Understanding the molecular genetic basis of virulence in the parasitic weed, *Striga asiatica*

A thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy by

Rachel Sarah Rose Greenhill

School of Biosciences, Faculty of Science,
The University of Sheffield,
Sheffield,
S10 2TN,
United Kingdom

March 2023

Acknowledgements

I would like firstly to thank my three supervisors, Professor Julie Scholes, Professor Roger Butlin and Doctor Roy Chaudhuri for accepting me onto this project and guiding me in this most interesting of research topics. I am ever grateful for all of their continued support throughout the past four and a half years, particularly with respect to the challenges brought to the project with COVID-19 and the adjustments we have made to the research ideas along the way. I also thank Julie and Roger for reading this thesis and for all of the suggestions put forward for improvements.

Next, I would like to thank my unofficial fourth supervisor Doctor James Bradley, particularly for his unwavering patience in aiding my Bioinformatics comprehension and being a constant source of valuable suggestions, despite the 4 hour time difference and busy schedule. Dr Bradley also kindly provided the *S. asiatica* secretome and raw sequence data for four S*. asiatica* individuals used in this study. From the Scholes lab I would like to thank everyone who helped with my many, many early morning infections, particularly David Rapley, Nat Phon-Or and Mamadou Cissoko, all of whom were integral sources of inspiration and wonderful friends and colleagues. From the Butlin lab I would like to thank Le Qin Choo for all of the fantastic advice and guidance in exploration of population genetics.

I would like to thank the UK Biotechnology and Biological Sciences Research Council for funding me on this project and the White Rose Doctoral Training Partnership, with whom I worked on the communications team, for providing excellent training courses. As part of this program I was also able to explore the horticultural industry, spending my 3-month PIPS placement with West Country Soil Improvement, coordinating nursery trials of a peat free compost and acting as a social media manager which helped greatly in identifying future career paths.

Outside of work I would like to acknowledge my soon-to-be husband, Laurence for his love and encouragement throughout the late nights, the tears, and the excitement, for always reassuring me of my worth and for answering my endless coding queries. I would also like to recognise my parents, David and Susan who helped shape me to be the person I am today and have offered every bit of their help with the PhD, even when they have understood nothing! Not forgetting my old dog, Rosie who has been a great companion throughout my analysis and writing from home, even if she did sleep most of the time. Finally, I would like to dedicate this thesis to the memory of my late younger brother, Christopher who is with me always and who was always my greatest supporter.

Declaration

No portion of the work referred to in this thesis has been submitted in support of another application for another degree or qualification at this or any other university or institute of learning.

Abstract

Parasitic weeds of the genus *Striga* are a major constraint to an efficient and profitable agricultural system in Sub-Saharan Africa, causing up to 100% losses in crop yields which disproportionately affect the poorest subsistence farmers. The mode of action for root parasitic plants such as Striga is in penetration of the host root cortex to form a xylem-xylem connection through which water and nutrients can be derived, while evading or suppressing the host immune system. A promising strategy for control of *Striga* spp. is the use of resistant crop varieties, but to develop a durable and broad resistance to Striga it is also necessary to understand the genetic basis of the parasite's virulence and how virulence can vary between and within populations.

The species *Striga asiatica* is particularly devastating in Madagascar, one of the major rice growing countries in Africa and a major focus for agricultural research. *Striga asiatica* is preferentially autogamous, and it is known that naturally selfing species tend to show a greater level of host adaptation to their sympatric hosts. Therefore, gene-for-gene interactions between virulence factors of *S. asiatica* and resistance genes in rice hosts in this evolutionary arms race were considered to be likely. In this study, *S. asiatica* accessions, based on field site location, were sampled across Africa with a focus on Madagascar. Rhizotron-based virulence screens showed that variation in Striga virulence was influenced by spatial separation of *S. asiatica* accessions and, to a greater extent, by host variety. However, this variation in virulence did not appear to reflect a race-like structure across the sampled *S. asiatica* accessions, as is seen with the autogamous species, *S. gesnerioides*.

Whole genome resequencing was performed on 47 individuals. Genotyping revealed well defined phylogenetic lineages between countries. Accessions within Madagascar contained multiple selfing lineages suggesting gene flow between sites. Selfing rates across the species within Madagascar were estimated to be at least 95%. A novel genotype-environment association study using redundancy analysis was employed to test associations of 400,000 genic SNPs with virulence phenotypes across rice hosts. A large number of candidate adaptive loci were discovered. Utilising homologous Arabidopsis annotations, highly enriched gene ontology terms for *S. asiatica* virulence candidates included several commonly found cell wall degrading enzymes such as pectin acetylesterases and glycosyl hydrolases.

Overall, this thesis has combined phenotypic and genotypic techniques to more clearly characterise the variation in *S. asiatica.* The major findings concluded that virulence is influenced by both spatial and host variety factors, suggesting that in the host-parasite relationship the virulence of *S. asiatica* accessions is not determined in a race-like manner. Whole genome resequencing identified distinct phylogenetic lineages among *S. asiatica* accessions with multiple selfing lineages, and candidate adaptive loci related to virulence, including cell wall degrading enzymes. This study effectively utilised the recent *S. asiatica* reference genome and redundancy analysis as tools for prediction of virulence-associated genes all of which can be built upon for future resistance breeding efforts.

Abbreviations

ANOVA	Analysis of variance
ETI	Effector triggered immunity
GWAS	Genome wide association study
IBD	Isolation by distance
MDS	Multidimensional scaling
NERICA	New Rice For Africa
PAMP	Pathogen-associated molecular pattern
PCA	Principal component analysis
PCWDE	Plant cell wall degrading enzymes
pRDA	Partial redundancy analysis
PTI	PAMP triggered immunity
QTL	Quantitative trait loci
SL	Selfing lineage
SNP	Single nucleotide polymorphism
SSA	Sub-Saharan Africa
WZA	Weighted z-score

Table of contents

Chapter 1

General introduction

1.1 Parasitism and parasitic plants

It is thought that the lifestyle of parasitism within plants has evolved independently at least 12 times in the Earth's history (Kujit, 1969, Westwood et al., 2010) with approximately 4,500 parasitic species of plants, which represent 1% of the dicotyledonous angiosperm species on the planet (Heide-Jorgensen, 2013). This lifestyle of parasitism has developed for many plant species to live, not merely alongside others, but through direct reliance on others, extracting all or part of their nutrients from a 'host' plant. The parasitic plants can be split into several categories based on their approach to parasitism, those are holoparasites such as *Orobanche aegyptiaca*, deriving all nutrition from their host plant or hemiparasites, which still retain some photosynthetic capability. Hemiparasitic species can be further subdivided based on their ability to complete their life cycle without the presence of a host, examples being the obligate hemiparasitic *Striga* spp. and the facultative hemiparasite *Triphysaria versicolor* (Figure 1.1) (Westwood et al., 2010). Often these facultative plants are almost indistinguishable from their autotrophic relatives, unlike the obligate parasites which have often lost many of the fundamental structures needed to exist as a free-living organism, including leaves for photosynthesis or roots for water absorption from the soil (Yoshida et al., 2016). However, while these plants may lose the capacity to photosynthesise they often still retain the genetic components with many photosynthetic proteins able to facilitate alternative signalling processes (Gu et al., 2021).

One common feature that is shared by all parasitic plants is the haustorium; a specialised organ which is used for invasion of host cells and enables subsequent connection between the vasculatures, allowing for multidirectional transfer of nutrients, genetic material and metabolites (Bouwmeester et al., 2021). While these parasitic plant structures have some functional analogy to fungal haustoria, since they are used for host invasion, they in fact are rather different organs. In the case of fungal (and oomycete) haustoria, these unicellular protrusions from hyphae or spores infect hosts intracellularly, invaginating cell walls and forming an extrahaustorial membrane through which small molecule transfer can occur (Koh et al., 2005). Parasitic plant haustoria on the other hand are larger, multicellular organs which persist intercellularly and attempt to grow inwards through host tissues in order to form a connection with host vasculature, unless an incompatible interaction is encountered.

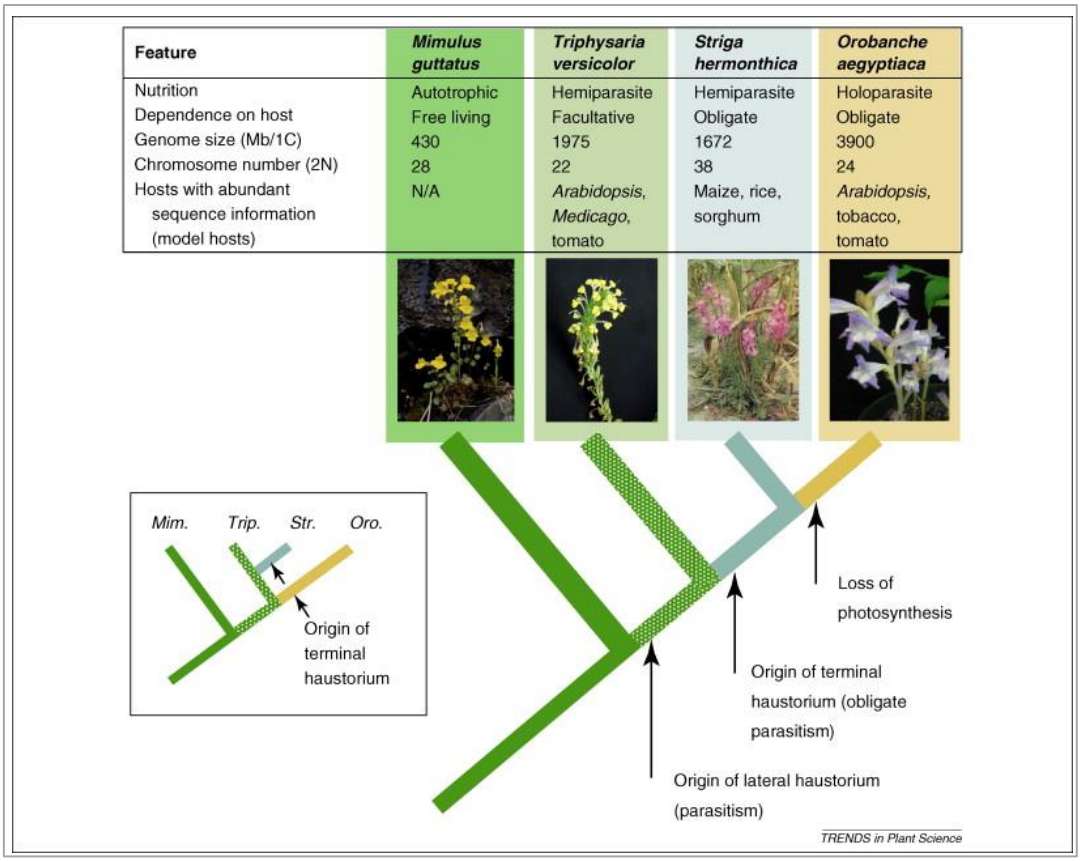


Figure 1.1 – Comparison of four different species of plant all from the Order Lamiales, which shows the changes seen from a non-parasitic lifestyle for *Mimulus guttatus* to an obligate holoparasitic lifestyle of *Orobanche aegyptiaca*. Figure taken from Westwood et al., (2010).

1.2 The Striga problem

Agriculture is one of the largest and most important industries in the world. Nowhere is it more essential than in sub-Saharan Africa, where around 57% of the population work in agriculture (The World Bank, 2017), relying on crop production for either subsistence or commercial farming. However, with a growing population and changing environmental conditions, crop production has not been optimal and many rural populations have struggled to overcome poverty and food shortage (Chauvin *et al.*, 2012). In particular, agricultural productivity losses attributable to biotic factors such as pests, weeds and infection pose a highly significant opposition to food security (Savary *et al.*, 2012). The population of this region is projected to almost double by 2050 (United Nations, 2022) and despite the percentage of malnourished individuals decreasing, the absolute number of people suffering with the consequences of famine has increased in the same period. This has led to food security being a continued concern due to low agricultural productivity (OECD and FAO, 2016).

Figure 1.2 – Images of the red-flowered *Striga asiatica* amongst fields of rice in Madagascar. Photos provided by Prof. Julie Scholes, University of Sheffield.

The plant family Orobanchaceae consist almost entirely of parasitic plants and many of which are particularly economically important due to their negative impacts on agriculture. These include the Broomrapes (*Orobanche* spp.), witchweeds (*Striga* spp.) and dodders (*Cuscuta* spp.). While *Striga* spp. are more prevalent in hot and dry climates, the greatest species diversity of *Cuscuta* are in more tropical, humid climates and *Orobanche* have a preference for dry temperate climates (Joel et al., 2007). The main focus of this study is Striga; it's devastating effects are often described as "the curse of poor farmers on poor land" (Radford, 2003) due to the parasite's preference for soil with low fertility in Sub-Saharan Africa (Bekunda *et al.*, 2010).

The distribution of Striga species is particularly widespread across the African continent and beyond, and has been reported in more than 50 Countries, the three species causing the foremost threat to agriculture being *Striga hermonthica* (Del.) Benth, *Striga asiatica* (L.) Kuntze and *Striga gesnerioides* (Willd.) Vatke. Both the pink flowered *S. hermonthica* and red flowered *S. asiatica* (Figure 1.2) are parasites of monocots, both cereal crops and wild grasses, while *S. gesnerioides* infects only dicots including crops such as cowpea (*Vigna unguiculata*). Geographic ranges for the three species overlap across Africa with *S. hermonthica* infecting crops in East Africa and West Africa. *S. gesnerioides* has a similar but more extensive distribution down to South Africa. *S. asiatica*, the most internationally spreading, infects crops in East Africa, West Africa and

Madagascar (Figure 1.3) as well as in the Arabian peninsula, India, Pakistan and a now-controlled outbreak (due to importation of contaminated maize seeds) seen in the USA (Kebede et al, 2018).

The origin of Striga, is theorised to be in the tropical savannah between the Semien Mountains of Ethiopia and Nubian Hills of Sudan; coincidentally this area is also known to be the same region in which sorghum and pearl millet were domesticated (Ejeta, 2007, Atera, 2011), two of the major host plants. The species' went on to proliferate throughout Africa and emerged into Asia as well as arising in the USA in the mid 20th century (Parker et al., 1993; Parker, 2012). As a consequence of introduction into the USA, an immense multimillion dollar effort was implemented to confine the parasites to one area and despite recurrences, it is no longer a threat to modern agricultural systems there. Studies have shown strong genetic uniformity within the USA population of *S. asiatica*, which is highly suggestive of a single introduction event (Werth et al., 1984).

In Africa, while Striga is regularly found in savannahs, parasitizing wild grasses (Mohamed et al., 2001), three of the key species (*S. hermonthica*, *S. asiatica* and to a lesser extent *Striga aspera* (Willd.) Benth.) also play a big role in decimation of agricultural cereal crop yields and have been estimated to cause around $7 billion in crop losses annually, causing up to 100% of the total crop yield to be lost for the most susceptible varieties (Jamil et al., 2012). The worst affected individuals are primarily small-holder farmers due to limited affordable treatment methods being available to them. In most arid or semi-arid areas across SSA there is erratic rainfall, which up to 90% of farmers rely on to water their crops (Araya et al., 2010; Sibhatu, 2016), as well as low organic matter levels, intensified land use and expansion of cereal production. This has led to higher competition with weeds such as *Striga* spp. for the scarce nutrients and an increase in the incidence of Striga infestation (Gethi et al. 2004; Parker, 2009). Soil fertilisation, land irrigation, and the utilisation of annually-acquired seed varieties of resilient cereal crops stand out as highly effective strategies. However, numerous farmers still depend on manual weed removal, which has minimal impact on reducing the soil seed bank. Furthermore, a significant portion of the harm to the host plants is irreparable and occurs prior to the emergence of striga from its subterranean phase.

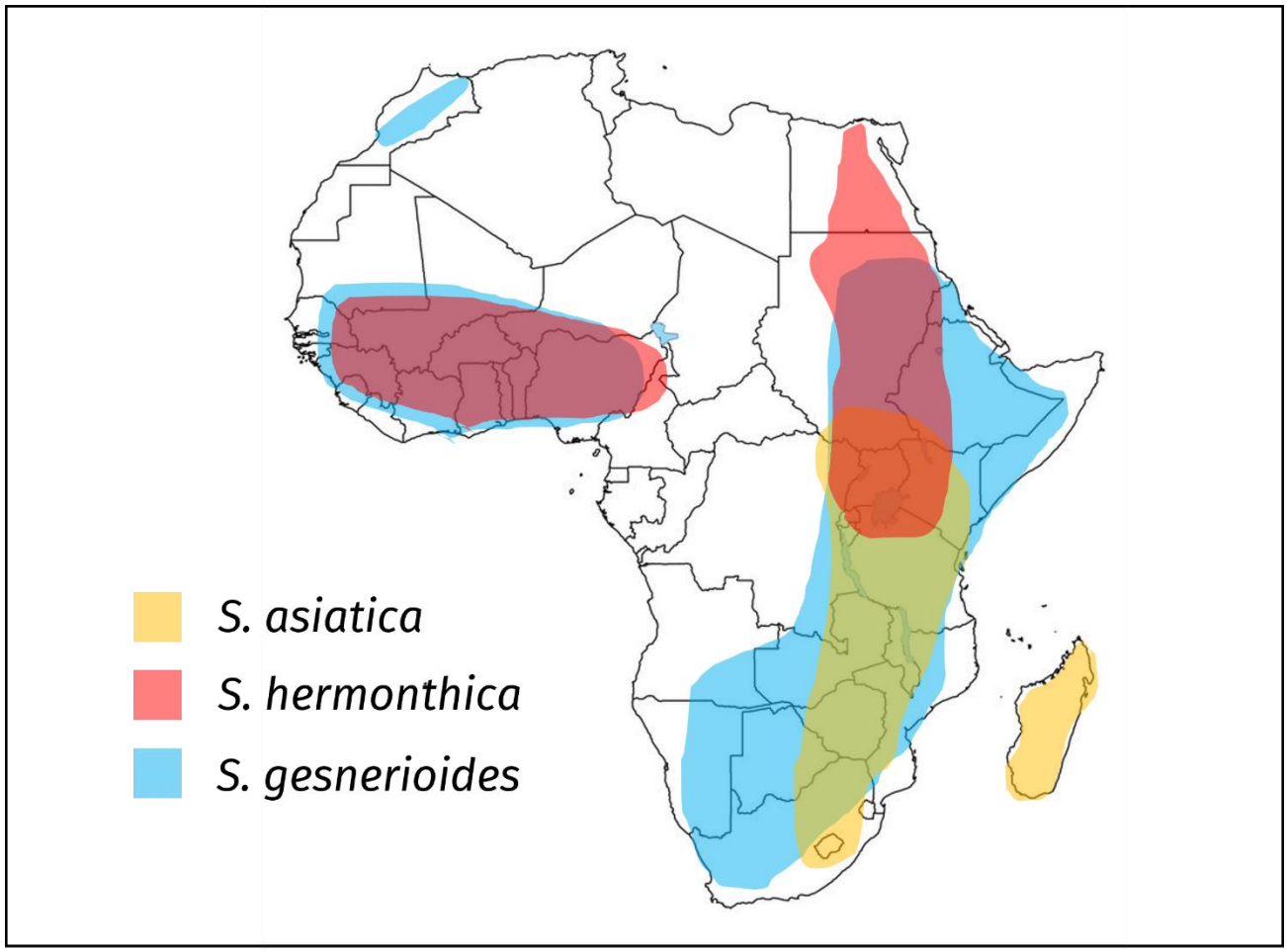


Figure 1.3 – Distribution of three of the most devastating species of Striga throughout Africa, *S. hermonthica*, *S. asiatica* which parasitise cereal crops and *S. gesnerioides* which parasitizes dicotyledonous crops. Adapted from Spallek et al. (2013)

1.2 The Striga life cycle

The life cycle of Striga species is wholly dependent on the host, to which the Striga must attach and form a vascular connection, in order to gain vital nutrients and water. The process can be divided into two phases: pre-attachment and post-attachment (Samejima and Sugimoto, 2022) as detailed in Figure 1.4. The seeds of Striga species are incredibly small, measuring around 200 micron in diameter (Spallek et al., 2013); an important factor in the parasite's success is due to the large numbers of miniscule seeds produced from every plant, leading to a rapid increase in parasite population density after a single generation (van Delft et al., 1997). Up to 500,000 seeds are released from the dried seed pods of each individual plant and lie dormant in the soil bank (Berner et al., 1995).

Seeds must overcome a first dormancy period where the embryo matures and seed coat becomes more permeable over time (Okonkwo and Raghavan, 1982). Next, seeds go through 'wet dormancy' which is

overcome through a conditioning period of moisture and heat; they can cycle through this rehydration and dehydration for a number of years and then it is only upon contact with germination stimulants (usually strigolactones) produced by a nearby host that some fully mature individuals will begin to germinate (Mohamed et al., 1998). This ensures that suicidal germination does not happen and a host is available to infect immediately as well as rendering the seed resistant to a number of herbicides that are effective post-germination (Runo and Kuria, 2018). The seed dispersal mechanism is varied, with such fecundity and their small size, *Striga* seeds can be spread by wind, water, farming machinery or animals, as well as being concealed amongst harvested grains (Csurhes et al., 2016). Unconditioned seeds may also remain in the soil seed bank for 20 years or longer; Studies by Bebabwi et al. (1984) found at least 10% germination of *S. asiatica* seed that had been buried for 14 years, deeper than 30 cm.

Following exposure to consistent moisture and warmth, seed become receptive to germination stimulants (Gbehounou, 2003). These stimulants are secreted as part of the root exudate of many host (and non-host) plants, delivered as complex communication signals out into the rhizosphere (Steinkellner et al., 2007), the best studied of which are the carotenoid-derived strigolactones. While the presence of strigolactones is thought to be rather ubiquitous across plant species, these compounds are thought to have two intended functions, as endogenous hormones controlling plant development and secondly as root exudates to promote branching of fungal hyphae in symbiotic interactions between the plants and soil microbes (Mishra et al., 2017).

Sensing of strigolactones by the seeds of parasitic Striga happens through receptors such as the *S. hermonthica* hyposensitive to light receptor proteins (ShHTL) identified by Tsuchiya et al. (2015) which have been seen to have a broad spectrum response to these stimulants (Wang et al., 2021a). Throughout their evolution, Striga species have had to adapt their signalling and perception apparatus for detection of potential hosts, dependent on their environment due to the wide range of strigolactones in the rhizosphere, often arising from different genotypes of the same host species (Mazaka et al., 2023). The presence of a suitable host and directional growth towards that host is coupled through this sensing, as a radicle grows towards the host root before tissues differentiate into the haustorium in response to host-derived haustorial-inducing factors (HIFs) (Runo and Kuria, 2018). Unlike the strigolactones, perception of HIFs is not a passive undertaking, instead Striga accumulate reactive oxygen species (ROS) to act as a partner in peroxidase action at the host cell-wall leading to production of HIFs such as 2,6-dimethoxy-*p*-benzoquinone (DMBQ) (Palmer et al., 2008; Wada et al., 2019). Suggestions have been made that lignin depolymerisation or degradation at the host cell wall by laccase is able to produce compounds that function as HIFs leading to parasite haustoria induction (Kamaya et al., 1981; Cui et al., 2018), despite it being well recognised that parasitic plants generally have weak lignin-degrading capabilities.

For the Striga haustorium, the primary objective is to penetrate successfully into the host root, overcoming host immune defences to reach the vasculature and form a bridge between parasite xylem and host xylem (Yoshida et al., 2016). The invasive phase of haustorium development begins following perception of HIFs, the parasitic radicle starts to show morphological changes including cell expansion and division and differentiation of haustorial hairs at the tip (Riopel and Musselman, 1979; Yoshida et al., 2016; Cui et al, 2018).

Secretion of a prominent surface adhesive from the parasite allows haustorial hairs to act as anchor points and attach to the host root (Baird and Riopel, 1985; Yoshida et al., 2016; Kokla and Melnyk, 2018). Following this, swollen protoplasmic cells at the apex of the radicle (pre-haustorium) will elongate becoming a column of intrusive cells and attempt to grow inwards through the root epidermis, cortex and endodermis layers before reaching the vasculature (Figure 1.5), dependent on host immune defences. A continuous xylem connection is made from host to parasite through further differentiation of certain contact cells (termed the osculum) which lose their cytoplasmic contents to become the water-conducting vessel elements (Ekawa and Aoki, 2017), building a bridge for nutrients and water to be siphoned away to further the parasite's subterranean growth stage (Yoshida et al., 2016).

In contrast to the hemiparasitic Striga, parasites such as *Cuscuta* and certain *Orobanche* achieve a connection with both xylem and phloem because they are not photosynthetically active and must also harvest carbon from their host (Ekawa and Aoki, 2017). One incredible facet of this invasion and the connection made with the host plant is that the parasite and host may often be distantly related, in stark contrast to traditional grafting techniques which will only be successful with closely related graft and host, or else be subject to rejection (Melnyk et al., 2015).

What is particularly devastating in Striga infection is that once a connection has been made with the host, the damage caused is usually irreversible (Jamil et al., 2010); it is also interesting to note that the biomass gained by the parasite is not proportional to the far greater losses suffered by the host which may be suggestive of strong toxicity or at least major negative metabolic influence of compounds reciprocally delivered into the host by the Striga (Gurney et al, 1999). Symptoms exhibited by the host can depend on the level of tolerance and resistance but in general hosts will suffer from chlorosis and stunting within days and eventually with loss of potential yield. While host varieties may exhibit similar levels of infestation, signifying their susceptibility, a strong level of tolerance can lessen the negative effects of the Striga and the impairment felt by the host (Rodenburg et al., 2006a). A tolerant host variety is defined as being able to achieve a higher grain yield in the presence of Striga when compared to another variety with the same level of Striga infection (in both cases as a % of the yield achieved in an uninfected plant) (Rodenburg et al., 2006b; 2017). Tolerance has been emphasized as a highly important area of Striga research which should be combined with resistance in breeding programs (Shaibu et al., 2021). Subterranean growth of the parasite continues until it emerges above

the soil and begins to photosynthesise and leafy growth shoots up before flowering. After 16 weeks Striga plants have matured, dried and set seed completing their life cycle (Musselman, 1980).

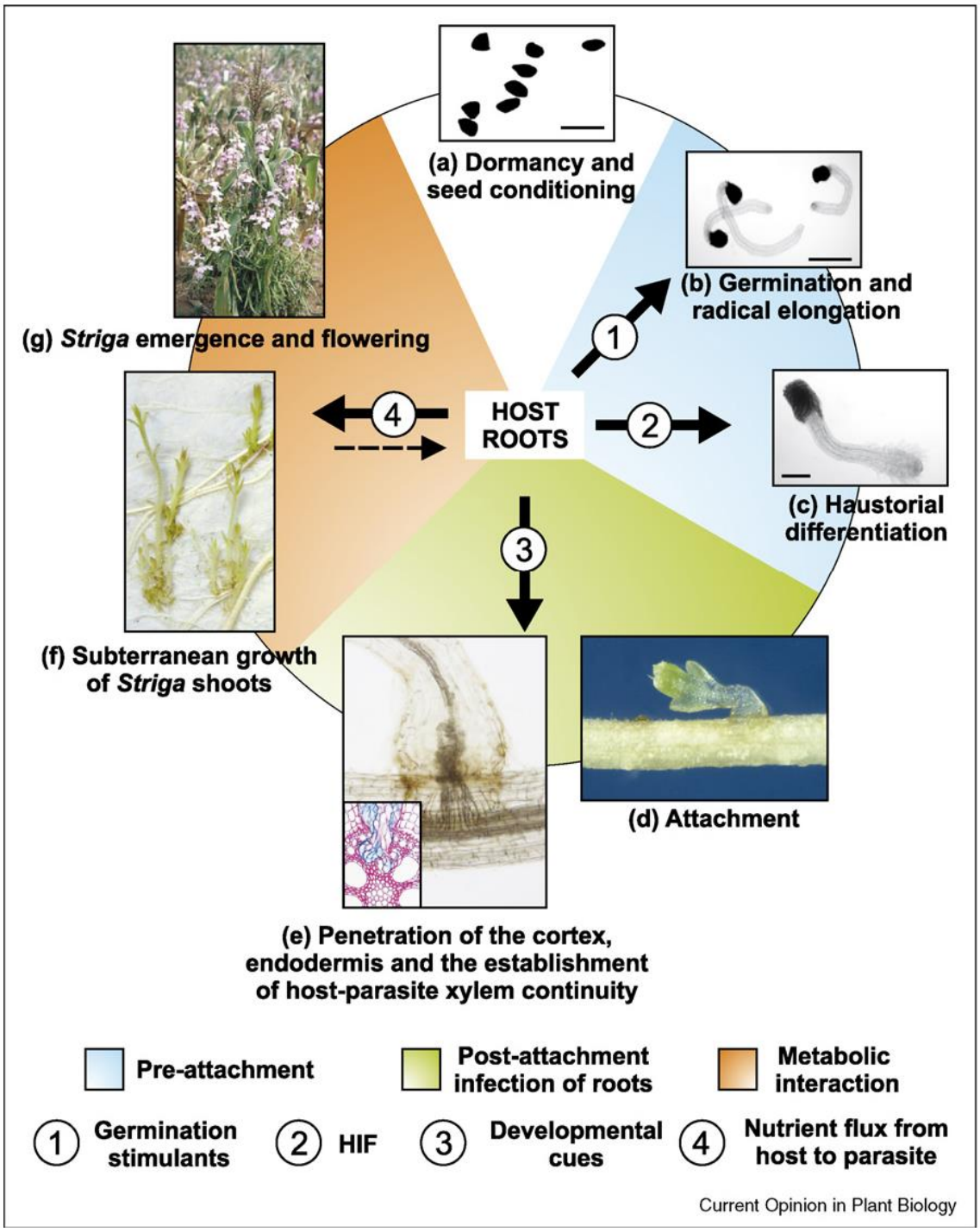


Figure 1.4 – The life cycle of *Striga* species can be split into two main categories – pre-attachment to a host and post-attachment. As obligate hemiparasitic root parasites, Striga spp. infect their host through penetration of host root tissue before establishment of a vascular connection between the two organisms and subsequent sequestering of nutrients to aid growth. Germination of the seeds occurs after perception of a nearby host germination stimulants, including strigolactones and haustorial inducing factors, when the tissue differentiates into the penetration organ, the haustorium. Much of the parasite's growth taking place below the soil, with detrimental effects on the host being seen before the Striga has emerged above surface level. Figure taken from Scholes and Press (2008). Copyright 2008 by Elsevier, reproduced with permission of the copyright holder via copyright clearance centre.

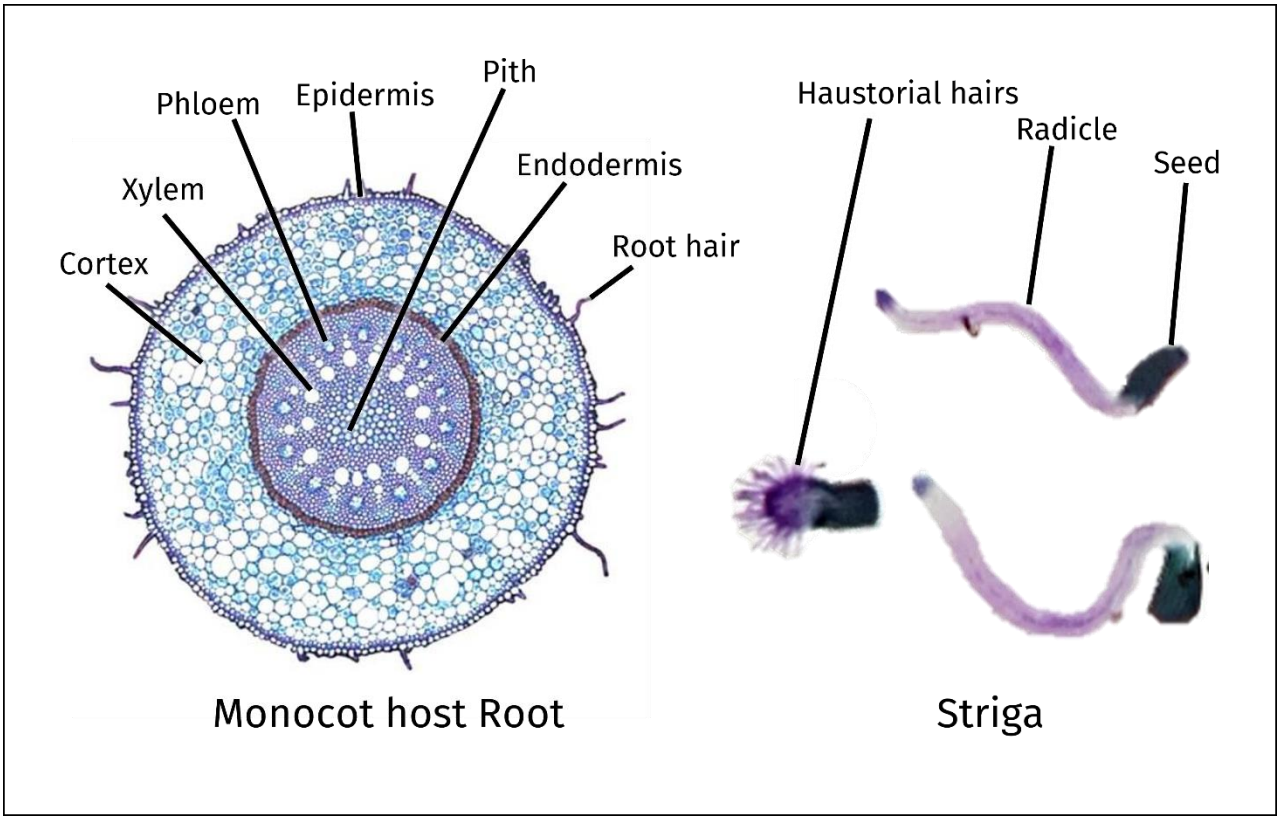


Figure 1.5 – A cross section of a monocot host root and a germinated Striga seed. Newly germinated Striga seeds will form a host-penetrating haustorium on the tip of a root-like radicle before a portion of which will attempt to breach the many layers of host tissue to form a vascular connection. Diagram not to scale.

