAQKDDRILLLNAPRLPLPDDKVAELIGDLHDYAHAHGLVMRLANDKMSSAVCQTTPLTLLPSPFPKNVF

DDAVRIQNLYASLYHYIAYDFDFLIDIHKNVVKTDEFTRNMVEILKKVKAQGLKQPITLAIQRSDYMCHK
DQFSAEYGLKQIEINNIASSMGAHAQRLTDWHIRVLKALEVPDDVIKRAIPENKPIAMIAEALFKAWSHF
NNPSAWVLVVVENVNQNQIDQRHVEYELEKLGVPMTCIIRRTLTQCYEQLSLNESSDLIIDGRPVAIVYF
RAGYSPDHYPTNKEWKARERMELSTAIKTPWIGLQVANTKKTQQVLSEDGVLERFVGKPREARDIRTSFA
GMWALENKDEITLKVVQGAQKHPDAFVLKPQTEGGAALHTGDEMVQMLQELPEEERGAYILMEKLRPMII
ENYLVVAQKPVAFAKAVSELGIYGYAFGAKDAPELKTAGHLLRTKPETTAMGGVAAGYAVVDTPFLYEFI

>XP_003372705.1 putative caspase recruitment domain protein [Trichinella spiralis]
MDDWQRQALDSNLIALADSLDLIELLPFLQQKGILREYHVDAIRKKTPYEARLEFVSIIKRRGPNAFEAL
CEGLARSGQTHLEQALRSCKSPNQPAEINNGSGGGGQLVECPEELNSNPQETKLQLVKTSAEQYYFAESI
LFSDVPMTDGLLSWLKVKNEGTIPADYNTVDCYKHSSKPKGLALIINNCAFTCDLPNRLGSDIDCKNIYR
LLTDLGYDVIKKHNLTAKGIVETMVKFSLNKEHERCDSAVVVILTHGLEGEIYGSDGGLVSVQKMIQLLD
AVNCPALKNKPKLFFLQACRGQRYDSGHDVIDSGDAIGSANARCKLNSLDENDAAALRRKVPTQADILIA
YSTTPGFVSWRNSLRGTWFIQAVCEVFSEYAWKEEVLHLFTRIHGLCIRLPGKYENSDEVQLAPFTLLPS
PFSKSCFDTAVNLHQAMMAVYHQIAFDYDFMEDALSPVMLSDVFVSKLFGIYRAVMKHGNMMSRLVLNIQ
RCDYMMQEEPNSSYSLKQVEVNHVAASFAGLGVRCRQWHRLMLSQMVENAKLDLLPENHALETVALGLLS
AWRAYGHSEAIILIMVEDELRNFADQRLVEYEIQSLAEFQHVPMRRFRLTDCPGNVAVDRRGRLMLNGVE
VAVVYYRTGYLPKHFPNEDVWSAFLQIELSEAIKCPWIGFHLAGMKRMQLLLSNSDTLRSVIDKTKLLVW
PVNLKQKILDDDDDEMYRQMKSVTVPMHDLDASKAGADELMKHVDANADNFVLKPHLEGGGNNFYGQNLI
KQLNQLCSNERSAYVLMERIRPPTFDNWIIRANIPAEQTSVVSELGTFGYSLANGQQIISCSRDGGFLLR
TKPSKEDEGGVMVGAAALDTPHLLLP

>EPB68233.1 putative glutathione synthase [Ancylostoma ceylanicum]
MAIHVKVAVSSRLRENIEVNNIASSMGAHAERVTKMHRRTMTELGYDKETIEKAIPKNEPIKLIAEALYK
AWELYSSSSAVILIVVEDQNQNQIDQKHVEYALEDLGVPVDQIVRRTLTLSPERHLFLSGSRVAVVYFRA
GYTPDNYPTEKEWAARLLIERSDAIKSPWIGLQVANTKKTQQVLAEDGVVERFVGHPRDAAAIRSTFAGL
WAINGDDPNAIAHPSRFVLKPQLEGGGGNFYGEKMAEKLQNLGKDELGAFILMERIQPLVAENYLVRAMQ
PVELTKVVSELGVYGYALGDRGMPEVRQGGHLLRTKGEKVDEGGVAVGFAVIDSPFLYELL

>PIC44069.1 hypothetical protein B9Z55_004565 [Caenorhabditis nigoni]
MAQKDDRVLLLNAPRLPLPEDKVKELAGDLHDYAHAHGLVMRLSTDKMSSEVCQTTPLTLLPSPFPKNVF
DEAVRIQNLYASLYHYIAYDFDFLIDIHKNVVKTDEFTRKMIEILKKVKEQGLKQPVTLAIQRSDYMCHK
DQSSAEYDLKQIEINNIASSMGAHAERLTQWHIRVLKALEVPDDVIRRAIPENKPIPMIAEALFKAWSHF
NNPDAWVLVVVENVNQNQIDQRHVEYELEKLGVPMTRIIRRTLTQCYEQLSLNESSDLLIDGKPVAIVYF
RAGYSPDHYQSDKEWEARERMELSTAIKTPWIGLQVANTKKTQQVLSEDGVLERFIGKPREARDIRRSFA
GMWALENSDEVTLKVVQGAQKHPEAFVLKPQTEGGAALHTGDEMVQMLKELPEEERGAYILMEKLRPMII
ENYLVFAGKPVTFAKAVSELGIYGYAFGAKDAPELKTAGHLLRSKPESTAMGGVAAGYAVVDTPFLYEFI

>PAV62300.1 hypothetical protein WR25_08118 isoform C [Diploscapter pachys]
MTSGGPPEGILLADYPKVPLAEELITKLVEDTHDYAHANGLVMRTREANTSSDVCQTCPIALLPSPFPRK
IFQQAVQVQDIIAQLYHEIAYDYDFLLKCHENVIQTDPFTKGLVDILKAVKEQGLAQETTLAIQRSDYMC
HKDPFTNEYCLRQIEVNNIASSMGAHAERATRLHRRTFAQLGYDKEFIDKALPEDNQPIALIAEALFLAW
KSFDNKNAVVLVVIEDENMNQIDQRHVEFKLEELGVPVDQIVRKTLTQCYECLTLSPERHLLYSGSRIAI
VYFRSGYGPQHYHSDREWEARKRMELSDAIKTPWIGLQVANTKKTQQVLAEDGVVERYIGDPRQAASIRA
TFAGLWSIEGNDPLTRKMVQGAISHPSQFVLKPQLEGGGGNFYSESMVNKLQILKPEERAAFILMERIHP
MRIEVGSTFSHFYCYIVILVKMF

>KHN87278.1 Glutathione synthetase [Toxocara canis]
MAMDAAGKSGRSHTSDTSSRYTFPVALDALSLKRMVEDAVDWAHGHGMVMRTPQHKDRSDVCQTAPFTLL
PSPFPRRIFQQAVDIQQATNLLYFRISWDYDFLIKSHADVVKTDDFTRHFVEILKRVHEAGVKQRKTLLI
QRADYMCDDRGDGEFRLRQVEVNNIAASMGWLSEMASRLHRRVLQDLNVPDDVIANALPQNRAIDTVAEG
IYDAWLDFGDQSALILFVVEEVNQNQLDQRHVEYRIDQLSSRRAKCIRLTLTQCAERLSLGGASGYDLMY
DARRRICIVYFRAGYLPDNYCSEREWNARLTMELSNAIKCPWIGLQLANTKKVQQVLACDGQLERFLPER
TADCDRVRATFAGLWGLENDDVQTQAIIKEAIEHPEQFVLKPQLEGGGGNYYGEEVAQKLREMSHDERAA
HILMERIQPMHVKNYLVRPFEDVSMGEVVGELGIYGCLYAEPALDAGNEKILKNVSHGHIIRSKAENVDK
GGVAIGAAVIDSPFLF

>OZC11777.1 glutathione synthase [Onchocerca flexuosa]
MDEVLLRNEDDSMLEYYPKLELIDRKLIKQLSDDAIDWAHAHGMVMRTAATANRSDICQTVPFTLFPSPF
PHSLFQEAMNIQQAFNFLYFRVSWDFDFLIKSHAEVTKTDDFIRHFVEILNAVRTSSFCQKKTLLIQRND
YMCHEDNHGNRSLKQIEVNNIAASMGSHAERVTCIHRRTLETLGIPNKIIKMVIPDNHPTITIAKGIYEA
WCDYGVPEAIILFVVEDVNQNQIDQRQVEYRIDELSSRNARCLRVTLTDGAKRLKLDETSHHLLLDNSLR
VAIVYFRAGYLPDNYPTEIEWTTRRIIELSDAVKCPWIGLQLANTKKIQQVIFRNCSQYLVLSEKDVLEK
YITDDNICARIRKTFVSLWGLESDDDKTLKIIQDAIAHPEKFVLKPQLEGGGGNYYGKDVAEKLKTMSRD
EMAAHIIMERITPMVVKNYLIRPQEEPVLMDVVGELGIYAYLYGSPAVSNTPEEHIMKNYVSGHIIRSKG
KNVDKGGIAVGAAVIDSPYLF

>ADY42284.1 Glutathione synthetase [Ascaris suum]
MNVDAERKQQQSDGSNIPNRYTFSLPLHELPLKDMLEDAVDWAHGCGMVMRTPQHKDRSDICQTAPFTLL
PSPFPRHIFQQAVDVQQATNLLYFRMSWDYEFLIESHAEVVKTDDFTRHFVDILKRVHEAGIKQTKTLLI
QRADYMCDGQRSDEFKLKQVEVNNIAASMGWLSEMASCLHRRVLQDLNVPDDIIANALPENRPIDTVAKG
IYDAWLDIGDQSALILFVVEEVNQNQVDQRHVEYRIDELSSRRAKCVRLTLTQCAERLSLGGRSGHDLML
DACRRVSIVYFRAGYSPDNYCSELEWNARLTMELSNAVKCPWIGLQLANTKKVQQVLACDGQLERFLPEL
KEDCERIRATFAGLWGLESDEEETQIILKEAIEYPERFVLKPQLEGGGGNYYGSEVAEKLKEMSRDERAA
HILMERIQPMRVKNYLVRPFEEVTLGEVVGELGIYGCLYAEPGFDRGCEKVYKNLAHGHIIRSKAANVDK
GGVAVGAAVIDSPFLF

>KFD57196.1 hypothetical protein M513_02081 [Trichuris suis]
MAFRIDSCGVVLTNDKLIDELLDDAKDWAATNGMCIRSSEQPWSSDVVEIMPFTLFPSPVPRKLFNEAMS
LQKVMNELYYRVAIDKEFITSCLADIAAADPFIAKLLDIYLAVMNSNDHEERIFLNIQRADYMFHKEEEN
INRSLVLKQVEVNNISAGLAAIGPVCTKLHERVMRKAEYACEAQYYCLPENSSENVVSVGLFEAWKVYGN
KKAAVIFMVEDHPRNIADQRLIEHQLERLSEYTAMVVRLKFSESPKRLLMVGSTLHLIPENVEIGVVYFR
TAYSPDQFDNDSVWSALRLIEMSSAIKCPWIGFHLAGIKKVQQTLSFPKNLNRFVQDAETRRRILSVTMP
MFGFDRSSSAADWETILSQVVKEPNDYVLKPSREGGGHNFYQDDMVELLRSCGPTERQAYILMKRIKPSV
HTNVFVKRNVESKLQACNSELGVFGYLLGNMRKIFHQRDGTFILRTKLHSENEGGLMHGTACVDSPFIYD
N

>KRZ48699.1 Glutathione synthetase [Trichinella nativa]
MDDWQRQALDSNLIALADSLDLIELLPFLQQKGILREYHVDAIRKKTPYEARLEFVSIIKRRGPNAFEAL
CEGLARSGQTHLEQALRSCKSPNQPVEINNGSGGGGQLMECPEELNSNPQETKLQLVKTSAEQYYFAESI
LFSDVPMTDGLLSWLKVKNEGTIPADYNTVDCYKHSSKPKGLALIINNCAFTCDLPNRLGSDIDCKNIYR
LLTDLGYDVIKKHNLTAKGIVETMVKFSLNKEHERCDSAVVVILTHGLEGEIYGSDGGLVSVQKMIQLLD
AVNCPALKNKPKLFFLQACRGQRYDSGHDVIDSGDAIGSANARCKLNSLDENDAAALRRKVPTQADILIA
YSTTPGFVSWRNSLRGTWFIQAVCEVFSEYAWKEEVLHLFTRMSTMFSKIKLPPGTTFDELIDSAKDWAQ
IHGLCIRLPEKYENSDEVQLAPFTLLPSPFSKSCFDTAVNLHQAMMAVYHQIAFDYDFMEDALSPVMLSD
VFVSKLFGIYRAVMKHGNMMSRLVLNIQRCDYMMQEEPNSSYSLKQVEVNHVAASFAGLGVRCRQWHRLM
LSQMVENAKLDLLPENHALETVALGLLSAWRAYGHSEAIILIMVEDELRNFADQRLVEYKIQSLAEFQHV
PMRRFRLTHCPGNVAVDRRGRLMLNGVEVAVVYYRTGYLPKHFPNEDVWSAFLQIELSEAIKCPWIGFHL
AGMKRMQLLLSNSDTLRSVIDKTKLLVWPVNMKRKILDDDDDEMYRQMKSVTVPMHDLDPSKAGADELMK
HVDANADNFVLKPHLEGGGNNFYGQNLIKQLNQLCSNERSAYVLMERIRPPTFDNWIIRANIPAKQTSVV
SELGTFGYSLATGQQIISCSRDGGFLLRTKPSKEDEGGVMVGAAALDTPHLLLP

>KRY14574.1 Glutathione synthetase [Trichinella patagoniensis]
MDDWQRQALDSNLIALADSLDLIELLPFLQQKGILREYHVDAIRKKTPYEARLEFVSIIKRRGPNAFEAL
CEGLARSGQTHLEQALRSCKSPNQPVEINNGSGGGGQLMECPEELNSNPQETKLQLVKTSAEQYYFAESI
LFSDVPMTDGLLSWLKVKNEGTIPADYNTVDCYKHSSKPKGLALIINNCAFTCDLPNRLGSDIDCKNIYR
LLTDLGYDVIKKHNLTAKGIVETMVKFSLNKEHERCDSAVVVILTHGLEGEIYGSDGGLVSVQKMIQLLD
AVNCPALKNKPKLFFLQACRGQRYDSGHDVIDSGDAIGSANARCKLNSLDENDAAALRRKVPTQADILIA
YSTTPGFVSWRNSLRGTWFIQAVCEVFSEYAWKEEVLHLFTRMSTMFSKIKLPPGITFDELIDSAKDWAQ
IHGLCIRLPGKYENSDEVQLAPFTLLPSPFSKSCFDTAVNLHQAMMAVYHQIAFDYDFMEDALSPVMLSD
VFVSKLFGIYRAVMKHGNMMSRLVLNIQRCDYMMQEEPNSSYSLKQVEVNHVAASFAGLGVRCRQWHRLM
LSQMVENAKLDLLPENHALETVALGLLSAWRAYGHSEAIILIMVEDELRNFADQRLVEYKIQSLAEFQHV
PMRRFRLTHCPGNVAVDRRGRLMLNGVEVAVVYYRTGYLPKHFPNEDVWSAFLQIELSEAIKCPWIGFHL
AGMKRMQLLLSNSDTLRSVIDKTKLLVWPVNMKRKILDDDDDEMYRQMKSVTVPMHDLDPSKAGADELMK
HVDANADNFVLKPHLEGGGNNFYGQNLIKQLNQLCSNERSAYVLMERIRPPTFDNWIIRANIPAEQTSVV
SELGTFGYSLATGQQIISCSRDGGFLLRTKPSKEDEGGVMVGAAALDTPHLLLP

>KRZ10293.1 Glutathione synthetase [Trichinella pseudospiralis]
MQYKHCFRPTDVTRTLNAMDDWQRQALDSNLIALADSLDLIELLPFLQQKGILREYHVDAIRKKTPYEAR
LEFVSIIKRRGPNAFEALCEGLARSGQTHLEQALRSCRSSNQPAEVNNGTGGGGGGGGQLMAWSEESNS
TSPQEMKSHLVTPSAEQYYFAESILFSDVPMTDGLLSWLKVKNEGTIPADYNTVDCYKHSSKPKGLALII
NNCAFTCDLPNRLGSDIDCKNIHRLLTDLGYNVIKKHNLTAKGIVETMVKFSLNKEHERCDSAVVVILTH
GLEGEIYGSDGGLVSVQKVIQLLDAVNCPALKNKPKLFFLQACRGQRYDSGHDVIDSGDTVGSANARRKL
NSLDENDANALRRKVPTQADILIAYSTTPGFVSWRNSLRGTWFIQAVCEVFSEYAWKEEVLHLFTRIHGL
CIRPPGKFENSDEVQLAPFTLLPSPFSKSCFDTAINLHQAMMAVYHQMAFDYDFMEEALSPVMLSDVFVS
KLFAIYRAVMKHGNMTSRLVLNIQRCDYMMQEEEANSSYSLKQVEVNHVAASFAGLGVRCRQWHWLMLSQ
MVENGKLDLLPENHALETVALGLLSAWRAYGRSEAIILIMVEDELRNFADQRLVEYKIQSLAEFQHVPMK
RFRLSDCPGNIATDGRGRLMLNGVEVAVVYYRTGYLPKHFPSEDVWSAFLQIELSEAIKCPWIGFHLAGM
KRMQLLLSNSDTLRSVINKTKLLAWPVGVKQKILHDDDEMYRQMKTVAVPMHDLDTSKAGVDELMKHIDA
NADNFVLKPHLEGGGNNFYGQNLIKQLNQLSGNERLAYVLMERIRPPSFDNWIVRANVPPEQTSVVSELG
TFGYLLATGQKIVNCSSDGGFLLRTKPSKEDEGGVMVGAAALDTPHLLLP

>EJW87969.1 glutathione synthetase [Wuchereria bancrofti]
MDEMPVRNDDNSVLKYYPKLEFIDGKLIKQLVEDTVDWAHAHGMVMRTAIATDRSDICQTAPCTLFPSPF
PYNLFQEAVDIQQAHYYFLSKLEILLYFRVSWDFDFLINSHAEVVKTDDFTRHFVEILNAVRTSNFCQKK
TLLIQRNDYMCHEDSYGNRSLKQIEVNNIAASMGSLAERVTCVHKRTLETLQLPNKIIERAILDNHPTVT
VARGICEAWYDYGVPEAIVLFVVEDANRNQIDQRHVEYCIDELSNRNARCLRITLTDGAERLKLNETNHH
LILDNILRVAVVYFRAGYSPDNYPTEMEWTARRVIELSDAVKCPWIGLQLANTKKVQQVLSENGILEKYI
TDDRMCARIRQTFAGMWGLENDDDKTQRIIQDAIAHPEKFVLKPQLEGGGGNYYGKEVAEKLKKMSRDEM
AAYIIMERITPMIVKNYVIRPQEEPVLMDVVGELGIYAYLYGSAAAAADNIPVENVMKNHVSGHIIRSKD
KSVDKGGVAIGAAVIDSPYLF

>Ppa-GSS-1 Pristionchus pacificus
MGDRKHSSVEEIPSFKEDFPSLPIDSSILNTLIEDAQDWAHANGLVMRSSENKSSSDSCIHAPFSLLPAPFPASL
YRQALQVQDATARLYHRIAYDTQWLLNAHENVIKTDEFTRNLCDILKKVTDEGLAQRKTLVIQRSDYMSHKDPFT
SEYTLKQVEVNNIASSMGAHAERVSSLHRRVLSLLGMEEGRIKAAIPHNKPVIMIARALFLAWKEFGRPNGVILV
IVEDVNQNQIDQRHIEYELTEQGVNPALIKRITLTQCHESVKLDSDRHLILDESTVSVVYFRAGYSPDHYHSHKE
WDARLTIERSDAIKAPWIGLQVANTKKVQQVLAEDGQLELFISNFAEAASVRQTFAGLWALDGNDPVAEKIIKIS
SHNWTTEMSSCVYKVLLYPLTYFIQHAQAKPEGYVLKPQLEGGGGNFYGDEVCDKLLHATAEERSAYILMEKLRP
LVVQSYLVRAHNSTQLAESVSELGIYGYAFGNDIDPPVVATGGHLLRTKGKLVLEGGVAVGASVIDSPFLYEYRG
E

>TMUE_1000003977 isoform 1 Trichuris muris
MMAFRIESSGVVLPDDKFVDELLDDAKDWAGTNGLCMRSREQPWSSDVVEIMPFTLFPSPVPRNLFDEAMSLQKV
MNELYYRVACDKDFIVACLADLAVADPFIANLLDIYLSVMNCAGHEDRVFLNVQRADYMFHQAEASANRLLVLKQ
VEVNNISAGLAAIGPVCSKLHQRTLRRARCPYETQGSLAVNSSINVVSIGLFEAWKAYGRSKAAVIFMVEDNPRN
IADQRLIEYQFDLLSKSTVAVFRLKLSESAERLQMTKSSLYLVPEDVEIAVVYFRTGYSPDHFDGDSAWRTLRLI
ETSKAIKCPWVGFHLAGTKKVQQMLCNPENLNRFVQDAQTRRRILSVTMPMFGFDKSTSNDDWNSIVSQVALEPT
GYVLKPSREGGGHNFYANDMVDLLKSCEQTERHAYILMKKIKPPEHRNVFVKSNVEWKLQSCASELGIFGFLLGN
MREIFDQRDGTFILRTKLHAENEGGLMRGTACVDSPFIY

>Cbn-GSS-1 Caenorhabditis brenneri
MAQKDNRILLLNAPRLPLEDAKLEELTGDLHDFAHAHGLVMRLANDKLSSEVCQTTPLTLLPSPFPKNVFEQAVK
LQDLYALLYHTIAYDFDFLVDIHKNVVKTDEFIRNMIEILKKVKAQGLKQPITLAIQRSDYMCHKDQASAEYGLK
QIEINNIASSMGAHAQRLTEWHIRILKALEVPDDVIKRAIPENKPIPMIAEALFKAWSHFNIPSAWVLVVVENVN
QNQIDQRHVEYELEKLGVPMTCIIRRTLTQCYEQLSLNETSDLMIDGRPVAIVYFRAGYSPDHYPTNKEWEARER
MELSTAIKTPWIGLQVANTKKTQQVLSEDGVLERFIGKPRDARDIRISFAGMWALENQDDVTLKVVQGAQKHPDA
FVLKPQTEGGAALHTGDEMVQMLRELPEEERGAYILMEKLRPMIIENYLVLAKKPVTFAKAVSELGVYGYAFGAK
DAPELKTAGHLLRTKPETTAMGGVAAGHAVVDTPFLYEFI