1.4 Population structure

The flower morphology of *S. asiatica* is highly indicative of its self-pollinating nature in which fertilisation occurs before the corolla is open (Nickrent and Musselman, 1979), whereas *S. hermonthica* morphology has been shown to favour outcrossing as well as exhibiting strong self-incompatibility and floral adaptations which visibly favour long-tongued insects (Safa et al., 1984). While one advantage of a selfing lifestyle lies in the short-term reproductive assurance, particularly in less auspicious environmental conditions, other advantages may include reduced outbreeding depression - as seen for less favourable hybrid crosses of *Arenaria uniflora* and *Arenaria glabra* in certain areas with overlapping pollinator ranges (Fishman and Wyatt, 1999).

One of the major impacts of self-fertilisation on genetic diversity is the increase in homozygous alleles which can lead to inbreeding depression (a decrease in fitness in offspring of related parents), however this is often temporary and any deleterious recessive alleles can be rapidly removed from populations (Garcia-Dorado, 2012). In addition, inbreeding causes a weakening of effective recombination due to more closely linked loci, influencing selection patterns (Raffoux et al., 2018). This leads to a reduction in effective population size and

so genetic drift (the frequency of an allele within a population no longer being independent or owing to chance) increases.

The mating system of a species is a major determinant of the neutral genetic structure of populations, with predominantly selfing biological systems exhibiting low intra-population and high inter-population variation (Charlesworth et al., 2003). Owing to these low levels of genetic diversity within populations, the effect of local adaptation can be very strong leading to high population subdivision (Loveless and Hamrick, 1984). For some selfing parasitic plant species this has led to the categorisation of genetically distinct races, such as the seven distinct races of *S. gesnerioides* on cowpea (Botanga and Timko, 2006; Ohlsen and Timko 2020), and six races of *Orobanche* known to exist on sunflower (Velasco et al., 2007; Chabaud et al., 2022). This is in stark contrast to *S. hermonthica* which has been seen to be highly genetically variable across a number of study populations with far less population structure seen with respect to *S. asiatica* and *S. gesnerioides* (Gethi et al., 2005; Unachuksonahwu et al., 2017; Welsh and Mohamed, 2011). Studies of population variability are important facets to resistance breeding programs in order to determine patterns of local adaptation or host specificity in order to find the best targets for appropriate sources of resistance.

1.5 Control strategies

A range of control methods have been used for dealing with Striga for a number of years, focusing on three separate aspects: (1) prevention of initial infestation (2) reduction of the Striga soil seed bank and (3) containment of infested fields. Many of these control methods are traditional cultural practices that have been undertaken for many years by farmers in developing countries, including crop rotation, labour intensive hand weeding and the use of trap crops (Kanampiu et al, 2007), though farmer acceptance is low for some newer methods, with research identifying information dissemination as a key area to improve (Murage et al, 2011). Simple solutions such as introduction of fallow years and crop rotation have been consistently shown to improve soil fertility as well as lower Striga infestation, as described in Cameroon (Ayongwa et al, 2009) and Madagascar (Scott and Freckleton, 2022), with the use of multiple crops in the rotation having a consistently negative effect on Striga weed density. In addition, 'push-pull' technologies - the utilisation of trap crops such as *Desmodium* spp. (Hailu et al., 2018) – have been shown to restore soil fertility and moisture content, while decreasing Striga soil seed bank through germination and subsequent stalling of growth through release of allelochemicals. This method is also successful at controlling other pests such as stem borer and fall army worm (Midega et al., 2010). While these methods tend to allow germination of Striga before lessening its effect, one strategy employed to avoid initial germination is through application of fertilisers. High levels of nitrogen fertiliser or chicken manure have equally shown a suppressive effect on Striga emergence (Woldemariam et al., 2020), while high levels of soil nitrogen may lead to seedling scorching or

toxicity, an abundance of rich soil around the host also leads to a reduction in strigolactone production by hosts due to the more optimal environmental conditions.

The use of herbicides is challenging, primarily due to complications in differential targeting of parasitic plant over host plant. Treatment of herbicide-resistant cereal seeds with the acetolactate synthase inhibiting herbicide, imazapyr, prior to sowing was initially introduced in the 1990s with reduction of Striga infestation seen frequently (Menkir et al., 2010; Chikoye et al., 2011). Despite the success, results are not consistent with seasonal effects such as increased rainfall being a limiting factor in efficacy (Kamara et al, 2020). Biocontrol has been another interesting topic in control. The fungal pathogen *Fusarium oxysporum* f.sp*. Strigae* (Fos) has been the subject of trials that have seen incidence of Striga infestation decrease in small scale trials (Marley et al., 1999; Yonli et al., 2010; Oula et al., 2020;Anteyi and Rashe, 2021), but is not yet widely adopted. The method by which Fos attacks its host is through hyphal colonisation and subsequent occlusion of vasculature. Biocontrol is a most promising and widely accepted treatment for Striga as shown in trials by Nzioki et al. (2016) where inoculation with Fos increased maize yields by up to 50% compared to non-inoculated fields across the fields of 500 smallholder farmers.

Despite the challenges in complexity, by far the most effective and sustainable control seems to be an integrated approach with a strong focus on the use of resistant and/or tolerant varieties alongside many of the other management practices previously mentioned (Rodenburg et al., 2010; Jamil et al., 2021). Resistance to Striga is defined here as the prevention of Striga infection of a potential host, with few or no attachments, as opposed to tolerance where hosts can still be susceptible to attack but are able to endure and still produce a smaller, yet viable yield (Rodenburg et al., 2006b). Development of resistant lines has been ongoing for many years, the major constraints being detection of genetic sources of both durable and broad spectrum resistance to the continuously evolving and geographically diverse *S. asiatica* and *S. hermonthica* (Mandumbu et al., 2019). Particularly in the case of *S. hermonthica*, this ability to overcome resistance within only a few generations is well known (Runo, 2019) due to the strongly diverse gene bank, and as such stacking of multiple resistance genes would be only essential for a successful breeding program (Rich and Ejeta, 2008).

Research on Striga resistance has differed in its focus on resistance type, from identification of crop varieties associated with reduced germination stimulant production (pre-attachment resistance) to those which display clear physiological barriers to parasite penetration (post-attachment resistance). While traditional breeding programs introduce new potentially favourable phenotypic variation into crop varieties through often empirical selection, marker assisted selection (MAS) determines resistance independent of environment and in an accelerated time frame (Badu-Apraku et al., 2021; Kushanov et al, 2021). With the advent and ever reducing-cost of genome sequencing and DNA marker technology, new opportunities to dissect the complexities of Striga polygenic resistance are constantly evolving (Timko et al., 2007), approaches

such as Genome-Wide Association Studies (GWAS) and Quantitative Trait Loci (QTL) analyses are becoming more widely used for gene discovery (Uffelmann et al., 2021). Using marker assisted breeding allows for the transfer and pyramiding of resistance-associated DNA alleles into locally adapted cultivars (Desphande et al., 2013).

The types of resistance to Striga can be split into two broad categories: pre-attachment and post-attachment and different suites of genes likely contribute to each. This phenotypic diversity of resistance was exemplified recently in an extensive study by Samejima and Sugimoto (2023) following pre- and post-attachment screening of 69 different World Rice Core collection cultivars where varying levels of resistance were seen to *S. hermonthica*. Successful germination of Striga from host root exudates was not correlated with successful parasitism. Physical characteristics associated with resistance to Striga infection and as such, targets for breeding programs include low production of germination stimulant (mechanistically the leading cause of pre-attachment resistance), physical barriers such as lignified cell walls, phyto-alexine synthesis, post-attachment hypersensitive reactions, necrosis, incompatibility and insensitivity to Striga toxins like iridoid glycosides (Makaza et al., 2023). Potential sources of variation for these traits may come from Landraces, inbred lines, wild relatives or species hybrids (Yacoubou et al., 2020).

Several crop varieties have been developed with good pre-attachment resistance (Dell'Oste et al., 2021); These include sorghum species with a mutated Low Germination Stimulant 1 (LGS1) locus which alters the stereochemistry of 5-deoxystrigol, producing instead the weak germination-inducing orobanchol (Gobena et al., 2017), as well as being replicated in rice (Jamil et al., 2011) and maize (Yoneyama et al., 2015). Elucidation of the mechanisms and genic apparatus of post-attachment resistance in hosts have been less well characterised due to their complex nature. However, in *S. gesnerioides* one such gene, termed RACE-SPECIFIC-STRIGA-GESNERIOIDES-3 (RSG3-301), has confirmed the presence of leucine-rich-repeat (LRR) immune receptors to Striga species (Li and Timko, 2009). This forms a monogenic resistance complex with an as yet-unidentified complementary *S. gesnerioides* gene product (effector) in a gene-for-gene response triggering a signal transduction cascade (Kaloshian, 2004), as seen in many pathogen-host interactions such as the widely studied Arabidopsis EFR pattern recognition receptor which binds to EF-Tu, a highly conserved protein amongst bacterial species (Piazza et al., 2021).

While *S. gesnerioides* resistance responses are expected to be highly qualitative and race-dependent (as *S. gesnerioides* is a predominately self-pollinating species), for *S. hermonthica* and likely also for *S. asiatica* resistance is more likely to be quantitative, complex and polygenic. Although polygenic resistance is often more long-lasting than monogenic resistance, with the additive impact of many genes, resistance is usually not as complete as it is in a monogenic gene-for-gene interaction between host and parasite. This monogenic gene-for-gene resistance model explains how new pathogenic races are able to form; as the avirulence gene

mutates to the point where it is no longer recognised by the host, the gene product becomes virulent in its actions against the host and resistance is no longer effective until a new mutation of a gene has arisen in the host (Sacristan et al., 2021). In both maize and sorghum, QTL analyses have identified a number of loci associated with resistance and tolerance traits such as decreased *S. hermonthica* emergence and increased grain yield (Haussmann et al, 2004; Yacoubou et al., 2020; Badu-Apraku et al., 2020). A similar picture was seen in GWAS studies on *S. asiatica* by Pfunye et al., (2021) who again identified 3 prospective loci on chromosomes 5, 6 and 7 associated with decreased *S. asiatica* emergence.

1.6 Striga virulence

Just as it is important to understand the mechanisms behind the host's resistance, we must also determine the mechanisms underlying the parasite's virulence as these are so closely intertwined. In general, for many phytopathogens the most general definition of virulence would be the reduction in host fitness caused by infection (Read, 1994), though fitness is difficult to measure empirically so in experimental terms is often quantified by measuring extent of colonisation (Cressler et al., 2016).

Therefore, in this study virulence is described as the success of the parasite, measured through the magnitude of infection. In field trials and pot experiments this has been determined by the number of Striga that emerge above the soil surface. In rhizotron experiments, parasite virulence is measured by either total biomass (dry weight) per host root system, cumulative length per host plant, or number of *S. asiatica* individuals per host plant (Gurney et al., 2006; Runo et al., 2012; Rodenburg et al., 2015; Rodenburg et al., 2017).

Variation in virulence is seen across many *Striga* spp., with regards to variation between crop species, between cultivars within a crop species as well as average virulence across hosts and specificity to certain host varieties. While this has been described, the mechanisms are not fully understood and a key question in Striga biology is to what extent genetic variability of the parasite plays a role in determining host range and virulence. Currently this is less clear for *S. asiatica* than it is for *S. hermonthica*.

For susceptible plants, a strong defence is vital for long term species survival and as such, plants have evolved a multitude of ways to combat both biotic and abiotic stresses that they may encounter. Host defence mechanisms exist alongside a range of parasite counter-adaptations that have co-evolved to evade or overcome host defences (Buckingham and Ashby, 2022). The parasitic component of these virulence/resistance complexes can be broadly referred to as either Pathogen Associated Molecular Patterns (PAMPs) or effectors, forming the first and second layer of host defence, a simplified model of which is shown in Figure 1.6.

PAMPs are considered to be highly conserved components of the parasite which can not be recognised as self-derived by the host. Commonly occurring PAMPs identified in other pathogenic species include bacterial

lipopolysaccharide (LPS), flagellin, peptidoglycan, oomycete glucans and fungal chitin, these are components of the parasite itself and not actively delivered into the host (Felix et al., 1993; Dow et al., 2000; Thomma et al., 2011). Recognition of PAMPS by host surveillance systems forms the first line of defence in plant immunity. Recognition follows binding of the PAMP ligand to a cell surface host Pattern Recognition Receptor (PRR) proteins (Block et al., 2008) such as a receptor-like-protein (RLP) or receptor-like-kinase (RLK) which leads to a PAMP triggered immunity (PTI) response.

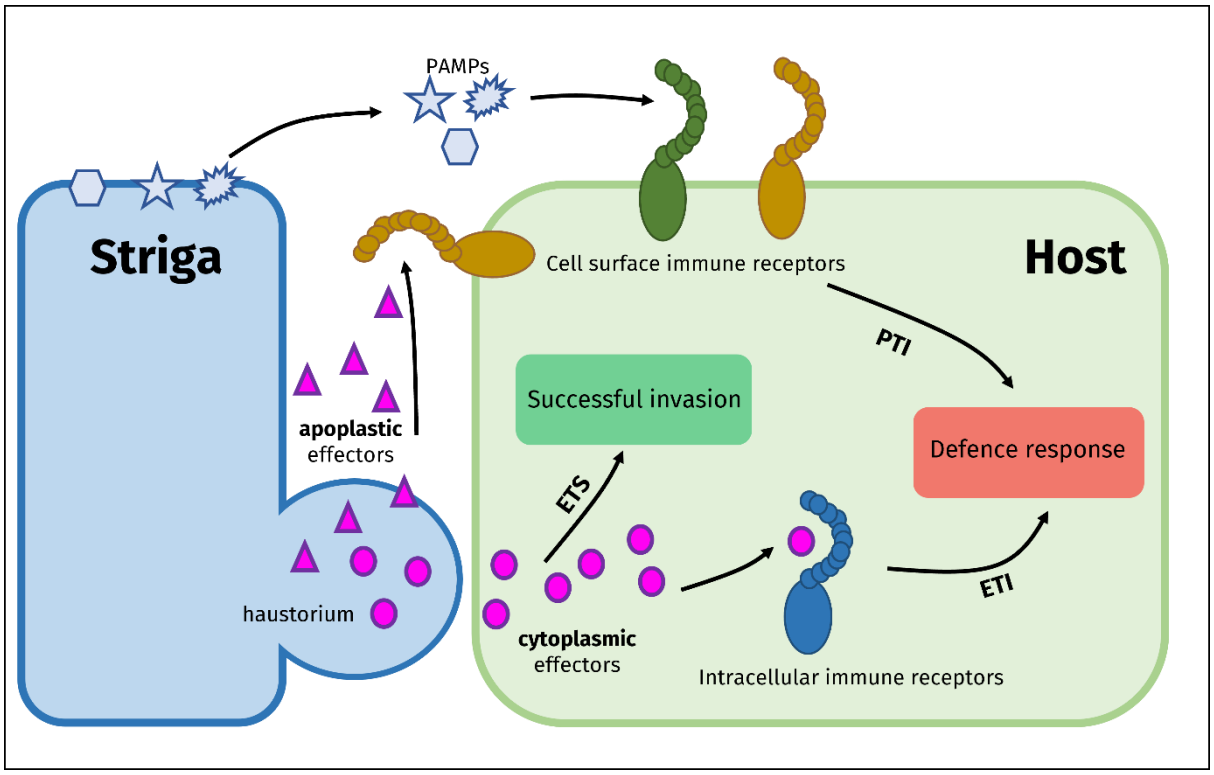


Figure 1.6 – Simplified model showing defence and resistance of *Striga* spp. infection. The Striga is first recognised by conserved pathogen associated molecular patterns (PAMPs), conserved parasitic signatures on the cell, by host pattern recognition receptors, eliciting a PAMP triggered immunity (PTI) response which is thought to confer partial resistance through signalling cascade responses. The next step in parasite-host interaction is the secretion of parasite effector molecules from the haustorium which can localise apopplastically or cytoplasmically into the host to facilitate colonisation. Depending on suitability of the host, these effectors will either be detected by nucleotide binding leucine rich repeat receptors (NB-LRRs) and activate the next level of host immunity called effector-triggered immunity (ETI), far stronger than PTI, or alternatively, evade the immune system in a process termed effector-triggered susceptibility (ETS). With ETS, effector molecules can travel intracellularly, disrupting cellular processes by further suppressing immunity or facilitating enhanced parasite growth.

One well-studied PTI response is triggered by the bacterial flagellin peptide, flg22 which elicits a significant immune response in Arabidopsis upon binding to the leucine rich repeat (LRR) PRR, FLS2 followed by formation of a physical barrier through callose deposition at the host cell wall (Sun et al., 2012). Such a response confers partial immunity. This has also been highlighted by a recently identified RLP receptor in tomato, CuRe1, which detects a PAMP from the shoot parasitic plant *Cuscuta reflexa* Roxb. (Hegenauer et al.,

2016; Hegenauer et al., 2020). The PAMP itself was found to be a short peptide motif called Crip21, a fragment of the larger glycine-rich protein, CrGRP. Recognition of Crip21 was found to lead to an upregulation in ethylene biosynthesis and cell death, a traditional PTI response.

The second group of parasitic elicitors of immune response are the effectors. Effectors are primarily proteinaceous biological molecules but small RNAs and toxic metabolites also play key roles in facilitating infection as well as manipulation of host metabolism (Collemare et al., 2019). Following secretion from cells of the parasitic haustorium, some effectors will localise to the apoplast while others will be translocated to the host cytoplasm. For the host, detection of apoplastic effectors is achieved at the cell surface by immune receptors such as RLPs whereas cytoplasmic effectors are detected by intracellular nucleotide-binding Leucine rich repeat (NB-LRR) proteins (Yuan et al., 2021).

Those effectors which are specifically recognised by the NB-LRR receptors go on to activate immune signalling cascades in the host and resulting in Effector Triggered Immunity (ETI) in a similar way to PTI, but instead with a stronger, quicker and more amplified immune response. Resistance may then be enacted through a hypersensitive response at the site of haustorial penetration or other acute reaction (Jones and Dangl, 2006). The gene-for-gene model predicts that these effectors, produced by a so-termed, avirulence gene (Avr) share a one-to-one correspondence with a matching host resistance (R) gene (Flor, 1971).

Success for the parasite is in evasion or disruption of host receptors by these effectors in an effector triggered susceptibility (ETS) interaction (Jones and Dangl, 2006). Effectors are highly varied in the roles they play in establishment of infection and much is still to be determined, with much of what is known, described in fungal or bacterial pathogens. Functions are thought to include suppression of reactive oxygen species (ROS), inhibition or induction of hormone signalling, inactivation of PRRs, inhibition of host enzyme activity and secretion of cell wall degrading enzymes (Tian et al., 2005; Chini et al., 2007; Gimenez-Ibanez et al., 2009; Hemetsberger et al., 2012 ;Albert et al., 2015; Gimenez-Ibanez et al., 2014). One such example is the corn smut (*Ustilago maydis*) effector Pep1 which acts upon its host to suppress immunity, in part through targeting peroxidase activity (Hemetsberger et al., 2012). The main function of peroxidases in the host is to facilitate dehydrogenation of substrates through ROS activity in order to synthesise protective cell wall components such as lignin, suberin and auxins (Lamb and Dixon, 1997).

Within parasitic plants, research of virulence effectors has predominantly, thus far, been limited to genomic prediction, with research many years behind other phytopathogens (Timko and Scholes, 2013). There has, however, been a strong focus on the actions of plant cell wall degrading enzymes (PCWDE) and their role as putative virulence effectors in facilitating cell wall modifications. These have included pectinesterase, polygalacturonase and glycosyl hydrolases, among others found in numerous transcriptomic studies (Zhang et al., 2015; Jhu et al., 2021; Lopez et al., 2019). Comparative transcriptome analyses by Yang et al. (2015)

revealed a core set of so-called parasitism genes across the parasitic species of *S. hermonthica, P. aegyptiaca* and *T. versicolor*. The most highly enriched GO-terms found from this study were Serine-type or Aspartic-type endopeptidase activity molecular functions and cellular responses to ROS biological processes. Similarly Lopez et al. (2019) identified oxidoreductases and lyases as the most common categories.

While functional validation studies of these effectors amongst plant parasites are scarce, two recent findings have given more detail to these aspects of virulence. Firstly a small secreted effector protein from *S. gesnerioides* race 4z causes ubiquitination and degradation of Cowpea VuPOB1 protein, which then suppresses immunity by facilitating the loss of hypersensitive response (HR) and programmed cell death (Su et al., 2019). In addition, a cysteine protease from *C. reflexa*, cuscutain, is assumed to play a role in weakening host structures through protein degradation (Bleischwitz et al., 2010). Expression of cuscutain has been identified as mainly haustorial, emphasising the association of the proteins with parasite invasion (Amini et al., 2018). These functional studies point to significant roles of effectors in parasitism in general but are still to be elucidated further in *Striga* spp.

1.7 Genomic and population genomics approaches to understanding virulence

With the relative affordability of genomic sequencing, we can now study more species using comparative genomics than ever before. Genomic studies provide diverse insights into populations by employing various approaches. Genome-wide association studies identify genetic variations and markers associated with virulence traits, evolutionary analyses enable the construction of phylogenetic trees and lineage comparisons, population genetic studies model gene flow and assess selection pressures, and functional genomics shed light on gene expression, protein function, and pathways contributing to virulence.

With that in mind, the recently published reference genome of *S. asiatica* has opened the door for new analyses, to identify different mechanisms across populations within the species and enable identification of novel virulence effector candidates (Yoshida et al., 2019). The high quality genome produced by Yoshida et al. was generated from a single individual of *S. asiatica* collected in North Carolina, USA, the genome was 600 Mbp in size with an N50 scaffold size of >1.3 Mbp, from which a total of 34,577 genes were predicted. This individual being from an introduced population of *S. asiatica* to USA in the 1950s.While the genome provides essential background to understand the origin of parasitism in Striga, with rapid evolution seen, it also allows for further analysis to assist in effector candidate discovery and eventual functional validation of those candidates.

The first step in identification of virulence related genes is in-silico prediction following genome sequencing. While other types are known to exist, broadly speaking virulence-associated effectors will be secreted proteins which translocate through secretory pathways into the extracellular space (Agrawal et al., 2010). The classical plant protein secretion pathway is highly conserved and utilises N-terminal signal peptides for

transportation of proteins from the golgi apparatus (Alexandersson et al., 2013). However, unconventional protein secretion does exist and in fact, a significant proportion of secretome proteins lack these signatures as seen for the well-studied Arabidopsis genome (The Arabidopsis Genome Initiative, 2000). Computational prediction tools rely on identification of positively charged signal peptides, often 20-30 AA in length, which are closely linked to a hydrophobic region and cleavage site at the C-terminus (Sonah et al., 2016). Effectors are also consistently small, to accelerate ease of movement as well as being rich in cysteine content, for example, in the effector protein SnTox1 from *Stagonospora nodorum*, which contains almost 20% cysteine content of the 100 amino acids, not including signal peptide (Liu et al., 2012). Genomic and transcriptomic sequencing can elucidate these features with a parasite genome, then alongside BLAST searches for homologous effector proteins – particularly to a well characterised and annotated genomes like *Arabidopsis*, complex models are used to predict the secretome using tools such as SignalP (Peterson et al., 2011).

The availability of parasitic plant genomes also opens the way to the identification of VFs using a population genomics approach. Qiu et al. (2022) recently identified suites of VFs associated with the ability of *S. hermonthica* (collected from Kibos, Kenya) to overcome resistance in a normally resistant rice variety (NERICA-17), by combining a genomics and population genomics approach.

Prior to embarking on the research described in this thesis, considerable effort was dedicated to overcoming the dormancy of *S. asiatica* seeds. Various experimental mechanical and chemical methods were employed, including techniques such as sonication and the use of different stimulants. However, despite these efforts, none of the attempted methods significantly accelerated the time required for the seeds to become receptive to germination. Consequently, the only Striga accessions available for use in this study were those that had naturally reached approximately one year post-harvest. Compounding the challenges, the accessibility of additional seed samples was hampered by the COVID-19 restrictions in place during the research period. As a result, the seed samples already in possession were the sole viable option for analysis in this project.

Furthermore, it is important to note that while the *S. asiatica* accessions collected from regions outside of Madagascar were not sampled in the same manner of field collection, their inclusion in this study provides some insight into the extent of virulence variation. However, it is crucial to exercise caution when interpreting the spatial significance of these findings. The populations in these regions were not necessarily natural populations in their respective habitats. For instance, the S. asiatica samples from the USA were introduced and quarantined, while the Ethiopian samples may have undergone outcrossing under controlled laboratory conditions. Moreover, the introduction of S. asiatica from mainland Africa to Madagascar may have occurred on multiple occasions, potentially leading to genetic exchanges between different lineages. Additionally, the presence of dormant seeds in the soil could have influenced the life cycle timings and overall dynamics of the

populations. These factors highlight the need to consider the limitations and potential complexities associated with the interpretation of spatial relationships in this study.

1.8 Aims and objectives of this thesis

The overarching objective of this thesis was to better understand the molecular genetic basis of virulence variation in *S. asiatica*, exploring the effects of population structure, genetic diversity and patterns of virulence to facilitate prediction of virulence-related genes. The specific aims for each chapter were:

I. **Chapter 2:** to determine the patterns of variation in virulence of *S. asiatica* accessions on a range of rice genotypes with varying levels of susceptibility and determine whether host variety or *S. asiatica* accession is the most important factor in explaining the variation in virulence.
II. **Chapter 3:** to determine the neutral population structure of *S. asiatica* accessions and assess the genetic variability and extent of selfing between groups based on SNP marker analysis.
III. **Chapter 4**: to take an experimental population genomics approach, utilising virulence data from chapter 2 to identify candidate virulence-related genes from the *S. asiatica* genome.

Chapter 2

Evaluating the extent of variation in virulence between *Striga asiatica* accessions on different rice hosts in Madagascar and determining whether host adaptation plays a significant role

2.1 Introduction

2.1.1 Background to Striga spp.

Parasitism is the relationship between two organisms, within which one party benefits to the detriment of the other. It is a widespread lifestyle across the Eukaryotes from animals to plants and fungi. The main commonality between these very different parasites is that they must maintain a great enough level of virulence to persist within their hosts but also ensure the host survives long enough for them spread effectively to others (Cressler et al. 2016).

Among the most notorious and widely studied parasitic plants are members of the Orobanchaceae family, due to the constraints they put on crop productivity in different parts of the world (Parker, 2009). Within this group, lies the noxious, parasitic weed species of the genus Striga, which are currently some of the most serious biotic constraints to cereal production in Sub-Saharan Africa (SSA) (Rodenberg et al., 2017; Runo and Kuria, 2018; Jamil et al., 2021). The preference of Striga spp. for arid and semi-arid conditions, along with low fertility, well-drained soils demonstrate just one aspect of their high level of adaptability. As obligate root hemi-parasites *Striga* spp. require a host to fulfil their life cycle but are also able to achieve low rates of photosynthesis in their above-ground developmental stages (Press & Stewart, 1987; Frost et al., 1997; Hegenauer et al., 2017; Jhu et al., 2022). Effects on hosts involve severe chlorosis, wilting and stunting (Spallek et al., 2013), all of which can happen before the parasite emerges above the soil surface, this reduction in host biomass is reflected in biomass increase in the parasite (Gurney et al., 1999). Other attributes to the severity of the Striga issue include high fecundity and long seed dormancy (Parker and Riches, 1993) in the soil seed bank; this can lead to many farmers to abandon fields once infested (Butler, 1994; Nambafu et al., 2014). These farmers are commonly growing for subsistence and lack significant resources to implement any effective control measures (Mohamed et al., 2001; Atera et al., 2012; Yohannes et al., 2015).

The three most devastating cereal-infecting spp. of *Striga* are, *Striga hermonthica, Striga asiatica* and to a

lesser extent *Striga aspera* (Runo and Kuria, 2018). Monocot hosts of these include maize, millet, sorghum and upland rice with recent estimates suggesting that up to 50 million hectares of land in SSA are infested with Striga spp. (Parker and Riches, 1993; Spallek et al., 2013). Each species has a somewhat different geographic range but there is a high level of overlap and all species are widespread across Africa. West African countries are normally infested by *S. aspera* and *S. hermonthica* while *S. asiatica* is more concentrated across East Africa and in Madagascar (Rodenburg et al., 2016). Another Striga species, *Striga gesnerioides*, causes similar devastating effects on agriculture through its infection of dicot crops such as cowpea (Rodenburg et al., 2010; Spallek et al., 2013; Fernandez-Aparicio et al., 2020; Albert et al., 2021).

Madagascar is a region of high priority for parasitic weed research (Rodenburg et al., 2016), with agriculture as a backbone of the national economy, accounting for approximately 30% of GDP (IFAD, 2021). The Striga species, *S. asiatica*, the focus of this thesis, is the only extensively observed Striga spp. in Madagascar. In contrast to many other *S. asiatica* infecting countries such as Tanzania and Ethiopia, Madagascar is unique in that rice is the principal crop grown by farmers (FAO and WFP, 2019). Recent assessments have suggested that the total rice-growing area infected by *Striga* spp. in SSA is approximately 887,000 ha with annual yield losses of 293,000 tons of milled rice (Rodenburg et al., 2016).

2.1.2 Striga virulence and host immunity

Gaining a deeper understanding of the mechanisms underlying Striga virulence is crucial for effectively addressing the Striga emergency and enhancing resistance in crop species.

While some Striga spp. are particularly virulent parasites on crop plants, the host's immune defence plays a vital role in determining the outcome of the interaction between host and parasite. A reciprocal adaptation phenomenon, often referred to as an evolutionary 'arms race' develops over many generations where the invading organism overcomes the host's defences through advantageous evolutionary changes but at the same time the host develops counter defences. This has been identified to exist for many parasites and their hosts including *Striga hermonthica* on sorghum, maize and pearl millet (Bellis et al., 2021), *Globodera pallida* on wild potato (Gautier et al., 2020) and *Cuscuta europea* L. on common nettle (Koskela et al, 2001). This can be observed indirectly where there are patterns of local adaptation evident (Schulte et al., 2011); In this scenario intra-regional genotypes would have a far higher fitness than exotic genotypes (Kawecki et al., 2004).

Incompatible reactions between Striga and its host can happen at different points in the infection process, both pre- and post-attachment. Pre-attachment defence mechanisms include low germination stimulant producing cultivars (see section 1.5) or the inability to form haustoria, a common occurrence for non-hosts as seen for *S. hermonthica* on Arabidopsis and cowpea (Yoshida & Shirasu, 2009). The induction of haustorial development is initiated by host exuded substances called Haustorial-Inducing Factors (HIFs). In *S. asiatica* one of these HIF constituents was initially identified as 2,6-DMBQ (Smith et al., 1996) but many other phenolic

and flavonoid substances have since been discovered as triggers for haustorial development (Riopel and Timko, 1995). Low production of haustorial producing factors may therefore be seen an advantageous genetic trait for potential hosts and one which can be bred into resistant crops (Mandumbu et al., 2019). However, this may not always be beneficial; many phenolic HIFs are precursors for lignin and in certain hosts phenolic HIFs can be incorporated into lignin polymers, to be used as a physical barrier deposited at Striga infection sites (Mutuku et al., 2019; Goyet et al., 2019).

A post-attachment resistance reaction which has been extensively studied within plant-pathogen systems is the Hypersensitive Response (HR), typically characterized by a necrotic lesion on the host tissue around the infection site, followed closely by a localised cell death (Balint-Kurti, 2019). This host response has been observed in sorghum infected by *S. asiatica* (Mohamed et al, 2003; Ejeta et al., 2007; Mohamed et al., 2010). Additionally, an HR response was observed on resistant cowpea cultivar B301 in response to *S. gesnerioides* infection, and ascertained to be mediated by an R gene, commonly found in other species subject to pathogen attack (Li and Timko, 2009). The presence of these R genes is highly indicative of the existence of opposing Striga effectors/virulence factors in *S. gesnerioides*, forming an evolutionary arms race.

Another related defence mechanism seen is the incompatibility response of resistant hosts where the host prevents further Striga tissue development, but does not itself suffer tissue necrosis. This has been reported in a range of hosts in a number of studies (Ejeta et al., 2007; Cissoko et al., 2011; Mutinda et al., 2018) where Striga showed stunted growth or development which did not reach further than primordial leaf, due to their inability to form a vascular connection with the host. In Orobanche species, callose deposition and lignification of cell walls are thought to form a barrier to entry of the parasite into the host xylem. (Fernandez-Aparicio et al., 2008).

2.1.3 Variation in Striga virulence and host adaptation

The outcome between many host-parasite interactions would usually be subject to the success of either the resistance genotype of the host or the virulence genotype of the parasite, but an interaction between the two opposing genotypes is also common (Hu et al, 2020). In this situation the so-called interaction between genotypes would mean the fitness or virulence of the parasite was subject to the host, and thus be associated with host specificity (Little et al., 2006). Nevertheless, this host specificity can break down when patterns of dispersal are disrupted (Okubamichael et al. 2016); for example if Striga seed was introduced into a new area with a new host, eventually selection would mean that over a number of generations a more virulent parasite genotype would be selected for. Generally the more inbreeding a species is, the less variation to exist naturally within populations and the longer it would take to establish a virulent population (Dick and Patterson, 2007). With parasites becoming host-specific they have become so morphologically, behaviourally, or physiologically

adapted to a particular host that being able to colonise new hosts successfully would simply be untenable (Tompkins and Clayton, 1999).

The common natural means of dispersal of *S. asiatica* seed is, within local areas, by wind, but is also subject to human-aided dispersal on a greater geographical scale through contaminated crop seeds, by farming machinery and by animals (Csurhes et al., 2016). For Striga species, variation in virulence has frequently been noted across different hosts, at both an inter- and intra-species levels. For example: *S. gesnerioides* on cowpea, where physiological groups were determined based on virulence profiles (Parker & Polniaszek, 1989) and *S. hermonthica* infecting 274 pearl millet accessions, where only 9 accessions were identified as resistant (Wilson et al., 2004). Parker and Polniaszek were able to show that in pot experiments of populations of *S. gesnerioides* across West Africa from Mali to Cameroon, the Striga accessions could be ranked on their virulence and showed distinct patterns of local adaptation. Early studies by Ramaiah and Werder (1987) also observed both *S. asiatica* and *S. hermonthica* parasites locally adapted to hosts. More recently, host-specific selection on parasitic genes has been identified in *S. hermonthica* populations from Western Kenya growing on maize and finger millet. Little evidence was found for variation in strigolactone perception but there was evidence of host selection on haustorial loci, particularly a locus encoding a chemocyanin precursor important for haustorial development (Bellis et al., 2022).

Previous research specifically on *S. asiatica* has shown variation in virulence among geographically separated populations. Botanga et al. (2002) ran cross-infestation experiments based in Benin that showed growth of Striga was only established when host and parasite had been collected from the same field location. This population-based differential virulence was supported by Spallek et al. (2013), who found that *S. asiatica* isolated from wild grasses were not able to parasitise common hosts such as sorghum and maize. Similarly, Mandumbu et al. (2016) identified differing virulence-associated effects of *S. asiatica* strains on sorghum hosts. However this has not been the same finding for all studies; Gethi et al. (2005) found little evidence for any significant population structure and suggested high gene flow between infested areas.

It is thought that *S. asiatica* entered the Island of Madagascar through multiple introductions at ports, likely from nearby African countries, with contaminated rice seed (Ejeta, 2007) and potentially originated from a variety of countries. Herbarium records suggest introductions into the Country began at the start of the 20th century around these coastline ports, and gradually spread further into the centre (Rodenburg et al., 2016; Scott et al., 2019). Being an island country has limited the introduction sites of Striga, the tiny seeds being unable to easily cross the ocean but also as crops are generally intra-regionally traded within Madagascar and a high level of subsistence farmers (Ralandison *et al*., 2018), there is likely less gene flow from outside of the country and to what extent Striga populations can be split within the country is undetermined.

One species of Striga, *S. gesnerioides*, endemic to West Africa and a parasite of dicots does exist within geographically separate populations, referred to as races. It was reported that the geographical isolation of these populations led to a host driven selection and formation of seven differentially infecting races of *S. gesnerioides* (Botanga and Timko 2006; Li & Timko. 2009). As another preferentially inbreeding species, this may also be the case for *S. asiatica,* but is still to be determined.

Previous studies using molecular marker analyses and field or laboratory screening have provided an unclear understanding of the connection between genetic and geographic distance, leaving the extent of virulence variation over larger geographic distances unknown. As seen with the different races of *S. gesnerioides*, geographic separation plays a key role in genetic differentiation, particularly in the case of inbreeding populations where gene flow is already limited and new virulence genes are infrequently occurring.

2.1.4 Aims and Objectives

This chapter aimed to investigate variation in virulence of *S. asiatica* accessions collected in different regions of Madagascar and East Africa against different rice genotypes. Specific objectives include:

I. Identify how the outcome of interactions between different rice hosts and Striga accessions vary.
II. Determine whether host variety or *S. asiatica* accession is the most important factor in explaining the variation in virulence.
III. Establish if there is a correlation between phenotypic variation in virulence of *S. asiatica* accessions on rice hosts compared with the geographic distance between the collection sites of the *S. asiatica* accessions.

2.2 Materials and Methods

2.2.1 Plant materials

The *S. asiatica* seed accessions used in this study were collected from 9 separate sites in Madagascar, 1 site in both Kyela and Tanzania, 2 sites in Ethiopia and 1 site in North Carolina, USA (Table 2.1). Each collection site was used for subsistence cultivation of cereal crops, including upland rice varieties.

Table 2.1 – *Striga asiatica* sampling site locations where seeds of the different accessions were collected.

Code	Site No	Commune Location	Host Crop collected from	Field coordinates (lat/lon)
Amb1	1	Ambalamiadana, Madagscar	Rice var. B22	-18.4349, 46.0118
Amb2	2	Ambalamiadana, Madagascar	Rice var. FOFIFA 3737	-18.4350, 46.0117
Anz	4	Anteza, Madagascar	Rice var. NERICA4	-18.9770, 46.4774
And	5	Androvasoa, Madagascar	Rice var. Chromrongdhan	-19.0246, 46.4377
Bel	8	Belanitra, Madagascar	Rice var. NERICA4	-19.7198, 46.6125
Eth1	9	Ethiopia (Collection1)	Unknown	Unknown
Eth2	10	Ethiopia (Collection2)	Sorghum	Unknown
Tan	11	Kyela, Tanzania	Rice	-9.35, 33.48
USA	12	North Carolina, USA	Maize	34.3834,-78.9460

Madagascan *S. asiatica* samples were collected directly from a number of field locations (sites) along two 100 km transects in the central region of the country, roughly following the course of the RN34 and RN1b roads (Figure 2.1). *S. asiatica* accessions were also collected from USA (North Carolina), Tanzania (Kyela district) and Ethiopia. All accession seed batches consisted of seeds from multiple *S. asiatica* plants collected from different host individuals within a specific field at the different sites. Maps of all sampling locations are shown in Figure 2.1.

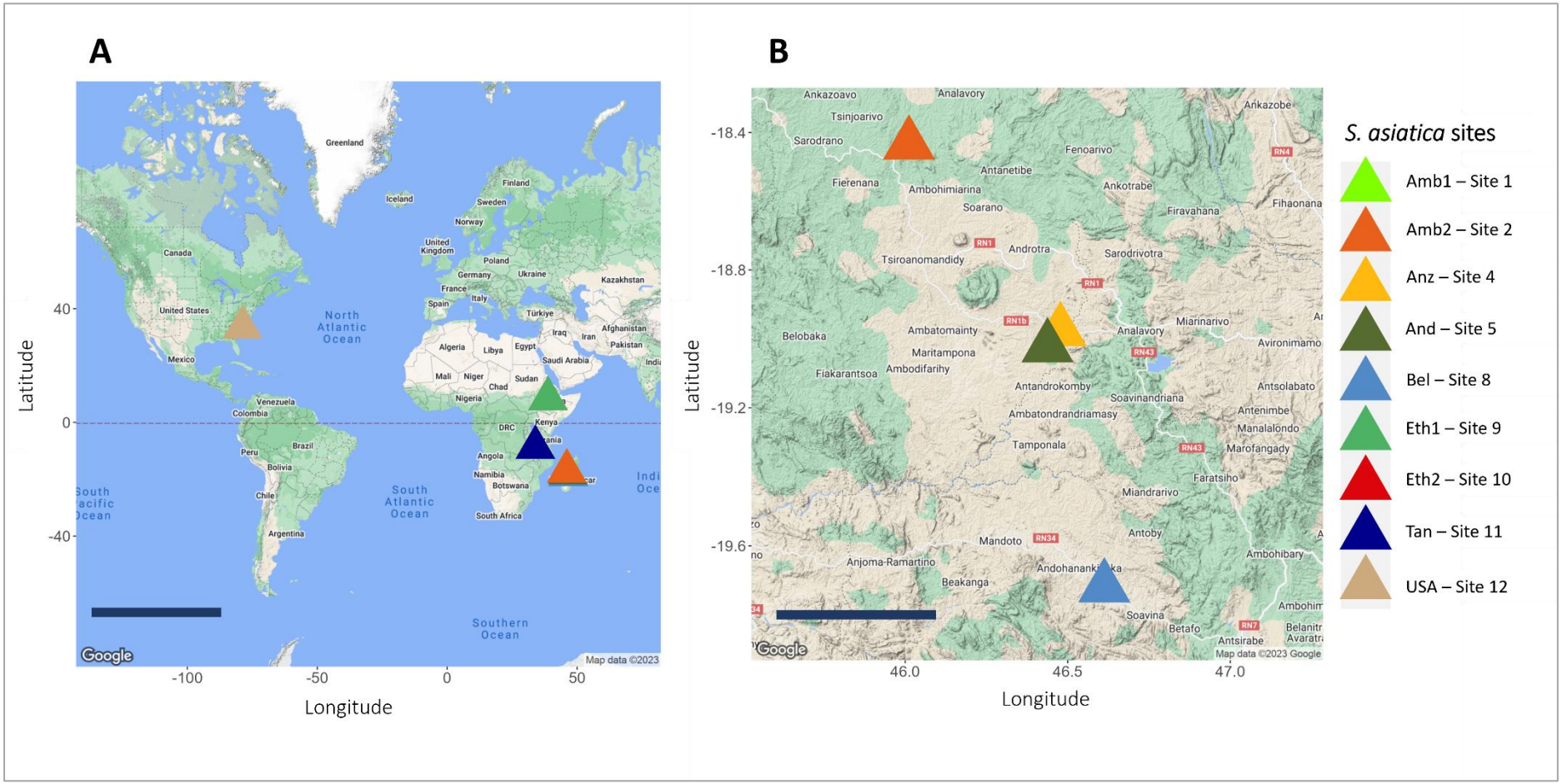


Figure 2.1 – Maps of the *S. asiatica* accession sampling sites. Map A shows the location of *Striga asiatica* sampling sites used to screen for virulence. From North Carolina, USA to Madagascar. The scale bar shows 4200 km. Map B shows *S. asiatica* sampling sites within Madagascar. The scale bar shows 50 km.

Sampling locations were selected across a widely spaced range of sites in order to test for patterns of local adaptation to host genotypes and whether variation in virulence could be seen at both smaller and larger spatial scales. Striga accessions from sites 1 and 2 were located in adjacent fields metres away from each other, whereas the distance from site 1 to site 9 was over 3000 km.

Striga seed collection in Madagascar began once *S. asiatica* seed pods had matured and dried in the field. Plants were harvested from multiple infected hosts within the field (site) being sampled. Thus each *S. asiatica* seed accession represents the genetic diversity within a field (site). Following harvest, seeds were dried and sent to the UK. Upon arrival in the UK, seed pods were further dried in an incubator at 50 °C for 1 month. After this, pods were picked from the stems and split open to release seeds. Seeds were then filtered through a 200 micron sieve, followed by a 150 micron sieve to remove plant debris. Seeds were stored at 4 °C in a sealed glass container with silica gel beads to absorb moisture and extend the storage life of the seeds.

Striga seeds from Ethiopia were collected in the 1980's from sorghum. Precise locations were not recorded. The two separate accessions (Eth1 and Eth2) may have originally arisen from different field locations, however it is unclear for how many generations the two batches have been successively bulked in the controlled environment chambers at the University of Sheffield and were expected to have lower genetic diversity among them. Seeds from Tanzania and USA were from original batches of field collected seed, the USA accession seeds found growing on maize and Tanzanian accessions found growing on rice. The location of sampling of the Tanzanian accession was from a rice field site in the Kyela district and the USA accession collected from near the border between North and South Carolina.

Nine rice varieties were used in this study (Table 2.2). This included 5 cultivars grown in Madagascar which have been readily adopted by farmers: B22 and Chromrongdhan (*O. sativ*a spp. *indica*), Fahita (*O. sativa*), NERICA-4 and NERICA-10 (inter specific crosses of *O. sativa* X *O. glaberrima*). Seeds were supplied by Dr Patrice Autfray (CIRAD, Montpellier, France) and Mr Jean-Augustine Randriamampianina (FOFIFA, Madagascar). The four other rice varieties used in this study were: Nipponbare and Koshihikari (*O. sativa* spp. temperate *japonica*) obtained from Professor Masahiro Yano at the National Agriculture and Food Research Organization (NARO), Japan and IR64 (*O. sativa* spp. *indica*) and Azucena (*O. sativa* spp. tropical *japonica*) which were provided by Dr Mathias Lorieux at the International Center for Tropical Agriculture (CIAT) Cali, Colombia. These rice varieties were chosen to study as their susceptibility/resistance to Striga species has been extensively studied previously (Gurney et al., 2006; Kaewchumnong & Price., 2008; Yoshida et al., 2009; Cissoko et al., 2011; Huang et al., 2012; Mbuvi et al., 2017; Mutuku et al., 2019; Dayou et al., 2021; Qui et al., 2022).

Table 2.2 – Rice varieties used as hosts in the *Striga asiatica* virulence phenotyping screen. Susceptibility to Striga species is reported.

Variety Name	Species/subspecies	Reported susceptibility
B22	*O.sativa* ssp. *indica*	Susceptible (Randrianjafizanaka et al., 2017)
Chromrongdhan	*O.sativa* ssp. *Indica*	Unknown
Fahita	*O.sativa*	Unknown
NERICA4	*O.sativa* ssp. *japonica* X *O. glaberrima* (CG14)	Resistant (Cissoko et al., 2011)
NERICA10	*O.sativa* ssp. *japonica* X *O. glaberrima* (CG14)	Resistant (Cissoko et al., 2011)
Nipponbare	*O. sativa* ssp. temperate *japonica*	Resistant (Gurney et al., 2005)
Koshihikari	*O. sativa* ssp. temperate *japonica*	Susceptible (Mutuku et al., 2019)
IR64	*O. sativa* ssp. *indica*	Resistant (Kaewchumnong et al., 2008)
Azucena	*O. sativa* ssp. tropical *japonica*	Susceptible (Kaewchumnong et al., 2008)

2.2.2 Growth of rice plants and infection with *S. asiatica*

In order to determine the extent of virulence of *S. asiatica* populations, a panel of rice varieties were infected with *S. asiatica* seeds. Rice seeds were germinated between strips of filter paper, sandwiched between horticultural rockwool blocks (Aquaculture, Sheffield, UK) and incubated at 27 °C in darkness for 7 days. Following the incubation period, a single rice seedling was transferred into a rhizotron, a 30 cm^2 X 3 cm deep, root observation chamber as described in Gurney et al. (2006) and Cissoko et al., (2011).

Each rhizotron contained horticultural rockwool, covered with 100 µm mesh (Plastic Group, Birkenhead, UK). Germinated rice seedlings were placed on top of the mesh. A lid was placed on the rhizotron which was covered in foil to prevent light penetration to the root system. Plants were grown in a controlled environment chamber with a photon flux density of 500 µmol m^{-2} s^{-1} at plant height, a 12 h photoperiod, relative humidity of 60%, and a day/night temperature of 27 °C /25 °C. Rice plants were supplied with 20 ml of 40% Long Ashton solution containing 2 mM ammonium nitrate 4 times per day, via a hydroponic system (Figure 2.2A).

Twelve days prior to infection, 8 mg aliquots of *S. asiatica* seeds (for each rice plant) were surface sterilised in a 10% bleach solution for 7 min and washed repeatedly with dH_2O. Sterilised seeds were then placed in 9 cm Petri dishes containing moistened glass fibre filter paper (Whatman, Cytiva, Amersham, UK) and sealed with parafilm before plates were incubated at 30 °C in darkness for 12 days.

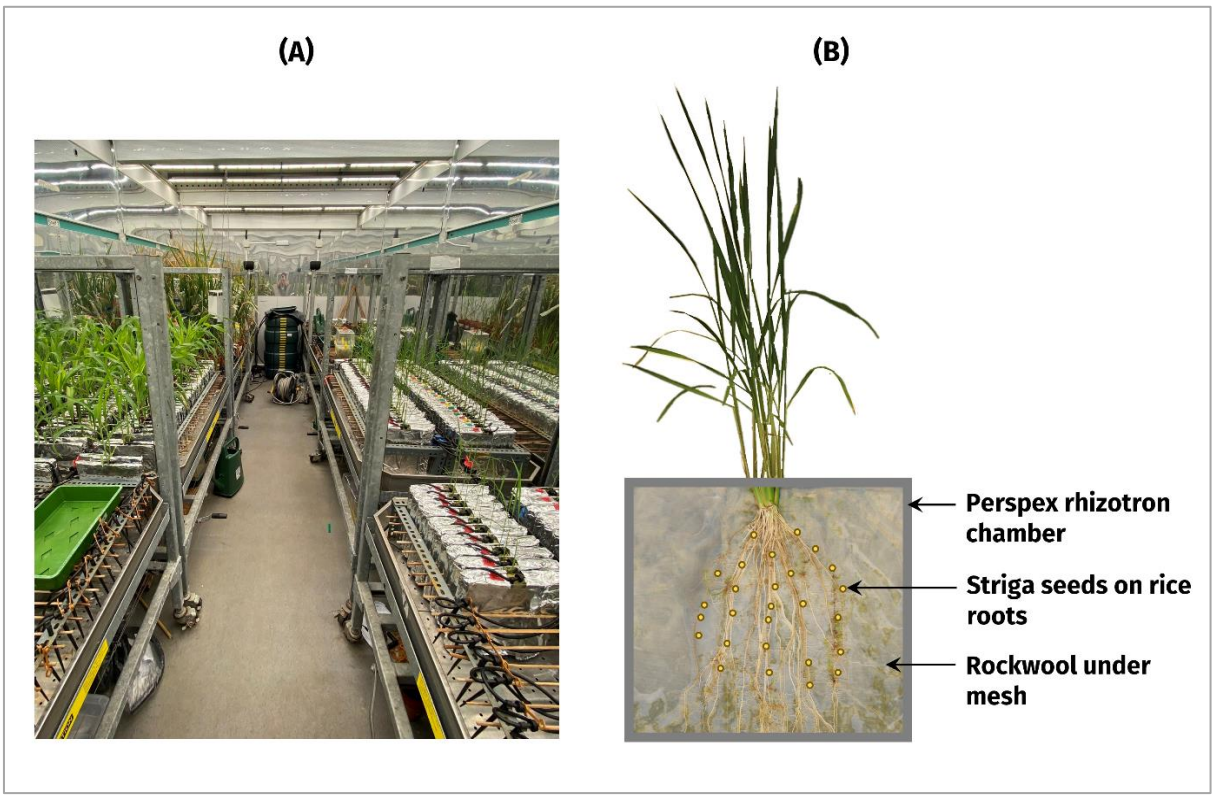


Figure 2.2 – *S. asiatica* virulence assays were performed by infecting rice plants growing in rhizotrons. (A) shows the chamber where the plants were cultivated. (B) shows the rice root systems which were painted with the pre-germinated Striga seeds before sealing inside the rhizotron for attachments to happen.

Germination of *S. asiatica* seeds within the Petri dishes was triggered through addition of an artificial germination stimulant (GR-24) at a concentration of 0.1 mg L^{-1} and left to incubate for between 15 to 18 h, depending on the germination time of the particular *S. asiatica* accession. Individual rice root systems were infected with ~8 mg of pre-germinated *S. asiatica* seeds using a paint brush (Cissoko et al., 2011)(figure 2.2B). Each batch of infections of rice plants was completed within 4 hours to ensure that the *S. asiatica* radicles did not become so long that they were no longer able to infect the rice roots. Infected rice plants were grown for 28 days within the rhizotrons before quantifying *S. asiatica* virulence/rice resistance.

On day 28, after infection of rice plants, a high-resolution (1200 dpi) scan of each root system was taken using a flatbed scanner (CanoScan 5400, Canon, Uxbridge, UK). The Striga attachments were harvested from each root system, place in a Petri dish and photographed using a Canon EOS 500D (Canon, Uxbridge, UK). *S. asiatica* plants were then dried in a 50 °C incubator for 3 days and before weighing *S. asiatica* biomass. Image J v. 146r software (Schneider et al., 2012) was used to measure the number and length of the S. *asiatica* plants.

2.2.3 Measures of *S. asiatica* virulence

The three measures of virulence used in this study were (1) dry *S. asiatica* biomass per root system, (2) total number of *S. asiatica* attachments per root system and (3) cumulative length of *S. asiatica* attachments per root system. These measures have been used in a number of studies previously (e.g. Gurney et al., 2007; Cissoko et al., 2011). By utilising three measures of virulence data can be used to interpret different nuances in infection patterns. After 28 days of infection there will be variation in in the size of *S. asiatica* individuals that have attached to host root systems in terms of width and length. For example, the same number of *S. asiatica* individuals may be present in both resistant and susceptible rice hosts but on susceptible hosts *S. asiatica* are likely to be larger and have a greater biomass. In contrast, in a resistant host, *S. asiatica* individuals may attach and begin to infect but are then prevented from growing and thus have a low biomass.

2.2.4 Determining variation in virulence of nine *S. asiatica* accessions against nine rice varieties

The purpose of this study was to determine the variation in virulence of *S. asiatica* accessions from Madagascar (sites 1, 2, 4, 5 and 8) and from Mainland Africa (sites 9-12) on 9 different rice varieties. To achieve this, virulence assays were carried out by infecting 9 rice varieties with the 9 Striga accessions, a total of 81 different combinations.

These virulence assays were carried out in three separate screens due to the large number of plants. Each screen contained a different combination of rice varieties and *S. asiatica* accessions, but with at least one of the same combinations, B22 x Eth1 (Site 9), being replicated in each of the 3 screens to act as infection controls. Each screen was further divided into 3 or 4 batches which comprised a maximum of 48 rhizotrons per batch. The germination of each batch of rice seeds was staggered by 7 days.

Across the whole experiment, for each combination of *S. asiatica* accession and rice variety, at least 6 biological replicates were established, two per batch. Infection of the plants with *S. asiatica* was performed by several people (infectors) as all the plants needed to be infected within 4 hours. The replicates for each combination were divided up among the infectors to ensure no one person infected all replicates within a combination.

2.2.5 Modelling virulence

A linear mixed effects model was fitted to the data during analysis to assess significance of virulence measures, taking into account the fixed effects of *S. asiatica* accession and rice variety. The random effects of **batch**, nested within **screen**, and fixed effect of **infector** were also taken into account in the model.

Using R version 4.1.2 (R Core Team, 2022) package ggplot2 (Hadley, 2016) boxplots were used to visualise the differences in virulence for each of the datasets. Classical multidimesional scaling (MDS) was also performed

on the *S. asiatica* weight and number datasets using the cmdscale function in R to represent the distances between the rice varieties and *S. asiatica* accessions, separately. Due to the presence of a high number of zeros (resistant plants), virulence datasets were significantly different from a normal distribution. Inspection of residuals from initial trial analyses with various transformations, including log and gamma, showed that parametric analysis was not possible. Therefore, the permmodels function of the predictmeans package (Luo et al., 2018) was used to provide permutation t-tests for coefficients of fixed effects and permutation F-tests to test the terms in the linear model. Dry biomass data were not transformed but square root was taken of total number data to try to stabilise variance and distribution. The p values of the permutation t-tests and F tests were obtained using 4999 permutations.

The central question for this study was to determine how rice variety, *S. asiatica* accession, and their interaction impacted the virulence of *S. asiatica*. To assess the variance of each virulence phenotype, using the R package lme4 (Bates et al, 2015), each dataset was fitted to a mixed effects linear model. The R package predictmeans (Welham, 2004) was used for comparing the best fit of different models, taking into account whether the random terms of "Screen", "Batch" and "Infector" were important in explaining variation and determining the simplest acceptable model through comparison of AIC values. The different models tested were:

Model 1 - Virulence phenotype ~ Rice variety + *S. asiatica* accession + infector + (1|Screen/Batch)

Model 2 - Virulence phenotype ~ Rice variety * *S. asiatica* accession + infector + (1|Screen/Batch)

Model 3 - Virulence phenotype ~ Rice variety * *S. asiatica* accession + (1|Screen/Batch)

Model 4 - Virulence phenotype ~ Rice variety * *S. asiatica* accession + infector + (1|Screen)

Model 5 - Virulence phenotype ~ Rice variety * *S. asiatica* accession + (1|Screen)

Boxplots comparing all virulence measurements by rice variety and Striga accession were plotted using R package ggplot2. To assess the impact of spatial separation of accessions on overall variation in virulence, a Euclidean distance matrix was produced in R for both coordinate data and median virulence per host for each Striga accession and plot using package ggplot2.

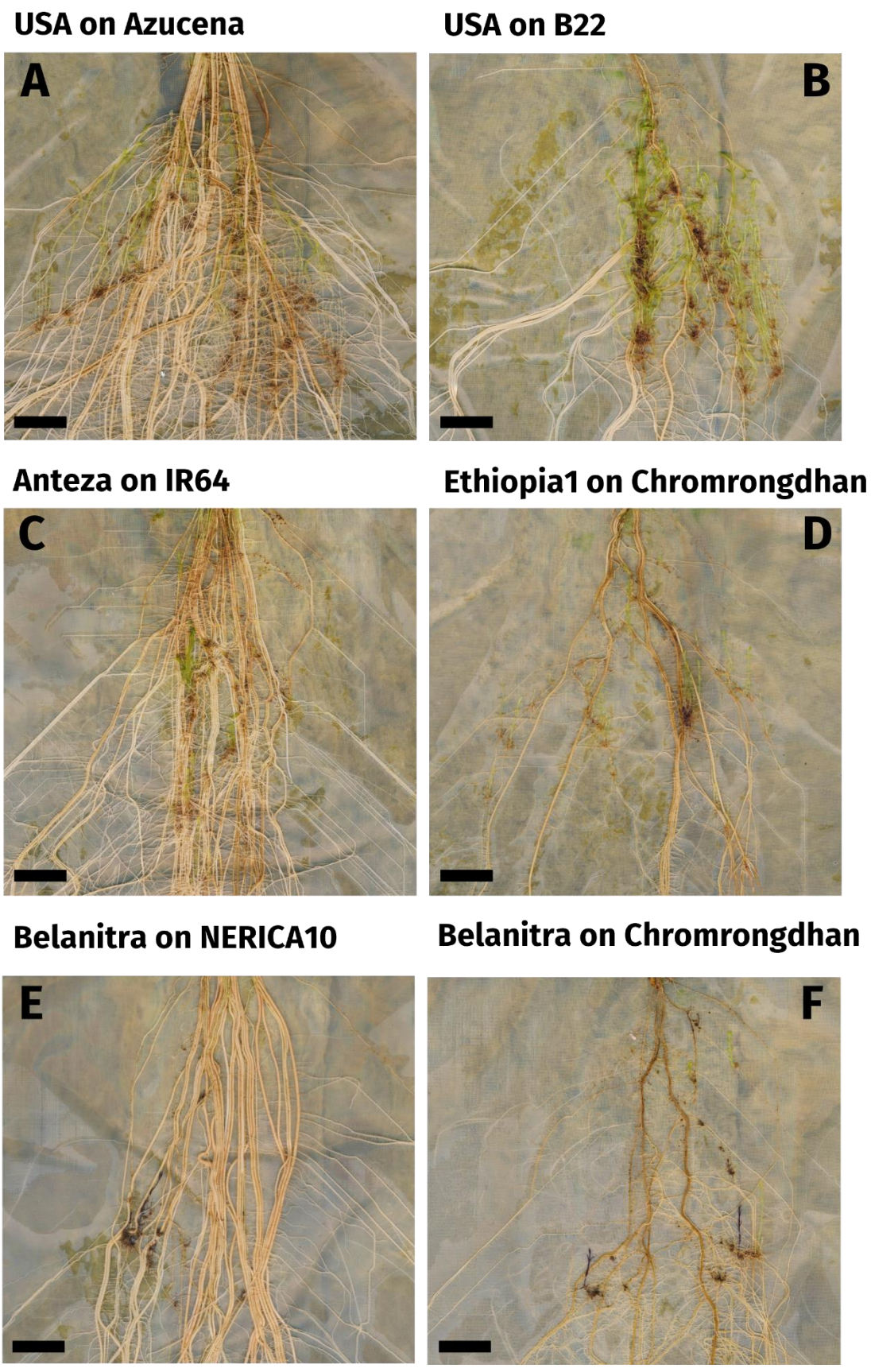
2.3 Results

2.3.1 How do the interactions vary between different rice hosts and *S. asiatica* accessions?

The classification of the interactions, whether resistant or susceptible was determined after 28 days, at which point the most successful (virulent) *S. asiatica* individuals would have grown to a height of several centimetres. Very small *S. asiatica* individuals, regardless, of whether host parasite xylem connection had been made, would not be compatible with further growth leading to reproductive success, thus indicating a resistant host- avirulent *S. asiatica* combination. In a resistant interaction host roots often exhibited dark brown areas, indicative of necrosis, around the sites of infection, while new roots remained white. This is particularly evident in Fig. 2.4 F, as *S. asiatica* (Bel) attachments were frequently surrounded by a dark necrotic ring of host tissue. The dark brown regions may indicate a hypersensitive response by the host, causing localised cell death.

A range of virulence/resistance phenotypes were seen across the 81 combinations of rice varieties and *S. asiatica* accessions in the virulence phenotype screening. These interactions could be broadly categorised into 4 groups as shown in Fig. 2.4. Fig. 2.4 A & B illustrate a susceptible interaction between the *S. asiatica* USA accession grown on rice host Azucena (A) and B22 (B), with many large *S. asiatica* plants visible. Fig. 2.4 C & D show a mostly resistant interaction of *S. asiatica* (Anz) on IR64 (C) and *S. asiatica* Eth1 on Chromrongdhan (D), with just a few small *S. asiatica* indivdiuals growing on the host roots. Fig. 2.4 E & F show host root systems with a small number of blackened, dead *S. asiatica* individuals indicative of a resistant interaction, commonly seen for *S. asiatica* (Bel) on Chromrongdhan (E) or NERICA-10 (F). Finally, Fig. 2.4 G & H illustrate highly resistant interactions with *S. asiatica* (Amb) (G) and (Anz) (H) on rice variety NERICA-4.

While the majority of *S. asiatica* accessions showed variable levels of attachments across the panel of rice varieties, one *S. asiatica* accession – *S. asiatica* (Bel) showed very low levels of attachments to any rice host. In addition, attachments that were made tended to blacken and die before reaching more than ~ 2 cm in height. This was despite the average percentage germination of the S. asiatica (Bel) accession being 65.2% across the screens, well within the range of percentage germination of the other Madagascan S. asiatica accessions (52.4-67.25%), none of which showed this avirulence phenotype. This phenotype was also found when *S. asiatica* (Bel) accession was grown in pots, as shown in Figure 2.4. Again some *S. asiatica* attachments grew on the host, albeit to a lesser extent compared to other *S. asiatica* accessions, but they then died long before reaching reproductive maturity. Figure 2.4 shows *S. asiatica* (Bel) within the pots at the different stages of growth, from (1) relatively healthy with some chlorosis, (2) to wilting and then (3) to finally blackening and dying before reaching maturity.