>Gpa-gss1
MSPSTNTDVVTPNYVADVVEMNENDDNHRQQQLTLLVEDALDWAHCFGLVLRTTDHKDRSDVCQAAPFALFPSPF
PRALFDEALAVQKALNLLYFRASWDLDFLTEAHRHVIPSDAFTRNVMDILVDVHREGVKQTITLLTQRADYMCHV
ASTDVGDEAGRQQKFELKQIEVNNIAVSMGGLAQHATVLHRRMLQKAGKVPSIGGTVLPENRPIDTLTEGIYIAW
CQFVKTFGDPNALLLVVVGEVDQNQFDQRFVEYELELKSTGQMKIVRLTLTQCAESLTLDSNDFTLRLGSRAVAV
VYFRAGYAPEDYPTQTEWEARRTIERSTAIKCPWIGLQVANTKKVQQVLDTPGAVERFFKEPNDAATVAAIRHVF
AGMWGLERDDDATNRVIQDAIANPDRYVLKPQLEGGGGNYFGEEIVTKLRAFTSHERAAHILMEKIRPLVVKNYL
VRPFQPSQLVNVVSELGIYGCLVGDGQGLSVCHNHAHGHILRTKSEHVNEGGVAVGAAVIDTPYLF
>Gro-gss1
MSPSTNTGVVTPNYVADVVEMNENDDNHRQQLSLLVEDALDWAHCFGLVLRTADHKDRSDVCQAAPFALFPSPFP
RGLFDEALAVQKALNLLYFRASWDLDFLTEAHRHVIPSDAFTRNMMDILVDVHREGVKQTVTLLTQRADYMCHVA
STDVGDEAGRQQKFELKQIEVNNIAVSMGGLAQHATVLHRRMLQKAGKVPSIGGTVLPENRPIDTLTEGIYIAWC
QFVKTFGDPSALLLVVVGDVNQNQFDQRFVEYELELKSTGQMKIVRLTLTQCAERLKLDTNDFTLRLDSRAVAVV
YFRAGYAPEDYPTQTEWEARRTIERSTAIKCPWIGLQVANTKKVQQVLDTPGAVERFFKEPNDAATVAAIRHVFA
GMWGLERDDDATNRVIQDAIANPDRYVLKPQLEGGGGNYFGEEIVTKLRAFTSHERAAHILMEKIRPLVVKNYLV
RPFQPSQLVNVVSELGIYGCLVGDGQGLSVCHNHAHGHILRTKSEHVNEGGVAVGAAVIDTPYLF
>Hav-gss1
MSPSTNIEIVTPNYVAEIVENADSSQQKLSLLVEDALDWAHCFGLVLRSSEYKNRSDVCQAAPFALFPSPFPRKL
FDEAMEVQKALNLLYFRISWDLDFLMEAHRLVIPSDTFTRNMIEILTDVHKDGVKQTFTLLTQRADYMCNATMTE
TQAEAGKYQTYELKQIEVNNIAVSMGGLAQRASLLHRRILQKSGKMTTIGDTEKSLPENRPIQTLTEGIHLAWKA
FGDLNALLLVVVGEVNQNQFDQRFVEYEMEQKTAGQMKIVRLTLTQCSHRLKLDPKLFTLHLDEHTVAVVYFRAG
YAPEDYPTQDEWEARRIIERSTAIKCPWIGLQVANTKKIQQVLATPGAVERFFKEPKDSATVAAIRHVFARMWGL
DRDDDETKRVMKDAITNPDRYVLKPQLEGGGGNYFGEEIVSKLCALTPAERAAYILMENIQPLVVKNYLIRPFQV
PCLSNVVSELGIYGCLVGDGRELSVSHNDAHGHILRTKSEHVNEGGVAVGAAVIDTPFFILKEHSSQWHLCISLS
HYYDI
>Hsc-gss1
MSPSTNTDLVTPNYIAEIVGVNETVAPQQQLNLLVEDALDWAHCFGLVLRTTEHKDRSDVCQAAPFALFPSPFPR
NLFDEAMAVQKDLNLLYFRISWDLEFLKEAHQHVIPSDAFTRKMLEILEDVHSGGVKQHITLLTQRADYMCHVTT
TDDQTETARQQQFELKQIEVNNIAVSMGGLAQRASVLHRRMLQKTSKTRVIEKIDSVLPENRPIDTLTEGIHNAW
KQFGDPNAILLVVVGEVNQNQFDQRFVEYEMEQKTTGQIKIVRLTLTQCSQKLKLDPKEYTLHLDAFKVAVVYFR
AGYAPEDYPTQTEWEARRTIELSTAIKCPWIGLQVANTKKVQQVLDTPGAVERFFKGPADEQKVAAIRHVFAKMW
GLDRDDAETNKVMQDAITNPQRYVLKPQLEGGGGNHFGEEIVSKLRTLTPAERAAFILMEKIQPLVVKNYLIRPF
RPPTLANVVSELGIYGCLVGDGRDLSVSHNNAHGHILRTKSEHVNEGGVAVGAAVIDTPFLF
>Lel-gss1
MLYDSAKLDQLLEKAKDWAQLNGLCLRTRERKNNSDYAQIAPFALLPSFVPKHLYNQAIALQQDMNLLYHQVGFD
YNFICTALSNVVKTDEFTRRLLEIYKYVYSGNGVLPSQPMVLTIQRPDYMFHQPPEMPALTSDTVTLQQVEVNQI
AVSLAGLSTTTYNLHRSMLREIGYKAETIAAHLPDNRAVHTVASGLAEAWRTYGDPDGCCLFVVEDESMNIFDQR
LVEYDFVQKTDYGGRVLRMTLTQVANELCLDENHGCRLCTKDGEKEISVVYFRAGYVPDHYHSDLEWQARLMLER
SKAIKCPWIGAQLAGTKKVQQLLSRPDVLRRFVTCAETFQKVSSTFVGLYGFDDYQIQHRLGLGGLGALKEETDD
VMTRLYRNPERFVLKPQLEGGGNNVYGQNIIDVLNQITQEQREAYIMMDRIFPMRHHNYLVRANEKCELSAVVSE
FGTYGYMLGSKDYVLKSFSGGHVLRTKSVATDEGGVMSGSSVLDSPYLVL
>Bxy-gss1
MQKIMQTERAKCAKVTVKNDHRKLKSSDVGKHSNVRDYVTGLIRNPEEKEELIEYAESYAHSIGLVSRTNERSFS
SEPAILVPIALLPSAFPRELYDQAVDVHATLAELYFRVACDHAFLVESFKDVCKTDAFTARMVGIVQAVHAEQNQ
GVRQPLTLSLQRADYLVHWEPQKDSFELKQIEFNIGPIGGPGCATQAAKLHAKMLDRLHAIHGSDVPMLAEAFTP
KVKARQKFARTLYQAWKLFGDPNAILLYITNSTNDPMCHFDGLQFVQFEVEKHGKRDGHLVEVVQMTLSKAAERL
TLDENGDFSLFVDGTKRVALAHITEGNMPEEFPTECEWHARTMLERSNAILSPNICTELSSSKKIQQILAMPGIL
ERFFTDEPDKCVALRRTFAGLWGLENDDEFTREIINEAIRSPHNYVLKCQLEAGKGNFFDDELVKKLGQMTLAER
GAFILQQKIKPMSVKNFLLRPFKPVELDDVIGELGIFGSLIGDQSTRKLLWNTVDGHVLKTRSASVNQAGVTAGF
GVVDTPLLFDASEFF
>Nab-gss1
MAHNINYIENSIKDKNQLAFLVEDALDWAHCFGLVIRTKEHRERSDVCQAVPFAILPSPFPSELFQQALAVQKSM
NLLYFKISWDYEFLLDAHREVVHSDYFTNKMVQILQDIHKQGNKQPITLLTQRADYMCHVRGESLTNFELKQIEV
NNIAVSMGSLSHRATLLHRRLFTKIGRDPNDIPDNDAVKSLTRGIYLAWEMFGNINGITLVIVADVNQNQIDQRF
IEYELESRSKGRMEIVRLTLTQCFEKLTIDEEYALKMGNKQISVVYFRAGYSPSDYPSEKEWEARLIIERSTAIK
CPWIGLQLANTKKVQQVLSRPRILERFFRNEDSEIISSIRVTFADMWGLENHDDETRSIIQDAIKNPQKYVLKPQ
LEGGGGNYFGQEISNKLKEFTTAQRSAHILMQRITPLIVKNYLIRPFEEPKLENIVSELGIYGSLIGNGQHLSVY
HNEAGGHILRTKPEHVDEGGIAVGAAVVDTPFLF
>Ppe-gss1
MNYVEADCNYSKSELKLLIEDAIDWAHSTGLIIRTTEHKDRSDVCQIAPFTLFPSPFPRRLFQQALDVHQSMQLL
YFKISWDYPFLVDAHKDVIKSDLFTQKMVNVLEQIHAEGIKQKLTLLTQRADYMCHYTGEGHAKESGDGKFQLKQ

IEVNNISVSMGGLAMRATNLHRRMLNKIRKQVGNEQVPDNDSISILADGLYTAWKLFGDQNALTLVVIGDVNQNQ
FDQRYVEYGLEDASRGQMCIKRLTLSECANSLIIDANSTLMFDGKAIGVVYFRAGYSPDDYPSEKEWTARLMIER
STAIKCPWIGLQLANTKKIQQVLSCDGVLKRFFPKTGDDQVLKLVKETFAQMWGLEKDDETTRIVIQDVISNPHK
YVLKPQLEGGGGNYFDQEIVDKLQAFSAEQRASHILMQKITPLVIKNYLIKAKALEARDNSSPNEPKLINVVSEL
GIYGCLIGDGNTMTVKCNKVGGHILRTKPEHVNEGGVAIGAAVIDTPFLF
>Min-gss1
MPLSVQKALNLLYFKVSWDYEFLKQSHQQVIKSDEFTRRLMSILDIVYKEGIKQPITLLTQRADYMCHYDGILTE
NVIDFDKFQLKQIEVNNIAVSMGGLAEKTTKLHRRVFKKMGMNVPNNDVMPLNEPIKTLCEGIYNACQNQFDQRA
IEYGLEELSEGKMIIIRLTFVECAERLVLNEKTFNLLLGDKTVGIVYFRTGYLPEDYVSENAWNARLLMERSTAI
KCPWIGLQMANTKKIQQVLSCSGILEKYLTPTKDLENVRSTFASLWGLERDDDETREIIKMRASHILMQRIRSLI
VKNCLIRPFDDDNSKKLQNVVSELGIYGSLIGIGGKEEVKFNSVGGHILRTKLENVNEGGIAVGASVVDSPFLF
>Min-gss2
MTSLNKQQDIDYISEIFVKNYQQLPSIIEDSMDWAHCNGLIFRTKEHKDRSDICQIAPFSLFPSPFPKRLFDQAL
SVQKALNLLYFKVSWDYEFLKQSHQQVIKSDEFTRRLMSILDIVYKEGIKQPITLLTQRADYMCHYDGIITENEI
NFDKFQLKQIEVNNIAVSMGGLAEKTTKLHRRVFKKMGMNVPNNDVMPLNEPIKTLCDGLYNACQNQFDQRAIEY
GLEELSEGKMIIIRLTFVEYYVSENAWNARLLMERSTAIKCPWIGLQMANTKKIQQVLSCSGILEKYLTSANDVE
NVRSTFASLWGLERDDDETREIIKMRASHILMQRIRPLIVKNCLIRPFDDENSKKLQNVVSELGIYGSLIGIGGK
DEEVKLNLVGGHILRTKLENVNEGGIAVGASVVDSPFLF
>Nab-gss2
MLNGIGKKKAFPAADDNFHYPRVVVENEEQLNELATFVRSWAQSNGLVFRAGSRENDVNRLITVPMTMLPSAFPE
QLFHKAMRVQKILNELYFRISWDWDFLVNAYKEVIRSDQLMQKFVDILKCLRAEGAHQKMTLMLQRTDYLVHQER
HNGEPELKQVEVNVGQIGCPGLGNRMSACHQRVADRLGLHKYGTIPENNCTHQFALALFQAWKQFDDQNALLLFV
NHAELCPYSYFDQWQVRDQLELVAQREGVRLDIIELTFAQCYKRLELLDDFSLVHGPDGRRVAVVHLWIGYLPEH
YPCEKSWKGRTMIERSSAILSPNIGQQLASTKKIQQLLSEPGCLEYFLPDQPESIALLRETYTNLWGGLDKSCMS
SLVKDAIKFPDKYVLKAQLDDGTGVYFDNELREKLNTLSEDELAAHILMKKLRPMSVPNYLLRGANAPELCNVIP
ELGVFGAFLGDGRGMRALHNNVIGYTFRTKRKGQNLGGILTGGTFDSALLVRSTAGAGDGEESDEYEGECG
>Gro-gss2
MASMNNGHVVQQQQKLKQAPQQSEALAKSNGAAATIIANNYVLAEVRNDDEMQMLAEYAVDYAHSIGLVSRSSEE
KYKNTNEMSVVPPLALLPSPFPRELYEQAIDVQQSLNELYFRVSCDHEFLMEAYKEVIKGDPFHAKLIEVEKRIQ
KEGIKQPLMLGLQRADYLSHWDEVAQKMELKQVEVNPGQIGGPGSATAVSKLHRKMLEKVEIVHGKKLPMLAKAV
VPENRPRDELAMTMYQAWKMFGDPNAMLLYVFQPDMFPVAYFEQLQFLMFAVEKLAKQDGNYVLVKRLSFIELRG
RLNLDEAGDHSLYLDGTKRIALVHMAYGYLPEHFHENDLDLRIMMERSTAIMSPTLRLQLAGTKKIQQVLSKPGV
LERFFPNEPQQVAKIRATFTELWGLGETDAITEAVVQNAMKNNKDYVMKSQMDGGHGVYFDDDITKMLQKLTKEE
RGAFILMKKIKPVVAKNFFVRPFEPPHQEDVNSELGIYGSLIGDQTTRQVLVNTVNGHIVRSKPVSQNMGGICAG
GGVFDSVLLFPSSEFH
>Gpa-gss2
MASMNNGHAAANGIQQQQQKLKQAQQSEALVKNNGGAETMIANNYAMAEVRSDEEMQMLAEYAVDYAHSIGLVSR
SNKYTNEVSVAPPLTLLPSPFPRELYEQAIDVQQSLNELYFRVSCDHEFLMEAYEEVIKGDPFHAKLIEVEKRIQ
KEGVKQPLMLGLQRADYLSHWDEAAQKMELKQVEVNPGQIGGPGSATCMSKLHRKMLEKVEIVHGKKLSMLAKAV
VPENRPRDEIAMTMYQAWKMFGDPNAMILTVNQPDLFPVCYFEQLQFLMFAVEKLAKQDGNYVLVKRLSFIELRD
RLTLDEAGDYSLYLDGTKRVAIVHLAYGYLPEHYHENDLELRIMMERSTAIMSPNLRLQLAGTKKIQQVLSKPGV
LERFFPNDPQQVAKIRVTFTELWGLDENDAITEAVVQNAMKNNKDYVLKSQMDGGHGIYFDDDITKMLKKLTTEE
RGAFILMKKIKPVVAKNFTVRSFEPPHQEDVNSELGIYGSLIGDQTTRQVLVNTVNGHLVRSKGASRNLGGICSG
GSVFDSVLLFPSSEFH
>Gpa-gss3
MASMNNGHVTANGIQQQQQKLKQVQQSEALAKSNGGAATMIANNYVLAEVRNDNELQMLAEYAVDYAHSIGLVSR
SSEEKYKHTNEMSVAPPLALLPSAFPRELYEQAIDVQQSLNELYFRVSCDHEFLMEAYKEVIKGDPFHAKLVEVE
KRIQKEGIKQPLMLGLQRADYLSHWNETARKMELKQVEVNPGQIGGPGSATSMSKLHRKMLEKVEILHGKKLPML
AKAVVPEDRPRDEIAMTIYQAWKMFGDPNAMILLVNQPDLFPVCHFEQLQFLMFAVEKLAKQDGNYVLIKRLSFI
ELRDRLNLDEAGDYSLYLDGTKRIALVHMAYGYLPEHYYENDLDTRIMMERSKAIVSPTLRLQLAGTKKIQQVLS
EPGVLERFFPNDPQQVAKIRVTFTELWGLSEHDAITEAVVQDAMKNNKNYVMKSQMDGGHGIYFDDEITKTLKKL
TAEERGAFILMKKIKPVVAKNFTARPFEPPQQEEVNSELGIYGSLIGDQSTRQVLVNTVNGYLVRSKAASRNLGG
ISSGGSVLDSVMLFPSSEFH
>Gro-gss3
MASMNNGHVTNGIQQQQQKLKLSEALAKSNGGAATMSANNYVLGEVRNDNELQMLAEYAVDYAHSIGLVSRSSEE
KYRYTNEMSVAPPLALLPSAFPRELYEQAIDVQQSLNELYFRVSCDHEFLMEAYEEVIKGDPFHAKLVAMEKRIQ
KEGIKQPLMLGLQRADYLSHWNETAQKMELKQVEVNPGQIGGPGSATSMSKLHRKMLEKVEILQGKKLPMLAKAV
VPENRPRDEIAMTIYQAWKMFGDPNAIILLVNQPDLFPVCHFEQLQFLMFAVEKLAKQDGNYVLIKRLSFIELRD
RLNLDEAGDYSLYLDGTKRIALVHMAYGYLPEHYHENDLDLRIMMERSKAIVSPTLRLQLAGTKKIQQVLSKPGV
LERFFPNDPQQVAKIRVTFTELWGLSEHDAITEAVVQNAMKNNKDYVMKSQMDGGHGIYFDDEITKMLKKLTAEE

RGAFILMKKIKPVVAKNFTVRPFEPPQQEEVNSELGIYGSLIGDQSTGQVLVNTVNGYLVRSKAASQNLGGISSG
GSVLDSVMLFPSSEFDQKINK
>Hav-gss2
MASINNGIAVGAVNGIQHQKRQQQQKQQTQFGQKNENGAAMVANVKANDYVLAQVRNDAELKLLAEYAVDYAHSI
GLVGRSGDEKYKYSNDVSEAPPIALFPSPFPRELYEQAIDVQQSLNELYFRIACDHEFLMEAYKDVIKGDPFHAK
LIELAKVIQKEGIRQPLMVGLQRADYLSHWNEAAQKMELKQVEVNPGQIGGPGAATAMSKNHRKMLDKVEIVHGR
KLPIMAKAVMPENRPGPGIALTIYQAWKMFGDPNAIILYVNQPDLFPVCHFEQLQFIMFQVEKLAKKDGNRVLVI
RQSFKDLRGRLSLDEAGDYSLYLDGTKRVALVHMAYGYLPEHYPEKDFDLRVMMERSTAILSPNLRLQLAGTKKI
QQILSKPGTLEHFFPNDPQQVAKIRNTFTELWGLEENDTTTKAVIVDAMKNCHDYVMKSQMDGGHGIFFDEEIPQ
MLVKLSLEERGAFILMKKINPVVAKNFMVRPFEAPHQEDVNSEMGIFGSLIGDQSTGQVLTNKMNGHLVRSKAAS
QNHGGVSCGSGVIDSVLLFPSSEFR
>Hsc-gss2
RFNYPSLRLICGSFPSKMASINNGHAAVGAANGIQQKKLLQQQQAHGTNMAQLPVNANDYALAAVRDDAELKLLA
EYAVDYAHSIGLVGRSGDDKYKHSNDVSVAPPIALLPSPFPRELYEQATDAQQSLNELYFRIACDHDFLMEAYKD
VIKGDPFHAKLIELAKVIQKEGIRQPLMVGLQRADYLSHWNEAAQKMELKQVEVNPGQIGGPGAATAVSKCHRKM
LDKVEIMHGKKLPIMAKAVVPENRPGPGIARTLYQAWKMFGDPNAIILYVNQPDLFPVCHFEQLQFVMFQVEKLA
KQDGNHVLVRRASFIELRKRLSLDEAGDYSLYFDGTKRVALVHMAYGYLPEHYPEKDFDLRVMMERSTAILSPNL
RLQLAGTKKIQQILSKPGTLEHFFPNDPQQVAKIRNTFTELWGLEENDDVTKAVIVDAMKNCHDYVMKSQMDGGH
GIYFDEEIPKMLATLSLEERGAFILMKKINPVVAKNFMVRPFEKPHQEEVNSEMGIYGSLIGDQSTGKVLVNSVN
GHLVRSKAASQNHGGVCSGGGVIDSVLLFPSSEFR
>Hsc-gss3
RFNYPSLRLICCSFPSICASKMASINNGHAAVGAANGIQQKKLLQQQQAHGTNMAQLPVNANDYALAAVRDDAEL
KLLAEYAVDYAHSIGLVGRSGDDKYKHSNDVSVAPPIALLPSPFPRELYEQATDAQQSLNELYFRIACDHDFLME
AYKDVIKGDPFHAKLIELAKVIQKEGIRQPLMVGLQRADYLSHWNEAAQKMELKQVEVNPGQIGGPGAATAVSKC
HRKMLDKVEIMHGKKLPIMAKAFVPENRPGPGIARTLYQAWKMFGDPNAIILYVNQPDLFPVCHFEQLQFVMFQV
EKLAKQDGNHVLVRRASFIELRKRLSLDEAGDYSLYFDGTKRVALVHMAYGYLPEHYPEKDFDLRVMMERSTAIL
SPNLRLQLAGTKKIQQILSKPGTLEHFFPNDPQQVAKIRNTFTELWGLEENDDVTKAVIVDAMKNCHDYVMKSQM
DGGHGIYFDEEIPKMLATLSLEERGAFILMKKINPVVAKNFMVRPFEKPHQEEVNSEMGIYGSLIGDQSTGKVLV
NSVNGHLVRSKAASQNHGGVCSGGGVIDSVLLFPSSEFR
>Hsc-gss4
RFNYPSLRLICGSFPSICASKMASINNGHAAVGAANGIQQKKLLQQQQAHGTNMAQLPVNANDYALAAVRDDAEL
KLLAEYAVDYAHSIGLVGRSGDDKYKHSNDVSVAPPIALLPSPFPRELYEQATDAQQSLNELYFRIACDHDFLME
AYKDVIKGDPFHAKLIELAKVIQKEGIRQPLMVGLQRADYLSHWNEAAQKMELKQVEVNPGQIGGPGAATAVSKC
HRKMLDKVEIMHGKKLPIMAKAFVPENRPGPGIARTLYQAWKMFGDPNAIILYVNQPDLFPVCHFEQLQFVMFQV
EKLAKQDGNHVLVRRASFIELRKRLSLDEAGDYSLYFDGTKRVALVHMAYGYLPEHYPEKDFDLRVMMERSTAIL
SPNLRLQLAGTKKIQQILSKPGTLEHFFPNDPQQVAKIRNTFTELWGLEENDDVTKAVIVDAMKNCHDYVMKSQM
DGGHGIYFDEEIPKMLATLSLEERGAFILMKKINPVVAKNFMVRPFEKPHQEEVNSEMGIYGSLIGDQSTGKVLV
NSVNGHLVRSKAASQNHGGVCSGGGVIDSVLLFPSSEFR
>Hsc-gss5
RFNYPSLRLICGSFPSKMASINNGHAAVGAANGIQQKKLLQQQQAHGTNMAQLPVNANDYALAAVRDDAELKLLA
EYAVDYAHSIGLVGRSGDDKYKHSNDVSVAPPIALLPSPFPRELYEQATDAQQSLNELYFRIACDHDFLMEAYKD
VIKGDPFHAKLIELAKVIQKEGIRQPLMVGLQRADYLSHWNEAAQKMELKQVEVNPGQIGGPGAATAVSKCHRKM
LDKVEIMHGKKLPIMAKAFVPENRPGPGIARTLYQAWKMFGDPNAIILYVNQPDLFPVCHFEQLQFVMFQVEKLA
KQDGNHVLVRRASFIELRKRLSLDEAGDYSLYFDGTKRVALVHMAYGYLPEHYPEKDFDLRVMMERSTAILSPNL
RLQLAGTKKIQQILSKPGTLEHFFPNDPQQVAKIRNTFTELWGLEENDDVTKAVIVDAMKNCHDYVMKSQMDGGH
GIYFDEEIPKMLATLSLEERGAFILMKKINPVVAKNFMVRPFEKPHQEEVNSEMGIYGSLIGDQSTGKVLVNSVN
GHLVRSKAASQNHGGVCSGGGVIDSVLLFPSSEFR
>Nab-gss3
MGVTACKTDCQPQQQVPKGKQLDNVQQRQNGDSQQQNGKMVDKAMDYVLEGVKDEQMLGELTQYAVDYAHIIGLV
AMWPERMDPQLSVQPPITLLPSPFPKICYEQAVSVQQTLSELYYRISLDHHFLMEAYKDVVKGDPFMARLVDMME
KVHAEGIHQKLTLAIQRADYMSNWEEQTGKMQLRQVEVNIGQVGGPGTAAKVTKLHRKMLDKVDSLLGAPLPLLA
NAHIQENQSSKNIARGIYQAWKLFGDPNAVVVFLTQADLFPVCYFEQLEFLQFGLEKHARQDGFRLNLVKMTLKE
APQRMRLDEQGDFSLIVDGSKRVALVHLAYGYLPEHYPTEAVWKARLDMERSTAIMSPNIRVQLSGTKKIQQVMA
KAGMMERFLPNAGKEKLAELRKTFAGLWGLENDDPGTQAVIRDAIQNPRKYVLKAQLGAGKGNYFDEKISEMLSK
MGLEERSAYILQEKIWPLVAKPPKLETVSSEMGIFGVLIGDSGKLLWNTVDGYYVRSKAEDVNQAKVGGGLGCVD
SMLLFPTEDLRR
>Nab-gss4
MGVTACKTDCQPQQQVPKGKQLDNVQQRQNGDSQQQNGKMVDKAMDYVLEGVKDEQMLGELTQYAVDYAHIIGLV
AMWPERMDPQLSVQPPITLLPSPFPKICYEQAVSVQQTLSELYYRISLDHHFLMEAYKDVVKGDPFMARLVDMME
KVHAEGIHQKLTLAIQRADYMSNWEEQTGKMQLRQVEVNIGQVGGPGTAAKVTKLHRKMLDKVDSLLGAPLPLLA
NAHIQENQSSKNIARGIYQAWKLFGDPNAVVVFLTQADLFPVCYFEQLEFLQFGLEKHARQDGFRLNLVKMTLKE

APQRMRLDEQGDFSLIVDGSKRVALVHLAYGYLPEHYPTEAVWKARLDMERSTAIMSPNIRVQLSGTKKIQQVMA
KAGMMERFLPNAGKEKLAELRKTFAGLWGLENDDPGTQAVIRDAIQNPRKYVLKAQLGAGKGNYFDEKISEMLSQ
MGLEERSAYILQEKIWPLVAKNYMLRPFQPPKLETVISEMGIYGVLIGTTEDGGKLLWNTVDGYLVRSKAASVNQ
GGISEGSGCVDSLLLFPTEDLRH
>Ppe-gss2
MVVTLPQQTATTQIEKKQKLDSSNPAAAAMGTQTSNNNANSNGVAGTGTAKHYVLEAIPNDSVLEELTNYAIDFC
HSTGNVALWAERKECNDLSVTSPIALMPSPFPAELFEKAKSVQQTLSELYFRISMDQDFLLEAYRDVAKADQWIA
KQIDMMKEVRAEGLHQKLYLQLQRADYMSHWNKEADRMELKQVEVNIGQVGGPGCATAMRKLHKRMVEKAENLHY
GKLPSLVNAKQPDNHSRRGMAETMLAAWKMFGDEKAVLVFLNQSDLFPVCHYEQLQFIQYDLEKIARKQNYHLNV
VRLSIKDLPERMHLDEKDFNLYVDGTKKVALIHMAYGYLPEHYPRPEEWKSRMDMERSTAIISPNIRLSLAGSKK
IQQVLAKPGVIERFLPNQEEKIAELRTTFAGLWGLENDDDRIRAIIADAIQNPRNYVMKAQLGAGKGNYFDDKMA
DMLREMSIEERGAYILQEKIWPVVTKNYMKRPFKAPTLENIVSELGIYGSFIGTEEDGGKVLWNRVEGFLVRSKP
HDVNQGGVSDGGGVVDSLMLFPEDELKH
>Ppe-gss3
MPVKNYVQELVEDEETLQELAKFAIDYSHSIGFVGLWSEHKLSYDLAITPPMTLLPSPFPLELFQKAVAIQETLN
ELYFRISMDHDFLVEAYKNVVKADKWVGRQMELLKRVNAEGIRQKYVLQVQRVDYMAHWECGKPMELKQIEVNMG
PGGPGFATKVAKLHRKTLQKVENLHGGKHPILAQAQMPPRQSIAEALYHAWRLFGDPKAVIVLVFDSSLYPIQHH
EQIQFVQFELENVGRKAGHPLNVIRMTLEDCARRMKLNEPGDFSLIIDGNKRAAVVYMVYGYLPEHYATEKEWRC
QLLMERSTAIVSPNIRLQLTGTKKVQQLLAKPGMVERFMPDQPKKVAALRSTFTGLWGLDGMDPATDALVSDAIE
HPQNYVLKALRDDGIGNFFDEELSQMLKAMSKQERSAFILQQKIRPIIVKNYTKRPFCPAQLENVVNELGIFGTC
ISSVEDGGKILWNRMDAHVSKTKGHNVNQGGIGEGSGVIDSLLLFPEADFH
>Min-gss3
MENLIISELQNGKTYSDNNHVINKHVNNHCIKNIANNYVVEVIKNEDMLKELTEYSINYAHSIGFVGLWSEQKKF
TDISVVPPMTLLPSPFPMELFQKATSVQSTLNELYYRISLDYDFLIEAYREVIKADKWVGRQVEMMKLVHAEGIR
QKFVLQVQRADYMTHCEDGQKIELKQIEVNFGPGGVGFAPKVTKLHKKMIEKVENLHGCTPNAMSEAALPKAKKI
AEALFQAWKVFGDAKAIVVLVFDSNLYPIQHHEQLQFVQFELEKIARNAGSILNISKMTLEECAERMSLNESDYS
LMIGESKRVAVVYMVYGYLPEHYTSEKEWNCQLNMERSTAIISPNIRSQLSGTKKVQQLLAKPGMLERFLADRPK
KIAELRSTFTGIWSLEGNDSFTQALVTDAISSPQNYVLKALRDDGVGNFFDEKLAEMLQTMTVQERSAFILQQKI
RPIANYLKRPFHPAKLEVVTNELGVFGTFLGTYDGKVLFNHVDGHFIKTKSHNSNQGGICEGSGVIDSALLFPEA
QFQKGIAK
>Min-gss4
MENLIISELQNGKTYSNNNHVINKHVNNHCIKNIANNYVVEVIDNEDMLKELSEYSINYAHSIGFVGLWSEQKKF
TDISVVPPMTLLPSPFPMELFQKATSVQSTLNELYYRISLDYDFLIEAYREVVKADKWVGRQVEMMKLVHAEGIR
QKFVLQIQRADYMTHCEDGQKIELKQIEVNFGPGGVGFAPKVTKLHKKMIEKVENLHGCTPNAMSEAALPKAKKI
AEALFQAWKVFGDAKAIVVLVFDSNLYPIQHHEQLQFVQFELEKIARNAGSILNISKMTLEECAERMHLNESDYS
LMIGESKRVAVVYMVYGYLPEHYTSEKEWNCQLNMERSTAIISPNIRSQLSGTKKVQQLLAKPGMLERFLADRPK
KIAELRSTFTGIWSLEGNDSFTQALVTDAISSPQNYVLKALRDDGVGNFFDEKLAEMLQTMTVQERSAFILQQKI
RPIANYLKRPFHPAKLEVVTNELGVFGTFLGTYDGKVLFNHVDGHFIKTKSHNSNQGGICEGSGVIDSAILFPEA
QFQKGIAK
>Gpa-gss31
MVVTLMTAEVASAQAKLDTLEMKQQLNVIRGEHGDLLLSNNDTHDENEQQQQQSNGTVKMNGHATNNGSANGTAP
PPLGMALSAIVGARSYVSDLVPDVQQAHELAEYGLDYAHSIGLCGRTLEHKFNSDIATAPPLALFPSPFPDSLYT
QAMEVQETLNELYFRISCDHEFLLEAYKDVIKGDPYIKRCVQIAQQIHEEGVHQPLAFSVQRADYLSHWDDQKQC
IELKQVEVNIGQIGGPGCATQANKYHRKMLEKLAIVRAGKGIEVLANAELPKNMPRHKMGQSLYEGWKLFGDPNA
VLLFVNQPDLFPLCHFEQVQFTTFQVEKLGIREGNHVEVIRMNLKECAERCRLDERDFSLYADGKRVALVHMAYG
YLAEHYPTEAEWQVRIAMERSTAILSPNIRLQLTGTKKIQQVLSKPGVLEYFFPKEPERVVALRDVFTDLWGLEN
DDDVTNEVICNAIQRPGNYVLKAQMGAGKGNYFDDEMVQKLRTMSLEERGAYILQKKIWPVAVKNYMLRPFQAPY
LENVVSELGIYGSIIADSSNGKVLWNSAEGYLSRSKPVSMNQGGVCEGSGVVDSVLLFPTNEFCAD
>Hav-gss4
MVVTLATAAAAAEMSTTTTQTQTLDMQEMKRQLNVIKKEHDGLLLTSSTENGVQAQNGENGTTNGQYVNGNANGT
TLTAPLQAKALLFVDARNYVPNLVADDQRMREFAEYGIDYAHSIGLCARTVEHKFKSDIASAPPLALLPSPFPRS
LYMKALEVQETLNELYFRISCDHEFLVEAYKDVIKGDPYIKRCLEIAQQIHDEGVHQALSFSVQRADYLSHWDEQ
KQCIELKQVEVNIGQIGGPGCATQTNKYHRKMLNKLDIVRAGIGGMEMLANSEMPVNKPRHKMGQTLYEGWKMFG
DANAVILFINQPDLFPLCHFEQLQFTTFEVEKLAKQDGNRVQVIRMTLTECTERCQLDENDFSLYADGKRVALVH
MAYGYLPEHYPTEAVWQVRIAMERSTAIISPNIRLQLTGTKKIQQLLSKPGVLEHFFPEETERIAALRDVFTGLW
GLENDDAVTNSVIEDAIQRPKDYVLKAQMGGGKGNFFDDEMVHKLKTMSLEERGAYILMKKIWPIAVKNYLVRPF
HVPYLENVVSELGIYGSIIADSSNGKVLWNSAEGYLVRSKPANVNQGGVCEGSGVVDSVLLFPDNEFGTN
>Hav-gss5
MVVTLATAAAAAEMSTMTTQTQTLGMQEMKRQLNVIKREHDELLLTNNILNGEQQQVGNISTTKGQHANGNANGT
TLAAPPEAKALLLVDARNYVPNLVADDQRMRELAEYGVDYAHSIGLCARTVEHKFKSDIASAPPLALMPSPFPRS
LYMQAFEVQETLNELYFRISCDHEFLVEAYKDVIKGDSYIKRCLEIAQQIHDEGVHQALSFSVQRADYLSHWDEQ

KQCIELKQVEVNIGQIGGPGCATQTNKYHRKMLNKLDIVRAGIGGMEMLANSEMPVNKPRHKMGQTLYEGWKMFG
DANAVILFINQPDLFPLCHFEQLQFTTFEVEKLAKQDGNRVQVIRMTLTECTERCQLDENDFSLYADGKRVALVH
MAYGYLPEHYPTEAVWQVRIAMERSTAIISPNIRLQLTGTKKIQQLLSKPGVLEHFFPEETERIAALRDVFTGLW
GLENDDAVTNSVIEDAIQRPKDYVLKAQMGGGKGNFFDDEMVHKLKTMSLEERGAYILMKKIWPIAVKNYLVRPF
HVPYLENVVSELGIYGSIIADSSNGKVLWNSAEGYLVRSKPVNVNQGGICEGSGVVDSVLLFPDNEFGTD
>Hav-gss6
MVVTLATAAAAAEMSTMTTQTQTLGMQEMKRQLNVIKREHDELLLTNNILNGEQQQVGNISTTKGQHANGNANGT
TLAAPPEAKALLLVDARNYVPNLVADDQRMRELAEYGVDYAHSIGLCARTVEHKFKSDIASAPPLALMPSPFPRS
LYMQAFEVQETLNELYFRISCDHEFLVEAYKDVIKGDSYIKRCLEIAQQIHDEGVHQALSFSVQRADYLSHWDEQ
KQCIELKQVEVNIGQIGGPGCATQTNKYHRKMLNKLDIVRAGIGGMEMLANSEMPVNKPRHKMGQTLYEGWKMFG
DANAVILFINQPDLFPLCHFEQLQFTTFEVEKLAKQDGNRVQVIRMTLTECTERCQLDENDFSLYADGKRVALVH
MAYGYLPEHYPTEAVWQVRIAMERSTAIISPNIRLQLTGTKKIQQLLSKPGVLEHFFPEETERIAALRDVFTGLW
GLENDDAVTNSVIEDAIQRPKDYVLKAQMGGGKGNFFDDEMVHKLKTMSLEERGAYILMKKIWPIAVKNYLVRPF
HVPYLENVVSELGIYGSIIADSSNGKVLWNSAEGYLVRSKPANVNQGGVCEGSGVVDSVLLFPDNEFGTN

>Gpa-gss4
MQKIMQTERAKCAKVTVKNDHRKLKSSDVGKHSNVRDYVTGLIRNPEEKEELIEYAESYAHSIGLVSRTNERSFS
SEPAILVPIALLPSAFPRELYDQAVDVHATLAELYFRVACDHAFLVESFKDVCKTDAFTARMVGIVQAVHAEQNQ
GVRQPLTLSLQRADYLVHWEPQKDSFELKQIEFNIGPIGGPGCATQAAKLHAKMLDRLHAIHGSDVPMLAEAFTP
KVKARQKFARTLYQAWKLFGDPNAILLYITNSTNDPMCHFDGLQFVQFEVEKHGKRDGHLVEVVQMTLSKAAERL
TLDENGDFSLFVDGTKRVALAHITEGNMPEEFPTECEWHARTMLERSNAILSPNICTELSSSKKIQQILAMPGIL
ERFFTDEPDKCVALRRTFAGLWGLENDDEFTREIINEAIRSPHNYVLKCQLEAGKGNFFDDELVKKLGQMTLAER
GAFILQQKIKPMSVKNFLLRPFKPVELDDVIGELGIFGSLIGDQSTRKLLWNTVDGHVLKTRSASVNQAGVTAGF
GVVDTPLLFDASEFF
>Gro-gss4
MQKIMQTERAKCAKITVKNDQRKLKSSDVGKHLNVRDYVTGLIRNPEETEELIEYAESYAHSIGLVSRTNERSFS
SEPAILVPIALLPSAFPRELYEQAVDVHATLAELYFRVACDHAFLVESFRDVCKTDAFTARMVGIVQAVHAEQNQ
GVRQPLTLSLQRADYLVHWEPQKDSFELKQIEFNIGPIGGPGCATQAAKLHAKMLDRLHAIHGSDMPMLAEAFKP
DVKARQKFARTLYQAWKLFGDPNAILLYITNSTNDPMCHFEGLQFVQFEVEKHGKRDGHLVEVVQMTLSKAAERL
TLDENGDFSLFVDGTKRVALAHITEGNMPEEFPTEREWHARTMLERSNAILSPNIRTELSSSKKIQQATIRSPQN
YVLKCQLEAGKGNFFDDELVKKLGQLTLAERGAFILQQKIKPMSVKILAMPGILERFFTDEPDKCVALRRTFAGL
WGLENDDEFTRGIINEAIRSPQNYVLKCQLEAGKGNFFDDELVKKLGQLTLAERGAFILQQKIKPMSVKNFLMRP
FKPVELDDVIGELGIFGSLIGDQSTRKLLWNTVDGHVLKTRSASVNQAGVTAGFGVVDTPLLFDASEFF
>Gro-gss5
MATIMNGNVHHAESTGAHVKNGTDLNGTKLVKPLLNTVAQTDACNNVKQYVLEAIRDKQELYNMEQYAIDYAHSI
GLVAPMPDQPKETFSNILAVPPPITLLPSPFPRELYEQAVDVQQQLSELYFRIASDHEFLMDSFKDVIKSDPFMA
RFVQIAKMVHEEGVHQPLAVQLQRSDYMTHLEPSDGTLALKQVEVNIGPLGGIGSVTGVSKLHRKTLDKVAIVRE
GRLAMLANAYAPVDRTRQNLARSFYQTWKLFGDPKAILLFLDTPDLMYFEQRQCIQFEVEKLGKQDGLLVVVLSL
PFVEASKRMSLDENGDFSLYMDGNKRVALVHITDGNAPDEFPTEREWTARTMMERSTAILSPNIRLQLSCTKKIQ
QVLAKPGMLERFIPNDSKLVAKLRSTFTGLWGLEVDDIATNEVIKAAIRSPRNYVLKSQMEAGLGNFFDEQVAEM
LLKLTKQDRAAYILQQRINPLVVKNFMMRQMKPAQMEDVVSELGIYASLIGNQSTGQILHNSVDGHTIRSKPSKV
NQGGVVFGGGTIDSALLFPAAEMLEAQ
>Gpa-gss5
MATIMNGNAHHAESNGVHAKNGIDQNGTKLVKPLNTVAQTDGSNNAKQYALEAIRDKQELYNLEQYVIDYAHSIG
LVAPMPDQPKGKFSNICAVPPPITLLPSPFPRELYEQAVDVQQQLNELYFRIASDHEFLMDSFKDVIKSDPFMAR
FVQIAKMVHEEGVHQPLALQLQRSDYMVHLEPSDGTLALKQVEVNIGPIGGPGFATGVSKLHRKTLDKVAIMREG
QLAMLANAYAPVDQTRQKMAYSLYQTWKLFGDPKAILLFLDTPNILRFEQLQFIQFEVEKLGKQDGHLVKVLSLT
FVEASKRMSLDENGDFSLYLDGNKRVALVQITDGNIPDEFPTEREWTARTMMERSTAILSPNIRLQLCCTKKIQQ
VLAKPGMLERFIPNNAKLVAQLRCTFTGLWGLEEDDKATKEVIEDAIRSPHNYVLKSQLEAGIGNFFDEEVAEML
QKLSKQDRAAYILQQRINPLVVKNYMMRQMKPAHMEDVVSEIGIYASLIGNQSSGQILHNSVDGHTIRSKPSTLN
QGGVGSGGGIVDSALLFPAAEMLEANSNGI
>Gro-gss6
MATIMNGNVHHAESNGVHAAKNGIDQNGTKLVKPLNTVAQTDGSNDAKQYVLEAIRDKQELYNLEQYVIDYAHSI
GLVAPMPDQPEGKFSNICAVPPPITLLPSPFPRELYEQAVDVQQQLNELYFRIASDHEFLMDSFKDVIKSDPFMA
RFVQIAKTVHEEGVHQPLAVQLQRSDYMVHLEPSDGTLALKQVEVNIGPLGGPGFATGVSKLHRKTLDKVAIVRE
GQLAMLANAYAPVDQTRQKMAYSLYQTWKLFGDPKAILLFLDTPNILRFEQLQFIQFEVEKLGKQDGHLVKVLSL
TFVEASKRMSLDENGDFSLYIDGDKRVALVQITDGNIPDEFPTEREWTARTMMERSTAILSPNIRLQLGCTKKIQ
QVLAKPGMLERFIPNNAKLVAQLRCTFTGLWGLEEDDKETKDVIEAAIRSPHNYVLKSQLEAGIGNFFDEDVAEM
LQKLSKQDRAAYILQQRINPLVVKNYMMRQMKPAQIEDVVSEIGIYASLIGNQSSGQILHNSVDGHTIRTKPSKV
NQGGVGSGGGTVDSALLFPAAEMLENFMMRQMKPAQMEDVVSEMGIYASLIGNQFTGQILHNSVAGYTIRSKPSK
LNQGGVDSGGGTVDSALLFPTAEMLEAKLI

>Hav-gss3
MATYQNGNVHRMKCNGTHATNGIDQNGTKMADHLQKSLEIGAKKQQRGRHTDVRQYVVEVIGDKDQLHTLEQYAI
DYSHSIGLVAPISDPPEGKFTNIFAVPPPIALLPSPFPRELYEQAVNVQQAMNELYFRVASDHEFLMDALKDVIK
GDPFMARLVQIAKIVHDEGVHQPLAVQLQRSDYMAHWEPSDRTMALKQVEVNIGPIGGPGFAYGLSKLHRKMVDK
LSIMHDGKPAILVNSEAPVNRSRQNMAFTLYHSWKLFGDPRAVLVFLDTPNITHFEQLQFIQFEVEKLGMQDGHF
VKVEVLTLTEASKRISLDESADFSLYLDGTKRVALAHMADGNVPDEFPTENEWTARTMIERSNAILSPNIRFQLS
STKKIQQVLAKPGMVERFFPNDPKRVALIRSTFTGLWGFEVEDEGTSEVIENAIRSPANYVLKSQLEGGLGNFFD
KQMAQKLQQMSKEDRGAYILQQRIKPLVVKNYLMRQPKSAELGDVVSEIGIYGSLIGDQSNGRILHNAFDGQRIR
SKLSNLNQGGVGCGSGAIDSPLLFPMAQMLDKSNGL
>Hsc-gss6
MAAFLNGNAHSSIECNGVVHAKNDIDQNGTKLADRLQHSVQIGTKKASVSRGSINVNKYVVEVISDTDQLHTLEQ
YAIDYAHSIGLVAPLSDPPEGKFTNIYAVPPPIALLPSPFPRELYDHAVNVQQAMNELYFRVASDHDFLMDAFKD
VVKGDPFMARLVQIAQMVHDEGIHQPLAVQLQRSDYMAHWDPSDGSMALKQVEVNIGPIGGPGFAFGVSKLHQKM
LDKLAIEHDGKPAILANSEAPLNRSRQNIAYTLYQSWKLFGDPKAVLLFLDTPNITHFEQLQFIQFEVEKLGKQQ
GNFVKVLVLSLTEASKRISLDESGDFSLYLDGTKRVALVHIADGNVPHEYPTEHEWTARTTIERSNAILSPNIRF
HLSSTKKIQQVLAKPGMVERFFPNEPKRVALIRSTFTGLWGLEVEDEATREVIEDAIRSPGNYVLKSQMEAGLGN
FFDDQMAQMLQQMSKEDRGAYILQQRIKPLVVKNYLMREAKPAELEDVVSEMGIYGSLIGDQSNGRILHNAFDGH
TIRSKRSDLNLGGVCCGGGAIDSPLLFPMAQMLDQSNGH
>Hsc-gss7
MAAFLNGNAHSSIECNGVVHAKNGIDQNATKCADRLEHSLEIGAKKSGDNRGEIHCNKYVLEVISDNDQLRMLEQ
YAIDYAHSIGLVCPISDPSDGKLTNIYAVPPPIALFPSPFPRELYDHAVNVQQAMNELYFRVASDHDFLMDAFKD
VVKGDPFMARLVQIAQMVHDEGIHQPLAVQLQRSDYMAHWDPSDGSMALKQVEVNIGPIGGPGFAFGVSKLHQKM
LDKVAIEHDGKPAILANSEAPLNRSRQNIAYTLYQSWKLFGDPKAVLLFLDTPNITHFEQLQFIQFEVEKLGKQQ
GNFVKVLVLSLTEASKRISLDESGDFSLYLDGTKRVALVHIADGNVPHEYPTEHEWTARTTIERSNAILSPNIRF
HLSSTKKIQQVLAKPGMVERFFPNEPKRVALIRSTFTGLWGLEVEDEATREVIEDAIRSPGNYVLKSQMEAGLGN
FFDDQMAQMLQQMSKEDRGAYILQQRIKPLVVKNYLMREAKPAELEDVVSEMGIYGSLIGDQSNGRILHNAFDGH
TIRSKRSDLNLGGVCCGGGAIDSPLLFPMAQMLDQSNGH
>Hsc-gss8
MAAFLNGNAHSSIECNGVVHAKNDIDQNGTKLADRLQHSVQIGTKKASVSRGSINVNKYVVEVISDTDQLHTLEQ
YAIDYAHSIGLVAPLSDPPEGKFTNIYAVPPPIALLPSPFPRELYDHAVNVQQAMNELYFRVASDHDFLMDAFKD
VVKGDPFMARLVQIAQMVHDEGIHQPLAVQLQRSDYMAHWDPSDGSMALKQVEVNIGPIGGPGFAFGVSKLHQKM
LDKVAIEHDGKPAILANSEAPLNRSRQNIAYTLYQSWKLFGDPKAVLLFLDTPNITHFEQLQFIQFEVEKLGKQQ
GNFVKVLVLSLTEASKRISLDESGDFSLYLDGTKRVALVHIADGNVPHEYPTEHEWTARTTIERSNAILSPNIRF
HLSSTKKIQQVLAKPGMVERFFPNEPKRVALIRSTFTGLWGLEVEDEATREVIEDAIRSTGNYVLKSQMEAGLGN
FFDDQMAQMLQQMSTEDRGAYILQQRIKPLVVKNYLMREAKPAELEDVVSEMGIYGSLIGDQSNGRILHNAFDGH
TIRSKRSDLNLGGVCCGGGAIDSPLLFPMAQMLDQSNGH
>Hsc-gss9
MAAFLNGNAHSSIECNGVVHAKNDIDQNGTKLADRLQHSVQIGTKKASVSRGSINVNKYVVEVISDTDQLHTLEQ
YAIDYAHSIGLVAPLSDPPEGKFTNIYAVPPPIALLPSPFPRELYDHAVNVQQAMNELYFRVASDHDFLMDAFKD
VVKGDPFMARLVQIAQMVHDEGIHQPLAVQLQRSDYMAHWDPSDGSMALKQVEVNIGPIGGPGFAFGVSKLHQKM
LDKVAIEHDGKPAILANSEAPLNRSRQNIAYTLYQSWKLFGDPKAVLLFLDTPNITHFEQLQFIQFEVEKLGKQQ
GNFVKVLVLSLTEASKRISLDESGDFSLYLDGTKRVALVHIADGNVPHEYPTEHEWTARTTIERSNAILSPNIRF
HLSSTKKIQQVLAKPGMVERFFPNEPKRVALIRSTFTGLWGLEVEDEATREVIEDAIRSTGNYVLKSQMEAGLGN
FFDDQMAQMLQQMSTEDRGAYILQQRIKPLVVKNYLMREAKPAELEDVVSEMGIYGSLIGDQSNGRILHNAFDGH
TIRSKRSDLNLGGVCCGGGAIDSPLLFPMAQMLDKSNGH
>Gpa_GSS6
MQLDAHGIPRPVEGVVHPMTQLLQFYANKGYRRVGKTTWFDPSNANYVVIPDDLCYLERMVKDVCAPSATTGGDG
AKRGREQESAEKSTKRMRKHQNSAYSAENYAKDLVKDEDDLNMLVEAAVDLAHDVRLIKRLEDSDSRSRRNSDVA
SIIPFTLFPSPYPRHIFQQALDVQTGLQKLYFRVSCDYAFLAEQFDDVIKTNNLMRKMAEIMHEIQFEGQKQAYT
LFLARSDYMVDLDKDGQEEYGLKQVEMNIGSVVGSTFGPRTAELHRQMLQKVGMDASNVPENRAYNTLAEGLYFA
WQKFGDPDAVVVFAVLQGSVHRFDERAIEYELQRISGGKLNVVRMSSKEAYHKLRLDNDFKLRLSADGRVVAVVY
SRSGPLPEWTDEEWQARRTIERSTAIKTSTMFSALSSSKKVQQVLAKPGMIERFLPDPEDKEIIVAIRKTFVGLW
GLEKDDDETRNLVQHAINNPGFYVLKPQSEGGGHNYFDEELREKLQQFTHEERAAHTLMERIQPMIVKNYLVRPL
EKPVLSDVVAELGVFGCLLGDKRDLSILHNKQHGYLVRTKPASSTESGITAGGVYDSLNLF
>Hav-gss7
MLTYKSVLLALFFASLANATPKERKTNTKEFPRKINLRNGQEVTIRLIRDEDINESANIIREAYLDDCQKIKHAT
SKQRDEMRNVPEDTTRLALRNFKAEKDSVLFVAEAEGNSGGTELKGCIRVKLNLTGNRKNDGPFAQIGPFATKLN
CHGTGIGRAMISAAEDYATDNWKVCETHLDAHGIPEPKKGAFHESPPLLQFYEKRGYQRIGKTTWFDPETANYVK
IEGSLCHLERMVKNTCLASEKDAKADKEKRKAPAEGESSSPKRRKMQMESSSSSKELIRNYALDAVKEQGDEIDT
LAEDIVDFAHDMGLIKRLEDDESRKRRFSIVASIQPISLFPSPFPRSVYQQAMDVHRGMQKLYFRVCCDYEFLAN
ASEQMVKTNELYGRMVKMMEQIQREGIKQPYQLFLTRSDYMVDNEQNGPQQKFGLKQIEMNIGSVVGSAMGARTS