USA on Azucena
A
USA on B22
B
Anteza on IR64
C
Ethiopia1 on Chromrongdhan
D
Belanitra on NERICA10
E
Belanitra on Chromrongdhan
F

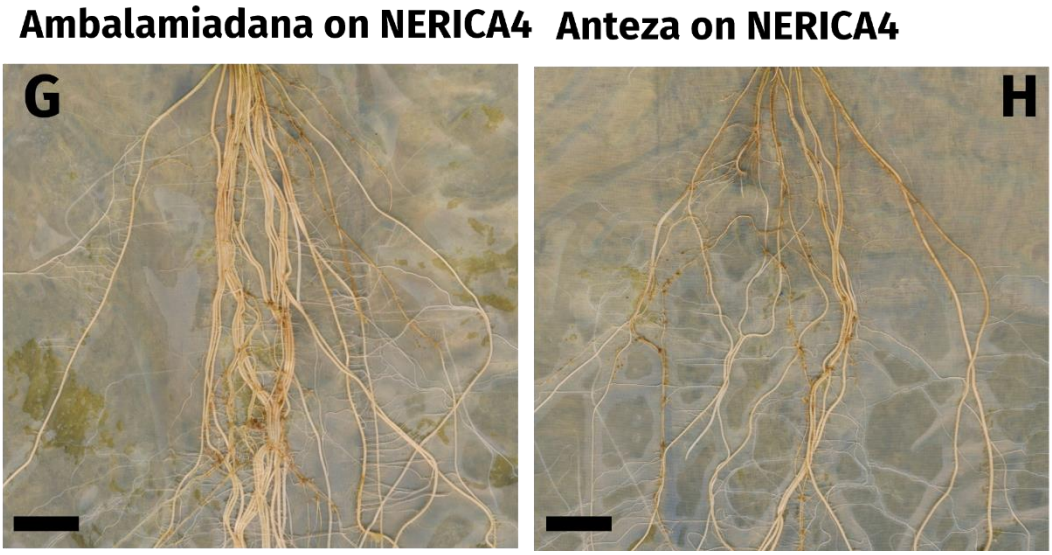


Figure 2.3 - Root scans of infected rice root systems taken 28 days after infection with *S. asiatica* seeds. A & B illustrate highly susceptible interactions with many large *S. asiatica* present, **C & D** are mostly resistant interactions with few smaller *S. asiatica* present, **E & F** are resistant interactions with few *S. asiatica* present and showing a blackened *S. asiatica* phenotype. G & H are very resistant interactions with zero successful *S. asiatica* attachments. A shows *S. asiatica* (USA) on Azucena, B shows *S. asiatica* (USA) on B22, C shows *S. asiatica* (Anteza) on IR64, D shows *S. asiatica* (Ethiopia1) on Chromrongdhan, E shows *S. asiatica* (Belanitra) on NERICA10, F shows *S. asiatica* (Belanitra) on Chromrongdhan. Scale bar represents 2 cm. G shows *S. asiatica* (Ambalamiadana) on NERICA4 and H shows *S. asiatica* (Anteza) on NERICA4. Scale bar represents 2 cm.

Figure 2.4 – *Striga asiatica* accession (Bel) growing on rice variety B22. The rice host, signified by the green arrow, surrounded by emerged *S. asiatica* plants at various stages of development from healthy (pink arrow), wilting (yellow arrow) to dead (blue arrow). This phenotype of *S. asiatica* accession from Belanitra (Bel) was seen across all rice varieties tested and in both pot and rhizotron experiments.

2.3.2 Is host variety or Striga accession the most important factor underpinning variation in Striga virulence?

Quantitative data for *S. asiatica* virulence were collated from the measurement of *S. asiatica* seedlings harvested from individual hosts, with 6 replicates per combination of different rice host varieties and *S. asiatica* accessions across all screens. Three screens were performed overall which were further subdivided into batches to facilitate effective experimental analyses. Substantial variation in virulence phenotype, exhibited by a wide distribution of *S. asiatica* growth was seen across the hosts (Fig. 2.5 – Fig. 2.7). Data were fitted to a mixed effects linear model to test which factors contributed to the observed variation in virulence. A number of different models were tested to see which of the terms were significant and model 5 (Virulence phenotype ~ Rice variety * Striga accession + (1|Screen)) was determined as the best fit for all virulence

datasets. The random effects of infector and the fixed effect of batch were not found to significantly affect the variation in virulence, while screen was found to be important.

For all three measures of virulence, virulence was significantly affected by rice variety, *S. asiatica* accession and the interaction between the two terms (Supplementary Tables 2.1, 2.3 and 2.5). The effect of rice variety did explain most of the variance for *S. asiatica* dry biomass ($F = 53.93$, $p<0.0002$, $df = 8, 511$) and cumulative length ($F=73.46$, $p<0.0002$, $df = 8, 511$). In comparison, the variance explained by *S. asiatica* accession was less than half that of rice variety, looking at both *S. asiatica* dry biomass ($F = 17.77$, $p<0.0002$, $df = 8, 511$) and cumulative length ($F=37.15$, $p<0.0002$, $df = 8, 511$). However, for the variance measure, total number of *S. asiatica*, *S. asiatica* accession ($F = 82.49$, $p<0.0002$, $df=8, 511$) explained more variance than rice variety ($F=70.17$, $p<0.0002$, $df = 8, 511$).

Fig. 2.5 illustrates the effect of the interaction between *S. asiatica* accession and rice variety on *S. asiatica* dry biomass, with *S. asiatica* (Amb1) far more virulent on rice varieties Azucena or B22 than on NERICA-4 or NERICA-10. A similar picture was seen for *S. asiatica* cumulative length with *S. asiatica* (USA) being highly virulent on Azucena compared with very low virulence on IR64 (Fig. 2.6). Overall the host varieties most commonly grown in Madagascar, NERICA-4, NERICA-10, Fahita and to a lesser extent, Chromrongdhan were particularly resistant to all *S. asiatica* accessions, not only those from Madagascar but also those from Ethiopia, Tanzania and the USA. Azucena, Koshihikari and B22 were the most susceptible rice varieties to the *S. asiatica* accessions.

The most striking difference in virulence of *S. asiatica* accessions was *S. asiatica* (Bel), which consistently showed low levels of *S. asiatica* biomass, total number and cumulative length per host (Supplementary Figures 2.7-2.9). To determine whether the variation in virulence among *S. asiatica* accessions was influenced by the differences between other accessions, rather than just the avirulent accession (*S. asiatica* (Bel)), datasets excluding this accession were fitted to a mixed effects linear model (supplementary Tables 2.2, 2.4 and 2.6). While the effect of rice variety remained relatively consistent, the effect of *S. asiatica* accession on *S. asiatica* virulence was greatly reduced for each virulence measure: *S. asiatica* dry biomass ($F = 4.47$, $p<0.0002$, $df = 7, 436$), cumulative length of *S. asiatica* ($F=19.57$, $p<0.0002$, $df = 7, 436$) and total number of *S. asiatica* ($F = 39.25$, $p<0.0002$, $df = 7, 436$). Despite this, effects of *S. asiatica* accession, rice variety and an interaction between both, remained significant even after the Bel accession was removed from the datasets.

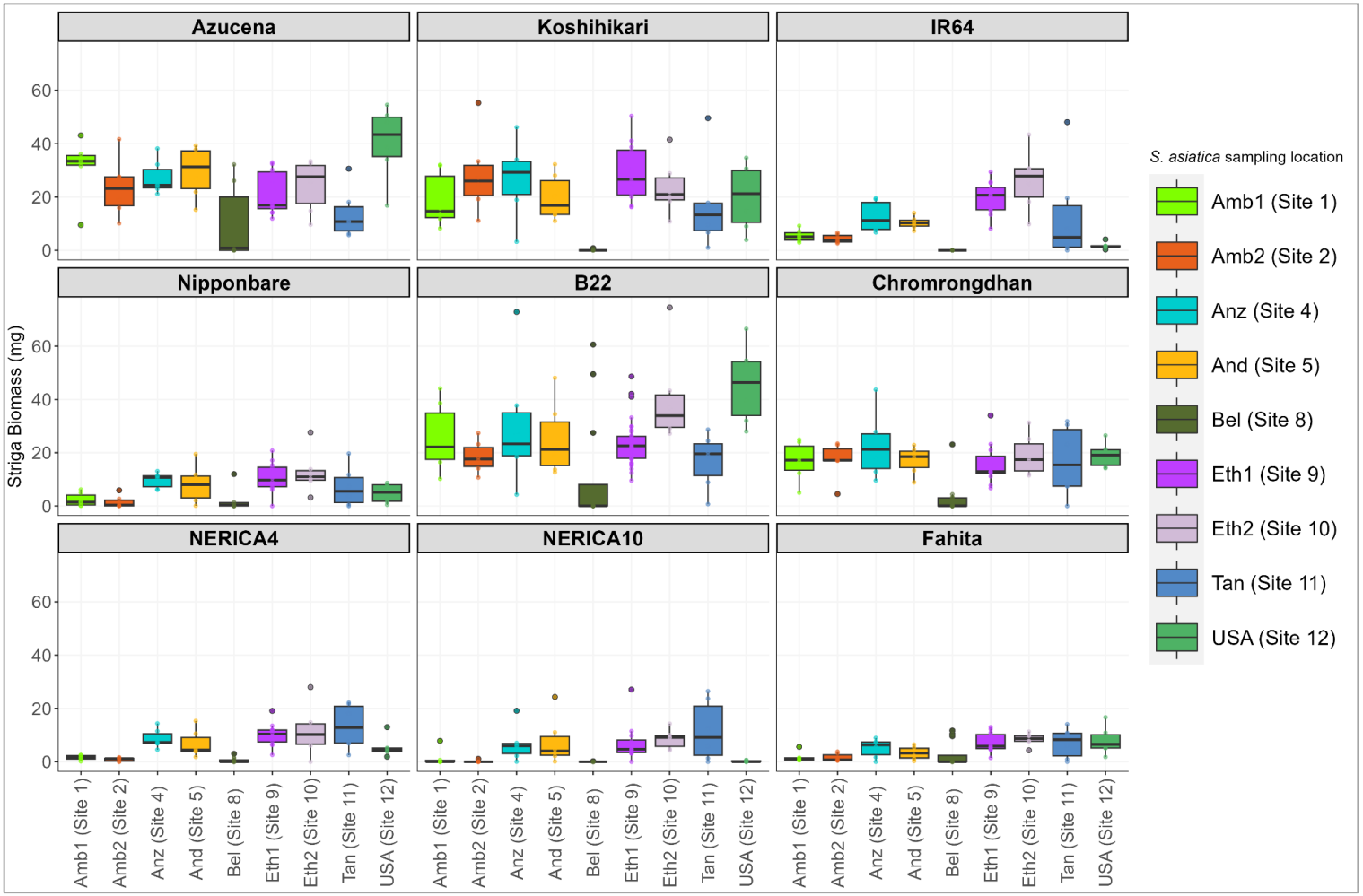


Figure 2.5 - Comparison of *S. asiatica* virulence profiles on a panel of rice varieties using *S. asiatica* dry biomass collected from the host root systems as a measure of virulence. *S. asiatica* sampling sites are given in the figure legend. Sites 1, 2, 4, 5 and 8 are from Madagascar, sites 9 and 10 are from Ethiopia and site 11 and site 12 are from Tanzania and USA, respectively. The plot shows a wide variation in *S. asiatica* virulence by host variety.

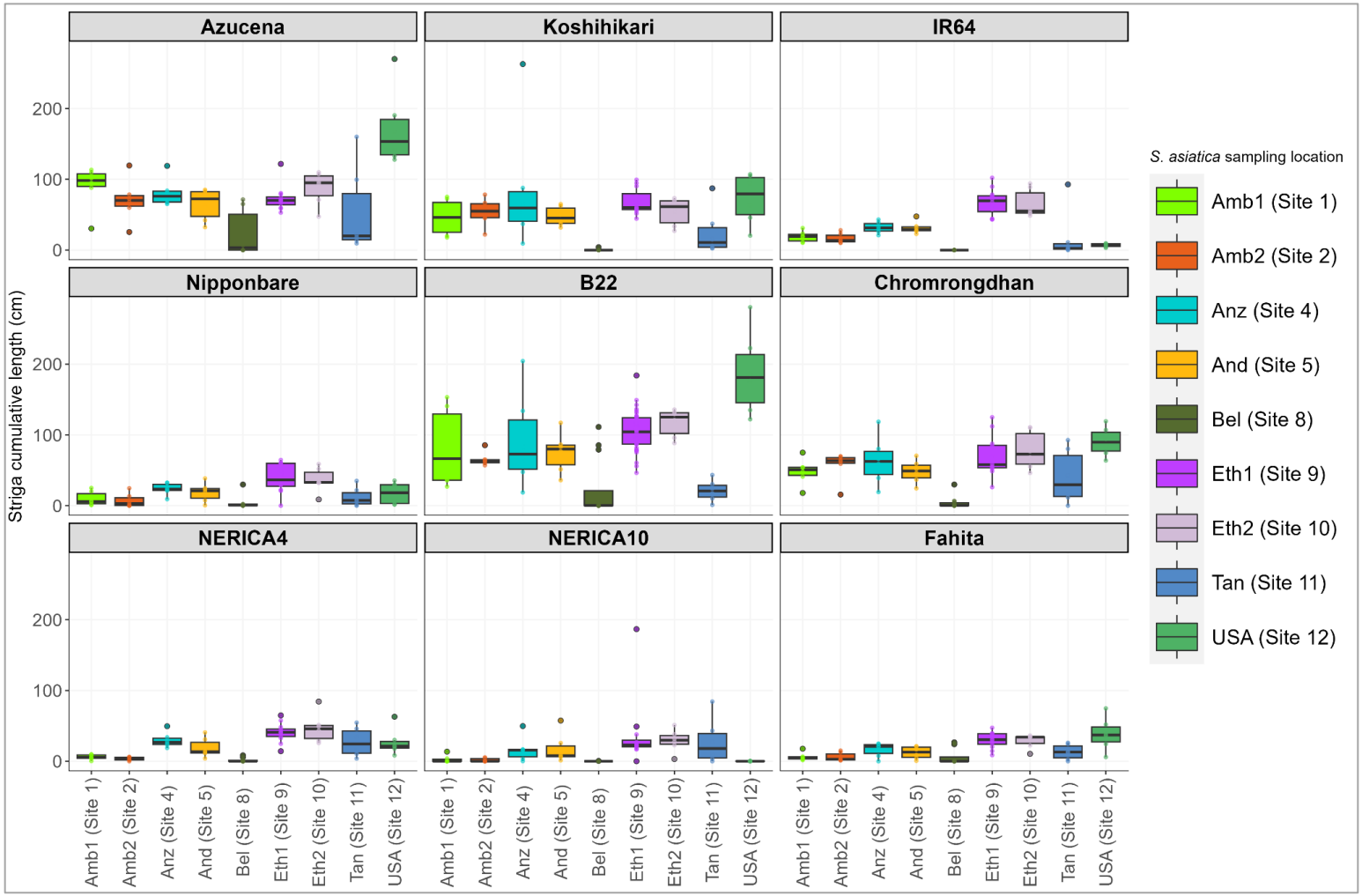


Figure 2.6 - Comparison of *S. asiatica* virulence profiles on a panel of rice varieties using cumulative length of *S. asiatica* collected from the host root systems as a measure of virulence. *S. asiatica* sampling sites are given in the figure legend. Sites 1, 2, 4, 5 and 8 are from Madagascar, sites 9 and 10 are from Ethiopia and site 11 and site 12 are from Tanzania and USA, respectively. The plot shows a wide variation in *S. asiatica* virulence by host variety.

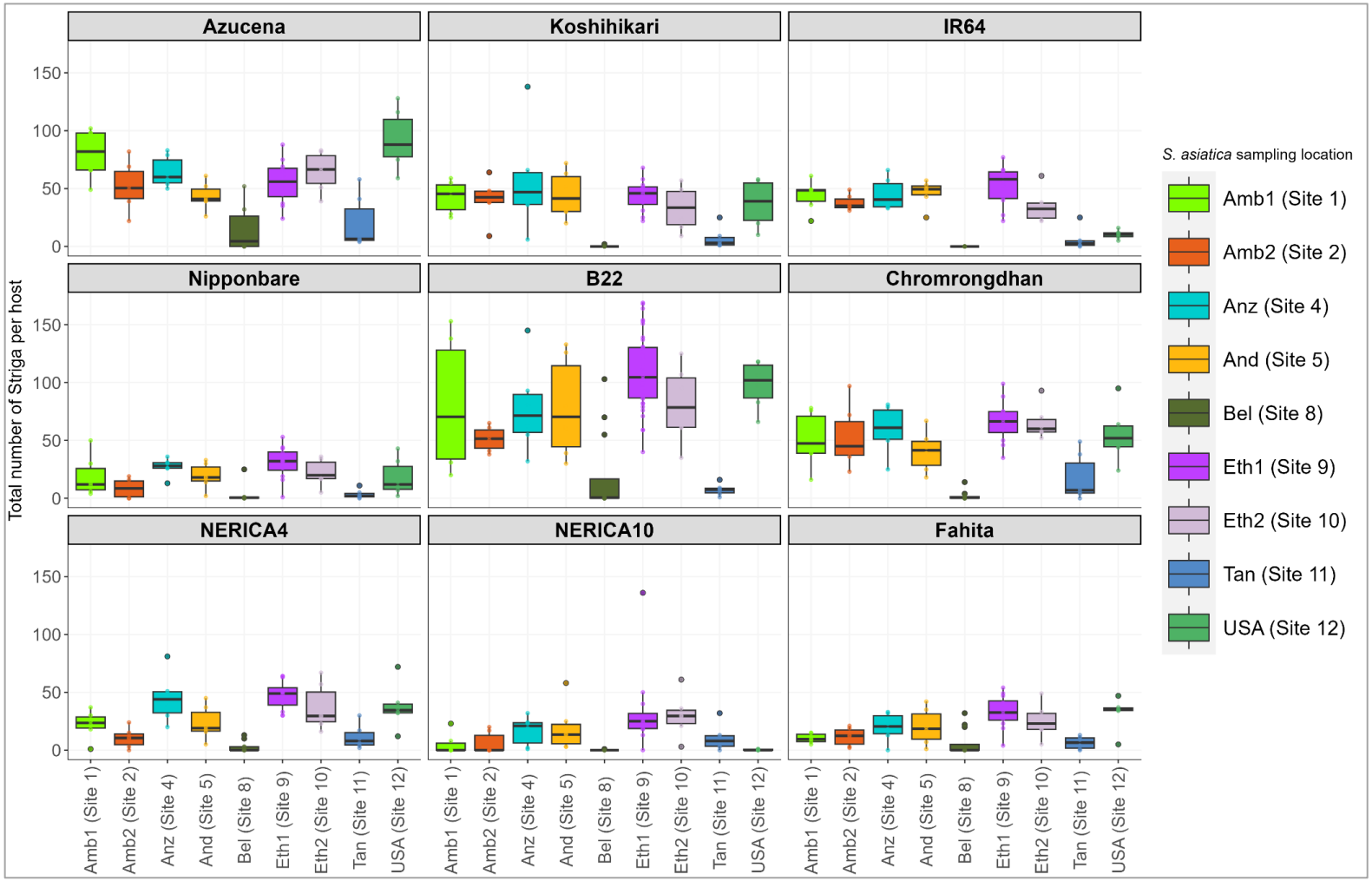


Figure 2.7 - Comparison of *S. asiatica* virulence profiles on a panel of rice varieties using total number of *S. asiatica* collected from the host root systems as a measure of virulence. *S. asiatica* sampling sites are given in the figure legend. Sites 1, 2, 4, 5 and 8 are from Madagascar, sites 9 and 10 are from Ethiopia and site 11 and site 12 are from Tanzania and USA, respectively. The plot shows a wide variation in *S. asiatica* virulence by host variety

2.3.3 What correlations can be seen in overall patterns of virulence?

The *S. asiatica* accessions that were investigated were from Madagascar, USA, Tanzania and Ethiopia. In order to identify patterns of virulence amongst the *S. asiatica* accessions with respect to host variety, multidimensional scaling (MDS) was carried out. The overall virulence of *S. asiatica* accessions across the panel of hosts differed depending on virulence measurement *(S. asiatica* dry biomass, total number of *S. asiatica* or cumulative length of *S. asiatica*, per host root system) (Figure 2.8), as did the overall susceptibility of the host varieties (Figure 2.9), though the pattern of host susceptibility to *S. asiatica* was more consistent between virulence measures which reflects the ANOVA results for the larger effect of host variety on *S. asiatica* virulence (Supplementary Tables 2.1 - 2.6). For these rice hosts, multidimensional scaling (MDS) clustered the rice varieties into groups, loosely based on susceptibility – the largely susceptible B22 and Azucena consistently clustering separately from the highly resistant NERICA-4 and NERICA-10 etc. and the intermediately susceptible Chromrongdhan and Koshihikari (Fig. 2.9). This was relatively consistent across all virulence measures.

A slightly different pattern was seen for *S. asiatica* virulence across the *S. asiatica* accessions depending on the measure of virulence. From observing the MDS plots (Fig. 2.8), the *S. asiatica* (USA) accession reliably clustered alone across all virulence measures showing a distinctly individual virulence profile across the hosts. This unique virulence profile can be seen in Supplementary Fig. 2.1, 2.2 and 2.3 as the USA accession showed a greater than average virulence on hosts such as Azucena and B22. However, this increased virulence was not universal across rice hosts, having lower than average virulence on rice variety IR64.

According to general patterns of virulence seen in the MDS plots (Fig. 2.8), the majority of *S. asiatica* accessions clustered relatively closely, though *S. asiatica* (Bel) clustered with *S. asiatica* (Tan) by cumulative length and number of *S. asiatica*, both appearing mostly avirulent when looking at these measures (Figs. 2.8B and 2.8C). However, when comparing biomass measures (Fig. 2.8A), the virulence of these Striga were very different: *S. asiatica* (Bel) producing a number of small non-viable Striga and *S. asiatica* (Tan) accessions producing few large viable attachments.

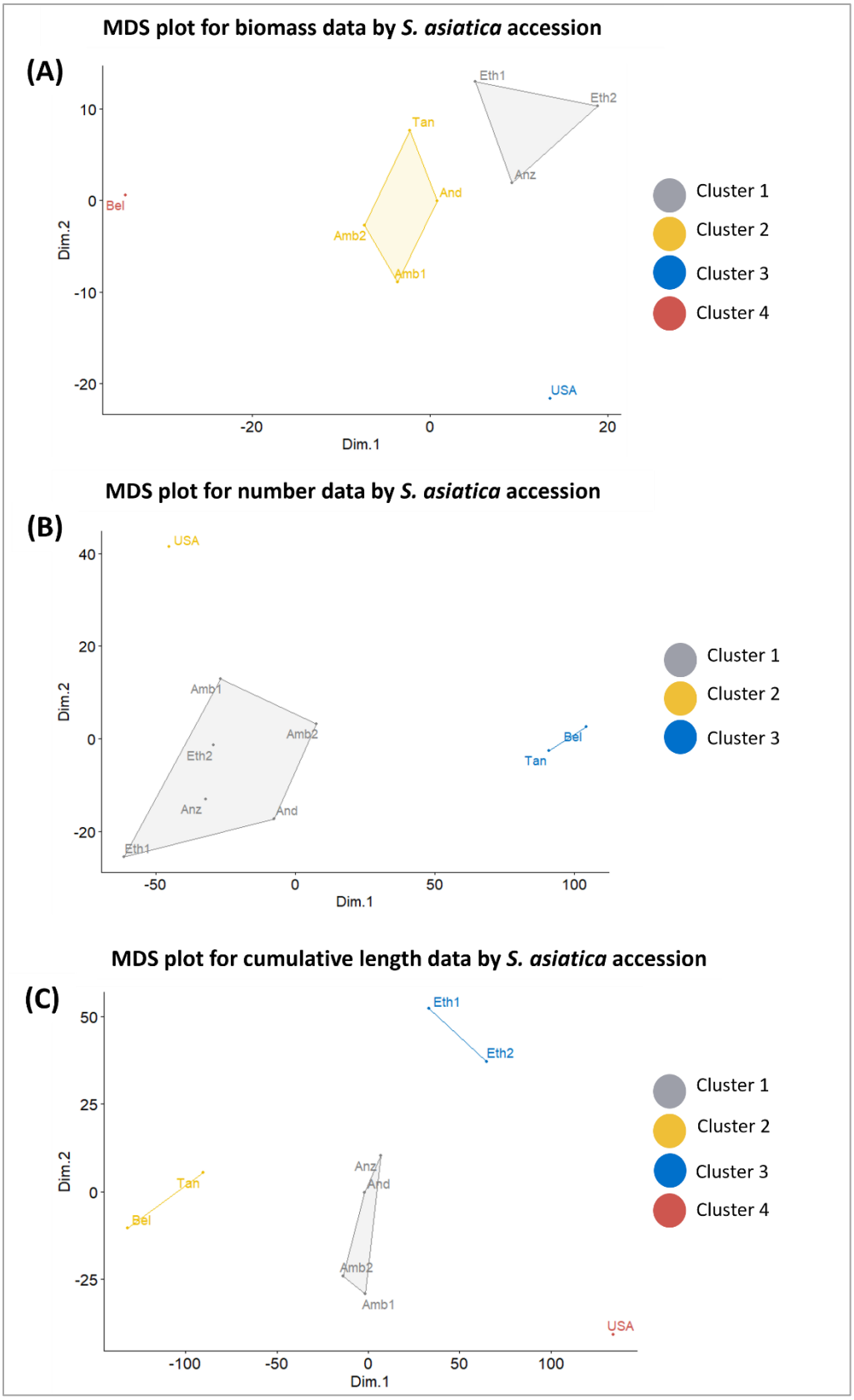


Figure 2.8 – Multidimensional scaling (MDS) plots show how *S. asiatica* accessions can be split into multiple phenotypic clusters by K-means based on median values of *S. asiatica* virulence across 9 rice hosts. Plots show distribution of *S. asiatica* accessions according to these general *S. asiatica* virulence patterns by (A) Biomass data (B) number data and (C) cumulative length data.

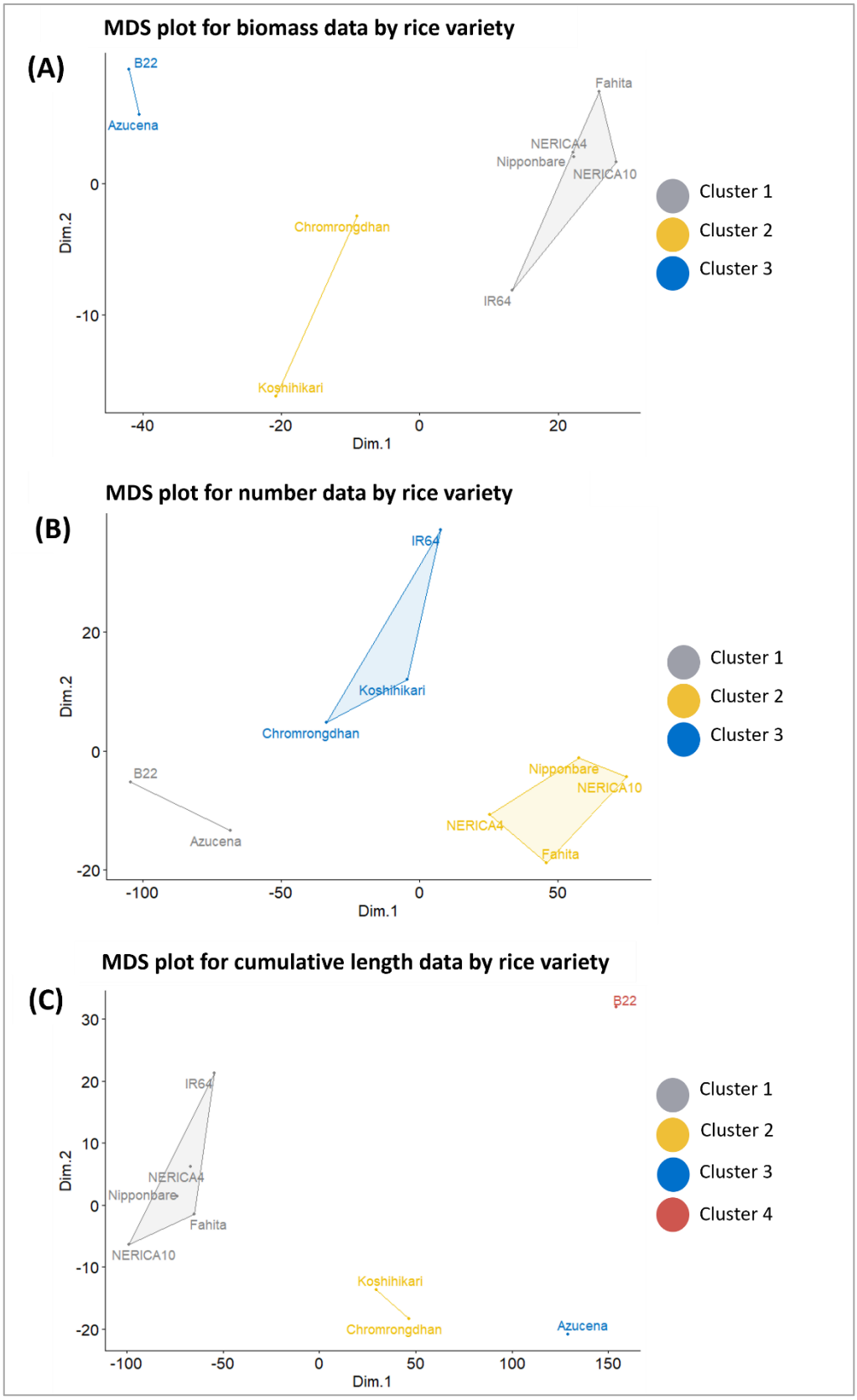


Figure 2.9 – Multidimensional scaling (MDS) plots show how rice varieties can be split into multiple phenotypic clusters by K-means based on median values of *S. asiatica* virulence across 9 rice hosts. Plots show distribution of rice varieties according to these general *S. asiatica* virulence patterns by (A) Biomass data (B) number data and (C) cumulative length data.

To identify whether geographically closer *S. asiatica* accessions shared a similar virulence profile across the hosts, the phenotypic distance between *S. asiatica* accessions was plotted against the spatial distance between them. The hypothesis was *S. asiatica* accessions closer together would share similar virulence profiles. Results indicated a very weak link between distance and *S. asiatica accessions,* with high variability of phenotype (phenotypic distance) amongst those *S. asiatica* accessions geographically closer (R^2 = 0.074, n = 72)(Figure 2.10). Following the exclusion of the atypical S. asiatica accession from Belanitra, the correlation in virulence variation became stronger (R^2 = 0.28, n = 56) as depicted in Figure 2.11. Conversely, after removing all but the four virulent *S. asiatica* accessions (from Anteza, Androvasoa and Ambalamiadana), the correlation weakened, indicating a reduced spatial separation between accessions (R^2 = 0.059, n = 43) as shown in Figure 2.12. Overall this suggested that virulence did vary by geographic distance, but to a greater extent when examining at a wider spatial scale, with clear differences in virulence between Madagascan and Ethiopian, Tanzanian and USA accessions, as opposed to a within-Madagascar scale.

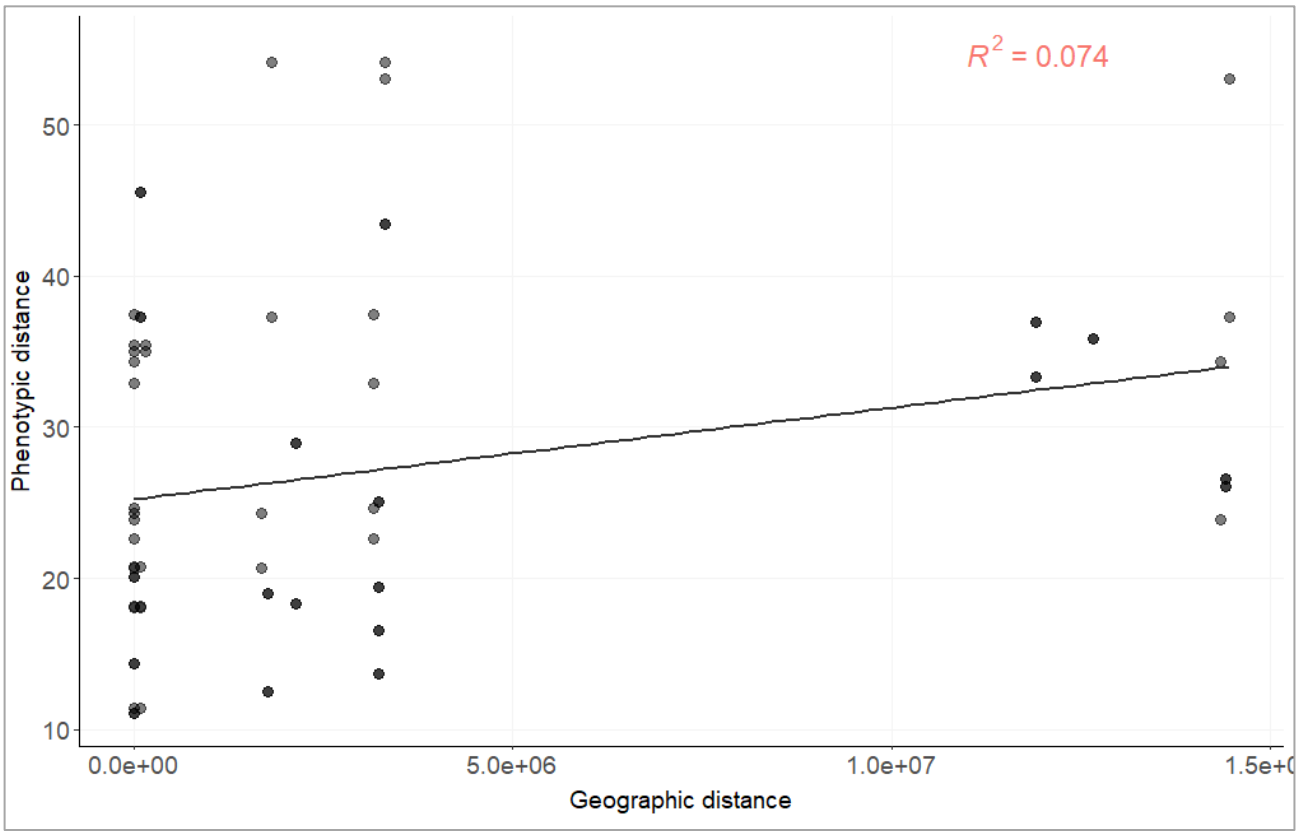


Figure 2.10 – Plot showing the correlation between differences in virulence of 9 different *S. asiatica* accessions on all 9 hosts compared with the spatial separation between the collection sites of the *S. asiatica* accessions. Phenotypic distance is shown as the difference between median *S. asiatica* dry biomass values between accessions, as a measure of variation in virulence. Distance is generated from the distances between *S. asiatica* accession sampling site coordinates and log transformed. The plot shows a weak positive correlation with greater differences in virulence with increasing distance between sampling locations.