EVHRRMLKKVGLDETNVPENKAYNTLARGLFQAWLHFGDPNAVVVFAVLQGSVHRFDERTIEYELQRISDYQVQV
VRMSSKDAYSKLFLDAQFDLRLSSDNRRVALVYSRSGPLPEWTEQEWQARLTIERSTAIKTSTMFSALSCSKRIQ
QLLAQPGMVERFLPEKEEQETVKAIRRTFVGLWGLENDDNETRQLVQSAIENPSLYVLKPQNEGGGHNYFDAQLQ
EKLQKFTHEERAAHTLMQRINPVTIKNYFVRPTDEKPVLDDVVVELGMFGYLLGNRSDQSIVKNEQHGYLVRTKL
ASSAEGGITAGGVYDSLNLF
>Hsc-gss10
MAHEVGLIKRLSDDDSRKRRNSDVASIQPISLFPSPFPRSAYQQAMDVHTGMQKLYFLVSCDFDFLVKATEGMDK
SNNLYGRMIEMMKEIHREGQRQPYTLFLTRSDYMVDSTTDERDGQQRFGLKQVEMNIGSVVGSAMGPRTAEMHRR
MLQRMRMDASNVPENRAFNTLARGLFQAWLRFGDPNAVVVFAVLQGSMHRFDERAIEYELQRISDWQVKVVRLSS
KEAYSKLQLDPNDFTLRLTADGRAVAVVYSRSGPLPEWTEEEWEARKRIERSTAIKTSTVFSALSSSKRVQQLLA
QPGMIERFLPEPKDRQMVEAIRQTFVGLWGLENEDKPTQKLIQHAIDNPHLYVLKPQNEGGGHNYFDDELKEKLL
QFTREERAAHTLMQRIWPVTAKNFLVRPMEEAVLDDTIVELGMFGYLLGDKRDRSIVRNKQHGYLVRTKPASSAE
GGISAGGVYDSLNLF
>Hsc-gss11
MFISNHLLFAFFVSALQFHSNANPNVSISNPAVEAVKSSDQLATLVEAAVDVAHEVGLIKRLSDDDSRKRRNSDV
ASIQPISLFPSPFPRSAYQQAMDVHTGMQKLYFLVSCDFDFLVKATEGMDKSNNLYGRMIEMMKEIHREGQRQPY
TLFLTRSDYMVDSTTDERDGQQRFGLKQVEMNIGSVVGSAMGPRTAEVHRRMLQRMRMDTSNVPENRAFNTLARG
LFQAWLRFGDPNAVVVFAVLQGSMHRFDERAIEYELQRISDWQVEVVRLSSKEAYSKLQLDPNDFTLRLTADGRA
VAVVYSRSGPLPEWTEEEWEARKRIERSTAIKTSTVFSALSSSKRVQQLLAQPGMIERFLPEPKDRQMVEAIRQT
FVGLWGLENEDKPTQKLIQHAIDNPHLYVLKPQNEGGGHNYFDDELKEKLLQFTREERAAHTLMQRIWPVTAKNF
LVRPMEEAVLDDTIVELGMFGYLLGDKRDRSIVRNKQHGYLVRTKPASSAEGGISAGGVYDSLNLF
>Hsc-gss12
MNSKLAQCLFLGIILILNNLLVIFGHSVENENDGTNGQTSEEFDKVTRNYVEQLVKSEKHLLSLRQFAVEWAHNN
ALIFRNKKVPPTTDAIYRSDVAVIAPFSLFPSPFPRHAFEHALAVQKALNLLYFRVGTDIDFLERAYSDLIKTDE
NFSNTMDVLRTVREEGIRQPITVMYQRADYMLNVVGGQDEAEPNYEIKQLEVNCGSVAGTSLDRRTAQLNHVLLQ
RAGFHPAPEDLPENWPDKAQIESIKMAWEAYNKSDAIVVILISPISETIFDANFFETELDRLSNGRIKVERITLN
DCVHRCKLDENFALRLDGREVAVVKSRYSVLGLRARGSELNILKNLRLMLERSLAIKIPSTFIGFSCSKKVQQLL
AEPGELEHFFPEESDAEMVKAIRKTFAGMWSLENTDENTEQKIQDAINHPENYVLKSNMECGGNNYFDEKIPIKL
TGITPTERSFHILMQKLRPMPIKNVMVHPNTKPKINEMVSELAVYGVLIGNMTTRTVSHNVQQGHLLKTKLATAN
EGGISTGSAVHDSPILF
>Gpa_GSS14
MNAMLAQLFAIIFILTNFGAIVAHPVDKENEAEASTNNNVDKVTRNYVEALVTDEEHLNSMRLFAVEWAHNNALI
FRTKKHPTKSDVSTIAPCSLFPSPFPRQPFEQALAVQKAMNELYFRIGTDFAFLQEAYKDVIEADDHFRNMMDML
KSVHEEGIKQPITVIFQRADYMLNVIKGQNGEEPTYEIKQLEVNCGSIAGTSLDRHAAELNHVMLKKAGFHAAPE
DLPENWPDKAQIESIKMAWELYGNPDAIVVIMISDHSQTVFDARFFETELDRLSDGKIKVARVTLNGCAIRCSLD
EEDSKLRLDGREVAVLNSRYSALGFMPGVHAMNARKMIERSQAIKIPSAFVGFSCSKKVQQLLAEQGMVERFFPN
ENDAETVKAIRQTFTGLWSLDKEDKATQDRIEDAIANPNNYVLKTNMECGACNYFDEQLANKLVEITPDQRPYYV
LMQKLRPMPIKNIMIHPFTASKIDTMVSELGVYGVLVGNMLTKEVKHNVQQGHLLKTKLETANEGGISTGTAVHD
SPILF
>Gpa_GSS15
MNAMLAQIFAIIFILTNFGAIVAHPVDKENEAEASTNNNVDKVTRNYVEALVTDEGHLNSMRSFAVEWAHNNALI
FRTKKHPAKSDVSTIAPCSLFPSPFPRQPFEQALAVQKAMNELYFRIGTDFAFLQEAYKDVIEADDHFRNMMDML
KSVHEEGIKQPITVIFQRADYMLNVIKGQNGEEPTYEIKQLEVNCGSIAGTSLDRHAAELNHFMLKKAGFHAAPE
NLPENWLDKAQIESIKMAWELYGNPDAIVVIMISDFSQTVFDARFFETELDRLSDGKIKVARVNLNDCAIRCSLD
ENSKLRLDGREVAVLNSRYSALGFMPGVHAMNARKMIERSQAIKIPSAFVGFSCSKKVQQLLAEQGMVERFFPNE
NDAETVKAIRQTFAGLWSLDKEDKATQHRIEDAIANPNNYVLKTNMECGACNYFDEQLANKLKEITPAQRPYYVL
MQKLRPMPIKNIMIHPFTASKIDTMVSELGVYGVLLGNMLTKEVRHNVQQGHLLKTKLETANEGGISTGTAVHDS
PILF
>Gpa_GSS17
MSNLTIFVVLFTFSASFFTCSTTNPSHHNSDEGENSSKKDEIKNIQNYASDVVKDEKHLGELALYAIEWAHNNGL
VLLSRETDIVEFAPISLFPSPFPRRSFEKALSVQKDMNLLYFRVASDYEFMAEAFKDLIPVDAHIAKLWQIVKEV
HEEGIRQPFTLLIQRADYMLNVVENPSAGEEQYQIKQVEVNGGSICGLGLKRRNSELHRQMLRKVGMDVSASPIN
QPDVALVEALHMAWKQFGDPNALFMFLAIKIPFVFDQTRIASELERVSNGKIEVIFMSLFDSAKNLHLDPEDFSL
RRNSDGRRVAVVYSNMSALGYAPKLHYDVQSEARKMIERSTAIKAPSLAIAISCTKKIQQLLTKPENLKRFFPRP
EDAETIKNIQSTFAGLWGLENDDQETQELIKDAMETPANYVLKPNRECGGNNYFDEQIPEAFQKFTPEERKAHIL
MQKLRPMAVPNYMLRPLQEPIEASVVPELGVYGFLLGNMVDGSVQHNVQQGYHFRSKFAHLNEGGITAGFGFYDT
AYLF
>Gro-gss7
MSNLTIFVVLFTFSASFFTCSTKNPSQHNSDEGENSTKKDEIKNIQNYASDVVKDEKHLHELALYAIEWAHNNGL
VMLSEQEDIAEFAPISLFPSPFPRKAFEKALAVQKDMNLLYFRVASDNEFMVEAFKDLIPGDAHIAKLWQIVQEV
REEGIRQPFTLLIQRADYMLNVVEVNGGAISGLGMKRRNSELHRQMLRKVGMDISASPVNQPDFALVEALHMAWK

QFGDSNALLMFLTPPKISYVFEQRHIASELERVSNGKIEVIFVSLFGTAKILHLDPEDFSLRRNSDGRRVAVVYS
NMSPLGYRPSSNYDMQMEARKMIERSTAIKAPSLAIGISCTKKIQQLLTKPENLKRFFPRPEDAETIENIRSTFA
GLWGLENDDQQTQELIKDAMENPSNYVLKPNRECGGNNYFDEQIPEAFQKFTPEERAAHILMQKLRPMAVQNYLL
RPLKEPKESSVVPELGVYGFLLGNMVDGSVQHNVQQGYHFRSKLAHLNEGGISAGFGYYDTAYLF
>Gpa-gss11
MSTINSNFPIVFVSVVFILCTFFGHAAPPHQHNADDKTASNDGAESLQNYIGNMALTEIELKKLAQYAIVWCHNN
ALILRPKGNEGRVDVAEFAPISLFPSPFPRDAFDRAMKVQKAMNLLYFRVARDHDFLMNAYKDTIRSDKFIAKLV
GIVKEVQEEGIRQPITLMLQRADYLLNVVEDNETKEVKYEPKQVEVNTGAVGGMGLKRRTTELHRRMIEKVGLDA
STERVPENRPDAALVNALHMAWELFGDSEAILVNLVASTSPFMFEGGYIYEELAKKSGDKIKVENYSLGDKSERQ
NSTKRLQLDMEDFSLRLDGRRVAVVYSGQSALGCYQIDEAGMEFRRIIERSTAIKVPSLAVAISSSKKVQQMLAM
PGALERFFPNPEDAATVADIRATFADLWGLEHNDEKTQRIIQDAIENPGNYVLKPNRECGGNNFYDEEIAEKLKE
FTPNERAAHILMQKLQPMVVKNYLVRPFDEPKLEDVIPELGVYGFLLGNLHDGRVLHNAHQGYHFRTKLSHVNEA
GISAGFGYYDTAYLF
>Hav-gss8
MTNTAILIVFGSFFMMISIRFGHSSPPQQQNVTENEIALINEQEIKRITRFAIEWCHNNGLITARIESKGRSDLS
GFPPITLFPSPFPRIAFEQALNVQKAMNLLYFRVASDFDFLMDAYKDVIESDKLIGKLVDLLKEVHEEGIRQPIT
LMLQRADYLLNVVENNETKEVKYEPKQVEVNTGAIGATGLKRRTTELHRRMVQKAGMDASKLRMPDNEPDYAKVE
ALYKAWLLFNDPTALVLFLLYEDTPFKYDFLYIEEELHRISGGKLKVERYLLADQSENLSHAKRLQLDPENLSLR
LDGRKIAVAYSSITTLGCKLDEHGLELQRIIERSTAIKAPSMFVALTGSRKVQQMLAMPGAIERFFPAPEDAETV
AQIRATFAAHWGLEKEDEKTQKLIEDAIANPGNFVLKPNRECGGNTFYDEKLVEKLRGFSPSERAAHILMQKLRP
MAVKNYVLRPYEDPQLAEVIPELGVYGFLIGDLKDGRVLHNVQQGYHFRTKLSHVNEAGISAGFGYYDTAYLF
>Hsc-gss13
IRLFVGIVLVISFFVICASSGDIAPSQHDDEDTPVEETELINEQEINRLAKFAIEWCHNNGLIMRQLLGSDHKIG
GSPPLEIIKERGDIATFPPLTLFPSPFPRSAFEQAMNVQKAMNLLYFRVACDFDFLMAAFKDVTKADYHIAKMVE
LAKEIHEEGIRQPITVMLQRADYLLDVVENNETNEVKYEPKQVEVNTGAIGATGLKRRTTEFHRRMLKKATGMDA
TTANIPDNKPDAALIDTFYMAWRKFDDPKAIMVCLIYNNDPFQYDLRYIAEELEKKSAGKMEVEIYSLADYSERE
NSTKRLQLDPEDFSLRLDGRKVAIVYSGQSALGCKFDELGMEFRQIIELSTAIKAPSLAVAISSSKKVQQMLAMP
GAIERFFPEPSDAATVAQIRATFANIWGLENLDDDTQKLIEGAIENPGNYVLKPNRECGGHNYYDDKLVEKLKGF
KQTERGAHILMQKLRPMVVKNYVLRPYEAARLEEVIPELGVYGFLIGDLKEGHVLHNVQQGYHFRTKLSHVNEGG
ISAGFGFYDTAYLF
>Hav-gss9
MIKVNGSSFIIVFLLTSFRGGNAVSHNHVVIKGTQTVQNYVEKVAKDENQLRELAQFAVDWSQTHALIIRNLYSE
GKLIAPEFRCDLAEFASVTLFPTPFPREAYNKIVHVQQAMNLLYFRVARDYEFMMDAYKEVVKTDSHIRALVNII
RDVHKEGIKQPYTVMIQRADYMLNVVGVHEYEVKQVEVNTGAIGSLALDRKITELHTAMLRKVGMNASKEVVPMN
KPDEELINVLYMAWKKFGDPNAIVVILTYIKYSPYKFDYTNIEMELARVSNGQIKVEYFSLSEGKKLTLDHETFK
LRLNDRVVAVVYSNLSGLGYQANAAEMETRRTIERSTAIKAPSLAVAISSSKKIQQLLAKPGVLERFFPRPSDVH
TIAAIRETFTGIWGMENDDYSTRKLIKNAIENPSNYVLKPNRECGGNNFFDEDVAQKLQQFTPEQRAAHILMQRL
RPMQVKNYFLRPFHEPKLSTTSGELGVYGFLMGNMVDGTVRHNVQNGHLLRTKLAHVNEGGVIEGASVGDTPYLF
>Gpa_GSS18
MIKINGSSFIIVFLLTSFRGNAVSHNHVAIEGTKEAMQDYIENVIKNERELRELSQFAIEWSHNHALIVRTSGTQ
ILDYKSDTAEFASVTLLPSPFPRKAFSQTLAVQTAMNLLYFRIASDYDFLMDAYKDVVKSDKHIRALVSIVKDVH
EEGIKQPYTVMIQRADYMLNVRDNHDYEVKQVEVNCGSIVSLTLDRKITELHRAMLKKVGMDASDRFVPVNKAAE
EFINVLYMAWKQYDDPNAIVVILTFIDFSPYKFDYTHIELELARVSDGQIKVEYLSLREGKKLSLDPETFTLRLN
GRVVAVVNSGTSALGYIANEAEMETRRTIERSNAIKAPSLAIAISSSKKIQQLLANAGVLERFFPHPEDAQTVAA
IRETFAGLWGLEHDDQQTQNRIKDAIENPRNYVLKPNRDCGGYNFFDEDVAIKLKEFTPDERAAHILMQRLHPMQ
YKNYFLRPFNEPEMNVVTGELGVHGFLMGNMLDGTVLRNVQQGHLLRTKLAHVNEGGVIEGAGVGDSPYNH
>Gro-gss8
MIKINGSSFIIVFMLTSFRGSAVSHNHVAIEGTKEAMQDYVENVTKNKMELRELSQFAIEWSHNHALIVRTSWKK
DPDYSTDVAEFASVTLLPSPFPRKAFSQTLAVQTAMNLLYFRIANDYDFLMDAYKDVVKSDKHIRALVSIIKDVH
EEGIKQPYTVMIQRADYMLNVIDNNDYEVKQVEVNCGSIVSLALDRKITELHRAMLKKVGMDASDRFVPVNKPDE
EFINVLYMAWKQYGDPNAIVVILTFITHSPYKFDYANIELELARVSDGQIKVEYLSLKEGEKLSLDPETFTLRLN
GRVVAVVNSGTSALGYLANEAAMETRRTIERSNAIKAPSLAIAISSSKKIQQLLANPGVLERFFPHSEDAQTVAA
IRKTFAGLWGLEHDDQQTQKIIKDAIENPSNYVLKPNRDCGGYNFFDEDVAKKLTEFTPDERAAHILMQRLHPMQ
FKNYFLRPFNELTLNVVTGELGVYGFLMGNMLDGTVLHNVQQGHLLRTKLAHVNEGGVIEGAGVGDTPYLF
>Gpa_GSS19
MLITLIRTFVLCLNSIDSTEITSMRNYPENSVQSEKELHELTQFTIEWAHNNGLILRASEYKTTSDIAEFAPVSL
FPSPFPHQAFDQIVAVHTAMQLLYFRVGNDLEFLLNAYNDVIETDRHIREMVKIVREAHEEGIKQPITLLIMRAD
YMLNSLKDSENDNEQQQQLEVKQIEVNTGAILALGIDHRTTELHRQVLKRAGLNTSNSPDNVGDSNLAESLFMAW
KAFDNPKALMVFYVSAFSPYKFDLHQLARKLKRLSNDQMDIEHVSLKDGPTQLQLGDDFSLLLNGKVVGVIYSCI
SALGTVLPAELLEVRRTIERSTAIKAPSLAHAISSSKKIQQLLAMPGAVERFFPDPADADKVAAIRETFAELWGL

DKNDDATERLIENAIEHPEKYVLKPNRECGGNNFYDEKLADKLRSLPQNERVSYILMQKLNPTTFKNYFLRPFHE
PKLSTVVGELGIYGTLMGNILDGSVQHNVQSGHLLRTKLAGVNEGGISVGTGVGDSPYLF
>Gpa_GSS20
MFIILSSIFLLCTKFTNSSFSIDSTEITSMRNYPENSVQSEKELHELTQFTIEWAHNNGLILRASEYKTTSDIAE
FAPAMQLLYFRVGNDLEFLLNAYNDVIKTDRHIREMVKIVREAHEEGIKQPITLLIMRADYMLNSLKDSENDNEQ
QQQLEVKQIEVNTGAIVALGIDHRTTELHRQVLKRAGLNTLNSPDNVGDTNLAESLFMAWKAFGNSKALMVFLTV
PSFGYKFDLHQLACKLKRLSNDQMDIEYVSHKDGQTQLKLGDDFSLLLNGKVVGVVYSCISALGYLITAASMEVR
RTIERSTAIKAPSLAHAISSSKKIQQLLAMPGEVERFFPDPADADKVAAIRETFAELWGLDKNDDATERLIENAI
EHPEKYVLKPNRECGGNNFYDEKLADKLRSLPQNERVSYILMQKLSPTTFKNYFLRPFHEPKLSTVVGELGIYGT
LVGNILDGSVQHNVQSGHLLRTKLAGVNEGGISVGTGVGDSPYLF
>Gpa_GSS21
MFIILSSIFLLCTKFTNSSFSIDSTEITSMRNYPENSVQSEKELHELTQFTIEWAHNNGLILRASEYKTTSDIAE
FAPVSLFPSPFPHQAFDQIVAVHTTDRHIREMVKIVREAHEEGIKQPITLVIMRADYMLNSLKDSENDNEQQQQL
EVKQIEVNTGAIVALGIDHRTTELHRQVLKRAGLNTLNSPDNVGDTNLAESLFMAWKAFGNSKALMVFLTVPSFG
YKFDLHQLACKLKRLSNDQMDIEYVSHKDGQTQLKLGDDFSLLLNGKVVGVIYSCISALGYVITAASMEVRRTIE
RSTAIKAPSLAHAISSSKKIQQLLAMPGEVERFFPDPADADKVAAIRETFAELWGLDKNDDKTERLIENAIEHPE
KYVLKPNRECGGNNFYDEKVSYILMQKLNPTTFKNYFLRPFHEPKLSTVVGELGIYGTLVGNILDGSVQHNVQSG
HLLRTKLAGVNEGGISVGTGVGDSPYLF
>Gro-gss9
MNFAKFIFFFLGIFLCANFAVCNDLEDYVEKSVHSETKLHELVDFAIEWAHNNGLIMRSKEIYDMAEFAPVSLLP
SLFPRDVAEFAPVSLLPSLFPRHAFQKAVAVQQAMQLLYFRVACDYEFMMDAYKDVVTTDNHLQQLVNIVKDAHE
QGIKQPITLLIMRADYMLNTLTSQTNDKEFELKQIEVNSGAIGGLIIDRRTTELHQQMLRKLGMDTSNSPVNNGD
SNLIKSLFMAWEAFGNKNALFVFLTHADSRYRFELRDMALQLEQMSNGQMKVEYISLKDGYEQLKLGEDFSLLLN
GKIVGVVYSRISALGYMANAQGMEARRTIELSNAIKAPSLAIAISSSKKIQQLLAMPGTLERFFPDPADADNVAA
IRETFAGLWALDKNDEQTKRVIKDAIENPGKYVLKPNRECGGNNFYDEALAEKLRTMPPTERALHILMQKLTPNA
TKNYFLRPFREPTLSVVVGELGVYGTLLGNMQNQNVWHNVQSGHLLRTKLEEANEGGISAGTGVGDSPYLF
>Gpa_GSS22
MNCDNAKFIIFFFFIIFLCANFAVCNELEDYVEKSVNSETKLHKLADFAIDWAHNNGLILRTKQFLNKSDVAEFA
PVSLLPSPFPRHAFEKAVAVHEALQLLYFRVACDYEFMMDAYKDVVNTDNHLRQLVNIIKDAHKQGIKQPTTLLI
MRADYMLNTLNSKGNDDEYELKQVEVNTGAIGGLGIDRRTTELHRQMLRKVGMDTSNSPANNGDSNMIESLFMAW
EAFGNKNALFVFLSHERLQYKFELRNIQCQLEELSNGQMKVEYVSLKAGYEQLKLGEDYSLLLNGEIVGVVYSTI
SALGHQANAREMEARRTIELSNAIKAPSLAIAISSSKKIQQLLTTPGTLERFFPSATEADKVAAIRETFTGLWGL
EKSDDQTERRIKDAIENPANYVLKSNGECGGNNFYDEALAEKLRTMPQAERASHILMQKLIPMATKNYFLRPFHE
PKLNVVVGELGVNGTLLGNLRDQSVRHNVQSGHLLRTKLREANEGGISVGTGVGDSPYLF
>Gpa_GSS23
MNCGNAKNVFFFFIIFMTANFAVCNEIEDYVEKSVNSETKLHKLADFAIDWAHNNGLILRSKQFLDKSDVAEFAP
VSLLPSPFPRHAFEKAVAVHKALQLLYFRVACDYEFMMDAYKDVVNTDNHLRQLVNIVKDAHEKGIKQPNTLLIM
RADYMVNTLNSKGNDNEFELKQIEVNTGAIGGLGIDRRTTELHRQMLRKVGMDTSNLPANNGDSNLTKSLFMAWE
AFGNKNALFVFLTHDRFQYKFELRNIECQFEKLSNGQMKVEYVSLKAGYEQLKLGEDFSLLLNGEIVGVVYSLIS
ALGHQANAQEMEARRTIELSNAIKAPSLAIAISSSKKIQQLLATPGTLERFFPSATEADNVAAIRETFAELWGLE
KSDEQTERVIKDAIENPRNYVLKPNGECGGNNFYDEALVEKLRTMSPTERASHILMQKLFPMATKNYFLRPFLEP
KLSVVVGELGVYGTLLGNMHNQSVWHNVQSGHLLRTKLEEVNEGGISVGTGVGDSPYLF
>Gro-gss10
MSLLYFRISRDFDFLKMAYKDVIQSDRSVRMYMKLLEDIKKEGIKQPLSILTQRSDYMIHVSTDPHTNEPEYQLK
QIEVNGGSIGTAGCIERLAKLHRRVLEKSGYSKRAIDNALPENRSGTALALTIFKAWEQLKDPKAIIVFMVVKRN
CWFFVHRYDEYELERLSGGRAKIVHLTLAECAKNLTLDDDFTLRLDGRRVGIVYINRVMLGVDYPPKLFAAIRMI
ERSTAIKATSLFFGMSASKKIQQLLAMPGMVERFFPDPSEAQMVADIRNTFAKMWGLENDDEQTRMVIEDAIAHP
ERYVLKPNKEGGSENFWGQDIADKLKTFTPSERAAHILMERLNPMITKNYMVFPFKQAKLTEVNNELGIYGYILG
NMETGTVLHYEQPGNMVRTKDMQKNEGGVSSGDGVLDSPFLY
>Hsc-gss14
MFWQFVCFCFLFSSTCFLDGALTETKTANQLGYDPESIHALVSDAIDWAHEVFLVLRVSGQKHRSDRAQFVPFSL
FPSPIPRKMYEQALSVQKAMSLLYFRIASDFEFLKMAYKDVIESDKSVRTLLGILEDIKKEGIKQPISIFLQRSD
YMITAETNSKSNQQNYQLKQIEVNGGSIGSAGCQDRLLSIHQRMLKHSGCSDQMINNALPKNRSGAAIAEIIYKA
WKLLNDPRAIILFVVVKDLSTWHFSKRYDEYELERLSGGRAKIVHLTTVECFENLKLDDDFTLRLDGRPIGIAYW
NLVRLGDDTFGHKSLAALRMIERSTAIKATSLFFELSTSKKIQQLLAKPGMVERFFTDPSEQQMVAAIRATFAKL
WGLENNDEETQKIIQDAIAHPERYVLKPNKEGGGGNIWGEEIAQKLSKFSRSELAAHILMERINPVTVKNFMVWP
FKRAEFAEVINELGIYGYLIGNMKTGEVLEYEQPGNMVRTKNMHNNEGGVSSGNGVLDTPFLY
>Gro-gss11
MTHDVQALVDDAIDWAHNLFMKMRTPDHYGRSDVAQFAPFTLFPSPIPRKFYDQAIAVQKAMSLLYFRIACDFDF
LKMAYNDVIESDASVRMYMKFLEEMKTEGIKQPLAIFLQRSDYMVHESYDNQTNKPKYELKQIEVNGGSVGTACM
SQQVRLLHARVLQKAGVPDAFIDSVLAKNQSSKALNRMLYQAWLTYGDPNAIILFMDNKTKSPWHFANYHDHYEL