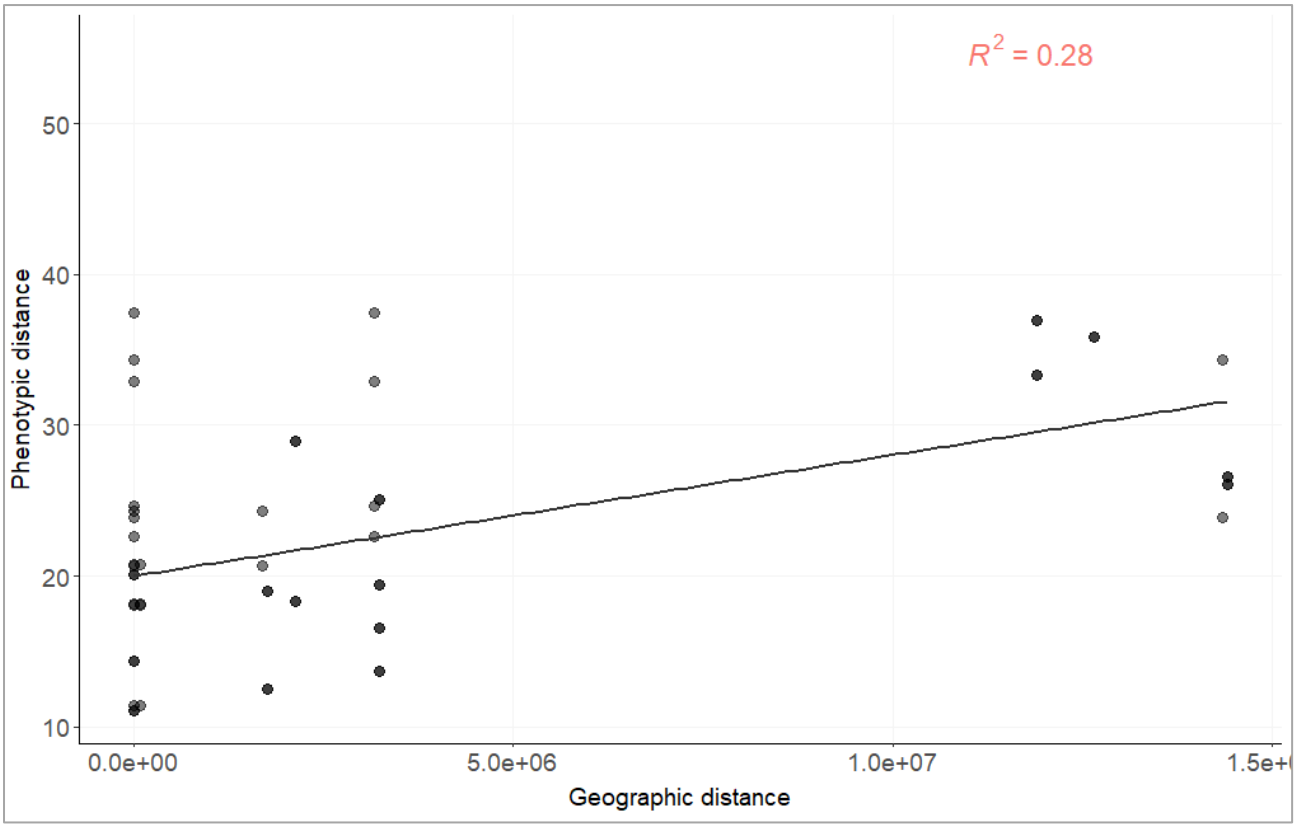


Figure 2.11 – Plot showing the correlation between differences in virulence of 8 different *S. asiatica* accessions on all 9 hosts compared with the spatial separation between the collection sites of the *S. asiatica* accessions, excluding the atypical, highly avirulent *S. asiatica* (Bel) accession. Phenotypic distance is shown as the difference between median *S. asiatica* dry biomass values between accessions, as a measure of variation in virulence. Distance is generated from the distances between *S. asiatica* accession sampling site coordinates and log transformed. The plot shows a weak positive correlation with greater differences in virulence with increasing distance between sampling locations.

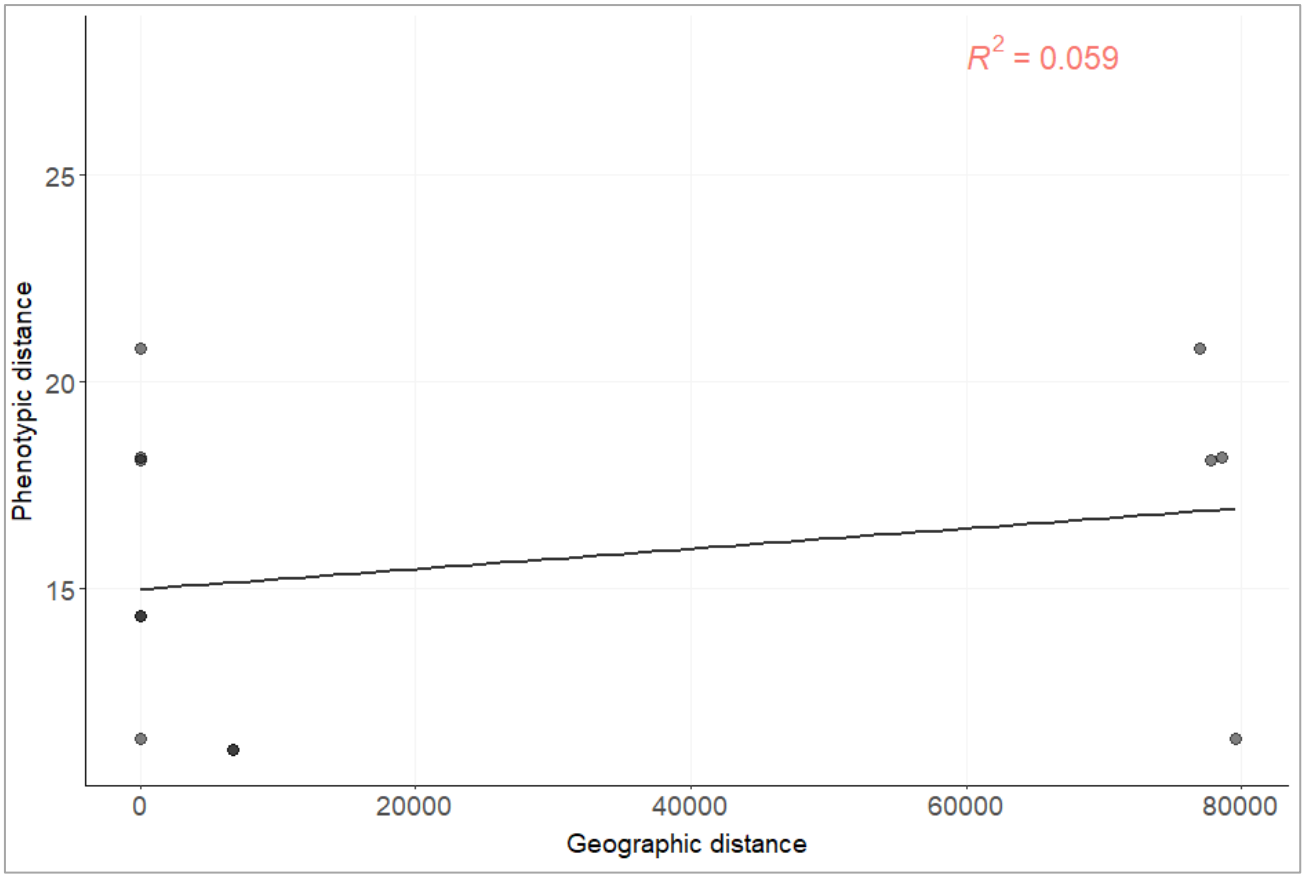


Figure 2.12 – Plot showing the correlation between differences in virulence of the 4 virulent Madagascan *S. asiatica* accessions on all 9 hosts compared with the spatial separation between the collection sites of the *S. asiatica* accessions, excluding the atypical, highly avirulent *S. asiatica* (Bel) accession, the introduced population of USA as well as the Tanzanian and Ethiopian accessions. Phenotypic distance is shown as the difference between median *S. asiatica* dry biomass values between accessions, as a measure of variation in virulence. Distance is generated from the distances between *S. asiatica* accession sampling site coordinates and log transformed. The plot shows a weak positive correlation with greater differences in virulence with increasing distance between sampling locations.

2.3.4 Does the choice of virulence metric affect the study conclusions?

The three measures of virulence used in this study were (1) dry *S. asiatica* biomass per root system, (2) total number of *S. asiatica* attachments per root system and (3) cumulative length of *S. asiatica* attachments per root system. *S. asiatica* biomass and total number have been used in a number of studies previously to study host resistance (Cissoko, 2011; Beardon, 2018). The correlations between these measures are shown in Fig. 2.13. The correlation between biomass and cumulative length (R^2 = 0.73, n = 594) and cumulative length and number of individuals (R^2 = 0.72, n = 594) were particularly strong, compared to a slightly weaker correlation between the biomass and number of individuals (R^2 = 0.4, n = 594).

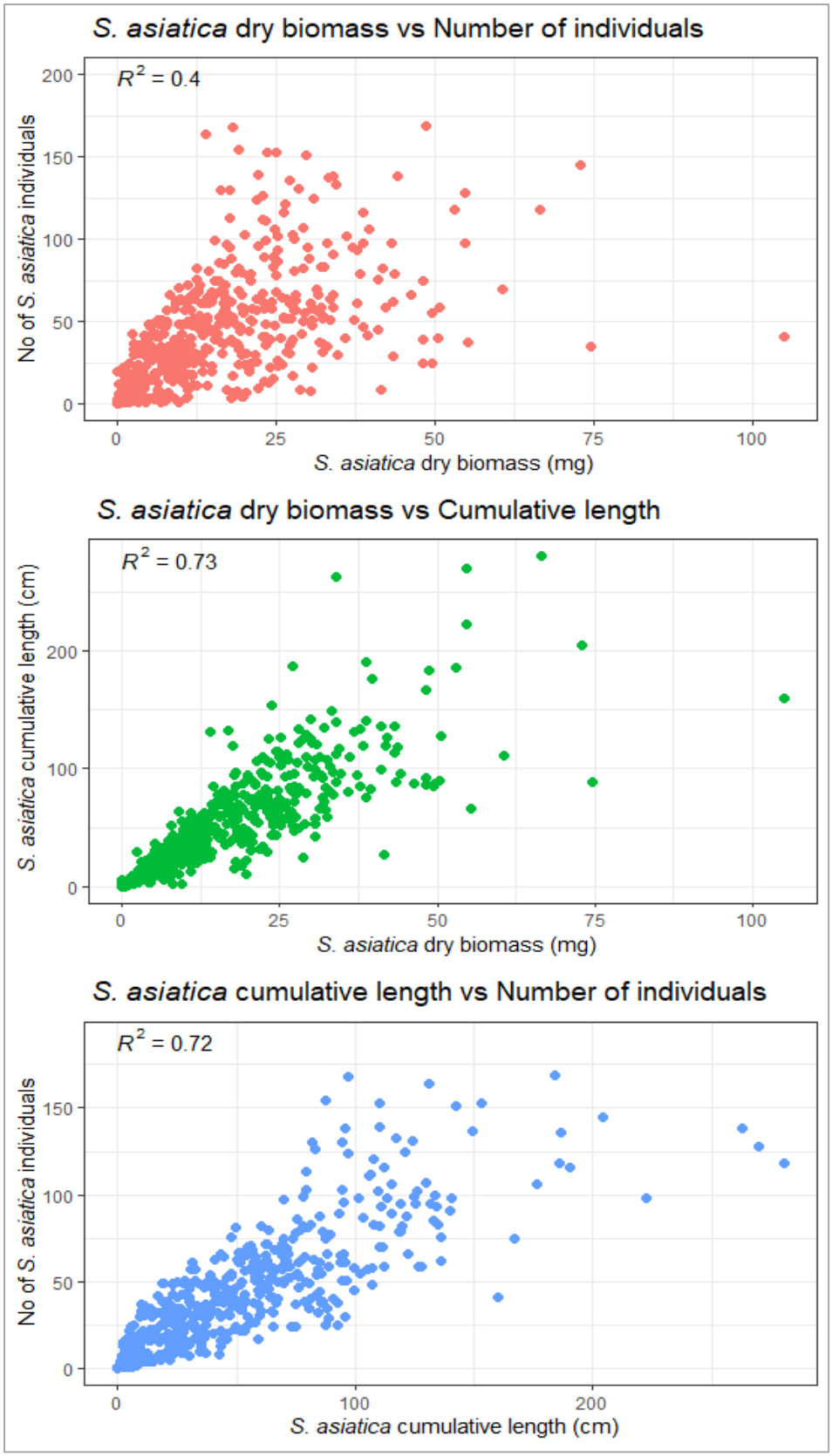


Figure 2.13 - The relationships between the different traits of *S. asiatica* virulence on rice. Correlation coefficients (R^2) are shown within the plots and show a strong relationship between biomass and length (B) as well as total number of individuals and length (C). There is a relationship between dry biomass and number of individuals (A), however this is not as strong.

2.4 Discussion

The interaction between *S. asiatica* and it's hosts is a longstanding one, though the dynamics are ever changing and evolving. It has long been known through research and development into resistant hosts, that no two Striga accessions behave exactly the same way, but the extent to which geographically separate populations differ with respect to virulence on a range of hosts is not well researched. It has been previously suggested that both host variety and geographic isolation are important factors in race formation (Botanga and Timko, 2006), but previous studies have generally focused on smaller geographical scales.

In this study we have confirmed that both rice host variety and *S. asiatica* accession have a significant impact on the virulence profile of *S. asiatica* by comparison of wide spatial ranges, though rice variety had the most significant effect. Distance between sampling sites of *S. asiatica* accessions was shown to influence the overall difference in virulence on a panel of hosts, likely due to differences in host adaptation between those sites which may have had different previous exposures to certain hosts. *S. asiatica* can infect many cereal species and many varieties of those species (Cochrane and Press, 1997; Rodenburg et al., 2017). This has a significant impact on virulence, with different hosts available in different areas and with virulence on each of those hosts varying. Despite this *S. asiatica* thrives as a species and is able to persist successfully in its wide environmental setting. This generalist lifestyle, infecting multiple different hosts, as opposed to a single host has been noted in pathogens to lead to a trade off for virulence, with varying levels of virulence seen for different hosts unless selective pressure necessitates it (Woolhouse et al., 2001).

From this study, the two *S. asiatica* accessions that differed most in virulence were those from Belanitra (Madagascar) and USA (Figure 2.7). The USA accession was derived from seed collected in North Carolina in the 1950s, which is thought to have originated from a single introduction into the USA on contaminated maize seeds which then spread 1750,000 ha (Eplee, 1983; Eplee 1992). This *S. asiatica* accession is highly monomorphic in nature (Werth et al., 1984) with no conclusions regarding the country of origin. In the current study, the USA accession was the most virulent accession on a number of hosts including Azucena and B22 (Figure 2.6), which are both highly susceptible cultivars to many accessions of *S. asiatica*. The geographic isolation of USA and in particular, lack of gene flow from other populations may be the reason for the differences in phenotype.

The Belanitra accession from Madagascar was similarly interesting, despite a very different pattern of virulence. *S. asiatica* (Bel) routinely showed a mostly avirulent phenotype on the panel of hosts and has lacked the possible virulence factors necessary to reach reproductive age in any pot experiment so far. The hypersensitive response consistently found on the hosts of *S. asiatica* (Bel) and the small size of the attachments that were grown suggest a post-attachment resistance mechanism. Due to the late-stage in

which the rice resistance/*S. asiatica* avirulence mechanism takes place, this could be thought to be the result of limitation in parasite ingress following an initial vascular connection. This same later-stage response was identified in *S. hermonthica,* where death of the *S. hermonthica* attachments happened after reaching at least 4 leaf pairs and was attributed to disruption of the flux of water, nutrients and carbohydrates (Amusan et al., 2007; Samejima et al., 2016). In the study by Amusan et al. (2007), a microscopy comparison of *S. hermonthica* attachments at 22 days after infection (DAI) on resistant maize genotype ZD05 and susceptible maize genotype 5057 identified clear differences. While at least four pairs of leaves of the newly emerged *S. hermonthica* were present on both host genotypes, those on the resistant line were more stunted and shoot tissue was purpling. Additionally, only a few lines of tracheids were identified connecting the two xylem, insufficient for growth of *S. hermonthica* at full potential. One reason for this *S. asiatica* avirulence/rice resistance being seen for most replicates of *S. asiatica* (Bel) across the trials may be due to different reasons. Firstly, a lack of sufficient variation in host species tested to properly screen for a full panel of responses which would exist in a natural environment, or alternatively, the particular environmental conditions used in this study did not favour growth for this *S. asiatica* accession.Ordinarily in a species with differentially infecting races, one would expect those with a history of adaptation to one host to have lost some virulence capability to others, such as is seen in *S. gesnerioides* (Botanga and Timko, 2006). So, given that it appears a more generalist species, *S. asiatica* may therefore possess virulence mechanisms which are general used to suppress host basal defences (Chen et al., 2012).

Within Madagascar accessions, Amb1, Amb2, Anz and And behaved in a similar manner to each other in this study on each rice host, though some variation was seen. *S. asiatica* (Amb1) and *S. asiatica* (Amb2), collected from fields metres away from each other generally showed very similar virulence profiles as did sites *S. asiatica* (And) and *S. asiatica* (Anz). This was particularly evident looking at the values of *S. asiatica* biomass and number for the most resistant host varieties (Figures 2.5, 2.6 and 2.7), though not a significant effect. This smaller variation in virulence of *S. asiatica* accession with closer geographic sampling sites was also seen in the correlation plots (Figure 2.10, Figure 2.11 and Figure 2.12). Figure 2.12 includes only the four virulent Madagascan accessions, providing the most reliable estimate of the correlation between virulence variation and distance. This exclusion to only field collected samples is justified because the USA accession originated from a small, introduced population, which may not accurately reflect natural geographic relationships and the Ethiopian accessions were derived from a seed batch of unknown history which may have included crosses with other accessions within the lab. Consequently, the geographic distance in Figure 2.12 is naturally lower compared to Figures 2.10 and 2.11, which include the USA accession. As a result, the correlation depicted in Figure 2.12 is expected to be weaker due to the reduced geographic range and variation.

Overall this suggests that the initial prediction of similarity between virulence profiles of *S. asiatica* accessions collected on a smaller geographic scale was confirmed. This similarity between individuals within an accession was seen at a genetic level by Botanga et al., (2002) using AFLP genetic variability analysis of *S. asiatica* accessions from Benin who saw strong genetic similarities and phylogenetic clustering of individuals within one of 2 lineages based on collection area.

In other *Striga* species the effect of geographic separation on virulence profile has been seen to play a greater role in *S. hermonthica* virulence (Bharathalakshmi et al., 1990; Koyama et al., 2000; Welsh and Mohamed, 2011) than for *S. gesnerioides*, for which virulence is very much influenced by host genotype driven selection (Botanga and Timko, 2005). The effect of host and geography is largely attributable to the neutral population structure and component of outbreeding, as discussed in later chapters. *S. asiatica* (Bel), also from Madagascar is the furthest from all other accessions however, taking into account virulence profiles for those accessions outside of Madagascar, it does not appear that this larger geographic distance between Madagascan accessions would lead to a highly isolated population with such a wildly different virulence profile. The lack of any statistically significant differences between the 4 Madagascan sites excluding site 8 suggests that separate populations within-country may not exist and gene flow between regions in Madagascar is continuous.

This study identified multiple levels of resistance/susceptibility of hosts to *S. asiatica* accessions (Figures 2.9 and 2.7). The two NERICA varieties used in this study NERICA-10 and NERICA-4 still showed good, broad spectrum resistance to *S. asiatica*, despite some growth of the *S. asiatica* (Tan) accession. The New rice for Africa (NERICA) rice varieties, developed by the Africa Rice Center through crossing of African locally adapted *Oryza glabberrima* and high yielding Asian rice *Oryza sativa*, were highly resistant crops and widely adopted for their different desirable productivity properties as well as Striga broad spectrum resistance (Rodenburg et al.,2015).

An intermediate level of resistance among varieties Chromrongdhan and IR64 show two different patterns. On IR64, Ethiopian varieties *S. asiatica* (Eth1) and *S. asiatica* (Eth2) showed higher virulence than others. While this would suggest some level of host adaptation in Ethiopia to IR64, where it is more commonly grown, Chromrongdhan, grown predominantly in Madagascar showed similar values of *S. asiatica* virulence across the accessions, whether Madagascar-based or not. One reason we may not see any higher virulence of Madagascan *S. asiatica* accessions on Madagascan rice varieties could be due to the excellent practices used in farming in Madagascar as well as it being a region predominantly focused on rice growing.

2.5 Conclusions

This study showed that both rice host variety and genotype of *S. asiatica* accession have a significant impact on the virulence profile of *S. asiatica*. While host variety had the most significant effect in determining *S. asiatica* virulence, this variation in virulence did not appear to reflect a race-like structure across the sampled *S. asiatica* accessions, as is seen with the autogamous species, *S. gesnerioides*. Some similarities were seen in virulence profiles between those *S. asiatica* accessions which were geographically closer to each other which is a common feature for the obligately allogamous *S. hermonthica*, however a wider spatial scale for future screening would be able to elucidate this effect further. The differential virulence profiles generated here will go on to form the basis of statistical correlation analysis between genotype and phenotype in Chapter 4, to elucidate potential virulence-associated gene candidates.

Chapter 3

Genetic diversity and population structure of *Striga asiatica* in Madagascar and mainland Africa

3.1 Introduction

As a devastating agricultural weed that parasitizes a multitude of the major staple food crops in sub-Saharan Africa, *Striga asiatica* is an important focus in recent food sustainability research. Not unlike the related species of *Striga hermonthica* and *Striga gesnerioides*, *S. asiatica's* mode of action in parasitism is to penetrate through a host plant's root cortex using a specialised organ called a haustorium and subsequently sequester water and nutrients. While the end goal is to derive these nutrients to further the parasite's own growth, there is thought to be much diversity in virulence mechanisms, how the parasite is able to overcome the host defences (Rodenburg, 2017), particularly when *S. asiatica's* hosts are wide ranging from maize, sorghum and rice to wild grasses and sugarcane (USDA,1957; King, 1966; Hosamani, 1978; Musselman, 1982).

3.1.1 Genetic diversity and inbreeding

In general, maintaining a high level of genetic diversity within a population remains central to allowing for adaptation to changing environments. For obligately parasitic plants such as *S. asiatica*, this is particularly important, as there is a constant need to evolve with the hosts, along with ever changing environmental factors. Neutral genetic variants make up the majority of genetic diversity within a genome, being the product of mutation, migration, genetic drift or mating systems as opposed to natural selection. Therefore, indirectly allows interpretation of adaptive genetic variation. Nevertheless, interpretation of the level and pattern of neutral diversity is not perfect and assumes a constant population size, which is often not the case in the history of populations with bottlenecks and large migration events (Teixeira and Huber, 2021).

Conversely, selection is the process by which certain genotypes become more prevalent due to the advantageous effect on the organism that carries it. The evolutionary response to this selection can happen very quickly. In this case, differences in genotype can reflect differences in phenotype leading to ecological consequences such as increased virulence of Striga on a particular host. It is beneficial to identify both the neutral and adaptive diversity of a species as they do eventually become interlinked, with their effects on structure due to genetic drift or migration (Vellend & Geber, 2005).

A key aspect of *Striga* diversity in adaptive loci is linked to its hosts. A reciprocal adaptation phenomenon has been identified to exist for many parasites and their hosts including *Striga hermonthica* on sorghum, maize

and pearl millet (Bellis et al., 2021), *Globodera pallida* on wild potato (Gautier et al., 2020) and *C. europaea* on common nettle (Koskela et al, 2001). This specificity for particular hosts develops over many generations where the invading organism overcomes the host's defences through advantageous evolutionary changes but at the same time the host develops counter defences, creating an evolutionary arms race.

A plant breeding strategy known to reduce effective population sizes and recombination rates and which therefore has a marked impact in decreasing genetic diversity is selfing (Burgarella and Glemin 2017). Self-pollination is the process by which plants fertilise themselves, transferring the individual's own pollen to the stigma. With selfing species, studies have found that within population diversity is generally decreased, while homozygosity exceeds the hardy-weinberg constant genotype frequency expectations, but among populations the genetic differentiation becomes greater, when compared to outcrossing species (Huang, 2019). The effective size of populations decreases and diversity within those populations is reduced.

While inbreeding may decrease genetic diversity, therefore leading to the widely-acknowledged theory that outcrossing is far more advantageous, inbreeding is actually very common and can be a successful short-term reproductive strategy in usually outcrossing plants, particularly when pollination by other organisms is scarce (Keller and Waller, 2002; Aguilar et al., 2006). For those long-term selfers, having quickly purged deleterious alleles, more energy begins to be directed towards seed production than investment into male function or floral attractants, in a type of reproductive assurance (Lloyd, 1979).

For many plant species, seeds naturally disperse over short distances from the mother plant, leading to many siblings in one area, and even if they do outcross, will be genetically similar due to the sharing of pollen between closely related individuals (Teixeira et al, 2008). There are however, many factors influencing seed spread. For *S. asiatica*, it is known that the parasite's miniscule seeds can adhere to crop seed such as rice which will be traded at markets, and further influenced by human behaviour through collection in farming machinery gear (Plantwise, 2016). While tackling the spread of *S. asiatica* is already a challenge, current estimates of the present climatically suitable areas for *S. asiatica* are around 14.53 million km^2 in Africa alone (David et al., 2022). This holds research into understanding the parasite in high regard in an ever changing climate.

While allogamous (used interchangeably with selfing) species tend to be more rapidly evolving in response to environments, though infrequently arising, virulent/pathogenic features can be more easily fixed into populations of autogamous species leading to periods of rapid parasite expansion (Pekkala et al., 2014). Nevertheless, as speculated by Stebbins (1957), the current absence of any large, long-lived clades of solely selfing species, ally to the now universal theory that these selfing lineages will eventually suffer from extinction, unable to adapt and becoming evolutionary dead-ends, unless this autogamy were incomplete.

3.1.2 Outcrossing, inbreeding and gene flow in Striga species

S. asiatica is one example of a predominantly selfing organism. These phytoparasites do have occurrences of outcrossing but are commonly self pollinating; this is in part due to the maturation of pollen before flowers open, with the style collecting the sticky pollen as it elongates (Musselman, 1982). Nevertheless, accurate estimates of selfing rate have yet to be obtained. Curiously, the two other most devastating Striga species, *S. hermonthica* and *S. gesnerioides* occupy opposite ends of the reproductive spectrum; The former, highly outcrossing and the latter highly autogamous (Musselman, 1980; Safa, 1984). One disadvantage of outcrossing and reason for development of this breeding strategy in many plants is that they are poor colonisers, being far more vulnerable to random extinctions (Jordal et al, 2001). Rapid expansion of Striga species into new regions and on new hosts as human-mediated agriculture has boomed, may have led to this adaptation from outcrossing. This rapid expansion of *S. asiatica* was seen by Yoshida et al. (2019) in comparison to *Mimulus*, which has a shared Lamiales ancestor. Through analysis of divergence patterns of synonymous substitutions (Ks) Yoshida et al. (2019) identified two significant whole genome duplication (WGD) events in the genome of *S. asiatica*, one of that was shared with *Mimulus*. The study identified many genes in *S. asiatica* which were likely acquired through horizontal gene transfer from hosts, promoting importance of haustorial formation through beneficial root development genes. This was seen as a very similar evolution pattern to parasitic *Cuscuta* spp. (Vogel et al., 2018), although *S. asiatica* retains more independence through photosynthetic capability.

For the highly inbreeding *S. gesnerioides*, the most widespread of the Striga species (Mohamed, 2001), the low but strongly structured genetic variability has led to the characterisation of a number of distinct races, all of which infect host varieties of cowpea differentially (Botanga, 2006). A study by Botanga (2006) identified molecular genetic markers which correlated the race-specific virulence with the accession groupings and while many studies have suggested that host driven selection had led to this race-formation, Botanga also proposed that geographic isolation may be a key factor for selfers.

Genetic variation is measured as the tendency of genotypes to differ from population to population and between individuals of the same population. A large component of this variation is down to spatial structuring, but how gene flow in space is structured can be dependent on the species (Fortuna et al, 2009). The term isolation by distance (IBD) refers to an increase in genetic variation between connected populations across a distance where those at opposite ends would be subject to more spatially restricted gene flow (Wright, 1978). These separated populations lack pollen flow between them but produce the same numbers of seed. In addition, selfing reduces effective population sizes by decreasing genetic diversity and limiting gene flow between populations as the contributions of unique genetic material from external populations is reduced. In this sense, selfing organisms would be subject to greater isolation by distance than outcrossing species. Concerning the population structure of autogamous species such as *S. gesnerioides*, they are often attributed

to host specificity, as with strong selfing, the whole genome will be shaped by selection on virulence loci, whereas the shaping of allogamous species' population structures, for example *S. hermonthica,* are more consistently credited to geographical separation (Musselman, 2001;Koyama, 2000;Welsh and Mohamed, 2011).

There is much information unknown about the genetic structure of *S. hermonthica* populations, primarily due to the great diversity in genetic material, hence increasing effective population sizes. This is seen in a study by Unachukwu (2017) which showed little to no population structure within the 254 Kenyan *S. hermonthica* sampling sites, approximately 70 km apart, but more structure across the 775 samples from Nigeria which had more host species' and a maximum of approximately 1000 km apart using 1576 SNP markers. A similar picture of a lack of structure and high genetic diversity is seen in other studies, particularly those that focus on within-country scales (Koyama, 2000; Welsh and Mohamed, 2011; Joel, 2018). However, the use of AFLP markers in many studies limits our understanding primarily due to their dominant nature and the fact that they cannot distinguish homozygous from heterozygous individuals (Savelkoul et al., 1999).

For *S. asiatica*, the studies that have been carried out thus far have shown mixed results. Gethi et al (2005) found no evidence of genetic diversification of Kenyan populations of *S. asiatica* while Botanga et al (2002) suggested that there was enough genetic differentiation to categorise *S. asiatica* from Benin into ecotypes (distinct geographic and phenotypic forms of *Striga*). Botanga also suggested a strong need for a large continent scale study of *S. asiatica* diversity to generate a more detailed map of accessions. However, both studies are early studies and used AFLP markers in genotyping, which are notoriously difficult to accurately infer population genetics parameters from due to an inability to easily distinguish between heterozygous and homozygous genotypes (Foll et al, 2010).

While the materials used for this study were sub-optimal, consisting of seed batches from a single inbred line sourced from Tanzania, Ethiopia, and a non-native population in the USA, the most important focus of this work was on multiple accessions from sampling locations across Madagascar. It is important to consider that the accessions analysed in the population genetic analyses do not represent a natural population and have likely been influenced to some extent by human activities. Nevertheless, this study provides a novel insight into the species, which has not been previously studied on a broader spatial scale, particularly in relation to Madagascar which is the rice-growing centre of Africa, and serves as a valuable starting point for future research.

3.1.3 Aims and Objectives

The use of high throughput single nucleotide polymorphism (SNP) genotyping by whole genome resequencing enables a unique insight into the genetic diversity of *S. asiatica* to achieve specific objectives:

i. Determine how genetic variation of *S. asiatica* populations is structured within Madagascar and within the wider context of the species.
ii. Establish how genetically variable different *S. asiatica* populations are within Madagascar
iii. Determine the selfing rate of *S. asiatica* in Madagascan and Ethiopian populations
iv. Confirm whether isolation by distance is acting in *S. asiatica* populations and over what scale.

3.1.4 Justification of methods

Whole genome sequencing (WGS) is a highly useful technique for studying populations and over the past decade or so has been used extensively as an unbiased alternative to probe into genetic variation and measurements of selfing rate and structure. The use of WGS also allowed for separation of neutral loci and loci under selection for use in later chapters in identifying virulence-associated loci. Being able to use these techniques to understand the extent of autogamy and genetic structure is incredibly important for breeding programs of resistant host crops as we can learn both how evolutionary processes have shaped the way in which the parasitic species' have so far adapted but also how they may respond in the future. Genetic variability among and within populations has a direct impact on how effective resistance expression can be (Kover et al., 2002) and just emphasises the need for an integrated control strategy, focusing on both post- and pre-attachment resistance, as well as good farming practices for the susceptible areas, such as trap cropping (Hussaini et al, 2003).

As previously mentioned in section 2.1, Madagascar is a unique environment for studying *S. asiatica*. Not only is it an agronomically crucial site for growing rice, an important subsistence crop in Africa, but also an interesting geographical location due to the lack of physical land borders. By comparing *S. asiatica* accessions at different spatial scales, both within accession and among accession, across several countries, identification of population structure may be more easily understood.

3.2 Materials and Methods

3.2.1 Sample collection and DNA isolation

The *S. asiatica* accessions used in this study were collected as seeds from the same areas as in section 2.2.1. This included field sites along two 100 km transects in Madagascar, as well as USA (North Carolina), Tanzania (Kyela district) and Ethiopia. All accession groups consisted of seeds from multiple *S. asiatica* plants growing on a number of individual host plants within a field at the different sites. For this study, the sample set was expanded upon from those used in chapter 2 to include three more accessions from Madagascar; the seeds of which were collected from along the same transect regions. The additional Madagascan samples were included to expand the spatial scales on which genetic and geographic diversity were able to be studied. Coordinate data and site codes are shown in table 3.1 and a map of the sampling locations can be seen in Figure 3.1. Further Madagascan seed collection information is detailed in chapter 2, section 2.2.1.

For generation of tissue for sequencing, Striga plants were grown in pots in a controlled environment chamber with a photon flux density of 500 µmol m^{-2} s^{-1} at plant height, a 12 h photoperiod, relative humidity of 60%

and a day/night temperature of 27 °C/25 °C. Striga seeds of the Madagascan accessions were directly collected from the field, whereas those from Ethiopia, Tanzania and USA had been grown for at least one generation in the lab to procure a greater abundance of seeds.

Table 3.1 – *Striga asiatica* sampling site locations where seeds of the different accessions were collected. These seed batches were sown into pots and grown-on in controlled-environment chambers to produce tissue for genome resequencing.

Code	Site No	Commune Location	Host Crop grown in field	Field coordinates (lat/lon)
Amb1	1	Ambalamiadana, Madagscar	Rice var. B22	-18.4349, 46.0118
Amb2	2	Ambalamiadana, Madagascar	Rice var. FOFIFA 3737	-18.4350, 46.0117
Ima	3	Imanga, Madagascar	Rice var. Chromrongdhan	-18.9389, 46.2964
Anz	4	Anteza, Madagascar	Rice var. NERICA4	-18.9770, 46.4774
And	5	Androvasoa, Madagascar	Rice var. Chromrongdhan	-19.0246, 46.4377
Ivo	6	Ivory, Madgascar	Rice var. SCRID	-19.5527, 46.41156
Ant	7	Antsakarivo, Madagascar	Rice var. Voanjolava	-19.6815, 46.5889
Bel	8	Belanitra, Madagascar	Rice var. NERICA4	-19.7198, 46.6125
Eth1	9	Ethiopia (Collection 1)	Unknown	Unknown
Eth2	10	Ethiopia (Collection 2)	Sorghum	Unknown
Tan	11	Kyela, Tanzania	Maize	-9.35, 33.48
USA	12	North Carolina, USA	Maize	34.3834,-78.9460

For each *S. asiatica* accession, two replicate 5 L pots filled with sand were impregnated with 50 mg of Striga seeds at a depth of 7 cm from the soil surface and kept moist and warm to stimulate seed conditioning (Figure 3.2). Pre-germinated rice seeds were planted into pots after 5 days to ensure roots reached the Striga layer around day 12 of the Striga seed conditioning period, at which point, the root exudates produced by the nutrient deprived rice seedlings would be perceptible to the Striga. A total of 3 rice seeds was used, of varieties B22 (*Oryza sativa ssp. Indica*) for Madagascan & USA accessions, and Supa-India for Tanzanian accessions.

Ethiopian accessions were grown on a number of different hosts including CT, IR64 and IAC165 (*O. sativa ssp. Japonica*).

At 21 days, from Striga seed sowing, a 40% Long Ashton nutrient solution containing 2mM ammonium nitrate was provided to the host plants to support Striga growth; 30 ml was supplied 4 times per day delivered via hydroponic system. Striga leaf tissue was harvested from 43 individual Striga plants that emerged from pots and were immediately flash frozen in liquid nitrogen. Younger leaves were preferentially collected due to their lower levels of phenolic compounds. Genomic DNA was extracted from the frozen *S. asiatica* leaves using the Qiagen DNeasy Plant mini kit (QIAGEN, Hilden, Germany) following manufacturers instructions. Approximately 400 mg tissue was needed per individual. In total, 43 high quality DNA samples across 12 different accessions were sent to BGI Tech solutions (Shenzhen, China) for DNBseq™ NGS whole genome re-sequencing on paired end libraries with target 15X genome coverage using 150 bp reads with 300 bp insert sizes.

In addition to the 43 DNA samples sequenced using DNBseq™, another 4 individual samples of *S. asiatica* DNA were sequenced using an Illumina HiSeq sequencer at BGI Tech solutions (Shenzhen, China). Genomic DNA from these samples was collected using an optimised CTAB method (Stewart and Via, 1993)

Figure 3.2 - *S. asiatica* seeds collected from field sites in Madagascar, Tanzania, Ethiopia and USA were grown in pots in controlled environment chambers to generate tissue for whole genome resequencing.

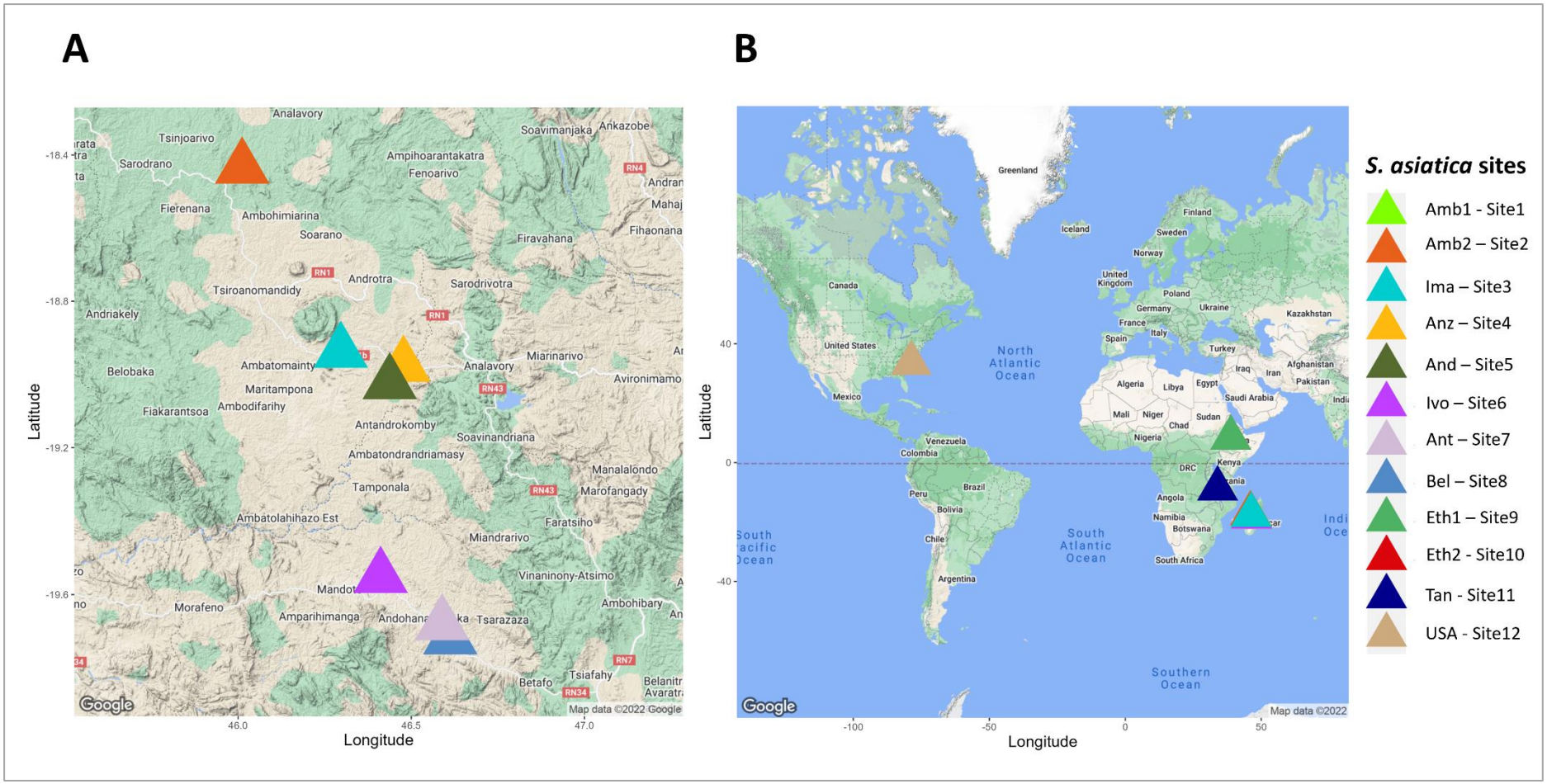


Figure 3.1 – Map of *S. asiatica* sampling sites used in this study. Map (A) shows the 8 Madagascan sampling sites. Sites 1 and 2 were located in adjacent fields in the region of Ambalamiadana (Amb1 and Amb2), not both visible at this scale. (B) shows all sampling sites including the introduced, non-native accession in USA. Terrain elevation and topography is detailed on the maps.

In brief, removal of polyphenolic compounds and polysaccharides was achieved through washing of ground plant tissue with a CTAB-free buffer (2% polyvinylpyrrolidone (PVP), 0.25M NaCl, 0.2M Tris-HCl, 50 mM EDTA), then following the protocol as described by Stewart and Via (1993). These were again sent for whole genome resequencing on paired end libraries using 150bp reads with 300bp insert sizes but with a target coverage of 30X. Raw sequence data from these four individuals was provided by Dr. James Bradley.

DNA integrity had been evaluated with agarose gel electrophoresis and quantity and quality measured using 260/230 and 260/280 ratios from Nanodrop ND8000 (NanoDrop Technologies, Wilmington, DE, USA) and the Qubit 2.0 flurometer (Thermo Fisher Scientific, Waltham, MA, USA). The *S. asiatica* reference genome (Yoshida et al., 2019) was downloaded from the datadryad repository (https://datadryad.org/stash/dataset/doi:10.5061/dryad.53t3574).

3.2.2 Alignment and sequence analysis

The 43 raw individual sequence sets from across 9 different accessions sequenced with DNBSeq are shown in table 3.2 along with sets of sequence data for 4 individuals sequenced with Illumina, the latter performed by Dr. James Bradley. The raw data from the 43 DNBSeq sequence sets were processed by the supplier, BGI using SOAPnuke (Chen et al., 2018) software. This processing discarded reads which (1) contained over 28% adapter sequence, (2) had a quality score lower than 20 for over 50% of bases and (3) had 2% or higher 'N' bases. These clean, processed reads were then each mapped to the reference genome of *S. asiatica* (Yoshida et al., 2019) using the Burrows-Wheeler Alignment (BWA-MEM 0.7.17). The raw reads from four other individual sequence sets, which had been sequenced with Illumina, also went through quality control, but using FASTQC v. 0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and CUTADAPT v. 4.4 (Martin, 2011) programs to visualise, filter and trim reads. All further filtering and alignment steps were performed in the same way across all 47 individuals. Unfiltered data accounted for around ~400Mb, coverage of approximately 70% of the *S. asiatica* genome.

GATK v. 4.14 MarkDuplicates (Poplin et al., 2018) was used for removal of PCR duplicates then Samtools v. 1.7 (Danecek et al., 2021) was used for selection of only properly paired reads with a MAPQ >=20 before GATK v. 4.1.4 HaplotypeCaller (Poplin et al., 2018) generated a number of potential variant sites for each sample against the reference. Coverage statistics for breadth and depth were generated by BEDtools v. 2.30.0 genomecov (Quinlan and Hall, 2010). To ensure all DNA sequences could be filtered along the same pipeline, only a subsample of reads was taken from the 4 extra higher-coverage individuals, to an average depth of 15X coverage. Summaries of coverage statistics are shown in figure 3.3.

Table 3.2 – Table showing the number of *S. asiatica* individuals used for whole genome sequencing. These individuals were clustered into the groups of Ethiopia and Madagascar with USA and Tan forming their own one sample groups or being combined in an 'other' group.

Code	Site No	Analysis Population/Group	Number of individuals sequenced
Amb1	1	Madagascar	4
Amb2	2	Madagascar	4
Ima	3	Madagascar	3
Anz	4	Madagascar	4
And	5	Madagascar	4
Ivo	6	Madagascar	4
Ant	7	Madagascar	4
Bel	8	Madagascar	5
Eth1	9	Ethiopia	12*
Eth2	10	Ethiopia	1*
Tan	11	Tanzania/Other	1*
USA	12	USA/Other	1*

*One sample from each of these accessions had previously been sequenced by illumina at 30X coverage but reads were subsampled to ensure coverage was comparable across all individuals.

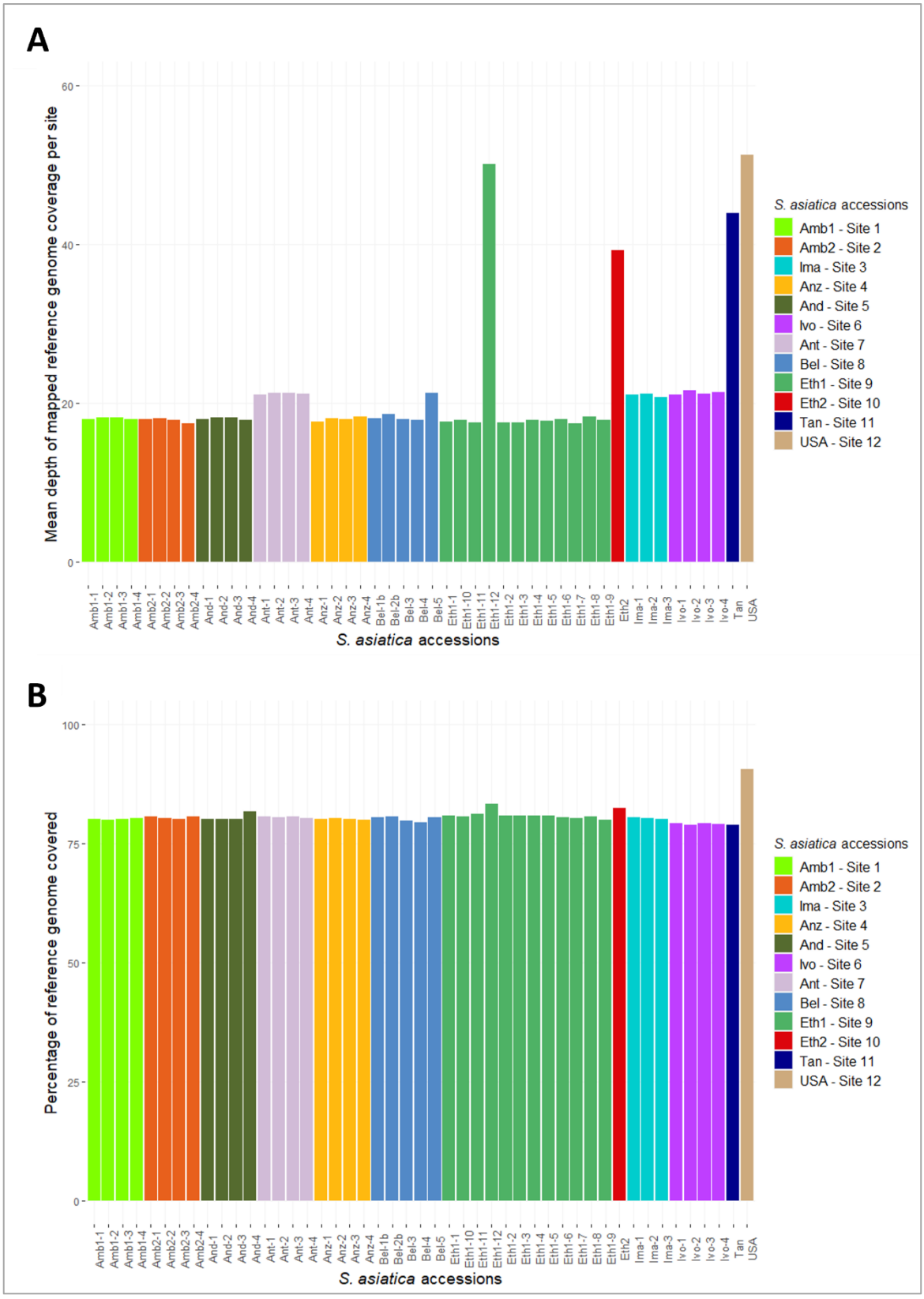


Figure 3.3 – Coverage statistics generated following alignment of resequenced individuals independently to the USA reference genome (Yoshida et al., 2019). (A) shows the mean depth of coverage of reads against the reference for each individual and (B) shows the breadth of the coverage and percentage of reference genome covered by reads per individual.

Table 3.3 – Table detailing the number of SNPs retained following the filtering procedures for SNPs used in population analysis.

Step number	Filtering parameter	Number of variants retained
1	Re-genotyped variant file	15,852,928
2	Selection of SNPs only	13,434,266
3	SNPs filtered by quality scores (INFO)	10,820,200
4	SNPs filtered by quality scores (FMT)	10,820,200
5	SNPs filtered by MAF and missingness	3,358,874
6	Excessively heterozygous SNPs removed	3,161,981
7	Complex SNPs removed	3,081,731

Following GATK best practice, GATK CombineGVCFs (Poplin et al., 2018) was utilised to combine all individuals within one file then GATK genotypeGVCF (Poplin et al., 2018) confirmed that each potential SNP site had at least one variant allele across the individuals. For further analysis, only biallelic SNP polymorphisms were selected with GATK SelectVariants (Poplin et al., 2018). Figure 3.4 details the bioinformatic filtering pipeline from this point and the numbers of variants retained after each step are shown in Table 3.3.

The re-genotyped files were then filtered with bcftools (Danecek et al., 2021) by mapping quality (FS>60.0, SOR>3, MQ<30, MQRankSum<-5.0, QD<2.0, ReadPosRankSum<-5.0, INFO/DP>2940), genotype quality & depth (FMT/DP<3, FMT/GQ<20), missingness and minor allele frequency (F_MISSING>0.1, MAF<=0.015). Due to a high proportion of missing data (around 10% per individual), as shown in figure 3.5, Beagle (Browning et al., 2021) was used to impute missing values across the dataset using default options.

An excess of heterozygosity, likely resulting from merged paralogs from sequencing was also removed from the files. This excess observed heterozygosity is shown as peaks at 1.0 in density plot for loci in figure 3.6. Bcftools (Danecek et al., 2021) was utilised to split combined genotype file into 'populations' of Ethiopia, Madagascar and Other (Tanzania and USA) before removing any loci with over 60% heterozygosity and remerging the files and intersecting only overlapping loci with full genotypes across samples. The total number of SNPs before bcftools filtering was 15.9 million and after filtering was 3.1 million biallelic SNPs. All further analyses were performed on this final, clean genotype data file.

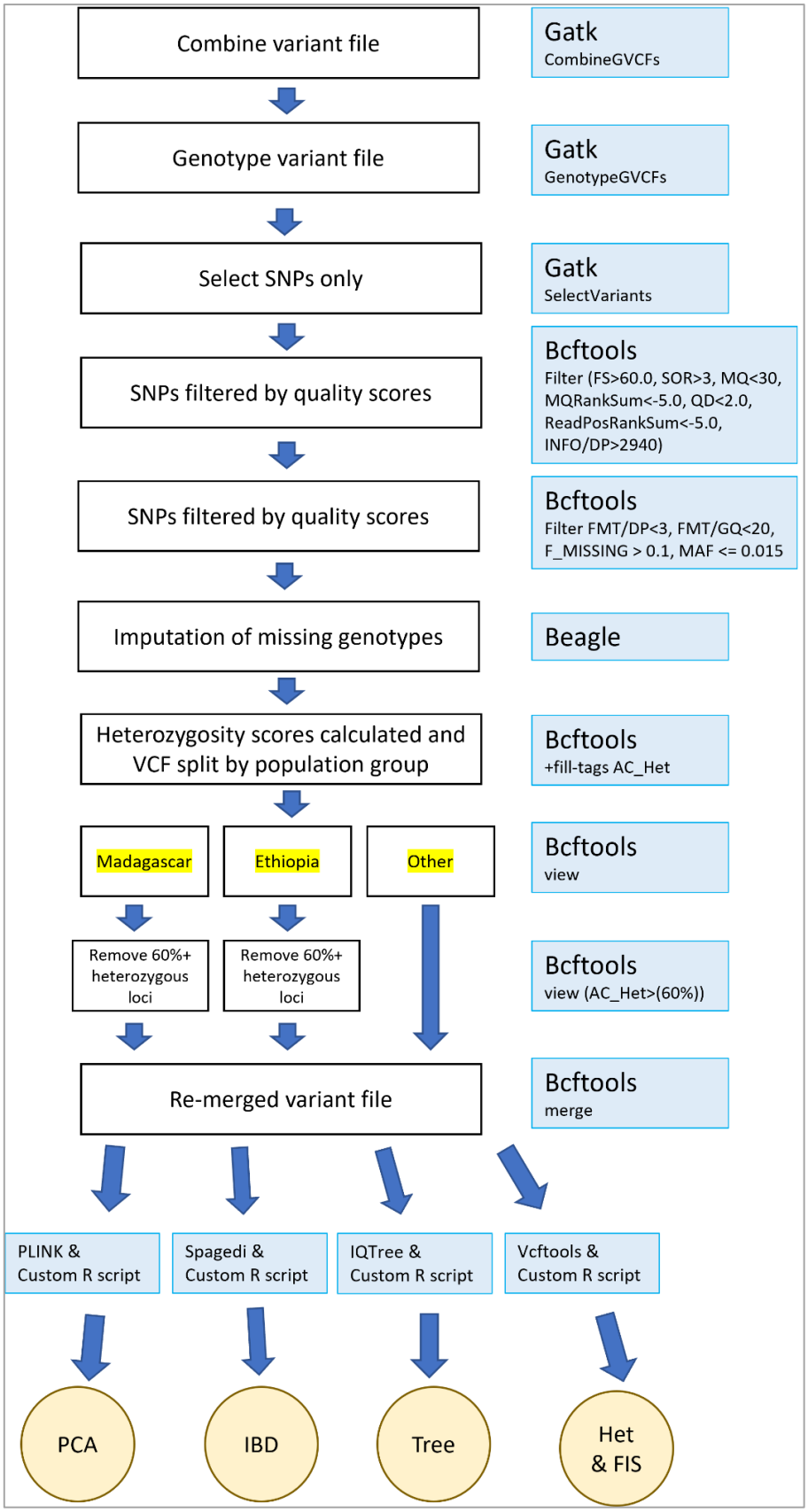


Figure 3.4 – Flowchart showing the bioinformatic pipeline for population and diversity analysis of *S. asiatica* accessions following SNP discovery with HaplotypeCaller, for all 47 sequenced *S. asiatica* individuals, separately.

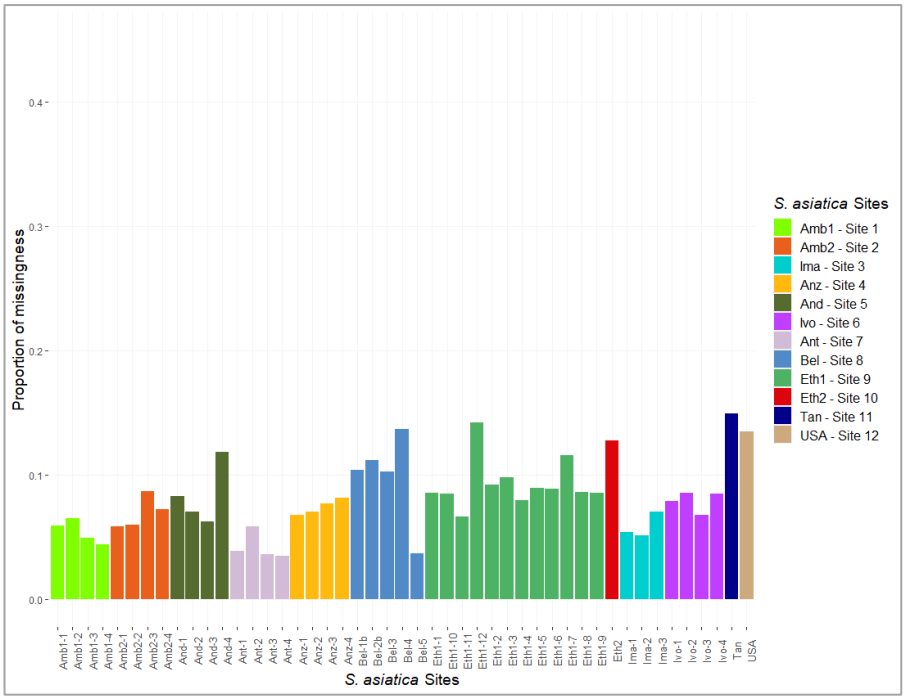


Figure 3.5 – Plot of missing data per *S. asiatica* individual following all filtering steps using bcftools. Due to the presence of over 10% missing genotypes for some individuals, Beagle was used for imputation of these values to complete the dataset.

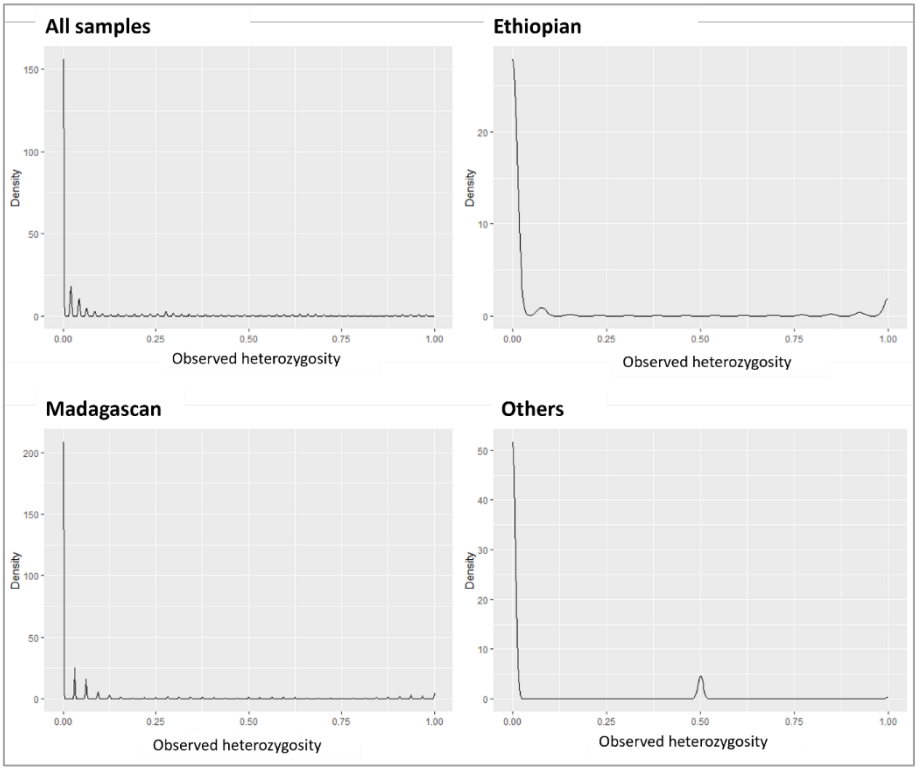


Figure 3.6 - Density plots of SNP loci observed heterozygosity (OHet) showed the prevalence of some excess heterozygosity - A peak at an OHet of 1.0 was present for (A) All individuals and could be seen in both (B) Ethiopian and (C) Madagascan subgroups. Y axes are different for each of the 4 plots.

3.2.3 Phylogenetic and principal component analyses

Phylogenetic trees were built in IQtree (Hoang et al., 2018) with model GTR+G and 1000 ultrafast bootstrap replicates to generate branch support values. The branch lengths shown do not take into account invariant positions or substitutions removed during filtering, but only within the filtered variable site dataset. PCA plots were built in R v4.2.2 (R core team, 2022) with package ggplot2 (Wickham, H., 2016) using eigenvalues produced by PLINK using `--pca` function (Purcell et al., 2007).

3.2.4 Genetic diversity and population structure

Heterozygosity, Inbreeding and population structure and diversity statistics including F_{IS}, the proportion of inbreeding within populations, π, the average pairwise difference of individuals to measure variability and Tajima's D, a signifier of selective sweeps, were calculated with vcftools across the loci and also across the individuals. R package ggplot (Wickham, H., 2016) was used to produce summaries of heterozygosity statistics. SPaGeDi (Hardy and Vekemans, 2002) was used to produce allelic richness and effective allele summaries for the groups.

3.2.5 Isolation by distance (IBD)

A subset of 15,000 SNPs were randomly selected with bcftools vcfrandomsample. Spatial distances between the field sites of accessions in this study were calculated in R using distm from the geosphere package (Hijmans et al, 2022). To ensure differentiation between individuals in the analysis, coordinate sampling site data for all individuals from the same accession was randomly generated around the actual recorded latitude and longitude values using rnorm with sd of 0.0001, correlating to roughly 10 m distance. Spatial distance and genetic distance matrices were generated in SPAGeDi (Hardy and Vekemans, 2002) with 999 permutations using Loiselle pairwise kinship as the statistic. R program ggplot (Wickham, H., 2016) was used for production of regression plots of genetic/spatial distances.

3.3 Results

3.3.1 How is genetic variation structured in *S. asiatica*?

To assess the evolutionary relationships between individual *S. asiatica* plants within different accessions across a number of African countries and determine how populations were structured, genotypic comparisons were made using SNPs. These polymorphisms for each individual were utilised to produce principal component analysis (PCA) plots and phylogenetic trees to display relatedness of individuals.

At a broad level, at least 4 selfing lineages (SL), containing several highly genetically similar individuals from a mixture of accessions, were identified from the *S. asiatica* dataset, as well as 3 separate outgroups containing individuals from single accessions (Figure 3.8). These outgroups consisted of all individuals of the Madagascan

S. asiatica (Ivo) accession, the single *S. asiatica* (USA) individual and the single *S. asiatica* (Tan) individual. The four selfing lineages SL1, SL2, SL3 and SL4 are likely old lineages, with individuals within these groups subject to little crossing outside of their lineage for many generations.

Branches separating the *S. asiatica* accession individuals within countries are relatively short. While the longest branches and likely the oldest, most distant divergences are those between the selfing lineages and from *S. asiatica* (Tan) and *S. asiatica* (USA). *S. asiatica* (Ivo) also appeared to have a rather long branch, distant and the other Madagascan accessions but with individuals within that accession each separated with relatively long branches from additional nodes, when compared to most individuals within an accession. Nevertheless it is worth emphasising that the phylogenetic tree is midpoint rooted and therefore distance between accessions cannot be too accurately described without a sequence of a distantly related individual which everything can be compared to, in respect to evolutionary timelines. In addition, in terms of true evolutionary distances, the scale for the phylogenetic tree shown in Figures 3.8 and 3.9 takes into account only polymorphic sites across the genome. In actuality the scale for Figure 3.8, if accounting for the whole genome would be roughly 0.004 as opposed to the 0.1 shown due to the presence of only 16 million variable sites out of 400 million mapped positions originally identified.

The four SLs of mixed individuals shown in Figure 3.8 are shown at a greater scale in Figure 3.9; These within SL views exhibit how individuals from accessions such as *S. asiatica* (Amb1) and *S. asiatica* (Ima), on a fine scale can be differentiated by accession, but genetic distances are very small with branch lengths between Ethiopian *S. asiatica* individuals at less than 0.02% of variable sites(only those identified as polymorphic), of which 16 million were identified (Figure 3.9D).