ERLSNGKAKIVHLALNEFKNLTMDEDFTLRLGGLPVAVAYKNLIFMGSILAPATFNMIRMIERSKAIKATSLFFE
FCTTKKVQQLLAMPGMVERFFPDPSEAQMIADIRNTFAKLWGLENDDEQTRMIIEDAIAHPERYVLKPNKEGGGK
NFWGQDIVDKLKTFNRSERAAHILMERLNPVPTKNFMVWPNKETQLSDVVNEIGIFGYIFGNLKDGTVVYYEQNG
NMIRTKLADSNEGGVSAGTGAFDTPYLYD
>Gpa_GSS24
MYAKSWHFCWLIAGIIHISLITRALSQTTHDIQALVDDAIDWAHNVFIIMRTPDHYGRSDVAQFAPFTLFPSPFP
RKFYDQAMAVQKAMSLLYFRIACDFDFLKMAYKDVIQSDKSVRQFMELLEEIKKEGIKQPLALFFQRSDYMIHES
YDEQTNKQKLELKQIEVNGTSIGTACLSQQVRLLHKRVLQKAGVSDAFIESVLPENQSSKAMDRMIYQAWLTYGD
PNAIILFMDGKKSSWHFVQSHEHYELERLTNGKAKIVHLDCNSEFKNLTMDEDFTLRLDGRPVAVAYKNMIFLGY
TSTPEEYHYIRMIERSKAIKVSSLFLELSTSKKVQQLLAMPGMVERFFPDPSEAQMVADIRNTFAKLWGLENDNE
QTRRVIEDAIAHPERYVLKPNKEGGGKNFWGQDIADKLKTFTQTERAAHILMERLNPLSTKNFLVRPNKETLFSD
VVNEIGIYGFIFGNLEHGTVVHYEQNGYMLRTKLADSNEGGVSAGSGALDSPYFYN
>Gro-gss12
MRTSAHYGRSDVAQFAPFTLFPSPIPRKFYDQAMAVQKAMSLLYFRIACDFDFLKMAYKDVIQSDKSVRQFMELL
EEIKKEGIKQPLALFFQRSDYMVHQSYDEQTNKPKFELKQIEVNGASIGTACLPQQTRLLHKRVLEKAGVSDAFI
ESVLPENQSSKTMNRMIYEAWLKYGDPNAIILFMDGKKSTWHFVQCYDHYELERLTGGKAKIVHLDCNSEFNNLT
MDEDFTLRLDGRPVAVAYKNMLFMGYTSSPEEFHFIRMIERSKAIKASSLFLEFSTSKKIQQKASSLFLEFSTSK
KIQQLLALPGMVERFFPDPSEAQMVADIRNTFAKLWGLENDDEQTRMVIQDAIAHPERYVLKPNKEGGGKNFWGQ
DIVDKLKTFTPSERAAHILMERLNPVSTKNFLVLPNKETQFSDVVNELGIYGFIFGNLVDGTVVHYEQNGNVIRT
KLAGSNEGGLSTGSGADAIAHPERYVLKPNKEGGGKNFWGQDIVDKLKTFTPSERAAHILMERLNPVSTKQNGNV
IRTKLAGSNEGGLSTGSGAVDSPYLYG
>Hav-gss10
MAAAFIFNLLIAILCCQQCTATNDTDAEKKAETQPKTSTSAQAINTNEVNYIFNDDDVQALLYDAIDYSHKVFLI
TRLPEERNKSDQSVFAPFTLFPSPYPREQFQQAIDVAKAMSLLYFRISRDFDFLKSVYKDLMETEPAIAQYMNMC
EEVKKVGIKQPLSVYLQRSDYFVHVNSEGKYELKQIEANGGSVGGANGLGPRVTQIHERVMKKAGFPNLPEDVLP
RNPNAKAGAAKAIVTAWKKYNKPSAIIVFLVVKATSFWHFLKRYDEYEIEQLTNHKAKVVHLTMGECMRDMTMDE
NFTLLLKGVPVAVVYTNIVLTGVKVSPKILATIQMIEQSTAIKAPSMFYDLSMTKKVQQVLCQPGMVERFFPKPE
EAPMVEAIRKTFAKMWSLDGEDEETNKVIEDAIAHPERYVMKPNKEGGGKNFWGNDIVEKLTTLTAKERGSYILM
EKLNPVTVKNFLVWPMSDEVTCDDVVMELGVYGFMVGNMVDGTVPYFDQPGHLVRTKLSSSNEGGISKGTGAFDS
VYLY
>Hsc-gss15
STAPTFVLDEESVQTLMEDAIDYSHKVFLVTRLPTAKDKSNMSVFAPFTLFPSPYPREMYKQAIDVAKAMSLLYF
RISRDLDFMKMVYKDVIASNSSIEQYMGFCEEMHAQGFNKQPLAIYLHRSDYFVHINKDGEFELKQIEFNSGSVG
GANGLAPRVTEIHERVMRKAGFPNLPEDVLPRNSNSKAGAAKILVAAWRRFKNPAAIIVSIVHKQYSYWHFLKRY
DEYEIDELTKNKVKIVYLTVFECAKYLTLDDDLTLRLKGEPVAVVYANVVMTGHKMLPETLALFKMIERSTAFYS
STVCADLSQTKIIQQVLTRPGMVEKFFPSPKEAPMVAAIRKTFAKIWALDNNDDDTKAIIADAIAFPDRYVLKPN
REGGGKNFWGQDIVDKLSQFTQKERASFILMEKLNPLTVKNYLVWPNRDEAAFDDVVMELGVYGFMLGNRVDGTV
PYFDQPGHLVRTKLASSNEGGISVGTGAFDSVYLY
>Hav-gss11
MNKIILLFAFLTFSSFLFVAINGTPTDNPSKSGTDEKVEPRPAGNNVMIDQNNIDTLVLDAIDWAHHIFLVMRPP
NNQHRSDLVQNVPFTLFPSPFFPREMFQQAVDVAKAMSLLYFRVSRDIEFLKMVYKDVIHTDVSIRNYLKICEEVY
NEGIKQPISIYLQRSDYMVHVSEGEDGNKKYELKQIEVNGGSIGGANGIPPRITQIHERVMKKAGFPNLPEDVLP
RNTEARSASAQMLVTAWKKFNNPKAIIVSLVIKDNSKWHFCKRYDEYEIDRITNNKIKVVYLSFFEAVKLLTMDD
DFTLRLEGKPVGVFYINMILIGANLHKQILEMLKMVERSTAIKSPSLFYEISISKKVQQVLAMPGMVERFFPNPE
EAPMVTAIRKTFAKLWGLENDDEETQRVIKDAIAHPERYVMKPNKEGGGKNFWEQELADKLRSFTPKQRAAYILM
ERLNQMTAKNYLIWPMEKVIYDEVATELGIYTYCVYNTKDGTLVQYSQPGQMTRTKLASSNEGGISVGTGVFDSL
YLY
>Gpa_GSS25
MKMINFNSLVIFIFLFGINNGANTNDPETTGKVEAITENPFDFDEENVQTLVEDAIDWAHNIFLVMRTPDHKDRS
DVVQNVPFTLFPSPFPREMFKEAVDVAKAMSLLYFRVSRDFEFLKMVYKDVRETDVSIRNYMKICEEVYNEGIKQ
PVSIYLQRSDYMVHVKEDGAEKKFELKQIEVNGGSIGGANGIPPRIAEIHARTLAKAGMPNLPEDVLPRSKEARS
ASASMLLTAWNKFNNPKAIIVSLVLKDRSKWHFCKRFDEYEIDRITKNRVKIVYLSISEAVQMLRMDEDFTLRLE
GKPVAVFYINIILIGTILSPRILEMLKMAERSTAIKSPSLFYELSISKKVQQVLTLPGMVERFFPDPKEAPMVQG
IRKTFARMWGLENDDQETREIIADAIAHPERYVLKPNQEGGGKNFWEKELADKLRTLTPKQRAAFILMERLNPLV
VQNYLIWPMEKAIYSDVVTEVGVYTYCVYNTKDGTLVHYTQPGQMNRTKLASSNEGGISAGTGVFDSLYLY
>Gro-gss13
MKMINFNSLVIFIFLFEINSGTEPKNPETTGKVEASTEKPFDFDEENVQTLVEDAIDWAHNIFLVMRTPEHKDRS
DVVQNVPFTLFPSPFPREMFKEAVDVAKAMSLLYFRVSRDFEFLKMVYKDVLETDESIRNYMKICEEVYNEGIKQ
PVSIYLQRSDYLVHVKEDGAEKKFELKQIEVNGGSIGGANGIPPRITEIHARMLEKAGMPNLPEDVLPRSKQARS
ASASMLVTAWNKFNNPNAIIVSLVLKDKSKWHFCKRFDEYEIDRMTNKRVKIVYLSISEAIKMLRLDEDFTLRLE
GKPVAVFYINIILIGVILSQKTLEMLKMAERSTAIKSPSLFYELSISKKVQQVLTLPGMVERFFPDPKEAPMVQG

IRKTFAKLWGLENDDSETREIIADAIAHPERYVMKPNQEGGGKNFWEKELADKLRTLTLKERAAFILMERLNPLV
VKNYLIWPMEKAIYSDVDGTLLRYTQPGQITRTKLASSNESGITAGTGVVDSLYLY
>Gpa_GSS26
MIAPQKYLIQIILFGIFLFNIFNAISAQPKKVDAKHGKIPAHNNGHISHKMPKTDADKSPTTTAEKHTDYLKSKT
RKRHYENGQNDHVHHHTSKRAKHYDVPNIGKNVNDLQVLIEDAIDWAQSNGLILRTLEHPDKSDIARIAPFTLFP
TPFPRQLFQQAVDVQKAMQMLYFRITADLDFLRKVHEDVIKTDPVVQSFMEIIEKVHEEGVQQPITLFIQRADYM
LHIKNGEATNEKEYELKQIEVNGCAAGGAGFSTHITNLHRRMLKKAGIYASRDNVPDNRPARMTAEALYEAWKQF
GNNNAVLLFLVNKTDLVQFDRRLIQYEFERVSRDKVDVVRLSLEECSEKLKLDPVDFSLRLVDDGRAVAVVFNQV
LMLGSTPTHMELAARLMIERSTAIKAPTLAFAMSHSKKIQQVLTRPGMVERFFSGPNEAHMAAQIRKTFAGLWGF
EADQTKNNELIQKATKNPERFVLKPIGEGCGTHFNYFDDDIPNKLAELSPTELTEFILMEKLKPKVYKNHLVRAL
RPTLFNTEVTPELGIYGSLIGDMRTGEILYNKQEGYTFKTKLATENEGGICSGTGVVDAPFLVDN
>Gpa_GSS27
MGAASKKCFIQIIIVSAIILLHSKQIGINAHPMNEASGIWHDEQSETGTITVFENEHSINMSLDNNNIEMSLNRP
TQNGHVYDVPSLVKNVNELQVLIEDAVDWAQSNGLLLRTRDHLDKSDVAQIAPFTLFPTPFPRQLFQQAVDVQKA
MQMLYFRITTDLDFLRKVHEDVIKTDKVMQSFMEIIEKVHEEGVQQPITLFIQRADYMLHIKNGEATNEKEYELK
QIEVNGCAAGGAGFSTRITDLHRRMLKKAGIYASKDNVPDNRSDRMTAEALYEAWKQFGNTNAVLLFLVNKADLT
QFDRRIIQYEFERVSRDEVDVVRLSLEECSEKLKLDPVDFSLRLVDDGRPVAVVFNQVLMLGASPTRMELAARLM
IERSTAIKAPTLAFALSHSKKIQQVLTRPGMVERFFSGPNEAHMAAQIRKTFAGLWGFEADQTKNNELIQMAIKN
PERFVLKPIGEGCGAHFNYFDDDIPKKLAKLSPTELTEFILMEKLKPKVYKNHLVRALRPTLFNTEVTPELGIYG
SLIGDMTTGRILYNKQEGHTFKTKLATENEGGICSGTGAVDAPFLVDN
>Gpa_GSS28
MKHLIQIYLVTLXFFSNYSVAESPLPKNGGSGSKEVLISSQHDKLNLSTNYTDNLLTILQRDRTKTELLIKKATE
WAQKIGLLIRKQSYPGYGTGVAPFTLFPSPFPRKFYEQAVNVQTALNLLYFRIMRDYPFLKEIYQKLIKFDQTLS
TALKIMEEIHSEGIKQPLTVLFQRADYMLCESNYEGNENPSYELKQVEVNGSAIGGLGFATKTSKLHQQILSEMG
LDLSNSVENNTRTMAVEAIYQAWQKFGDPKAIIILIFDERYAFFFYERSNLYFELKNKFEGQTEIVALNLNQCAK
LLKLDPHDFTLRYDDKIVAVVFNQETMISADTKKMEARRTIERSTAIKAPSLAAALAHTKKVQQVLAKPGMVERF
FPNPEEAPLIDAIRNTYVNFWTVEEDNNEYTQIIQAVKKNPHNFVLKKTEYALNNQNLNPIYFGEEIVKSIANFT
PPEPYILMEKLQSTIVKNHIVKTMFDTKQNVPTIFDIVVKTIFDKKKNVPPSILEDVTPELGIFGTLLGNIVDGQ
VLHNAQMGSKIKTKLASENEGGLDRGQSAYDSAYLVD
>Gpa_GSS29
MINLIQFSFTLFLLNYSVVESAPPKKVSGKEVLISSQHDELNSSTNYTDILTTFLQRDHAKTQLLIQEAKDWAQN
IGLIMREPKFDANLLWWFYQTVVAPFTLFPSPFPRKFFEQAANVQTALNLLYFRIMRDYPFLKEVYRNLIKHEQP
LSSALQIMEEIHLEGIKQPLTVLFQRADYMLCESNYEGNEKPSFELKQIEVNGSAIGGLGFSTRTSKLHRQILSK
TGLDLSKSVENNTSTLTVEAIYQAWQKFGDPKAIIILIFDEAYAFVYYERTGLYFDLADKFEGKTEIIALNLKHC
AEFLKLDPHDFTLRYDNKIVAVVFNQDVMLSTDPGKMEARRTIERSTAIKAPSLVAALAHTKKVQQVLAKPGMVE
RFFPNPEEAPLIEAIRKTYANLWTIEEDDNNDYPQIIQAVKKNPHKFVLKKIEYAQYRNRNLERIYFKEEILKSM
TKFTPIERSAYILMEKLQPIIVKNHIVKTIFDENMNVPPSTFEDVTPELGIFGTLLGNIVDGKVLHNVQLGHQLK
TKLASENEGSIALGKSVYDSAYLVD
>Gro-gss14
MINLIQLSFLYTLFLLNYSVVESKNVSGKEVLISSQHDELNSSTNYTDILTTFLQRDDAKTQLLIKEAKDWAQNI
ELIIREPKFDAKWFYQTVVAPFTLFPSPFPRKFFEQAANVQTALNLLYFRIMRDYPFLKEVYQKLIKNEQPLSSA
LQIMEEIHLEGIKQPLTVLFQRADYMLCESNYEGNEKPSFELKQIEVNGSAIGGLGYSTRTSKLHRQILSKTGLN
LSNSVENNTSALTVEAIYQAWQKFGDPKAIIIFIFDEAFFAYYERIGLYFELADKFEGKTEIIALNLNLKLDPHD
FTLRYDDKIVAVVFNQDNMLSTDPKKMESRRTIERSTAIKAPSLAAALAHTKKVQQVLAKPGMVERFFPNPEEAP
LIEAIRKTYANLWTIEEDDNNEYPQIIEEVKKDPHKFVLKKIEYAQYQNRNLARIYFEQEILKSMTNFTPIERSA
YILMEKLQPTIVKNHIVKTMFDENMNVPPSIFEDVTPELGIFGTLLGNIVDGKVLHNVQLGHQMKTKLASENEGG
IARGKSAYDSAYLID
>Hav-gss14
MKIQIYAIIFSCFCFINIGSATPITTEDEESQQHNQQYCVSDIEHDPQILREQGLDAKDWALSNGLVMFVNASCT
SCGKQSKSIITQHVPVAMYPSPFPKKLFQQAVELQKAMLLLYFRASNDFQFLKEAHHQLLEMEGPNKTKRLVEGL
EGLYKEGIRQPLAMFCQRTDYVASKSDHNEYVLKQVGVTTGAVDSFAISPRVSELHQRMLKNAGIDATDEVTPLS
TTDHMIAETLYQAWLQFGNPEAVIIVLHQSKHSNLMLESRQIEHQLEQISPVAIECRFITINDGLTRLKVDPFDF
TLILDDKYVVAVVFNRVVNEELSNEEADLAFAFERSTAIKTPPFVFALSHTDKMQQYLTKPGNVERFFTHPKEEH
MAEEIRTVQTKRWALGDDKNEAEEIKKKALENPKEYVLLKTDQSGQSTGQTMFFDEDIPKELARMTPAEHNYYFI
MEKMRPMVIKNHFVRPNVEPLLNVEATSELGIFGCLIGNTVTGQVTYMSSTESYIMKTKMANVNEHSDLREKSVA
DSVYLV
>Hav-gss15
MMKFEYKIIFLFYIIHIASAKPSDDDSDQEQAHEQYCVSDIEHDPKILLEQSLDAKDWSLSNGLIKLQKLICGDC
SKKTKNLKNIVADSIPISLYPSPFPGKLFHQAMEVHKAMLLLYFRVANDFQFLKEAHHQLLESEEKNKMKTIIPK
LETLNKEGIRQPVAMFCQRSDYMASQNDHGEYVLKQVEVNTGAIGGIGACSRVSQLHKRMIKNAGIDASESVMPL
DQTDLLFAETLYESCIKVDPVDFSLILDNKFVVAIVFDRLGGLITREEAALNFEFERSTAIKTPPYMFAVSHTKK

MQQYLTKPGMVERFFTDPKEAHYVEAIRKVQTMGWSLGDGKEEAEEAKKRALANPEKYVLKSNECGPNIHPKMFF
NEDIPKKMEKLTPVEHHYFFLMEKLRPMMVKNHFVRPNTGLSLNVDVTPEIGIYGCLIGNTDTGEVSYISRFGHT
MKTKLAEVDEGGILRGASVGDTPYLI
>Hav-gss16
MMKFEYKIIFLFYIIHIASAKPSDDDSDQEQAHEQYCVSDIEHDPKILLEQSLDAKDWSLSNGLIKLQKLICGDC
SKKTKNLKNIVADSIPISLYPSPFPGKLFHQAMEVHKAMLLLYFRVANDFQFLKEAHHQLLESEEKNKMKTIIPK
LETLNKEGIRQPVAMFCQRSDYMASQNDHGEYVLKQVEVNTGAIGGIGACSRVSQLHKRMIKNAGIDASESVMPL
DQTDLLFAETLYESWKQFGNPEAVILFVHGEESPILLDSRKIQYQLEKISSERIMCRFITLKEGLTRIKVDPVDF
SLILDNKFVVAIVFDRLGGLITREEAALNFEFERSTAIKTPPYMFAVSHTKKMQQYLTKPGMVERFFTDPKEAHY
VEAIRKVQTMGWSLGDGKEEAEEAKKRALANPEKYVLKSNECGPNIHPKMFFNEDIPKKMEKLTPVEHHYFFLME
KLRPMMVKNHFVRPNTGLSLNVDVTPEIGIYGCLIGNTDTGEVSYISRFGHTMKTKLAEVDEGGILRGASVGDTP
YLI
>Gpa_GSS30
MANIFLILFLFCFINFGNATPTTHNQDKTDKEEQYCVPNIEQDPQILLEQSLDAKDWALSNGLVKFVDVPTCPEC
GKKTKKMMTQFLPLSLYPSPFPRKLFQQAVDVQKAMLLLYFRASCDYEFLKEAHREVLNSELDNGIKKLVKRLDG
MLSDGIRQPVAMFCQRADYMASQEDDGQYVLKQVEVNTGAIGSFGTTPRFSRLHRRMVSNAGIDASESVMPSDQT
DTMAAETLYQAWLEFGNAEAVILFLHGSPNSHLMLESRQITHQLESISTERIKCRFITITEGLNRLKRDPNNFSL
ILDDKFVVAVVFDRLGGAVTKEEMDLNFVIDHSTAIKTPPYIFALSHTKRMQQVFTKPGMVEKFFNNPEEEHMAE
AIRKVQTKGWAIGKDEDLTEDIIKKATENPHRYVLKNNGCSSNAADMFFNEDILKKLKTMAPADRDFYYLTEKLR
PMVIKNHFVRPNMAPTLNLDATPELGIFGCLLGNMETGKVSYFSRTGHMMKSKLANVDEGGVWKGFSVYDSPYLV
>Gro-gss16
MNMANILLILFLFCFINFGNATPTTHNQDKTDKEQQYCVSNIEQDPQILLEQCLDAKDWALSNGLVKFVVVPTCP
ECRTKTEKPNVMTQFSPLSLYPSPFPRKLFQQAVDVQKAMLLLYFRASCDYEFLKEAHREVLNSNNANYIKTIAT
NLDGMFSEKIRQPVTMFCQRADYMASKNDDEQYVLKQVEVNAGAIGCFAIASRFTRLHRRMVSNSGLDASEAVMP
SDQTDALVAETLYQAWLEFGNAESVILFLHGGPNSHLMLESRQIQHQLESISTEGIKCRFITIKEGLNRLSLDPD
NFSLILDNKYVVAVVFDRIGVFLHKDEVDLLQITIARSTAIKAPSLALALTHTKRMQQVFTRPGMVEKFFNNPEE
AHMAKAIRQVQTKAWAIGEDKDVFTRPGMVEKFFNNPEEAHMAEAIRQVQTKAWAIGEDKDKATENPQRYVLKSN
ECSSEIAGNFFNEDIPKKLATMAPADRDFFLLTEKLRPMVVKNHFIRPNTEPALNVDATPELGIFGCLIGNMETG
KVSYFSRTGHMMKTKLASVDEGGVWKGYSVGDSPYLV
>Gro-gss17
MQLRRRTIRIKQTKNNQYCVSNIEQDPQILLEQCLDAKDWALSNGLVKLVVVPSCPECGKKMDKPNVMTQFSPLS
LYPSPFPRKLFQQAVDVQKAMLLLYFRASCDYEFLKEAYREVLNSEIENGVINLVPKLDKMFSEKIRQPVTMFCQ
RADYMASENVDGQYVLKQVEVNTGAIGCFAIASRFPRLHRRMVSNAGIDASESVMPLDQSDTMVAETLYQAWLAF
GNAEAMILFLHGGPNSHLMLESRQIQHQLENISTEGIKCRFITLKEGLNRLKRDPNNFSLILDDKFVVAVVFDRL
GVALSKEENDLNIEIDRSTAIKTPSIIFALSHTKRMQQVFTKPGMVEKFFNNPEEAHMAEAIRKVQTKGWAIGKD
EDLTEDIIKKATANPHRYVLKNNGCRLMSEDIFFNENIPKKLESMEPADRDFFYLTEKLRPMVVKNHFIRPNMEP
ALNVDATPELGIFGCLIGNMETGKVSYFSRTGHMMKTKLASVDEGGVWKGYSVGDSPYLV
>Hsc-gss16
FLFAICFVFLPPLAEATKNLEDIPSGSTNNEAEEDAQLMDLQFLMGQIEEAKKWALANGLTQRVGLFGAFNFAPF
SLFPSPFPRALFHKAVDVQKSLQLLYFRAMRDFDFLKEMHRDIIETNEKFRQMVDLTENCYKDGFKQPLIWFCQR
ADYMTHQSEEKLELKQVEVNAGPIGGLGASSRVTMLHQHVLSMANADTSPSALPPNHPDTMVAKTLHMGWNAFGN
SEAVILFIHAHSFDPRLNESHQVANEVERISNGQTKCVFLLLSEAVERLTRHPENFSLILDGQILVAVLHDCYTA
SRATPDQLKLIFEIIEQSTAIQNSYHLAMAHTKRMQQLFTLPGVVERFFPRSEETHMVKAIREVLTKSWSIGEGD
EEAEEIIKKVKMNPENYVMKWNTCGSPMSGKSIFFGDEIIKELDRMTNFERNNFIIMEKLRPMQVKNHFIFPDSA
HLNVAATPELGIFGCLLGNIEDGTVLQQFSGEAHQMKTKLASENEGGIWNGKSVYDSPLLV
>Gro-gss15
MAQLVPFSLFPSPFPRELFQQAMDVQKALNLLYFRAARDFGFLKEMHRDELKTNPSFKEQVDFLESMNRDGIRQP
LTLFCQRADYMTHESFAENNVKKVELKQVEVNTGPIGGFGTSSRVTALHRRFLTMLGVDASPSVVPENFTDTMMG
QALYRAWLQFGDREAAIIFLHSSRQDPRFIESRQVQHELERISKGQIKCIFLTLSKAINRLKLDPNNFSLILDDK
FVVAVALHRYSSATRAELMFSREIVRRSTAIQGTYFLLMAHTKRMQQIFTKSGVVERFFGAPGEAQMVTAIRNVL
TKSWSIGQGDEEADEILRKVKMNSERYVMKWNECGAGRRGPDIFFGADILRKLDNMTSAERNNFIIMEKLRPTVV
NNHFVRPDTKPLLNVEVTPELGVFGCLLGNMVDGTVLQHFGNASQMKTKLASEDEGGIWNGKSVYDSPYLV
>Hsc-gss17
LPPLSPATPTEQQPQQSADNVATTYCVENVEKDEVLLKQYVRDAKDWALAHGLVLPTMPFRCKNEECKEANKALN
IVSVDVVKIAPFTLFPSPFPRELFVQAQEVHKAIQLMYFRAASNFEFLKDVHKGMLKTDPNFIDTFKSIESRHLE
GIKQPLTLFCNRADYMTSKVIDEETNEEKYVLKQVETNHGAMGGHGTSPRITALHRRMLSIAGVDSSLSVVPKNE
TGKMFATALYKAWEAFGNNTKAVILFLHNPYRDYLLIEAREVQHDLERIAGGKVQCFFLTLKQAKERLTFPDDSD
SAILDNKYVVAVALNRYTVYATRDEIVTARALRRSNAIQVPSSVFILAQSKKMQQVFTNPGVVEKFFTGPGEAHL
AAEVRKVLTRSWSIGEDEEKAEEIIEMVKANPEKYVMKWNECGSRVEQTKFFGDKIPAKLESLSMEEREKFFIME
RLEPMVVKNHFVRPGTEVALNVNVTPELGIHGCVLGNILDGTVMEYFWPESQMKTKLAEEEEGGVMKGKSVFDSP
YLV

>Hsc-gss18
MNGLRPSKTFLILLSISICWQLSVVSSSPTTHDIAQSAEVQSESEEQLPPLSPATPTEQQPQQSADTVATTYCVE
NVEKDEVLLKQYVRDAKDWALAHGLVLPTMPFRCKNEECKEANKALNIVSVDVVKIAPFTLFPSPFPRELFVQAQ
EVHKAIQLMYFRAASNFEFLKDVHKGMLKTDPNFIDTFKSIESRHLEGIKQPLTLFCNRADYMTSKVIDEETNEE
KYVLKQVETNHGAMGGHGTSPRITALHRRMLSIAGVDSSLSVVPKNETGKMFATALYKAWEAFGNNTKAVILFLH
NPYRDYLLIEAREVQHDLERIAGGKVQCFFLTLKQAKERLTFPADSDSAILDNKYVVAVALNRYTVYATRDEIVT
ARALRRSNAIQVPSSVFILAQSKKMQQVFTNPGVVEKFFTGPGEAHLAAEVRKVLTRSWSIGEDEEKAEEIIEMV
KANPEKYVMKWNECGSRVDQTKFFGDKIPAKLESLSMEEREKFFIMERLEPMVVKNHFVRPGTEVALNVNVTPEL
GIHGCVLGNILDGTVMEYFWPESQMKTKLAEEEEGGVMKGKSVFDSPYLV
>Hav-gss13
MSEWSSNPTKCSFLLLFFITSICRFLNATDGSTGKYCVENIEKDETILKEQIRDAKDWALTNGLTFTVPKRIVKC
GGCEKEVEGIVLDDMAQMVPFSLYPSPFPRELFQEAMDVHKAFQLLYFRAMLAPNFDFVKEVHRNVLKVEKPLRE
FAASIDQMHVQGTRQPLILYCNRADYFASKVTNEETKEIKYELKQVEMNGGVVGGYGAPPGVSALHQRMLTKAGI
DTSPEVAPINRTNVMVATALHRAWLEFGNSEAVVLFLHAKDRNFLLIESHAVQYELERISNYKTKCFFLTLGEGV
HRLKLDSNSHTMVLDGKYVVAVSISRNSMATNAAEQRVTMEMRQSTIIQVNPRVVIMAQSKKMQQALSDPGIVER
LLNGPNEAHLVAAVRKVLTKSWSISENEAQTEEIIRKIKEQPEKYVMKWNEVTDLKKTDKLVYFGEEIVPKLDAM
TRKEREQYFIMERLQPMVVKNHFVRAHMESLLNVDATAELGIHGCLLGNLADGTVVDQFWPEFQMRTKLAEVNEG
GVIKGNSVVDSLYLV
>Hav-gss12
MNKWSFTIFFLFLVGNFQLNFVNSTFTEGETCDGKYCVENIEKDEKILMEQIRDAKDWALTNGLTFPVQRSKMNC
AECNKKVENVVLDDLAQLVPFTLYPSPFPRELYHQALDVNKALMLMFFRASLPSNFEFVKDLHKSLLPISRSVTN
IADTIERKHKEGIRQPLMLICTRVDYMASETETDEKNNQKKFELKQIEVNGGSIGGYGSPPPLTKLHRRMLSNAG
IDASPSVVPENRTSEMIAMALYRTWQKFGDSEAVILFLHSKVRIFLLVEARAVQHALERISDGKPKPKCFFLTLT
EGIERLKLDPTTSAMVLDDKYVVAIAMSRNATDDATAEEWALARTIRQSTAIPLYNTLFMLAHSKKVQQTLSKPG
VVEHFFRLPEEAHLAEAVRKVVTKSWSIGADEEDAEQIIEMVKANPHNFVMKWNELTAIKKGKLIFFGDEITEKL
DSMTKEERETFVIMEKLRPMVVKNHFVRVHSEPLLNVDVTVELGVHGCLLGNIVDGTVFDYFWPETLIKTKLANV
NEGGIMKGNSVFDSPYLV
>Hsc-gss19
FLILFFLTINQLKFVNSTFTQDTEEKYCVENIERDEKVLLERIRDAKDWALTNGLTFPVPRRKLKCDGCDNINES
VLLDDMAQIVPFTLYPSPFPRELFHQAMDVNKTLLLMYFRASLPHNFEFLKELHKSVLAVSPSLRSTADLIERKH
KEGIRQPLMLICIRTDYMASEEIDEKSNEKKYALKQIELNGGSIGGYGTPQPLTALHRRMLSNVGIDNSTSVMPE
NQTSEMLATAMYRAWQEFGDPKAVILFLHPKLRIFLLVEARAIQHAMERIFEGKPKPKCVFLTLEEGIDRLKLNS
DNFLILDGKFTVAVSMARNATDDLTTKANFPVWRAIKLSKAIQVWNTLFLLAHSKKVQQELSKPGVVEHFFRMSD
EAHLAAEVRRVMTKSWSIGADEEEAEKIIRMVKANPDNFVMKWNEVTPLKKGSKNVYFGDEIIEKLDSMDKKERD
TFFIMEKLRPMVVKNHFVRVHDKPLLNVDVNIELGVHGCLLGNVVDGTVIDHFWPENTLKTKLASENEGGIIKGH
SVVDTPYLV
>Hsc-gss20
FLILFFLTINQLKFVNSTFTQDTEEKYCVENIERDEKVLLERIRDAKDWALTNGLTFPVPRRKLKCDGCDNINES
VLLDDMAQIVPFTLYPSPFPRELFHQAMDVNKTLLLMYFRASLPHNFEFLKELHKSVLAVSPSLRSTADLIERKH
KEGIRQPLMLICIRTDYMASEEIDEKSNEKKYALKQIELNGGSIGGYGTPQPLTALHRRMLSNVGIDNSTSVMPE
NQTSEMLATAMYRAWQEFGDPKAVILFLHPKLRIFLLVEARAIQHAMERIFEGKPKPKCVFLTLEEGIDRLKLNS
DNSLILDGKFTVAVSMARNAAHDANEAGMEVWKALKRSTSIHVFNTLFMLAQSKKVQQALSKPGVVEHFFRMPEE
AHLAAEVRKVMPKSWSIGADEEEAEEIIRMVKANPDNFVMKWDEVTPLKQGAKNVYFDDEIIEKLDSMDKKERDT
FFIMEKLRPMVVKNHFVRVHDKPLLNVDVNIELGVHGCLLGNVVDGTVIDHFWPENTLKTKLASENEGGIIKGHS
VVDTPYLV
>Hsc-gss21
FLILFFLTINQLKFVNSTFTQDTEEKYCVENIERDEKVLLERIRDAKDWALTNGLTFPVPRRKLKCDGCDNINES
VLLDDMAQIVPFTLYPSPFPRELFHQAMDVNKTLLLMYFRASLPHNFEFLKELHKSVLAVSPSLRSTADLIERKH
KEGIRQPLMLICIRTDYMASEEIDEKSNEKKYALKQIELNGGSIGGYGTPQPLTALHRRMLSNVGIDNSTSVMPE
NQTSEMLATAMYRAWQEFGDPKAVILFLHPKLRIFLLVEARAIQHAMERIFEGKPKPKCVFLTLEEGIDRLKLNS
DNSLILDGKFTVAVSMARNAAHDANEAGMEVWKALKRSTSIHVFNTLFMLAQSKKVQQALSKPGVVEHFFRMPEE
AHLAAEVRKVMPKSWSIGADEEEAEEIIRMVKANPDNFVMKWNEVTPLKKGSKNVYFGDEIIAKLDSMDKKERDT
FFIMEKLRPMVVKNHFVRAHDKPLLNTDVNIELGVHGCLLGNVVDGTVIDDFWPGTVIKTKLASENEGGVMKGHS
VVDTLYLV

>Gpa_GSS7
MLDINIVSLKFCLFILINIIIGVVTSPQPSNQPSNNNNAQNEEEMENLDVQALVEDAIDWAQNISLVWLPTYNTR
SDVTQFVAFTLFPSPFPRKLFEQGQKLQHAYNLLYFRISHDYDFLAKAYEEVGKTNVPIQRLLNILNAVKAEGIK
QKISLLLTRSDYMCHVEKNKENDEQHYELKQVEFNAGQIGGISVSRRIPNLHRRMMWKASRKWTQNEMPDSEGDL
SFAEALYEAWHAFGDPNAIILIVANKRSKNRLGQRHIEYEIERLKNRKVKAVRIGEVERAALLKDGRLTLDPNDF

SLRMDGTTISVVYQITDPVEREQTEHEAAAQLLIERSTAIKSPTVGLNLASQKKVQQLLARPGMLEHFLPEPKYK
DMIDDLRSTFAGLWDLNDNDEDTLMAIKDAIANPEDYVLKPNMEGGGHNIWGEKIAEKLRTFTPDERAAHILMKR
VHPMVIKNFIVRPGKANAYFGRMTSELSIVGWLLGDAARDFHVLKADQRGHFMRTKMENVNEGGISVGTGAFDSP
YLI
>Gro-gss18
MFDLKLLFIIALLVLVESAPKQPTGSDNEKLEGISAKENVLEAKNEAESKNENEEINIQAIVDDAIDYAHKISLI
TLPTNHRERSDLAVITPFSLFPTPFPRDLFELAQKVQKAYNLLYFRISNDFDFLVETFEPVAVTNESVRTMLNIL
KSVHAEGIKQTKSMVLARSDYMCNVVTNKETNEQHYELKQVEMNAGQIGGLSVSSRITEMHRRTLQKAGLIATNE
HVPDNEPDLGTAEMLYAAWQAFNNPKAIILFVASRTTTNRFGQRHIEYEIERISNRAVKVVRIALPQCNEMIEAE
RLILADDYSLRLDGHTVAVVYLATYSNKVNNWAARHLFERSTAILSPTIALDLASLKKVQQLLAKPGMIERFLPE
DPSTVAMLRSTFAGLWSLNDKDERSQKAIKDAIEHPENYVIKPNMEGGGHNFFDQQVREKLLSFTDDEREAHILM
QKLQPMVIKNYMVRSLKEPIFGEMTTELGIFGFLVGDSRDKTVAHNVQKGHFLRTKLASVNEGGVSLGTGVFDSP
YLIP
>Gpa_GSS8
MKIFLISFLLALLTVPANSTQPDQKPTTDQDEGGGDVQVIYEDAVDFAQNLSLTFLPKLHKGRGDVAEIVPFTLF
PTPFPKNLFEEAKQVQKAYLLLYFRISNDFDFLIENFEEVAKTNTNVKNYLDILKTVHTEGIKQKNVLLLGRSDY
MCHEVIDDETNDGHRYELKQIEFNTGQLGGIHVSRHMTQLHRRTLLKAGLEASKEQVPDNPGDIAVAEALYMAWS
AFGDPEAIFLFAASTTSRNRFGQRLIEYLLEEKSKGKMKVIRISLPDCAEAMKIGGLTLDPEDSTLRLYGQKVAV
TYIATEPPNPSAGEWAVRLLFERSTAIKSPTIGQDLANQKKIQQLLAKPGMVERFLPEPENAANVDAIRRTFAGL
WAIHDKEDELSQQKINDAIQNPDNYVLKPNREGGGHNIWGKEVKKKLLTFTPEEQNAHILMERLNPMVFKNYMVR
LEKLDYTEMTTELSIIGYLFGNAHDSSVQKNVQKGHFLRTKMASENEGGVSLGTGAWDTPFLF
>Gro-gss19
MKIFLISFLLAFLTVCANSTQPDQKTTTDEDEGGGDVQVIYEDAVDFAQNLSLTFLPSDHEGRGDVAEIVPFTLF
PTPFPKHLFEQAKQVQLAYLLLYFRISNDFDFLIENYEEVAKTNTNVKNYLHILKTVQAEGIKQKNVLLLGRSDY
MCHKVMDDETNDGQHYELKQIEFNTGQLGGVHVSRNLTQLHRRTMLKAGLDPTKEQMPDNPGDMAIAEALYMAWT
AFGDPEAIFLFAASKTSRNRFGQRLIEYLLEEKSKGKMKVIRISLPDCADAMKTGGLTLDPEDSTLRLYGQKVAV
TYIATEPPNPSRGEWAVRLLFERSTAIKSPTIGQDLANQKKIQQLLAKPGMVERFLPEPENAANVAAIRSTFAGL
WPIHDQDDKQSQQIIQDAIQNPDKYVLKPNREGGGNNIWGEKVKEKLLTFTPEEQNAHILMERLNPMVVKNYMVR
PRKVIKEKKETEKIDYAQMTTELSIIGYLFGNAHDLSVQKNVQKGHFLRTKMATENEGGVSFGTGAWDTPFLF
>Hav-gss17
QAVQPAEGQGQAVQPAEGQGQAVQPAEGQDISEEFQAIYQDAVDYLQSISLTCLPMKHKGRGDVAVIQPVTLFPT
TFPRHLYEEAKEVQTAYLLLYFRISNDFDFLIEHYEITATTNAHVRNYLNILKTVKSEGIKQRKTLMIGRSDYMC
HVVEDGKTDDGQPKYELKQIEFNTGQIGGVHIARILTDFHRRMLQKAGLQPTKDQLPDNGGDYVIAEALYTAWIE
FGDPKALFMFVASKTSRNRFGHRHIEYLLEQISKGKMKIIRISMPACSHAMNVGQLTLDPEDSTLRLYGQKVAVT
YIATEPPNPTPGEWAVRLLFERSTAIKSPTVGQDLANQKKVQQLLAKPGMLEHFLPEPENATKVEALRRTFTGLW
GLHDEDEKTQRAIQDAIENPDKYVIKPNREGGGHNFWGEKVKEKLLSFTPEERHAHILMEKLNPMVIENYIIRPL
KGLLYGKMTTELSIIGYAFGNVGDPSVKKNVQKGHFLRTKMADVNEGGVSFGTGAWDIPFFF

>Gro-gss20
MANNLFKFVLVLPFLVVIFPAYYSTPTGPKNAGKSSSVSNEGFQLGKRLEECDKLETLIQDAVDWAHTVSLVNRV
REHRERSDVVEIVPFALFPSPFPRRLFEEAQAVQKTLQLLYFRVSHDYAFLKETLREAGETDNYLRHMLDILDDV
NERGVKQPITLILQRSDYMCHVNSETGEYELKQVEVNLGAIGGNARTQGVSKVHRRVFSKLGLTNDNLPLNESCT
ATGEALTKAWKYFGDPLAIIVFMSYTKVQGIFDQRLVEYEIEKFSKQQIKIVRLTLEECGKKLILDPNDSSLSYN
GRKVAVIYQRNFVYEWDWPTEKEWDIRRKLERSTAILTTKVGSNLAASKKVQQVLAEPGMLERFLPDVKEEMIQS
VRKTFAGLWGLNKDDAETRAVIKRAIEHPEKFVLKANRDGVGNNFWDEQLAKKLRTLSQKERAGLILMEKLEPLR
VTNYSIRPRGGTSQFESMVSELGINGYFLGNAKTMGTLDNVPRGHMLRTKPVDAREGGVGIGIGVHDSPFLF
>Gro-gss21
MSAINAYLFLQKTFITLTFLLVLECQVSATPVSINAIPENVLNSVATRVKGDEQLQELVEDAVDWAHHIGMAWRA
DKKIKRSDNCVFVPFTLFPSPFPRALYQQAVDIQTDIQLLYFRVSNDYDFMFKTLEPVAKTDYAINKWLKVYTTI
HKEGNHQPLTLLLTRSDYMGHLNKNNQRNEQNYELKQVEVNIGQIGGMAIANRTTDLHRRMLSKVGDDNLNNQLP
PNDAEGIVAQGLYEAWKAFGIDDAIIIAVAGSTGRNIEKFQVGVRVEQLSGNKIKIVKLSLVECDDKLELDENDY
SLRYKGQLIAVVFYQTTVEAPPAKYVSARLKIERSTAIKSPTIGVELTGAKKVQQALSMPGVLEHFLPEPENAKK
IERIRNTFARMWGLEKNDDETEKIIKDAIANPDNYVLKPSKECGGNNFWGQEIAEKLRTFEPSERAAHILMERLR
PPVVKNYMVRPAEEVHEISNVVSELSIYGYLLGNSTDMSVLLNKREGYMVRSKGENSNEGGVQAGGGAHDSPYLV
>Hav-gss18
MNLLYFRVSVNHTFLEDTLRSTIDTNSYVRALFDILKKSTEEGNKQPITFILGRSDYMAHANKNEQNGETNLELK
QIEVNIGQMGGPARADRITNIHRHFMRRAGHSLENMPANGASEMVAEALFAAWKAFGDPNAIIVVVVGRMYQNYE
QHQLIEGVQRMSDYKIKIANLSLEECNELLTLDENDFNLHYQGKVVGLVYHKTVVVNQTPQQFDARLKIERSTAI
KCPSVALELTCTKKVQQVLALPGVLESFFTDQEAEIAASIRSSFAGLWGLEKEDEETRHIIEDALANPDNYVLKP

SQEGGGNNFWGKRIAEKLITFTPSERAAHILMQRLKPIPVKNFLVQSYRPTQLEDVVSELSIYGFLLGNADDMSI
ELNERRGFMLRTKLESTNEGGIGAGGGVHDSPYLF
>Gpa_GSS12
MTSRKFVTFAALLLFIKLGNQNASSTHVETSTSPKSETKSDSNIQEDGGLSAQDLDELIEQAKIMAMNVGLITRT
KAHKDKNDEAEFVPFTLFPSPFPHKLFLQAYNIQKEMNLLYFLISLDHKFLEETLQTVAETNASVRGILEIFKKV
NEEGNKQPITLVLERSDYMAHVNNENEQNELKQIEVNIGQMGGPVRADRITKIHRHLMKKNGHDLTNMPENGAST
IIAKGLYTAWKSFGDPEAIMVVVAGRMYQNFEQHQLSQKIREMSDYKIKIAHLSLAECGEKLTLDENDFILRYNG
KAVALVYHKTLVIKPSQEQLNARLKIERSTAIKCPTVALELACTKKVQQALALPGVLERFFTNNEVVDSIKSTFA
GLWGLEKDDEKTRQRIEDAIANPDNYVLKPSEEGGGNNFWGEEISQKLRTFEPSERAAHILMQRLYPLPTKNFLV
RPFKPVTLEEVVSELSIYGFLLGNAREKSVQRNECRGFMLRTKLEKTTEGGIGAGGGFHDSLYLY
>Hsc-gss23
MNSLFLFAIFCTIFSNFLVTPQPESLNNVDAAPSLRDLEVFIEYAKFWAHHLGLIIHQKDKLTKKDAAVIKSFSL
FPSPFFPRKLFEQAIKVQKAMNLLYFRVSQDHAFLIETLEPLAETSFHIRVWLELLREIQSEGIHQPISLILMRSD
YMSHIKNNEHEIKKMDYELKQVEINIGPYGGAAHGGHMTKFHRKMMEKAGRRVKSDSMPDNEGAEQLAEGLYEAW
KLFKNPNAIVLIVADTMNRTYEMSQIDQILIQLAKNDNYKLKIVNLALHECDKRLTLDEAGDFSLHLGDQIVGVV
HFKTFCFHPNKAQKDSRRKIERSTAIKCPDVGSDLSNMKKVQQAIAMPGTLERFFPDPNEAEMIVELRASFAGMW
GLGNEDEDTKNIINDAIENPGNYVLKPSKEGGGNNTWGDEIAEKLKAFTKKQLEAHILMQRLKPIVGKNYLVYAH
RDVVYTDTTSELSTFGYLLGDVPNMKVLHNVSKGHMMRTKPESVNEGGVEAGGGVHDSPYLI
>Hav-gss19
MMNNNNFKLFFTAYLSAFTLLGWVNVEANDETKQENKGELDVQAIVDDAIDRAHQIGLIIRTKEHREKSDIAEFV
PFALFPSPFPRKCYAKAKELQKAMSLLYFRISQDYDFLHETLGQVAETNAIIRKQLEILRQVQEEGAKQPYSFVL
GRSDYMCHVNDNETDEQKKYVLKQIEMNIGPIGGYGRATRATKLHRRIMDKAGRDVSYDSMPPNNPEALVAESLY
LAWKNFGDPNAILLIVVGRAFQTFEQKQVEHLTDKMSNRKMKIIQLSLAECSEKLILDENDFSLRLDGRLIGVVY
FRTIIVTASHPIIVAARRKIERSTAIKCPTVAMELASSKKIQQVLSLPGTVERFFPNEEDADTVAAIRSTFAGMW
ALDRDDVESNRIIEDAIKNPQNYVLKPSEEGGGNNYWGDKIPEKLKTFTKQQLSAYILMERVKPIVIKNYLIRPL
SPPVKLVNAVSELSTFGYIIANGNAVEQNVAEGHMVRTKPEHITEGGIGAGGGVHDSPYLF
>Gpa_GSS9
MLSNLLNLAICVEFFVVFTAKFLTNASPTEIDHKSETLDVNIEAMVEDAINWAHIHGLVVRTKEMKLKNDIAMFL
PFALFPTPFPREIFEQARDVQTAMQLLYFRVASDFKFLREHLQPVAETNEVLQSLLEILQKVHDEGIKQPLTVVL
MRSDYMCNVDKNEHTGEPVYGLKQIEVNIGQIGGLFNAPCITDLHRRTMAHAGLDTSNVFMPINEPDAMVVDALI
MAWKAFGDKDAIVLIVAGKLYQTFQQYKMNYLLEKVSNNKIKIVQLSLLEAGEILTLDDDFSLRFGTQKVAVAFY
RSITNLKNSKLFAARLMIERSTAIKIPTIAQGLAGQKKIQQVLVLPSMVERFFPSSNEADTVVAICKTFAGMWGL
DDPEDEATKSVIQNAIDHPDKYVLKPCREGGGNNFWGKEIPEKLREFSPAELGGHILMQKLTPLAVPNLLVRPLQ
DVQFENVVSELGIFGFLLGNVHTKSVQHNVQRGHYARSKSQDAQEGGVYGGAGVVDSPLLF
>Gpa_GSS10
MLSNLLNLAICVEFFVLFAAKFLTNASPTEIDHKSETLDVNIEAMVEDAINWAHIHGLVIRTKEMKLKNDIAMFL
PFALFPTPFPREKFDEARAVQTAMQLLYFRVASDFEFLREHIQSVAASEDCIRRLLEISQIVHDEGIKQPLTVVV
MRSDYMCDFDKDEHTGEPIYGLKQIEVNIGQIGGFFNAPCITNLHRRTMAHAGLDTSNVFMPINEPDAMVVDALI
MAWKAFGDKDAIILIVANKLYQTFQNYKMNYLLEKVSKNKIKIVQMPVAEVGEIMTLDDDFSLRFGTQKVAVAFY
RSQTGLNDPKHFAGQLMIERSTAIKIPSSTQILSAQKKIQQVLALPGMLERFFPSSNEADMVVAIRKTFAGLWGL
DNPEDEATKSIIQNAIDHPDKYVMKPCREGGGNNFFGEKIPKKLREFSRAELGAHILMQKLTPFAAPNIMVRPLQ
DVQFENVVSELGIFGFLLGNVQTKSVQHNVQRGHYTRSKSQEAQEGGIYGGEGVVDSPLLF
>Gpa_GSS11
MLSNLLNLAICVEFFVVFTANLTNASPTEIDHKSETLDVNIEAMVEDAINWAHIHGLVIRTKEIKPKNDIAMFLP
FALFPTPFPREKFDEARVVQTAMQLLYFRVASDFEFLREHIQSVAASDNCIRRLLEIAQIVHDKGIKQPLTVVVM
RSDYMCDFDKDEHTGEPIYGLKQIEVNIGQIGGFFNAPCITDLHRRTMAHAGLDTSNVFMPINEPDAMVVDALIM
AWKAFGDKDAIILIVANKLYQTFQQYKMNYLLEKVSKNKIKIVQMPVIEATEVLTLDEDFSLRFGTQKVAVAFYR
SQTSLQNPKLFAGQLMIERSTAIKIPSATQILSAQKKIQQVLALPGMVERFFPSSNEADTVATIRKTFARMWGLD
DPEDETTKSIIQNAIDHPDKYVLKPCREGGGNNFFGEKIPQKLREFSPAELGAHILMQKLTPLAVPNFLVRPLQD
VQFENVVSELGIFGFLLGNVHTKSVQHNVQRGHYLRSKLQEAQEGGIYGGEGVVDSPLLF