On a wide spatial scale any individual(s) within the Madagascan, Ethiopian, USA and Tanzanian accessions clustered closely on a 'Country' level. Within the Ethiopian and Madagascan accessions, many individuals from within one accession such as *S. asiatica* (Amb2) and *S. asiatica* (And) clustered tightly together, indicating more genetic similarities for these individuals within an accession. However some accessions contained multiple lineages such as *S. asiatica* (Anz) and *S. asiatica* (Bel), within these lineages the distance between sampling locations of the mixed accessions was up to 150 km, as seen in SL3 (figure 3.8B)

The most striking exceptions across the samples were the individuals within the *S. asiatica* (Ivo) accession which formed an outgroup away from the other Madagascan accessions and the four individuals within this group were far more genetically variable than any other group of individuals within an accession. The long branch suggests an old selfing lineage but the comparative dissimilarity of individuals within the accession suggests there has been a recent cross.

A similar picture of the population structure is seen with PCA analysis (Figure 3.9), where all Madagascan individuals, except those from *S. asiatica* (Ivo), cluster together while USA, Ethiopia and Tanzania all clustered independently. The Madagascan Ivory accessions appear as a loose cluster on the PCA plots and on the phylogenetic tree show long branch lengths indicating a greater level of diversity within the accession than for any other Madagascan accessions. The relatively tight clusters and short branch lengths within accessions show a high degree of genetic similarity for all other sites. This is again seen in the PCA plot for only Madagascan individuals (Figure 3.10). Here several clusters are formed which show the same split of accessions with *S. asiatica* (Bel) and *S. asiatica* (Anz) spread across two groups.

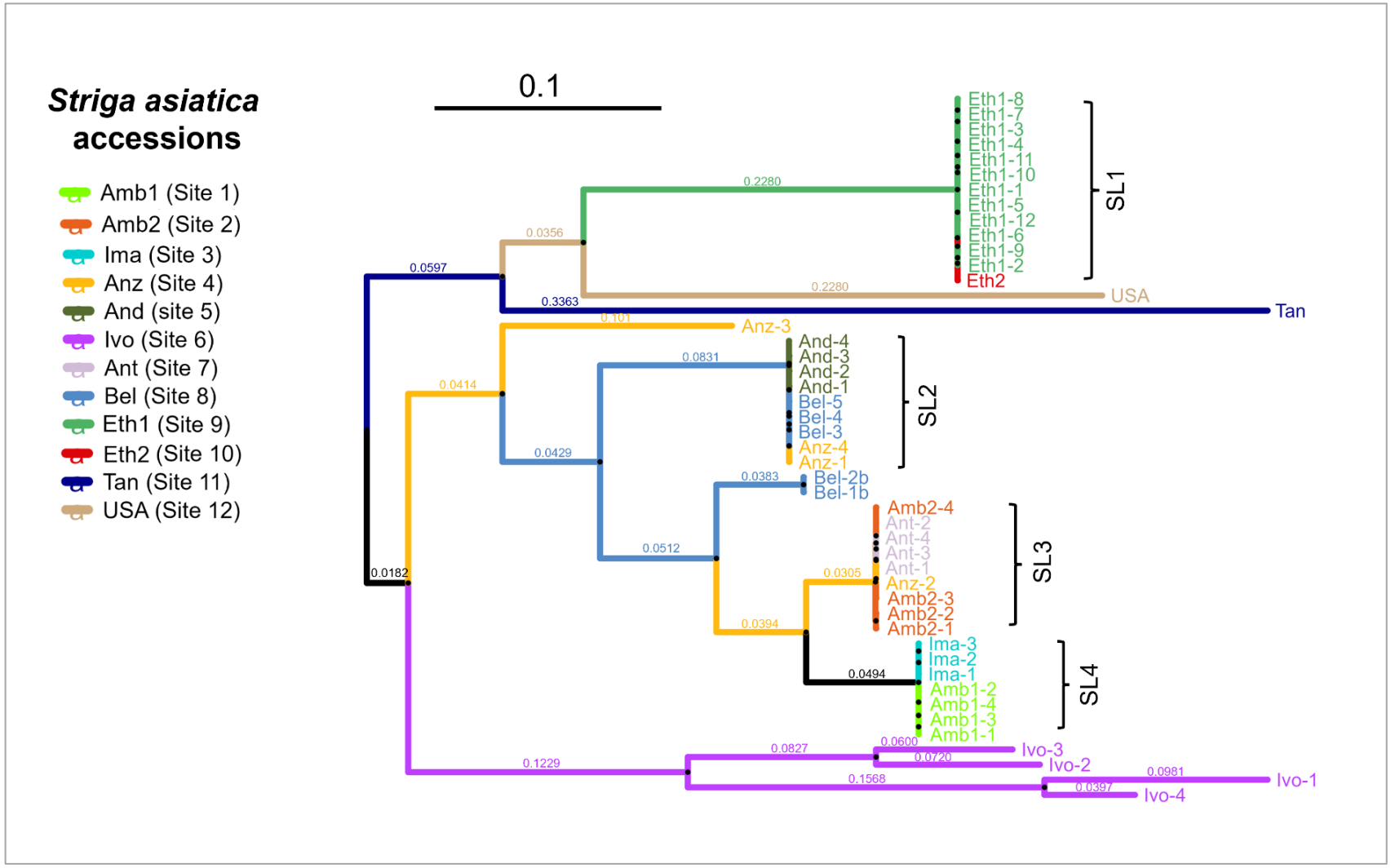


Figure 3.8 - Phylogenetic relationships between individuals of *S. asiatica* from 12 different accessions inferred from polymorphic SNPs. Tree is midpoint rooted and GTR+G substitution model used to calculate maximum likelihood. Branch lengths are shown and scale bar represents 10% genetic distance taken from polymorphic SNPs for these accessions, taken from across the genome. Node support values are not shown but were consistently high. Four selfing lineage (SL) groups are shown along with the independent outgroups of USA, Tanzanian and Ivory, Madagascar accessions

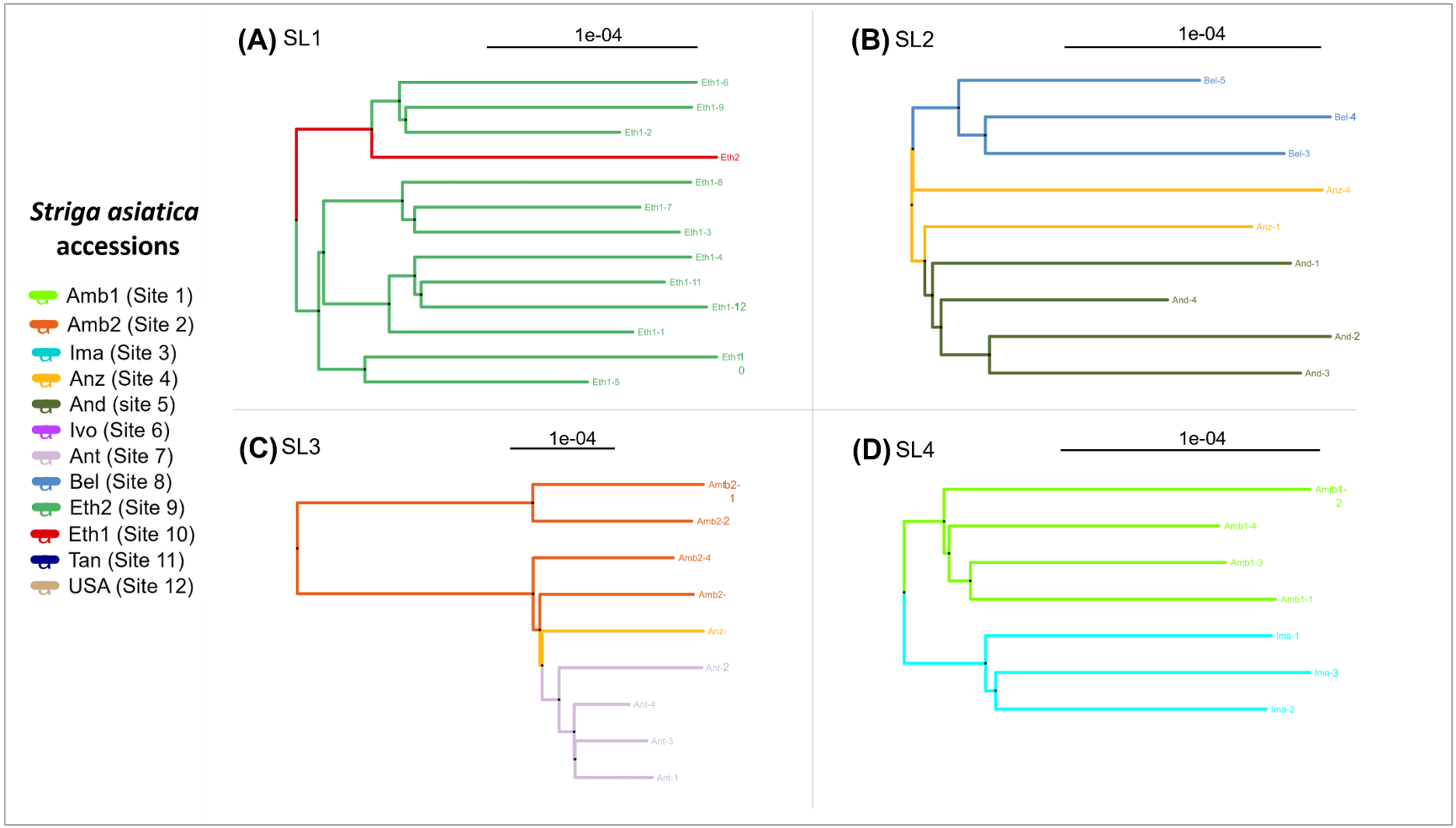


Figure 3.9 – Enhanced look at phylogenetic relationships between individuals of *S. asiatica* within the 4 closely grouping clades . Tree is midpoint rooted and GTR+G substitution model used to calculate maximum likelihood. Branch lengths are not shown but scale bar represents 0.001% genetic distance, taken from polymorphic SNPs for these sites, taken from across the genome.

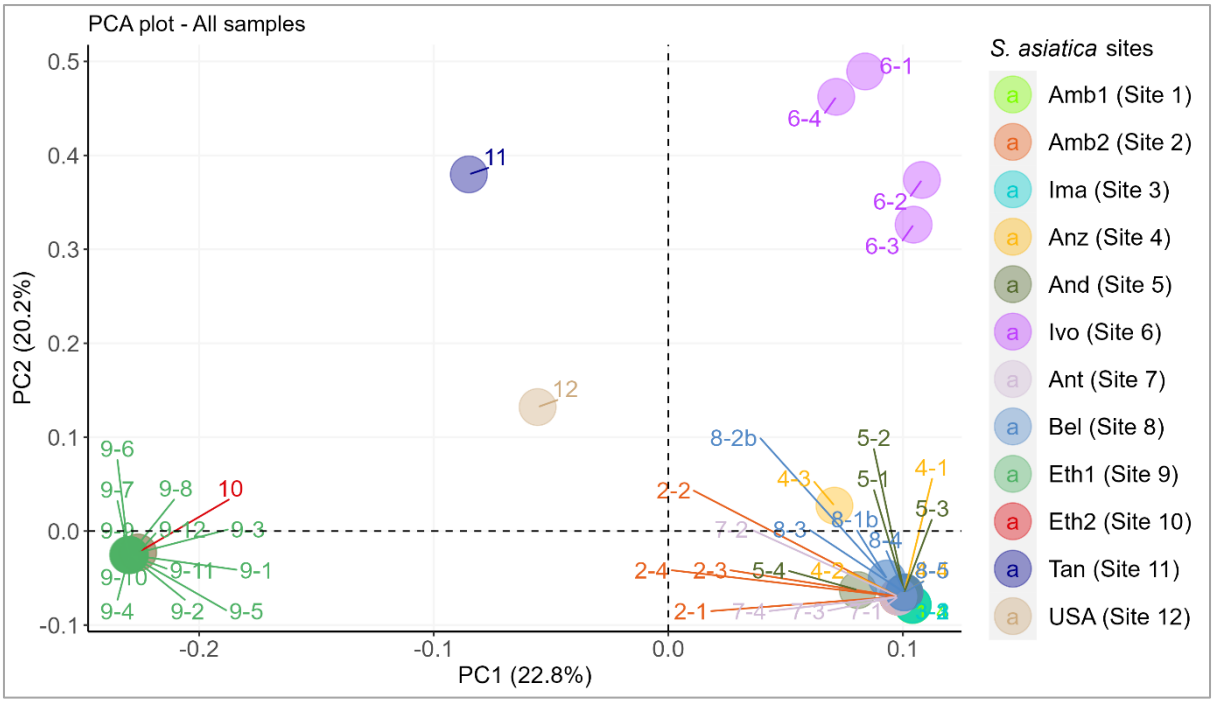


Figure 3.10 – Genetic variation of all *S. asiatica* individuals across all sites structured as a principal component analysis across the first two axes. Individuals are coloured and labelled according to accession collection location site.

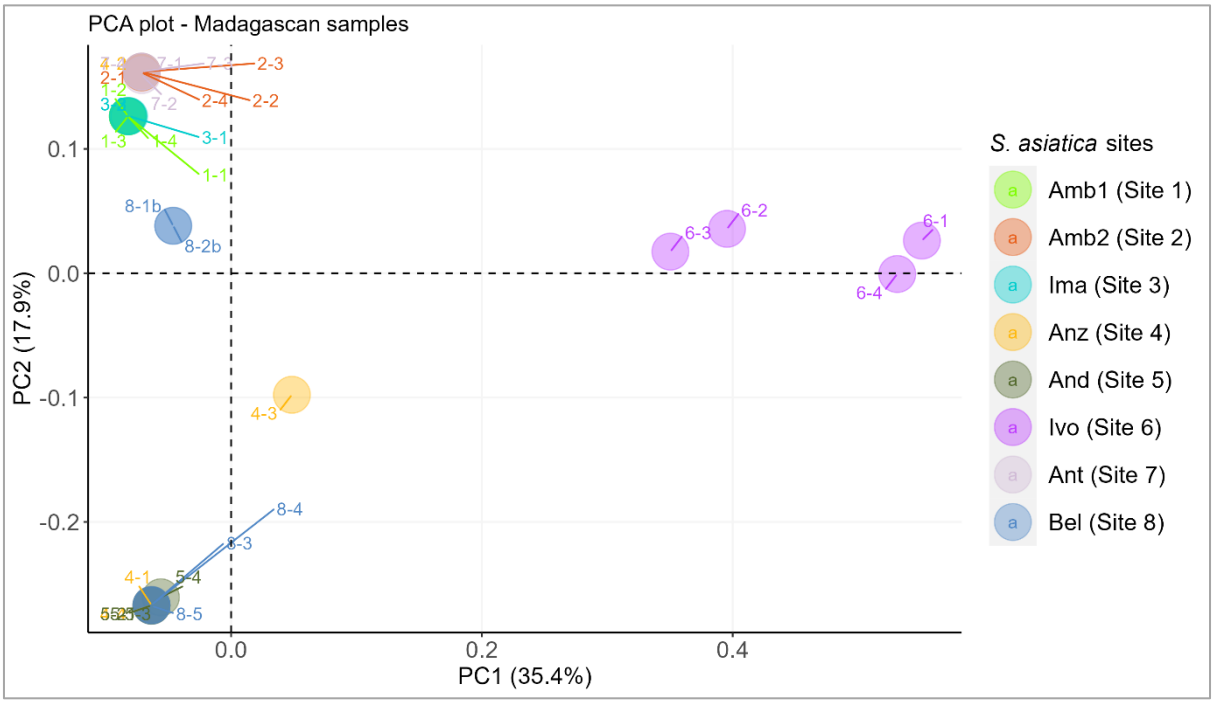


Figure 3.11 - Genetic variation of only Madagascan *S. asiatica* individuals structured as a principal component analysis across the first two axes. Individuals are coloured and labelled according to accession collection location site.

3.3.2 How genetically variable is *Striga asiatica*?

One commonly used measure of nucleotide diversity and genetic variability is Pi (π), the average pairwise difference of individuals within a sample, measuring the degree of polymorphism within a given population. π was estimated for the native populations of Ethiopia, Madagascar and Tanzania separately as well as a total measure of π encompassing 1 randomly selected individual from each of those three regions. Mean values are given in table 3.4 and density plots showing the distribution of π scores

are shown in Figure 3.12.

Table 3.4 – Pi nucleotide diversity (π) for *S. asiatica* within the 3 native country populations of Madagascar. Ethiopia and Tanzania separately, as well as for all samples combined, across all genome SNP sites.

Population name	Madagascar	Ethiopia	Tanzania	All samples	One of each*
Number of individuals (N)	28	11	1	40	3
Nucleotide diversity (π)	9.85×10^{-4}	0.45×10^{-4}	2.09×10^{-4}	12.56×10^{-4}	17.65×10^{-4}

*One of each means one randomly selected Madagascan, Ethiopian and Tanzanian individual were chosen to be used as a representative individual for a more accurate calculation of overall π for *S. asiatica* as a species.

The mean π value given for all samples is one estimate to measure the total nucleotide diversity within *S. asiatica*, this was estimated to be 0.0013 (Table 3.4). However, this significantly underestimated the diversity within the species, due to oversampling of Madagascan and Ethiopian individuals compared to Tanzania. Therefore one randomly selected Madagascan, Ethiopian and the Tanzanian individual were also used for an estimation of genome wide π. This was assessed to be 0.0017, 1.4X greater (Table 3.4). For the three population regions of Madagascar, Ethiopia and Tanzania, mean values were lower and the patterns of π distribution, as seen in Figure 3.12, appeared heavily influenced by selfing with one large peak and a long tail, indicating that the individuals within each group have come from a very similar lineage. The distribution of π for Madagascan individuals, while still low, was more varied than for the Ethiopian individuals, which was predicted due to a wider spatial scale of sampling sites within Madagascar.

For a measure of how the two population groups of Madagascar and Ethiopia are evolving, we used the same loci to test for Tajima's D. Tajima's D investigates departures from neutrality for allele frequency distributions across the genome by comparison of polymorphic sites. This comparison of nucleotide diversity is estimated from the number of polymorphic sites (a simulation of populations that have remained constant in size) against nucleotide diversity estimated from actual allele frequency of polymorphic sites (Carlson et al., 2005).

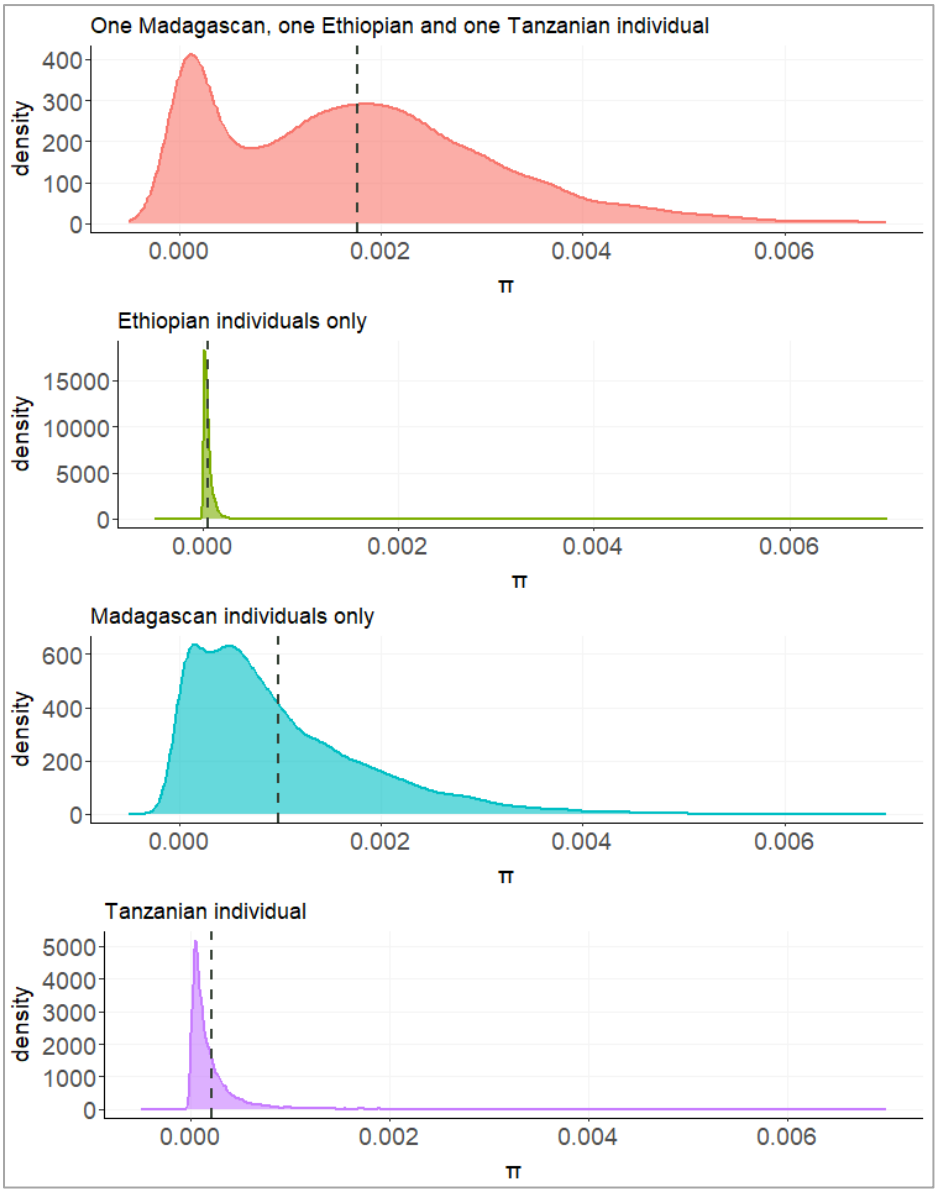


Figure 3.12 – Density plots showing the distribution of nucleotide diversity (π) for *S. asiatica* within the three country populations of Madagascar. Ethiopia and Tanzania separately, as well as for all samples combined across all polymorphic genome SNP sites. The dotted line signifies the mean value of Pi for each test.

Under neutrality, with purely randomly evolving loci in a population of constant size, Tajima's D is 0. Negative values of Tajima's D are associated with excess rare variants due to expanding populations or directional selection. Conversely, positive values of Tajima's D are associated with excess variants with intermediate frequency due to population contraction or balancing selection. It has been predicted that coevolving loci of hosts and parasite will show signatures of balancing selection due to the successive selective sweeps of interacting loci through each generation of the host-parasite 'arms race' (Tellier et al., 2013). Distributions of Tajima's D for the closely related individuals for each of the two populations are shown in Figure 3.13 and means given in Table 3.5.

Table 3.5 – Mean Tajima's D values for within Ethiopian and Madagascan of *S. asiatica* separately, across all polymorphic genome SNP sites.

Population name	Madagascar	Ethiopia
Number of individuals (N)	28	11
Tajima's D	-0.0235	-1.497

For Madagascar, the distribution of Tajima's D is far more broad; the large negative peak seen again highlights the likelihood of a single or very narrow genetic lineage for the individuals within the group, possibly through great expansion of the Madagascan population from a small introduction. However the distribution of Tajima's D also reaches into the positive side of the spectrum which may suggest multiple lineages coming from those 4+ selfing lineages seen in Figure 3.8, though this could be due to a relatively recent introduction of *S. asiatica* into the island of Madagascar compared with other countries.

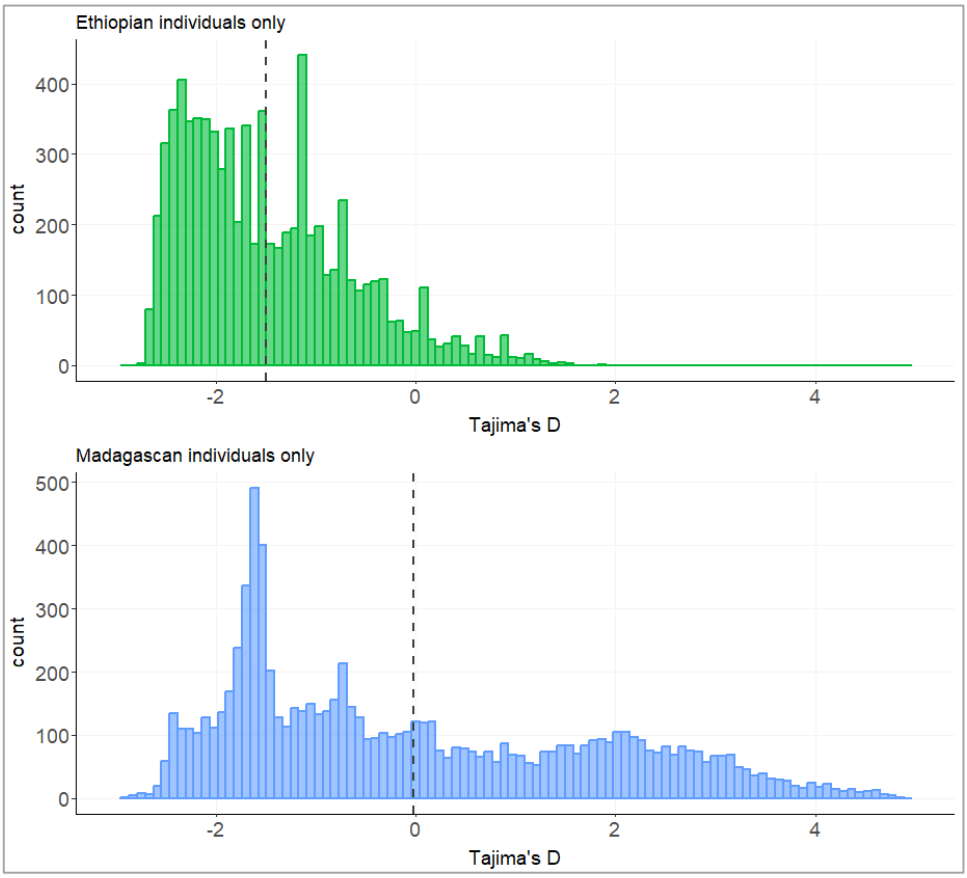


Figure 3.13 - Genetic variation of only Madagascan *S. asiatica* individuals structured as a principal component analysis across the first two axes. Individuals are coloured and labelled according to accession collection location site. The dotted line gives the position of mean values.

3.3.3 What is the level of inbreeding in *S. asiatica*?

Genetic variability was assessed to look at allelic patterns for at a broad spatial scale for Madagascan, Ethiopian and overall groups. This scale was chosen to utilise a larger number of individuals for each group. In preliminary investigations, two Ethiopian individuals, *S. asiatica* Eth2 and Eth1-12 had an excessively low F_{IS} score in comparison to other individuals within this group (Figure 3.7) despite their strong relatedness to other individuals within the group as seen in the phylogenetic tree (Figure 3.8). These were excluded from the overall 'Ethiopia' group heterozygosity calculations presented in this work, as were Madagascan accessions from *S. asiatica* (Ivo), due to their more variable nature and their presence as an outgroup from other Madagascan samples as seen in the PCA (Figure 3.10) and phylogenetic tree (Figure 3.8).

First, to explore inbreeding, mean allelic patterns and heterozygosity estimates for each of the Madagascan and Ethiopian groups are shown in Table 3.3. For Madagascar, the observed heterozygosity was very low and F_{IS} values reflect a highly inbreeding species. This was a different story for the Ethiopian group which had far higher observed heterozygosity and a slightly negative F_{IS}, indicating that this population had undergone much more outcrossing. The number of effective alleles and allelic richness was higher in Madagascar than for Ethiopia, potentially reflecting the greater number of sampling sites for the Madagascan group. Derived from F_{IS} ($S= 2*F_{IS}/(1+F_{IS})$), selfing rate (S) was determined across the individuals within the countries of Madagascar and Ethiopia as well as for all samples together, including the individual from Tanzania. Selfing rate was estimated to be almost 80% for the species as a whole. For the Madagascan group this rate was 0.967 (97%) where for the Ethiopian group this rate was very low (-0.134). Being derived from the F_{IS}, this value would also be indicative of outcrossing, however, as this group was not directly field-derived, instead being generated in controlled lab conditions, the overall selfing rate for *S. asiatica* may well be much higher than 80%.

Table 3.3 - Mean allelic patterns across populations of *S. asiatica* collected in Madagascar and Ethiopia. The category 'All samples' includes the single Tanzanian individual plus individuals from the Madagascan and Ethiopian accessions.

Population name	**Madagascar**	**Ethiopia**	**All samples**
Number of individuals (N)	28	11	40
Observed heterozygosity (Ho)	0.018	0.203	0.009
Expected Heterozygosity (He)	0.288	0.190	0.246
Inbreeding Coefficient (FIS)	0.937	-0.063	0.661
Selfing rate (S)	0.967	-0.134	0.795
Number of effective alleles (Ne)	1.28	1.01	1.39
Allelic richness (Ar)	1.17	1.01	1.23

3.3.4 Is there isolation by distance (IBD)?

Assessment of IBD was key to establish how gene flow was behaving across the *S. asiatica* individuals and how spatial scales influence population structure. On a wide geographic spatial scale, comparing all native accessions of *S. asiatica* from table 3.1 classified by country of origin, excluding only the introduced USA accession, pairwise kinship was plot against ln(distance) following Rousset (2000) (Figure 3.11).

An R^2 value of 0.539 based on 999 permutations of the data matrices showed a highly significant (p=<0.001) relationship between genetic pairwise kinship and geographic distance. On a smaller geographic scale, focusing solely on the relationships between Madagascan individuals for which we have complete geographic data, a comparison of the pairwise kinship between Madagascan accessions against geographic distance was also made. This produced an R^2 value of 0.041 based on 999 permutations of data matrices (p=<0.003).

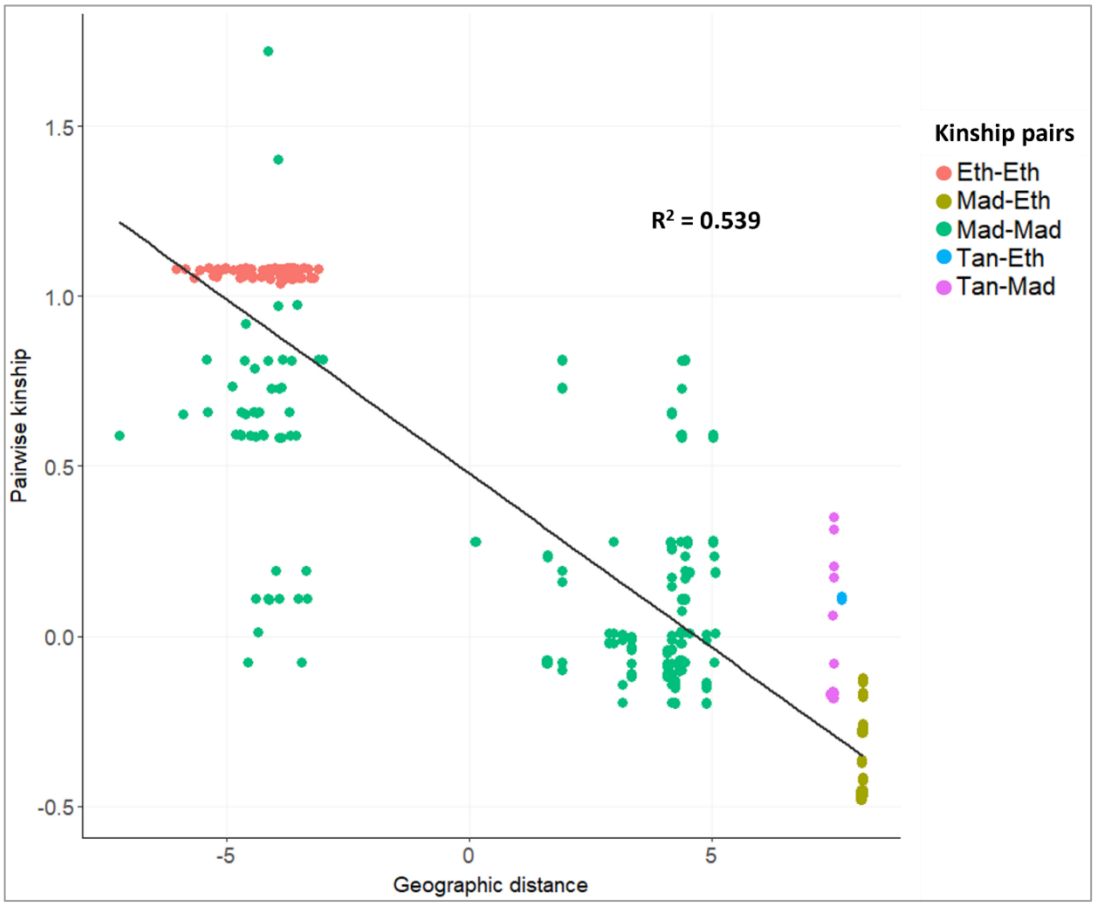


Figure 3.14 - Regression analysis of Pairwise kinship and Geographic distance of *S. asiatica* individuals across Tanzania, Ethiopia and Madagascar. Pairwise kinship was calculated using Loiselle kinship coefficient (Loiselle, 1995) with SPaGeDi. The real geographic distance between Madagascan and Ethiopian individuals is over 2000 km with Geographic distance on x-axis being ln(km).

These results highlight that there is isolation by distance with a decrease in genetic similarity the larger the distance between individuals. However, between accessions, as presented within the island of Madagascar, geographic distance is much weaker.