>Hsc_gene_18451.t1
MSPSTNTDLVTPNYIAEIVGVNETVAPQQQLNLLVEDALDWAHCFGLVLRTTEHKDRSDV
CQAAPFALFPSPFPRNLFDEAMAVQKDLNLLYFRISWDLEFLKEAHQHVIPSDAFTRKML
EILEDVHSGGVKQHITLLTQRADYMCHVTTTDDQTETARQQQFELKQIEVNNIAVSMGGL
AQRASVLHRRMLQKTSKTRVIEKIDSVLPENRPIDTLTEGIHNAWKQFGDPNAILLVVVG
EVNQNQFDQRFVEYEMEQKTTGQIKIVRLTLTQCSQKLKLDPKEYTLHLDAFKVAVVYFR
AGYAPEDYPTQAEWEARRTIERSTAIKCPWIGLQVANTKKVQQVLDTPGAVERFFNGPAD
EQKVAAIRHVFAKMWGLDRDDAETNKVMQDAITNPQRYVLKPQLEGGGGNHFGEEIVSKL
RTLTPAERAAFILMEKIQPLVVKNYLIRPFLPPTLANVVSELGIYGCLVGDGRDLSVSHN
NAHGHILRTKSEHVNEGGVAVGAAVIDTPFLF

>Hsc_gene_9280.t1
MSTILIIFLIVFTSFFVICASSGDVAPSQHDDEDTPVEETELINEQEINRLAKFAIEWCH
NNGLIMRQLLGSDHKIGGSPPLEIIKERGDIATFPPLTLFPSPFPRSAFEQAMNVQKAMN
LLYFRVARDFDFLMAAFKDVTKADYHIAKMVELAKEIHEEGIRQPITVMLQRADYLLDVV
ENKETNEVKYEPKQVEINTGAIGATGLKRLTTELHRRMAEKAGMDASEAHIPENKPDKAK
VKALYEAWRLFNDPNAIMLFLGYPYYPLNYDLQHIEEDLHKMSGGVMKIHHYALFDLLEY
HGITERLKLDPKDFSLRLDGKKVAVVYSGLTFLGCRLGSHGLKMRRMIERSTAIKAPSLF
VALTGAKKVQQMLAMPGAIERFFPEPSDAATVAQIRATFAGLWGLENEDNDTEKLIEDAI
ENPGNYVLKPNKECGGNNFYDDKLVKKLKEFRQIERSAYILMQKLRPMVVKNYLLRVNEI
PKYADVIPELGVYGFLIGDLKAGKVLHNVQQGYHFRTKLSHVNEGGISAGFGFYDTAYLF

>Hsc_gene_21951.t1
MTHFSSAFVFFLVPLLIFHLPLSFATSTTVPIRAEESVEPNYESDADFETIVQEAVDWAH
TVNLVNRVKEHLERSDVVEIVPFALFPSPFPRRLFEEAKAVQKTLQLLYFRVSQNYEFLK
RTLGEAGKGDAYIGHLLDILDDVQQRGNKQPISLILQRADYMCHLNAESGEYELKQVEVN
MGAIGGNARTEGVSKVHRRVFTKLGLSTANLPPNESCAGAAEALVKAWKQFNDPLSVIVF
MSYTKVQGIFDQRLVQYEIERISGNKIEIVRLTLQECGEKLILDPNDSSLSYNGRKVAII
YQRNFLFEKDWQTEKEWDIRRKLERSNAILTTNVRIDLAGTKKVQQTLALPGMLEHFLHD
QKAETIAAVRKTFAGLWGLDKHDEETSAIIKEAIEKPERFVLKTNRDGGGNNLWDEQLAD
KLRTMTRKERGGLILMEKLEMLQVTNYSIRAREKPKMYAMTSELGIVGYFLGNAKTMATI
DNVQRGHMLRSKAAEAREGGVRIGIGLHDSPYLF
>Hsc_gene_9216.t1
MNNIFKTFIFTAIPFVAILLLNFVHVEANDDDKKGNEEGDHDIQALVDDAIDWAHHIGLI
KRCLRHMAISDVAEFVPFALFPSPIPRKCFTKAKELQTAMNLLYFRISQDYDFLRETLQQ
VAETNAIIRDQLDILRQVQEEGGVKQQYSFVMGRSDYMCHVDDTETDGQNYSLKQIEMNI
GAIGGYGRATRATNLHRRTMAKSGANFSPDAMPSNNSEGIIVDSLYLAWQKFGDPNALLL
IVVGQAYQTFEQKQVEYLLLEKSNKQMKIVQLSLKDCDEKLVLDENDFSMRLEGKLIGVV
YFRSIIVAPKPEQISARRKIERSTAIKCPNVAMELASSKKIQQVLSLPGVVERFLPNAED
EGTVVAIRSTFAGLWALDRDEAESNRIIEDAIKNPENYVLKPSEEGGGNNFWGEKITEKL
QTFTKQQLSAYILMQRLKPLKIKNYLVRPHKRPVKLVEAVSELSTFGYLIAHGDTVLENV
SDGHMVRTKQEHITEGGIGAGGGVHDSPYLF
>Hsc_gene_24468.t1
MANIHLKTKIVFVLFLLFGLFEPKIGVLTSPSKSNKENEASTTQEEKIKNYAVKIFEEYD
ENTQLESFEDLVTDAIDWAHNVGLVNRYCDPRQRGATEIVPFHEIGMANRIRDHRNRSDF
AEIAPFSLFPSPFPRRLFEHAQTIQKTLQLLYFRVSQNYEFLKETLGEAAKTDSFLRHQL
DILEDVQRRGSKQPITLLLQRSDYMCHVNGESGEYELKQVEVNIGPIGGNVSAQAVRQVH
GRVFSKLGLSEDNLPENRASAAIGEALAKAWELFRDPSAVVVIMSIKNNHGHFALRHIQY
EIEQASAYKIKVIRLTLAECDEKLILDPNDSSLRFNGRKVAVVYQRTYLSEKDWPTEKHW
DIRRKIERSTAIVPFNVNIHLAGSKKVQQTLALPGMLEQFLPDVDEGMISSIRKTFADLW
ALDKEDEATEAVIRKAIENPENYVLKTNRDGGGNNFFGEEIAKKLSDLPRDERSSMILME
KLKPMEVKNFPIRRLRDHTVRNMSSELGIVGYFLGNGQTMATIANVQQGHLLRTKFAESN
EGGVGLGVAVHDSPFLF
>Hsc_gene_9268.t1
MRTILNLFLIGFTSFFVICASSGDVAPSQHDDEDTPVEETELINEQEINRLAKFAIEWCH
NNGLIVRQLLGRDHKIGGTPLIEIIQERADIAGFPPLTLFPSPFPRSAFEQAMNVQKAMN
LLYFRVACDFDFLMDAFKDVTKADNHIAKLVELVKEIHEEGIRQPITVMLQRADYLLDVV
ENKETNEVKYEPKQVEINTGAIGATGLKRRTTELHRRMAEKAGMDASEAHIPENKPDKAK
VKALYEAWRLFKDPNAIMLFLVYPDGSFNYDVRYIEEELHKMSDGTMKIDHYSLADLSEY
RGITERLKLDPEDFSLRLDGKKVAVVFSGLTFLGCRLGSHGLKMRRMIERSTAIKAPSLF
VALTGAKKVQQMLAMPGAIERFFPEPSDASTVAQIRATFANIWGLENLDDDTQKLIEDAI
ANPGNYVLKPNKECGGNNFYDDKLVKKLEEFTPTERGAHILMQKLRPMVVKNYLLRVNEI
PKQADVIPELGVYGFLIGDLKEGHVLHNVQQGYHFRTKLSHVNEGGISAGFGFYDTAYLF

>Hsc_gene_9835.t1
MNRKFTQNLLVLSIALALANFVTIFGHPVDHDNEEDSEPNSQNVDKVTRNYVHKLVKDEE
HLSSMREFAVEWAHNNALIFRTKKNPTRSDVSVFAPVSLFPSPFPRHPFEHALSIQKALN
ELYFRVGTDFDFLERAYSDLVKTDEHFRDTMQLLKKVHEEGIKQPITVVFQRADYMLNVV
EGQTEEEPTYEIKQLEVNCGSVAGTSLDRRATQLNHLMLKKAGFHAAPEDLPENWPDRAQ
IESIKMAWKAYDNPDAVVLLTISPISQTAFDARFFETELDRLSDGRIKMVRSSMGDCARR
CKLDENFVLSLDGREVAVVYSRYSVLGKADPKSMALILDARFTIERSRAIKIPSAFIAFS
CSKKVQQLLAEPGQLEHFFPKESDAEIVNDIRKTFAGMWSLENIDEKTEERIKDAIKHPE

GYVLKSNLECGGNNYFGDKIPQKLQEIKQDERPFHVLMQRLQPMSIENIMVQPDKVSKID
TMASELGVYGVLMGNMRTGEVNHNVQQGHLLKTKLATSDEGGISAGFAVHDSPILF
>Hsc_gene_8950.t1
MNRKFTQNLLVLSIALALANFVTIFGHPVDHDNEEDSEPNSQNVDKVTRNYVHKLVKDEE
HLSSMREFAVEWAHNNALIFRTKKTPLRSDVSVFAPVSLFPSPFPRHPFEHALSIQKALN
ELYFRVGTDFDFLETAYSDLVKTDEHFRDTMQLLKKVHEEGIKQPITVVFQRADYMLNVV
DGQTEEEPTYEIKQLEVNCGSVAGTSLDRRATQLNHLMLKKAGFHAAPEDLPENWPDRAQ
IESIKMAWEAYDNPDAVVLLTISPISQTAFDAHFFETELDRLSDGRIKMVRSSMGDCARR
CKLDENFVLSLDGREVAVVYSRYSVLGKADPKSMALILDARFTIERSRAIKIPSAFIAFS
CSKKVQQLLAEPGQLEHFFPKESDAEIVNDIRKTFAGMWSLENIDEKTEERIKDAIKHPE
GYVLKSNLECGGNNYFGDKIPQKLQEIKQDERPFHVLMQRLQPMSIENIMVQPDKVSKID
TMASELGVYGVLMGNMRTGEVNHNVQQGHLLKTKLATSDEGGISAGFAVHDSPILF
>Hsc_gene_13399.t1
MMDKKDAAKFLSFTLFPSPIPRRLLTKAVQLQKAFNELYFRVSQDHKFLIEKLKPLAHSN
GRVRAWLNILEEVQSEAIVRQPITLLLARSDYMFHQNEDKFTNDAAYELKQIEFNSGNYG
GPAFGVRATELHRHIMELSGRKMSDEFLPNNVSEQHMAIGIYEAWKSYGDPNAIVLMVSN
KFSRWHEFIQIDLLIEEISKKKIKIIYLSLFECAEKLSLDDDFTLRLDNIYSVGVVHFKN
LFSSNPSPQVITARRIIERSKAISTPTVAQDLASTKKIQQVLAMPGMLERFFPEPESAEI
VTELRSTFAGMWGLDKGDDVTTAVIQEAIKNPANYVLKPSQEGGGHNFWGEEIAEKLLSF
SHEERAAHILMQRVKPLIVKNILVRAYENIQYENVISELSIFGYLIGNVNEMKVLQNMGD
GHMMRTKPSDSDEGGAGKGNGVIDSPFLF
>Hsc_gene_17104.t1
MKVFVLLICLLFITKAESVEEKQPNDDDQDFDKEDIQAIFEDAVDYAQALALVTLPHKHE
GRGDEALIHPFTLFPTPFPRQLYEQAIELQWAYNLLYFRISNDFDFLIEHYEIAAKTNAH
VRHYLNIMKEVKKEGIKQKKTLLLGRSDYMCHVVKDENTEKGERYELKQIEFNTGQLGGV
HLARLLTQLHRRTMQKAGLEASKDQLLNNGSDFVIAEALFTAWTEFGDQKAVFLFVASKT
SRNRFGQRHIEYLLEKISNYKMKIIRISLPACARQMKLGQLTLDPQDNTLRLYGQKIAVT
YIATEAPNPSDDEWKVRLLFERSTAIKSPTIGQDLANQKKVQQLLTKPGMLERFLPEPEH
AAKVEAVRRSFAGLWALHDEDEQTERAIQDAIRNPQNYVIKPNREGGGHNIWGEELKKKL
LTFTKDERNAHILMEKLNPMVIHNYIVRPVPNNYKYGEMSTELSIIGYAFGNVDEMEVKK
NVQKGHFLRTKLAHVNEGGVSFGNGAWDIPFLI
>Hsc_gene_25975.t1
MTAMPRIYCCPCRPLYASLLLILIVINCATTINGSPALISENIEQDNAIPLTEVDENGNI
GNNQIDQISIIDEAKAILLTEADQNGNTENDQIVQSNGILVPIENGNVQPLVDHHHHHHH
HNLVKTPDVVATENYQDRLRPRTGKSKSKKSLNGFSHNFLLPKKIKKLRRVEHYDSSQVV
KNANELQVLIDYAIDWAHSNGLILRTRGHLNTSDLAEFAPFSLFPTPFPRKIFNQALNVQ
TAMQLLYFRISKDFEFLKTVHQDIIKTDKVVKNFMEIVEKVYEEGIHQPITLFFQRSDYM
LHSTKNDQSEDNYALKQIEVNGSALGGAGLVTRVTRLHRRMLKKAGIEAPKSNVPDNGSD
VMTAKALFHAWQLFDKADAVLLFLVDTNADILQFDRRNIQYEFERVSKDQVDVVRLNLMQ
CTAKLMLDPVDFSLRLVDDGRVVAVVFNQVLMLGSSPSHMELDARLMIERSTAIKAPSLV
FAMSHSKKIQQVLTRPGMVERFLPDPSEAHLVEKIRQTFAGLWGFEADKEKNEQLIQMAT
KHPDRYVLKPIGEGCGAHFNYFDEDIPKKLAQLSPIELNDFILMERLKPKAYRNHFVRAF
LPPMINAEVTSELGIYGCLLGNISTGQVLLNRQEGHVSKSKLLSSNEGGICSGTGMIDTP
YLVDID
>Hsc_gene_21707.t1
MSNFALLFIATLFIHFSNATNNSPPAEEEGTDQRQGENTSKNVDLKSIKSYAVNAVKDKK
HLDELALYAIEWAQNNGLVLRKYDPKASADTAEFAPISLFPSPFPRKAFEKALSVQKAMN
LLYFRVANDHEFLMESFKDLVPMDEHIAKMVEIVKEVREEGIRQPITLLIQRADYLLNVV
TDQSSGEEKYEIKQVEVNSGSVAGLSLKRRNSELHRQMLRQVGMDTAPSPDNQPDAALVE
TLHMAWELFNDPNAVVLILSTTFIPYKFDQRQIATELEQISEGKIECIFYSLQGTMENLH
LDPNDFSLRKNSDGRRVAVVYSNMSALGYRPTFLKTYEMQMEGRRMIERSTAIKAPSLAI
GISCTKKIQQLLTKPEVLRRFFPREEDEETIEQIQSVFAGLWGMEKDDQKTQDLIKDAME
RPENYVLKPNRECGGHNYFDEKITEALQKFTQQEKAAHILMQKLRPMTVENYTLRPLAEP
QKATLVPELGVYGFLLGNEVDGTVMANVQQGYHFRSKLAHLNEGGIGAGLGVYDTAYLF
>Hsc_gene_9271.t1
MAKFNFAFGSALLVVLFEICCVNFADATQPNENDNEAKDQLLAIGEFELKRMAQYAIEWC
HNNALILRRKGDEGNRGDAALFPPLTLFPSPFPRDAFEQALKAQKAMNLLYFRVARDYEF
LMNAYKDVIKGDEHIAKLVDILKKVHEEGIRQPITVMLQRADYLLDVVENNEMNEVKYEP
KQVEVNTGAIGATGLKRRTTEFHRRMLKKATGMDATTANIPDNKPDAALIDTFYMAWRKF
DDPKAIMVCLIYSNDPFQYDLRYIAEELEKKSSGKMEVEIYSLADYSERENSTKRLQLDP
EDFSLRLDGRKVAIVYSGQSALGCKFDELGMEFRQIIELSTAIKAPSLAVAISSSKKVQQ

VLAMPGAIERFFPEPSDAATVAQIRATFAKIWGFENEDDDTQKLIEDAIENPGNYVLKPN
RECGGHNYYDDKLVEKLKEFTRTERGAHILMQKLRPMVVKNYVLRPYEAARLEEVIPELG
VYGFLIGDLKAGKVLHNVQQGYHFRTKLSHVNEAGISAGFGFYDSVYLY
>Hsc_gene_9767.t1
MAFCSYSKYLLIIILLWQLLYVVISNPTTEELSKNEDGGGNADDIQMLVEEAIDMAQNLSM
IISPKALVGRGDTAEIVPFSLFPSPFPRKLFEQAKNVQMDFNLLYFRISNDYKFLVETYE
ELAKTNEQIRMHLQILKKTHEEGIKQTKSIMLARSDYMCHESTNNDGTSKDEAKYELKQI
EFNAGQIGGTSVSSRLYQIHRHTVKKAGLSVPDDNLPAGAGDTGIAEVLEKAWYAFGDPK
GIILFVTHRRNRNRFAQRHIEYELWRLTKNKAETVRIGYPECTEMMKNGRLTLDDNFKLR
FDNRTVAIVYFVTDFFKPSEDDWEMRLMIERSTAIKSPTSGLQLASMKKMQQVLAQPGMV
ERFLPENPKKAAAIRETFAGLWSLTDQDDAAIEATKDAIVNPSNYVLKPSKEGGGNNIWD
DEIADKLLNFTSEERFAHILMQKLQPMFFKNYMVQPNRDNALFAKMSTELSIFGFLFGDS
KDNRVFYNKQNGHFMRTKLAGENEGGVVHGTAVFDSPFLF
>Hsc_gene_11699.t1
MVLTLASATAEMSTTTTQTQTLDMREMKRQLNVIKGEHGELLLTSSNENGNKQHQQNGQN
GTTNGHHANGNANGTTANTTAPAQPTALLTVDARNYVPNLVADDQRMRELAEYGLDYAHS
IGLCARTVEHKFESDIATTPPLALMPSPFPLSLYTKAFEVQETLNELYFRISCDHEFLLE
TYKDVIKGDPYIKRCVEIAQQIHDEGVHQPLALSVQRADYLSHWDEQKQCIELKQVEVNI
GQIGGPGCATQTNKYHRKMLDKLAIVRAGTGGMEMLAHTEMPVNKPRHKMGRTLYEAWKL
FGDPNAVLLFVNQPDLFPFCHFEQLQFTTFEVEKLAKRDGNIVQVIRMTFKECAERCHLD
ENDFSLYVDGKRVALVHMAYGYIGEHYPTEAEWQVRIAMERSTAIISPNIRLQLTGTKKI
QQVLSKPGVMEKFFPDEPQRVAALREVFTGLWGLENDDAVTNAVIEDAIQRPRDYVLKAQ
MGAGKGNFFDEEMVEKLKTMSLEERGAYILMKKIWPVAVKNYLMRPFQVPYLENVVSELG
IYGSIIADSSNGKVLWNSAEGYLSRSKPANLNQGGVCEGSGVVDSVLLFPDNEFVGTN
>Hsc_gene_8166.t1
MSQIFWTISFLLIIGSAISNGTDEVKRVEISLQDLVEDAINWAHHIGLAWRADKKFPRSD
TCVFVPFTLFPSPFPRKIYNEALEIQKAMQLLYFRVSNDFDFLTKTLEPVAATDVTIRSW
LKLYREVHSLPIISQPLTLVLTRSDYMCHLSKSNGTDEEEYQLKQVEVNIGQMGGPAIAN
RTTIIHRQMLAKVGYEAANLPDNDAEGIVAKGLYQAWMAFGVDDAIVVVVARRAARNIEQ
FQLEVRLEQLSGNKIQIVKLSIDECDDQLFLDPKDNSLRYNGQLVAVVYYKTIIVNPALK
SYNARLKIEKSTAIKSPTISLELACAKKVQQALSEPGVLEHFFPEPEYAQMVNDIRKTFA
KMWSLDQENDEIMKIISDAIENPGNYVLKPSQEGGGNNFWNEEIAKKLRTFKPKERAAHI
LMERLRPLVVKNYMVRPYVKEVPQLSNIVSELSIYGYLLGNSSNMAVLRNERDGYMLRSK
REDDTEGGIHAGGGVHDSPYLF
>Hsc_gene_9207.t1
MAAFLNGNAHSSIECNGVVHAKNDIDQNGTKLADRLQHSVQIGTKKASVSRGSINVNKYV
VEVISDTDQLHTLEQYAIDYAHSIGLVAPLSDPPEGKFTNIYAVPPPIALLPSPFPRELY
DHAVNVQQAMNELYFRVASDHDFLMDAFKDVVKGDPFMARLVQIAQMVHDEGIHQPLAVQ
LQRSDYMAHWDPSDGSMALKQVEVNIGPIGGPGFAFGVSKLHQKMLDKLAIEHDGKPAIL
ANSEAPLNRSRQNIAYTLYQSWKLFGDPKAVLLFLDTPNITHFEQLQFIQFEVEKLGKQQ
GNFVKVLVLSLTEASKRISLDESGDFSLYLDGTKRVALVHIADGNVPHEYPTEHEWTART
TIERSNAILSPNIRFHLSSTKKIQQVLAKPGMVERFFPNEPKRVALIRSTFTGLWGLEVE
DEATREVIEDAIRSPGNYVLKSQMEAGLGNFFDDQMAQMLQQMSKEDRGAYILQQRIKPL
VVKNYLMREAKPAELEDVVSEMGIYGSLIGDQSNGRILHNAFDGHTIRSKRSDLNLGGVC
CGGGAIDSPLLFPMAQMLDQSNGH
>Hsc_gene_22122.t1
MASINNGHAAVGAANGIQQKKLLQQQQAHGTNMAQLPVNANDYALAAVRDDAELKLLAEY
AVDYAHSIGLVGRSGDDKYKHSNDVSVAPPIALLPSPFPRELYEQATDAQQSLNELYFRI
ACDHDFLMEAYKDVIKGDPFHAKLIELAKVIQKEGVRQPLMVGLQRADYLSHWNEAAQKM
ELKQVEVNPGQIGGPGAATAVSKCHRKMLDKVEIMHGKKLPIMAKAVVPENRPGPGIART
LYQAWKMFGDPNAIILYVNQPDLFPVCHFEQLQFVMFQVEKLAKQDGNHVLVRRASFIEL
RKRLSLDEAGDYSLYFDGTKRVALVHMAYGYLPEHYPEKDFDLRVMMERSTAILSPNLRL
QLAGTKKIQQILSKPGTLEHFFPNDPQQVAKIRNTFTELWGLEENDDITKAVIVDAMKNC
HDYVMKSQMDGGHGIYFDEEIPKMLATLSLEERGAFILMKKINPVVAKNFMVRPFEKPHQ
EEVNSEMGIYGSLIGDQSTGKVLVNSVNGHLVRSKAASQNHGGVCSGGGVIDSVLLFPSS
EFR
>Hsc_gene_26058.t1
MIKINGSTFIIVFILNSFRGNAVSHNHVAIEGNKTVLNYVENAVKNENQLHKLSQFAIEW
AINHALVVRTNKRFAKGKVIAVIPPVEYRSDCTEFASVTLLPSPFPREAFNKVVAVQEAM
NLLYFRVANDYEFMMDAYKDVVKTDLHVRALVNILKEVHATGIKQPYSVMIQRADYMVNV
VGENNYEIKQVEVNCGAIASLALDSKITDLHTAMLRKVGMNASKDVVPVNKPDQEFINML

YLAWQKFGDPNAIVVILHFINSSPYNLDYTNIEMELARVSNGQIKMDFVTLKEGKRLNLD
PETFTLRLDGRVVAVVNSGTSALGYLANRAEMETRKTIELSTAIKAPSLAIAISSSKKIQ
QLLVKPGVLERFFPLPSDARTIDAIRETFTGMWALETDDHLTEKIIQDAIQNPGNYVLKP
NRECGGNNYFDDEIPKKLQQFSLEQRAAHILMQRLHPLQVKNYFLRPYAKPTLCNTSGEL
GVYGFLMGNMKDGTVDRNVQSGHLLRTKLAHVNEGGVIEGAGVGDTPYLF
>Hsc_gene_17207.t1
MLPIKLIRLSVFLGILFDFVFIGVKCNAPKETNEDVKNKQNEYLDTKALVEDAIDWAQNI
SLVWLPKQHMGRSDVTEIVPFTLFPSPFPRKMFEFGQKVQIAYNLLYFRVSNDYEFITKA
YEKVAETNIAIRRLLTILKAVKAEGIKQTKSLVLARSDYMCHVATEHAEMEQQKYELKQI
EFNAGQIGGLSVSRRITNLHRRTMWKASRKWTQKQMPNSEGDTAIAKALYHAWQAFDDTE
GIILIIANKRNLNRIGQKHIEYEINKLSGGKVKTKRIGEPERALLVKKGSLTLHPEDFSL
LLEGKRVSVVYQITDPQEDEQNADEAAAQLLIERSTAIKAPTVSLNLASMKRVQMLLAKP
GMLEKFLPEPEYKEMVAELRSTFAGLWDLEDNDEDTTRAIEDAIENPHNYVLKPNMEGGG
HNFWGDKIREKLRTFTPSEKAAHILMKRVQPLVIKNFMVRSQQTKTQYGRMTSELSIIGW
LLGDANGNRVLDNVQSGHFMRTKLEKVDEGGISVGTGAFDSPFLI
>Hsc_gene_19136.t1
MILSYKLTHNHIAQFAPVSLLPSPFPREAFEKAIAVQEAMQLLYFRVGCDFDFLFEAYKD
VVKTDKQIKQMVDILREVQKQGIKQPQTLLIMRSDYMLNKVGSNTENGTDHYEIKQIEAN
TGAIGGLGNDRRTSELHQRLLKRIGMDPTNAPQNEGDAHLINSLFMAWKAFDNSDALLVI
LSHVKFSYKYELRKIEDELDRLSEGQLRVAYVSLNDGYYDFKLAADSSLLLNGKVVGVVY
SLISALGYKANEEAMEARKMIELSTAIKAPSLAIAISSSKKIQQLLASPGGVERFFPDPA
EADKVKAIREIFTGLWGLEGHDEETERIIADAIENPSNYVLKPNGECGGNNYYDDQLVEK
LRTMTNNQEERAAHILMQKLHPMITKNYFLRPTILPRLGVVVSELGVYGTLMGNMPDRTV
SYNAQSGHLLRTKLAGANEGGISVGTAVGDSPYLF
>Hsc_gene_24944.t1
MILRYDIRHNFIAQFSPVSLLPSPFPREAFEKAIAVQEAMQLLYFRVACDLDFLLDAYKD
VVKTDKHIRQLVDILREVKEQGIRQPKTLMIMRSDYMLNTVKSSKKDGKDRYEIKQIEAN
TGAITGLRIDRRTTELHQRLLKRIGRDPTNAPQNEGDTNLINSLFMAWESFGNSDALFVI
LSANWNKYKFELRNIEDELERLSGGKLRVEYVPLLEGYTNFSLADDNSLLLNGKIVGIVY
SGLSALGYQANEKEMHTRKIVELSTAIKAPSLAIGISSSKKIQQLLASPGILERFFPDPA
EADKVKAIRETFTGLWGLEKNDEQTERLIAGAIEHPSNYVLKPNGECGGNNYYDERLREK
LLTMTREERSAHILMQKLHPMTTKNYFLRPFYAPKFGLVVSELGVYGTLMGDLLTRDVSS
NVQRGHLLRTKSAGVNEGGIGVGTAVGDSPYLF
>Hsc_gene_26047.t1
MNSLFLFAIFCTIFSNFLVTPQPESLNNVDAAPSLRDLEVFIEYAKFWAHHLGLIIHQKD
KLTKKDAAVIKSFSLFPSPFPRKLFEQAIKVQKAMNLLYFRVSQDHAFLIETLEPLAETS
FHIRVWLELLREIQSEGIHQPISLILMRSDYMSHIKNNEHEIKKMDYELKQVEINIGPYG
GAAHGGHMTKFHRKMMEKAGRRVKSDSMPDNEGAEQLAEGLYEAWKLFKNPNAIVLIVAD
TMNRTYEMSQIDQILIQLAKNDNYKLKIVNLALHECDKRLTLDEAGDFSLHLGDQIVGVV
HFKTFCFHPNKAQKDSRRKIERSTAIKCPDVGSDLSNMKKVQQAIAMPGTLERFFPDPNE
AEMIVELRASFAGMWGLGNEDEDTKNIINDAIENPGNYVLKPSKEGGGNNTWGDEIAEKL
KAFTKKQLEAHILMQRLKPIVGKNYLVYAHRDVVYTDTTSELSTFGYLLGDVPNMKVLHN
VSKGHMMRTKPESVNEGGVEAGGGVHDSPYLI
>Hsc_gene_15421.t1
MNSKLAQCLFLGIILILNNLLVIFGHSVENENDGTNGQTSEEFDKVTRNYVEQLVKSEKH
LLSLRQFAVEWAHNNALIFRNKKVPPTTDAIYRSDVAVIAPFSLFPSPFPRHAFEHALAV
QKALNLLYFRVATDIDFLERAYSDLIRTDENFSNTMDVLRTVREEGIRQPITVMYQRADY
MLNVVGGQDEAEPNYEIKQLEVNCGSVAGTSLDRRTAQLNHVLLQRAGFHPAPEDLPENW
PDKAQIESIKMAWEAYNKSDAIVVILISPISETIFDANFFETELDRLSNGRIKVERITLN
DCVHRCKLDENFALRLDGREVAVVKSRYSVLGLRARGSELNILKNLRLMLERSLAIKIPS
TFIGFSCSKKVQQLLAEPGELEHFFPEESDAEMVKAIRKTFAGMWSLENTDENTEQKIQD
AINHPENYVLKSNMECGGNNYFDEKIPIKLTGITPTERSFHILMQKLRPMPIKNVMIHPN
TKPKINEMVSELAVYGVLIGNMTTRTVSHNVQQGHLLKTKLATANEGGISTGSAVHDSPI
LF
>Hsc_gene_5333.t1
MECDGVVHAKNGNDQSGTKFADRLEHSLEIGTKKSGANKGEHHCNKYVLEVISDNDQLRT
LEQYAIDYAHSIGLVTPISDPPDGKLTNIYAVPPPIALFPSPFPRELYDHAVNVQQAMNE
LYFRVASDHDFLMDAFKDVVKGDPFMARFVQIAKMVHDEGIHQPLAVQLQRADYMAHWDP
SDGDGTMALKQVEVNIGGIGGPGFGSAISKLHQKVLDKFAIEHGGQPAILANSEAPVNRS
RQNIAYALYQAWKLFGDPKAVLLCLGTMDMTHFEQLQFIQFEVEKLGKQQGDLVNVLVLS
LTDASKRISLNESGDFSLYLDGTKRVALVHITDGNVPDEFPTEHEWTARTMIERSNAILS

PNIRFHLSCTKKIQQPGMVERFFPDDPKRVALIRSTFTGLWGLEVEDEATREVIEDAIRS
PANYILKSQVEGGLGNFFDHQMAQMLQQMSKEERGAYILQRRINPLVVKNYLMRQKMPAL
LEDVVSEMGIFGSLIGDQSNGRILHNAFDGHLVKSKRSDVNHGGLCCGGGLIDSLLLFPM
AQMLDKSNGH
>Hsc_gene_9160.t1
MNLLYFRVSQDFDFLIETLQPLAETTEHIRVWLELLREVKSEGVHQPINLLLIRSDYMCH
VNNGEAMEKNDYELKQVEVNIGNYGGAGYASHLTQFHRKMMETAGRNVLAGTLPDNGSDE
QLAEALYEAWKLFGDPKAIMLIVANKYNRTFEMSHIDQILIKLAKNDNHQIQIVIFSLAE
CVERLTLDEDNFSLRLDGQIVGVVHLKTTCFKPTPEQIASRRMIERSTAIKCPTAAADLA
SMKKVQQVLAKPGVLERFFPNSDDAELIAALRPTFAEMWALDKEDEHTKSIIKDATENPM
KYVLKPSQEGGGHNFWGEEIGEKLRTFTKEELAAHILMQRLKPMVGKNYLVHSFRDVKFT
ATTSELSTFGYLLGNGRENTVSHNVSKGHMMRTKPEHINEGGVIAGEGVHDSPFLI
>Hsc_gene_19574.t1
MQKLYFLVSCDFDFLVKATEGMDKSNNLYGRMIEMMKEIHREGQRQPYTLFLTRSDYMVD
STTDERDGQQRFGLKQVEMNIGSVVGSAMGPRTAEMHRRMLQRMRMDASNVPENRAFNTL
ARGLFQAWLRFGDPNAVVVFAVLQGSMHRFDERAIEYELQRISDWQVKVVRLSSKEAYSK
LQLDPNDFTLRLTADGRAVAVVYSRSGPLPEWTEEEWEARKRIERSTAIKTSTVFSALSS
SKRVQQLLALPGMIERFLPEPKDREMVEAIRQTFVGLWGLENDDEQTQQLIKHAIANPSL
YVLKPQNEGGGHNYFDDELKQKLLQFTREERAAHTLMQRIWPVTAKNFMVRPMEEAVLDD
TIVELGMFGYLLGDKRDRSIVRNKQHGYLVRTKPASSAEGGISAGGVYDSLNLF
>Hsc_gene_9276.t1
MNLLYFRVARDYEFLMKAYKDVIKGDEHIAKLVDILKKVHEEGIRQPITVMLQRADYLLD
VVENNETNEVKYEPKQVEVNTGAIGATGLKRRTTEFHRRMLKKATGMDATTANIPDNKPD
AALIDTFYMAWRKFDDPKAIMVCLIYNNDPFQYDLRYIAEELEKKSAGKMEVEIYSLADY
SERENSTKRLQLDPEDFSLRLDGRKVAIVYSGQSALGCKFDELGMEFRQIIELSTAIKAP
SLAVAISSSKKVQQMLAMPGAIERFFPEPSDAATVTQIRATFAKIWGLEHEDDDTQKLIE
DAIANPGNYVLKPNRECGGHNYYDGKLVEKLKEFTRTERGAHILMQKLRPMVVKNYVLRP
YEAARLEEVIPELGVYGFLIGDLKAGKVLHNVQQGYHFRTKLSHVNEAGISAGFGFYDSV
YLY
>Hsc_gene_25977.t1
MQLLYFRISKDFEFLKTVHQDIIKTDKVVKSFMEIVEKVYEEGIHQPITLFFQRSDYMLH
STKNDQNEDNYALKQIEVNGSALGGAGLVTRVTRLHRRMLKKAGIEAPKSNVPDNGSDVM
TAKALFHAWQLFDKADAVLLFLVDTNADILQFDRRNIQYEFERVSKDQVDVVRLNLMQCT
AKLMLDPVDFSLRLVDDGRVVAVVFNQVLMLGSSPSHMELDARLMIERSTAIKAPSLVFA
MSHSKKIQQVLTRPGMVERFLSDPSEAHLVEKIRQTFAGLWGFEADKEKNEQVIQMATKH
PDRYVLKPIGEGCGAHFNYFDEDIPKKLAQLSPIELNDFILMERLKPKAYRNHFVRAFLP
PMLNAEVTSELGIYGCLLGNISTGQVLLNRQEGHVSKSKLLSSNEGGICSGTGMIDTPYL
VDID
>Hsc_gene_21504.t1
MQKLYFLVSCDFDFLVKATEGMDKSNNLYGRMIEMMKEIHREGQRQPYTLFLTRSDYMVD
STTDERDGQQRFGLKQVEMNIGSVVGSAMGPRTAEMHRRMLQRMRMDASNVPENRAFNTL
ARGLFQAWLRFGDPNAVVVFAVLQGSMHRFDERAIEYELQRISDWQVKVVRLSSKEAYSK
LQLDPNDFTLRLTADGRAVAVVYSRSGPLPEWTEEEWEARKRIERSTAIKTSTVFSALSS
SKRVQQLLALPGMIERFLPEPKDREMVEAIRQTFVGLWGLENADEQTQQLIKHAIANPSL
YVLKPQNEGGGHNYFDDELKQKLLQFTREERAAHTLMQRIWPVTAKNFMVRPMEEAVLDD
TIVELGMFGYLLGDKRDRSIVRNKQHGYLVRTKPASSAEGGISA
>Hsc_gene_26289.t1
MFISKNLLFVFFVSALQFHSNANPNVSISNPAFEAVKSSDQLATLVEAAVDMAHEVGLIK
RLSDDDSRKRRNSDVASIQPISLFPSPFPRSAYQQAMDVHTGMQKLYFLVSCDFDFLVKA
TEGMDKSNNLYGRMIEIMKEIHREGQRQPYTLFLTRSDYMVDSTTDERDGQQRFGLKQVE
MNIGSVVGSAMGPRTAEMHRRMLQRMRMDTSDVPENRAFNTLARGLFQAWLRFGDPNAVV
VFAVLQGSMHRFDERAVEYELQRISDWQVKVVRLSSKEAYSKLQLDPNDFTLRLTADGRA
VAVVYSRSGPLPEWTEEEWEARKRIERSTAIKTSTVFSALSSSKRVQQLLAQPGMIERFL
PEPKDRQMVEAIRQTFVGLWGLENEDKPTQKLIQHAIDNPHLYVLKPQNEGGGHNYFDDE
LKEKLLQFTREERAAHTLMQRIWPVTAKVIKMGNHALIN
>Hsc_gene_25072.t1
MEKILKDNNAKIELVIDEAKLIALDVGLTMRHKSMKSDAIIPERIPAPFTLFPSPFPRQF
FQQAYNIQTALNLLYFRVMRDHNFLSAVYRNLLKYDQYFRNAFQIVQQVHAEGIKQPYTV
LFQRTDYMLCGKSADKQQNSDYVLRQLEVNGGAIGGISFSARTSALHRQILSKFGLDLSN
AVEVQTNKGIVEALYRAWLKFGNPKAIVLMIFNDVPSSLYYERNSLFNDLILKFFGKAQI
VSLTLAECSAFLTLDPNDFSLRFGDKIIAVVFNQQMMISANQAEMEARRMIERSTAIKAP

SLAAALAHTKKVQQVLAKPGILEYFFKNPNEAPLIEPIRNTFTKIWRIDDENTQYNQELI
EKVKQQPNQFVLKKIEYAFTEKLVGRTYFGQEILSRLSSFTAAERASYILMEKLQPITVK
NHMVWPRKKPQQLLQDEEHENHQSFFEEVTPELGIFGTLFGNVVNGQVFYNVPLGHLLKT
KLANENEGGVIRGNSAYDSAYLVD
>Hsc_gene_21874.t1
MQHELKKDFDQRFRLRLLFSWAWWSPRPITLVEQFTLFPSPFSRTYFEQAYNVQTAFNLL
YFRVMRDYEFLREVYQGIIQHNHLLAKAFDIMDQVHSEGIKQPITLLIQRADYMICEKNS
NDNDVSKTYELKQIEVNGTSVGGLASSELTTKLHKRILSKLGLNLANSIENPTLSNTVEA
LYKAWTQFNNPKGIILLIDKMCYRDNESNLIKEQLVGKYKINVFVLALQHCVKRLTLDEN
DFSLRLDQRHDVAVVFNQENMLNKRKNLMKVRRIIERSTAIKAPSLIAALAHSKIVQQVL
SEPGMVERFFPNPEEAPLIKAIRDTYAKMWRIEENENNEQFSEIIERVKKQPNNFVMKLT
EYALWNAFNKDEVKKIYLGEEILETLSNFEADKRRAYILMEKLRPKSVKNHIVWAEDEEG
SSGGDFFEEVTPELGIFGTLLGNIANGEVLHNAQIGHKLRTKLASANACGIENVKTAYDT
AYLVD
>Hsc_gene_21955.t1
MFIYKTFLCVLFVFPLIIHSNVANETLKKYKPLNQNDEVIIRLIRDEDIEETAEIIQESY
LQDCLKIKHATPEQCEQMKKVPLERTRMALSTFKAEKDGAVIVAETTAPTEAGSSKGGAR
LMGCIRVKLNSQGNTKDNGPFAQIGPFATRVDVHGLGIGTMMLQVAEEYAVNRWNVCEMI
LDAHGIPEPVKEVVHPPMTPLLKFYEKRGYRRIGKTNWFDPAESNYVRIPNSLTHLERMV
KDTCLAKETSRKGTKRGREEEKGKAPEGESIGKNKRMHQQHESTELVRNYAAEAVKNEEE
LAELVEAAVDLAHEFGLIKRLSDDESRRRRTTDLASIQPFSLFPSPFFPRSLFHRATAVHK
GMQKLFFRVSCDYEFLAKATEQLAKTDKTYERMVGLMDQVRREGHQQPYTLLLTRADYVM
DNNSITEQPNGQQRQFGLKQTGMTIGTVGSVAMSPRAAEVHRQMLQKLGMDASNVPPNRA
VNTMARGLFNGWLRFGDADALVVFIVSPEDKFRFDERAIENELQQISDGQIEVERMTSEE
AFAKLFLDESDFTLRRSSDRRAVALVHSSSNGHLPEWTDDEWEARRRIERSRAIKTSTVF
SDLSTSKAVEQLLAQPGKIEQFMRDEEDEQMVETIRRTFVEQWPLEKDDEPTRQLIQLAI
ANPGLFVLKAQNEEGTPNYVDEELREKLQQFTHEERAAHELRHRIQPVTAKNFLVRPLEG
AVPGDVAVELGIFGFLLGDTRDGSIVRNTQQGFIARSSTKLANGTEEEKDEVYDSLNLI
>Hsc_gene_21160.t1
MVDSTTDERDGQQRFGLKQVEMNIGSVVGSAMGPRAAKMHRRMLQRMRMDTSNVPENHAF
NTLARGLFQAWLRFGDPNAVVVFAVLQGSMHRFDERAIEYELQRISDWQVEVVRLSSKEA
YSKLQLDPKDFTLRLTADGRAVAVVYSRSGPLPEWTEEEWEARKRIERSTAIKTSTVFSA
LSSSKRVQQLLAQPGMIERFMPEPKDRQMVAAIRQTFVGLWGLENDDEQTQQLIKHAIAN
PSLYVLKPQNEGGGHNYFDDELKQKLLQFTREERAAHTLMQRIWPVTAKNFMVRPMEEAV
LDDTIVELGMFGYLLGDKRDRSIVRNKQHGYLVRTKPASSAEGGIGAGGVYDSLNLF
>Hsc_gene_8164.t1
MCHLSKSNGTDEEEYQLKQVELNIGQMGGPAIANRTTIIHRQMLAKVGYEAANLPDNDAE
GIVAKGLYQAWMAFRVNDAIVVVVAGRADRNIEQFQLELRLEQLSGNKIQIVKLSIFECD
DQLFLDPKDNSLRYNGQLVAVVYYKTIIVNPALKSYNARLKIEKSTAIKSPTISLELACA
KKVQQALSEPGVLEHFFPEPEYAQMVNDIRKTFAKMWSLDQENDEIMKIISDAIENPGNY
VLKPSQEGGGNNFWNEEIAKKLRTFKPKERAAHILMERLRPLVVKNFMVRPYVKEVPQLS
NIVSELSIYGYLLGNSNNMAVLRNERDGYMLRSKREDATEGGIHAGGGVHDSPYLF
>Hsc_gene_25978.t1
MRDYQFLREVFQETITHNQMLAKAFEILGQVHAEGIKQPFTLLFQRADYMVCEKNSDENE
ETKLYQLKQVEVNGASTTGFAFSDLTTKLHKNVLSNLGLNLANHVESDTISLTVDALYHA
WQKFADTKAIVLIIVNNIRPNYYESWIIKKKLHDKYAINSVVLSLFQCAKMLTLKIKDFS
LRLNKQIVVAVVFNQQTMLSKNTNEMKARLIIERSTAIKAPSLIAVLAYSKKVQQALAQP
GMVERFFPNPEEAPLIKAIRDTEAKMWTIERDNDKFAEMMKKIENQPNNFVLKRIESTLQ
DSDTKKIYFGEEIIENLAIMDEEERWAFILMEKLQPMSVKNHIVWSESKSKEESSGGDFF
EEVTPELGIFGTLFGNIANGEVERNAQLGHWLKTKMANDNEGGIATGHSAYDSAFLVD
>Hsc_gene_21788.t1
MRDYQFLREVYQGVIGHDQIISSAFDIMDQVHSEGIKQPITLLFQRADYMICEKNRNDNE
DSEIYELKQIEVNGSAIGGLAFSELIANLHKRILSKLGLNLENSVENLTLSNTVEALYKA
WVKFGKQNAVILIIEFPLYNHNESLIIKKQLNNNYGINVFVWSLEYCAKKRQQKLTLDEN
DFTLRLGQRHDVAVVFNQQNMLDERPELLEVRRIIERSTAIKAPSLIAVLAHSKKIQQVL
SEPGMVERFFPNPEEAPLIKAIRDTYSNMWRIEKNENNEQFSAIVDRVKNQPNNFVMMQS
LYDLWDALFKTIRDTYSNMWRIEENENNEQFTEIIERVKKQPNNFVLKKTEYALWDAWIK
RDVKKIYFGQAILETLEKFGADKRWAYILMEKLLPMSVKNHIVWAKDKEGSRRGDFFEEV
TPELGIYGTLFGNISNGEVLYNTQLCHKLKTKLASENEGGIASGNSAYDSAYLVD
>Hsc_gene_16013.t1
MLTLGDDHSLRYGTQKVAIAFYRSLSSLRDERIFNARLMIERSTAIKIPTVANGLASQKK

IQQILAKPGMVERFFPHPNEADKVAAIRKTFTGLYGLDDPDNENTKRVIQDAIAHPDNYV
MKPSREGGGNNFWGDEIPKKLREMSRAELSGHILMQKVHPFSAPNYMVRPNDGVQHGNVV
TELSTFGTLLGHVKTKAVLHNAQQGHYARSKPEGATEGGVYGGGGVVDSPFLF
>Hsc_gene_6080.t1
MESVFIRKHLLMYYGKDAVVLSLQNCVERLTLDENDFSLRLDQHHDVAVVFNQQTMLNEN
PDLMEVRRIIERSTAIKAPSLVAAFAHSKKMQQVLSEPGMVERFFPNPEEAPLIKAIRYT
YANMWRIEENENNEPFSEIIERVKKQPNNFVLKKTEYALWNDDNAMKIYFGQEILENLAN
FNADKRRAYILMEKLRPVSVKNHIIWAEDGEESSGGDFLEEVTPELGIFGTLLGNIANGE
VLYNAQLGHQLKTKLASENEGGMATGNSAYDSAYLLD
>Hsc_gene_21890.t1
MSSRTKATARKKYEIKQFEVNSGSVAGLSLKRRNSELHRQMLRQVGMDTAPSPDNQPDAA
LVESLHMAWKLFNDPNAVVLILSTTFIPYKFDQRQIATELEEISDGKIECIFYSLQGIME
NLHLDPNDFSLRQNSDGRRVAVVYSNMSALGYRPTFLKTYEMQMEGRRMIERSTAIAPSL
AIGISCTKKIQQLLTKPEVLRRFFPREEDEETIDRCLPGCGAWKRTTKKRGI