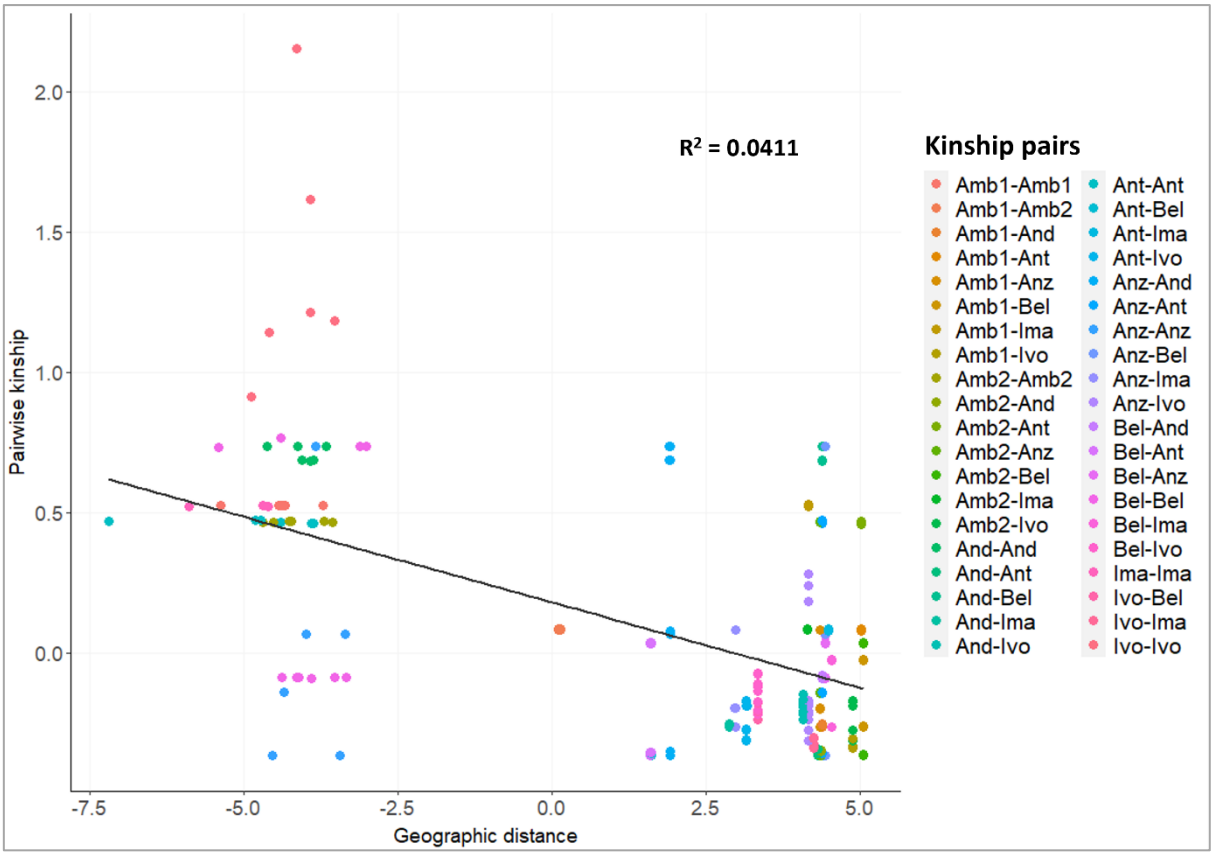


Figure 3.15 - Regression analysis of Pairwise kinship and Geographic distance of *S. asiatica* individuals determining relationships between 8 accession sites within **Madagascar**. Pairwise kinship was calculated using Loiselle kinship coefficient (Loiselle, 1995) with SPaGeDi. The real geographic distance between furthest accessions was over 150 km with Geographic distance on x-axis being ln(km).

3.4 Discussion

It is important for us to understand the genetic diversity of *S. asiatica* populations in order to correctly implement control methods and incorporate resistance into host breeding programs in a way which is durable and broad spectrum. This study has broadened the knowledge base in discerning how *S. asiatica* populations are structured, and importantly on a large geographical scale.

3.4.1 *S. asiatica* has low levels of genetic variability

Estimation of genetic variability has been challenging due to the presence of only two large populations of *S. asiatica* (Madagascar and Ethiopia) which were sampled in very different ways and reflect different spatial

scales. As shown in Table 3.3, the ~3,000,000 filtered variable sites used to distinguish between individuals account for less than 1% of the actual genome, compared to the whole genome size of 600Mbp (Yoshida, 2019). Therefore, this thesis presents variability in a very small portion of the genome.

Despite this, the overall genetic variability of the species of 0.0018 (Table 3.4) that is presented here should provide a reliable measure based on the 400Mb mapping coverage. In part, this is due to being generated using the comparison of the three separate native *S. asiatica* accessions, with only one individual from each being used in order to avoid underestimating the diversity due to the effects of inbreeding. A summary of population and species level π based on breeding systems from Ai et al., (2014) suggests a π_T (species-level diversity) of 0.00427 and π_S (population-level diversity) of 0.00176 for selfing plants or a π_T of 0.00633 and π_S of 0.00360 for mixed mating plants, while outcrossing species were usually greater. Qiu et al. (2022) reported previously that *S. hermonthica* overall mean π was estimated to be 0.11, suggestive of a markedly heterozygous genome.

Soybean provides an interesting comparison to *S. asiatica* due to its similarly low levels of genetic variability. Many studies have looked into the genetic variability of the soybean genome, an important food crop heavily cultivated in the Americas, in order to best identify how to improve breeding programs to introduce more diversity. Current nucleotide diversity levels of ancestors, wild and cultivated populations of Soybean are comparable to *S. asiatica*, ranging between 0.001 – 0.002 (Hyten et al, 2006). Hyten and co-workers report that, although genetic bottlenecks have occurred frequently for Soybean, introduction of wild relatives into breeding programs could easily improve the crop.

For *S. asiatica*, despite currently having very low levels of genetic variability due to being highly preferentially autogamous, any genetic introductions made through transfer of seed between populations could quickly create highly virulent accessions in new regions. The Tajima's D distribution as shown in Figure 3.13 shows that despite a negative mean, the Madagascan population did show both rare and intermediate alleles within the population. This may indicate some old single lineages dominate the gene pool but that new genetic variation is still being introduced into populations. The Madagascan accession from Ivory possibly demonstrates this, forming an outgroup from the other Madagascan accessions.

3.4.2 *S. asiatica* populations are structured into multiple lineages

The population structure of *S. asiatica* appeared to resemble that of a highly inbreeding species. Phylogenetic analyses identified low genetic distance between individuals of different accessions sites and in general, even less genetic distance between individuals within those accessions. It was also identified that clustering of individuals was not necessarily by accession and instead a number of mixed accession selfing lineages were characterized, with even less genetic distance between individuals within that lineage. For this reason, genetic diversity measurements of within and between 'accession' (individuals collected from a single sampling field

site) was not a particularly accurate measure. Instead identifying the structuring of selfing lineages by number, distribution and relationships would yield more useful information.

The PCA and phylogenetic analyses showed that for *S. asiatica*, even at this low level of genetic distance, there is still differentiation between groups. With the exceptions of *S. asiatica* (Ivo) and *S. asiatica* (Anz), individuals from accession sites followed the same lineage (Figure 3.8). However, the presence of Ivory and Anteza individuals across separate groups demonstrates the presence of multiple selfing lines per accession. In particular this would be the case for *S. asiatica* (Anz) one of the middle points between Ambalamiadana and Belanitra, both of which sites Anteza individuals shared genetic similarities with.

The phylogenetic clades containing multiple individuals from different accession sites also suggest that classification of accessions based on their single field collection site may be too small a scale and gene flow is higher than expected between sampling sites. Nevertheless some within Madagascar grouping definition was seen and the analysis performed in this work was able to definitively separate lineages by country – Ethiopia, Tanzania, USA and Madagascar. Research by Gethi et. al. (2005) on *S. asiatica* populations within Kenya also showed less defined lineages than expected and in fact, individuals from accession sampling sites exhibited even more variability than seen in the current study. In their study two individuals each from the accessions Mtepeni and Takaungu were seen to be split across multiple branches of the phylogenetic tree and regression analysis of genetic and geographic distance revealed linear relationship. Gethi et al. suggested that the lack of genetic differentiation seen in the sampling region may reflect a potential race of *S. asiatica*, though phenotypic virulence traits were not investigated.

Concerning *S. asiatica* in Madagascar, multiple introductions are thought to have been made throughout the 20th century in these Madagascan regions (Rodenburg et al, 2016; Scott et al., 2020) which may explain the difference in lineage differentiation seen in this study of Madagascan accessions compared with Gethi (2005). However, as the accession sampling sites used in the Gethi study were at most 80 km apart, in contrast to the present study which used almost double that distance, purely for within Madagascar studies, it would be unsurprising that less population differentiation was seen. Overall, comparing both studies, the individuals of the separate experiments (focusing on Madagascar vs Kenya) cluster among few, long branches of the phylogenetic trees suggesting long-lived selfing lineages which have arisen from a small number of founders but the mixing of accessions among these long-lived lineages suggests that substantial gene flow exists, for the Madagascan population in the current study this is validated by the selfing rate seen of over 95% and sharing of selfing lineages among distant localities.

Another founding study of *S. asiatica* in Benin assessing genetic diversity across a large number of sites using AFLP markers (Botanga et al., 2002) found more defined population structure than we have seen here within Madagascar. Phylogenetic analysis by Botanga et al. demonstrated stronger clear genetic differentiation and

separate lineages between the two accession sites of Massi and Gbegourou, compared with the mix of accessions within genetically differentitated selfing lineages seenin the current study, however the findings were congruent with the notion of *S. asiatica* genetic population structure within and among populations. Botanga also examined populations by morphotype and geographic distance, as opposed to looking for neutral population structure, not influenced by ecotype which may have increased the finer population structure seen when compared to the current study. The existence of ecotypes of *S. asiatica* based on physiological characteristics and often differentiated by geographically separated accession sites has been widely recognised previously (Bharathalakshmi & Jayachandra, 1979; Rodenburg et al 2015).

3.4.3 *S. asiatica* is a highly inbreeding species

Assessment of inbreeding for two populations (Madagascar and Ethiopia) in this study was performed on over 2,000,000 SNPs; Table 3.3 summarises these findings. On initial inspection the two populations of *S. asiatica* appear quite different; though both were expected to be inbreeding populations, Ethiopian individuals had a greater proportion of heterozygous sites, though genomes were less variable with fewer polymorphic sites and with a lower F_{IS} inbreeding score than for the Madagascan population. Opposite trends are also seen for expected vs observed heterozygosity, with a greater expected than observed heterozygosity for Madagascar and greater observed than expected heterozygosity for Ethiopia. It is likely that the Madagascan population is truly inbreeding while the individuals from the Ethiopian population derive from a single inbred lineage. One likely explanation for the behaviour of the Ethiopian population is that they arose from a recent outbreeding event followed by subsequent 2-3 generations of self-pollinated bulking within lab conditions.

Nevertheless, it is important to emphasise the influence of ascertainment bias in the current study as, while the Madagascan accessions contained fewer individuals per accession, the Ethiopian accessions were likely to have arisen from the same batch of seed collected in a far smaller area, therefore reducing the amount of genetic variation that will be seen. The loci used in genotyping in this study were selected based on their polymorphic nature, amongst all individuals, but due to there being far more Madagascan, there would likely be less polymorphic regions detected for Ethiopia, compared with Madagascar.

Observed heterozygosity is calculated from the genotypes taken from individuals within populations while expected heterozygosity is estimated taking into account allele frequencies. Additionally, If the observed heterozygosity is lower than expected, we can attribute the discrepancy to forces such as inbreeding. Whereas when heterozygosity is higher than expected, we might suspect an effect such as the recent mixing of two previously isolated populations (Ritland, 1996). Taking this into account, it is important to note the background of the two populations: Madagascan individuals were collected directly from field sites across known varied locations while Ethiopian individuals were bulked on for multiple generations under

lab conditions. For the Ethiopian accessions (Eth1 and Eth2), they are likely to have arisen from the same sampling area. In a preferentially selfing population, one would expect a highly inbred pool of individuals with low levels of heterozygosity, but this is not exactly the case. One theory may be that crossing occurred several generations ago in the lab environment, perhaps in part due to strong ventilation systems, or alternatively an outcrossing event happened in the sampling area just prior to collection of seed from Ethiopia. This event may have affected the heterozygosity and therefore altered the FIS value to suggest a non-inbred population. Alternatively, there are simply a large number of rare mutations in occurrence amongst a background of mostly fixed loci.

Despite this, using the more natural Madagascan population, which was field collected, we can visualise inbreeding rate of the species. An F_{IS} value of 0.937 or selfing rate of 0.967 for Madagascar is exceptionally high in comparison to a number of other short-lived parasitic plants such as 0.2 for populations of *Agalinis strictifolia* (Dieringer and Werth, 1994), as low as -0.1 for *Orobanche foetida* (Boukteb et al, 2021), 0.1 for *Rafflesia tuan-mudae* (Barkman et. al., 2017) and up to 0.621 for *C. europaea* (Mutikainen and Koskela, 2002). In terms of inbreeding, *S. asiatica* is indeed more comparable to *Striga gesnerioides*, estimated to have a mean inbreeding coefficient (F_{IS}) of 0.939 (Sawadogo et al, 2020), *Orobanche cumana* of close to 1.0 (Pineda-Martos et al, 2014) and *Phelipanche ramosa* with a mean F_{IS} of 0.863 (Le Corre et. al., 2014).

While this current study and many other studies mentioned use crude estimates for selfing rates, it is also possible to assess more accurately, for example through assessment of runs of homozygosity, which are continuous long stretches of homozygous genotype, and offer a better estimate of autozygosity at the genomic level, in order to identify specific Identical by descent (IBD) regions which can be shared among multiple individuals within a genome (McQuillan et al, 2008).

3.4.4 Strong isolation by distance exists across *S. asiatica* accessions

In order to better understand the processes behind the differentiation of accessions of *S. asiatica*, isolation by distance (IBD) was discerned with a Loiselle pairwise kinship test. Significant relationships between genetic distance and geographic distance were observed at both a within-Madagascar (Figure 3.12) and across Africa (Figure 3.11) level. On the wider spatial distance scale which looked at pairwise kinship between Ethiopian, Madagascan and Tanzanian individuals, an R^2 of 0.54 indicating a strong correlation between geographic and genetic distance on this scale. Colonisation history will have had a great influence on the IBD, in particular for the Madagascan island population.

Isolation by distance had also previously been calculated in the study on Benin *S. asiatica* populations by Botanga et al. (2002) where the R^2 between the Benin accessions was 0.61. As a comparison, the similar geographic scale of Madagascan accessions in the current study had a far lower R^2 of 0.041, though still statistically significant correlation. This difference between the two regions of Madagascar and Benin in the

Botanga study may be due to a number of reasons but particularly: (1) the introductions of *S. asiatica* are more recent in the island of Madagascar than in the heavily bordered country of Benin and so local populations have not yet fully formed as genotypically distinct. (2) Sampling was across too small an area to sufficiently see within-Madagascar population differentiation. Introductions into Madagascar have been noted as widespread across the country, yet sampling was only feasible along two 100km transects (Cochrane and Press, 1997; Rodenburg et al., 2016). In addition, further causes of a weak correlation in IBD may be due to a higher than expected dispersal within the country or to larger effective population size, there may also be a difference in terms of continuity of populations through less patchy host cultivation in Madagascar compared to Benin. It is also evident that there is more mixing of accessions seen within the selfing lineages from phylogenetic analysis.

3.4.5 Conclusions

The genetic variability of Striga accessions seen in this study reflects a highly inbreeding species with some defined population structure. The lessons learnt from this data can go on to assess how the variation in these null loci can reflect virulence against their hosts and how this virulence can be dealt with to ensure broad spectrum resistance.

Chapter 4

Identification of candidate virulence-associated loci in *S. asiatica*

4.1 Introduction

In Chapter 2 variation in virulence was identified among *S. asiatica* accessions across geographically close and distant locations in Africa, with variations in *S. asiatica* virulence seen on different host varieties as well as being dependent on the *S. asiatica* accession which was infecting. In Chapter 3, the neutral population structure for these same locations was determined and *S. asiatica* was identified as a strongly inbreeding species with multiple selfing lineages found within Madagascan accession groups, and genetic differentiation between individuals from the three countries of Madagascar, Ethiopia and Tanzania. This chapter's focus was to combine that knowledge and determine whether candidate loci could be identified that are associated with this variation in virulence.

For the hemi-parasitic weed *S. asiatica*, changing environmental situations such as adaptation to the often diverse physiologies and molecular phenotypes of available local hosts, plays a key role in forcing phenotypic plasticity and development of varied virulence mechanisms. Being a major cause of the loss of millions of dollars in crop yield annually (Rodenburg et al., 2016), a key area of study for *S. asiatica* research is to identify and understand this genetic variation underlying parasite virulence, to effectively enhance crop breeding efforts (as for *S. hermonthica*; Qiu et al., 2022). One approach to achieving this goal is by identifying locally adaptive genotypes and associating them with phenotypic variation (Zhang et al., 2019).

4.1.1 The *S. asiatica* infection process and role of the haustorium

One commonality among parasitic plants of all species is the specialised organ, the haustorium, which connects the parasite to the host vasculature via root or stem (Kujit, 1969). A primary function of this organ is the acquisition of water, nutrients and macromolecules, unsurprisingly, to the detriment of the host plant (Kokla and Melnyk, 2018). With respect to root parasites, the action of the haustorium is to invade through the host root epidermis and cortex via elongated 'intrusive cells' which stretch along the interface between the two organisms (Yoshida et al., 2016). Upon reaching the host vasculature, the adjacent parasitic cells differentiate into tracheary elements forming a direct xylem-xylem bridge between host and parasite (Neumann et al., 1999; Yoshida et al., 2016). The process of this is well theorised to be by way of physical

pressure and enzymatic degradation of cell walls in a number of parasitic plant species such as *Orobanche aegyptiaca* (Yang et al., 2015), *Phtheirospermum japonicum* (Kurotani et al., 2020), *Orobanche ramosa* (Gonzalez-Verdejo et al., 2006)*, S. hermonthica* (Qiu et al., 2022) and *S. asiatica* (Yoshida et al, 2016*)*. Though the process of penetration is not entirely understood, the parasite intrusive cells are the site at which this happens. The intrusive cells were first noted by Baird and Riopel (1984) as having dense cytoplasm, indicative of high levels of cell wall hydrolytic activity.

Plant cell walls act as both structural support for cell shaping as well as barriers to biotic and abiotic stresses (Underwood, 2012). They are primarily made up of polysaccharides such as cellulose, hemicellulose and pectin along with a very small percentage of catalytic, structural or signalling proteins (Calderan-Rodrigues et al., 2019). The plant proteome contains a number of protein classes found in all plant species such as proteins acting on carbohydrates and oxido-reductases as described in Table 4.1. In dicots these proteins account for around 10% of the cell wall mass, but only 1% in the monocot hosts of *S. asiatica* (Caffall and Mohnen, 2009*)*. Cell walls are made up of a number of layers as described in Figure 4.1. The primary cell wall is present in all growing cells, while the secondary cell wall, an added layer of rigidity found in cell-types such as tracheids, is only produced once cell expansion has ceased. In general, primary cell walls contain cellulose, pectin and hemicelluloses whereas secondary cell walls contain cellulose, hemicellulose and lignin (Zhong et al., 2019).

Table 4.1 – Functional plant proteome classes Proteins constitute a small percentage but vital component of the plant cell wall. Adapted from Albenne et al. (2013)

Proteome Class	**Protein family**
Proteins acting on carbohydrates	Glycoside hydrolases, carbohydrate esterases
Oxio-reductases	Peroxidases, multicopper oxidases
Proteases	Asp proteases, Cys proteases, Ser proteases
Proteins with interaction domains	Lectins, enzyme inhibitors, leucine-rich repeats proteins
Proteins involved in lipid metabolism	Lipases GDSL, lipid transfer proteins
Proteins involved in signalling	Arabinogalactan proteins, receptors
Structural proteins	Leucine-rich repeat extensins, glycine-rich proteins
Proteins of unknown function	Purple acid phosphatases, phosphate-induced proteins, DUFs

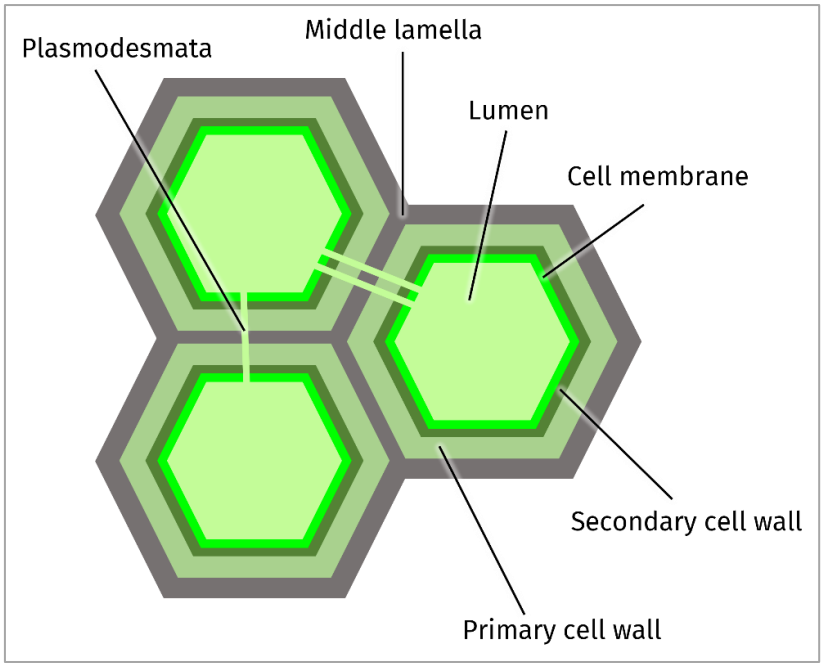


Figure 4.1 – Simplified diagram of plant cells to illustrate cell wall and membrane layers. Individual cells are surrounded by the middle lamella while the inner lumen of cells contains the organelles. Plasmodesmata allow cytoplasmic continuity and are essential for intercellular trafficking. Degradation of host cell walls is a primary target for parasitic plants for colonisation success.

4.1.2 Plant surveillance systems

Plants are constantly bombarded by potential threats, not only in the form of parasitic plants such as *S. asiatica* but herbivorous insects, bacterial or fungal pathogens and parasitic nematodes, all aiming to assimilate plant-derived nutrients for their own growth (Fatima and Senthil-Kumar, 2015). For the plant under attack, the first line of defence is in recognition of the opponent in question, usually before a vascular connection has been made, in the case of parasitic plants. The first layer of host immunity is through pattern triggered immunity (PTI), where pattern recognition receptors (PRR) on potential host cell surfaces attempt to detect non-self pathogen associated molecular patterns (PAMPs)(Zhang and Zhou, 2010). Among many pathogens these PAMP signatures can be highly conserved and therefore easily recognised, however distinguishing more closely related invading plant tissue is far more challenging, such as in parasitic plant-host interactions. Nevertheless, some parasitic plant PAMPs such as Crip21 from *C. reflexa* have been identified and exist as distinct parasitic signatures detected by hosts despite homologous protein sequences existing in other non-parasitic species. Studies have found that Crip21 from a number of parasitic *Cuscuta* spp. will elicit an immune response from tomato but a tomato-derived Crip21 homolog will not (Hegenauer et al., 2016; Hegenauer et al., 2020).

Upon recognition of PAMPS, PRR receptors can dissociate from their partner proteins, stimulating downstream immune signalling and the upregulation of basal defences. This is described in more detail in

Chapter 1, section 1.6. This can involve accumulation of reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2), rapid bursts of Ca2+, activation of mitogen-activated protein kinases (MAPKs), production of phytohormones such as salicylic acid or transcriptional changes (Miller et al., 2008; Ma et al., 2017; He et al., 2020; Flury et al., 2013; Tsuda and Somssich, 2015;Mitsumasu et al., 2015). All together these changes can result in full or partial restriction of parasite/pathogen growth through abiosis/localised cell death, prevention of vascular development or formation of physical barriers such as callose deposition or cell wall stiffening (Wang et al., 2021b; Qiu et al., 2021).

However, in parallel, the invading parasite is able to circumvent resistance by developing independent offensive strategies, effectively targeting components of the host's cellular immune system and manipulating or supressing it (Keller et al., 2016; Ngou et al., 2022). This layer of the immune system is based on 'effectors' or 'virulence factors'. These are proteins produced in the invading parasite which may be secreted directly into the host via the vascular connection made between host and parasite xylem, or via secretion into the host apoplast. In turn, the next layer of the host's surveillance, the Nucleotide-binding leucine rich repeat (NB-LRR) receptors can recognise these pathogen effectors, leading to an immune response or otherwise the effector molecule can reach its target, increasing virulence.

4.1.3 Cell wall modification

With virulence of plant parasites such as Striga, different layers of virulence factors must be involved for successful colonisation of hosts to overcome both physical barriers of host cell walls, ensuring a stable vascular connection, as well as permanent suppression of host immunity to limit any opposition to invasion. Some of the most strategic virulence factors involved making way for the establishment of parasitic connections are the plant cell-wall degrading enzymes (PCWDEs)(Fig. 4.2). Pectin degrading enzymes such as pectin methylesterases, polygalactonurases and pectin lyases have been identified as virulence-associated 'gateway' enzymes used in initial stages of host attack for many parasitic and pathogenic species due to their effect on the structural polysaccharide, pectin (Lionetti et al., 2012; Wang et al., 2021b). Pectin itself is a major component and one of the more complex macromolecules in the cell wall. As well as being crucial in maintaining porosity, pH and ion balance within the cell wall to help with structural integrity (Voragen et al, 2009), pectin is also known to activate defence responses such as phytoalexin accumulation (Nothnagel et al., 1983; Ferrari, 2013). Pectin activates these defence responses when broken down into oligogalacturonides, which are some of the best characterised damage-associated molecular patterns (DAMPs).

To combat the threat of pectin degradation, hosts have developed defensive strategies such as modulation of pectin methylesterification (PME), which offers some protection from the action of parasite pectate lyases and polygalactonurases, maintaining stability of the cell wall (Bellincampi et al., 2014). In potato hosts, this defence response has been shown to be activated by a jasmonic acid signalling pathway, when targeted by

Dickeya dadantii bacterial infection (Taurino et al., 2014) and in *Arabidopsis* hosts infected with *Botrytis cinerea* through the activity of PME inhibitors (Lionetti et al., 2017).

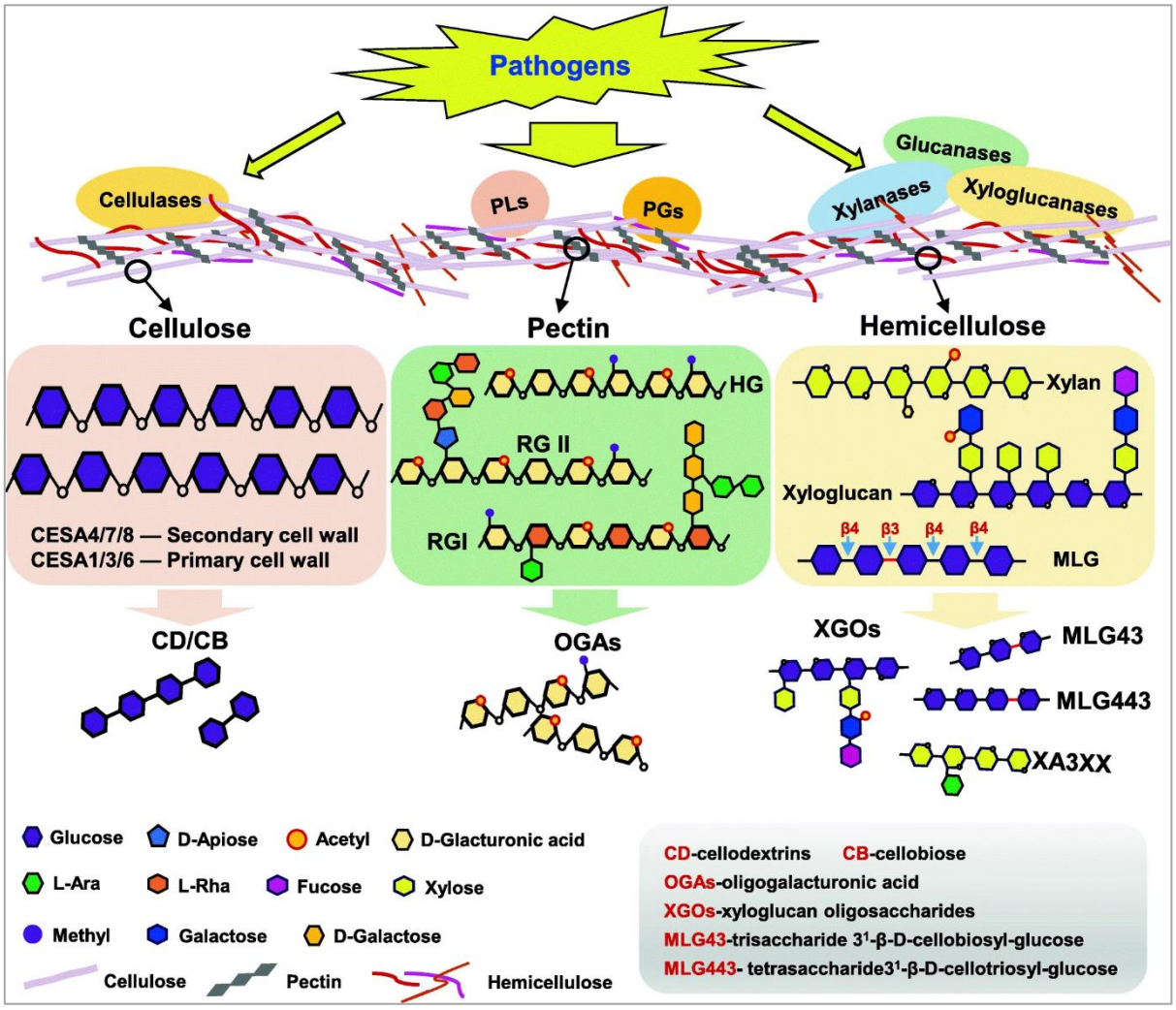


Figure 4.2 – The structure of the plant cell wall polysaccharides, cellulose, pectin and hemicellulose. During parasitic plant invasion of host, plant cell wall degrading enzymes (PCWDE) from parasitic plants act upon these host cell wall components to break them down and enable successful colonisation of the host plant. These PCWDE include cellulases pectin lyases (PL), polygalacuronase (PG), xylanases, glucanases and xyloglucanases. Figure taken from Wan et al. (2021).

Following degradation of pectin by invading organisms, the remaining components of the cell wall, the hemicellulosic chains, may be broken down by cellulases and xylanases (Murashima et al., 2003; Harholt et al., 2010; Rao et al., 2020). Many PCWDE have been identified through transcriptomic analysis as being heavily involved in virulence of parasitic plants due to upregulation of genes within the haustorium. These include a β-expansin which causes loosening of the cell wall seen in *Triphysaria versicolor* infection of maize (Honaas et al., 2013) and β-1-4-glucanase, a cellulose degrading enzyme which can remodel the cell wall in *Phtheirospermum japonicum* infection of *Arabidopsis* and may be implicated in successful tissue adhesion between species (Kurotani et al., 2020). Research by Yang et al (2015) showed upregulation of four glycosyl

hydrolase genes and five pectin lysate genes in the haustorium of at least two species of Orobanchaceae, including *S. hermonthica*. In an extensive study by Qiu et al. (2022) on *S. hermonthica* infecting rice, cell-wall related proteins and proteases accounted for a large proportion of the enriched GO terms amongst the secreted proteins. Qiu and colleagues were able to study differences in allele frequency of genes between pools of *S. hermonthica* DNA collected from susceptible (NERICA-7) and resistant (NERICA-17) hosts. They identified 38 secreted candidate virulence factors, which were clustered into 6 classes: cell wall related, proteases, kinases, lipases, repeat containing proteins and unknown proteins. Examples of these include expansin, B-glucosidase, peroxidase and glycosyl hydrolase all of which were up/down-regulated at different time periods during the infection process. Eight of the 38 genes identified were annotated as cellulases which were hypothesised to facilitate invasion by *S. hermonthica* intrusive cells through the host root, in order to form a xylem bridge.

4.1.4 Suppression of host immune response

While the parasite may be able to break down the cell wall barriers, to persist within a host suppression of the host immune system is another virulence-enhancing strategy. With this, significant roles are likely to be played by the salicylic acid (SA) and jasmonic acid (JA) signalling pathway as identified by Mutuku et al. (2015) who found that knocking down the *O. sativa* WRKY45 transcription factor, involved in the SA pathway, enhanced parasite virulence. In the aforementioned study, the *O. sativa* knockout mutant demonstrated a reduction in expression of downstream SA signalling genes OsWRKY62 and OsPR1b following *S. hermonthica* infection. These signalling pathways have also been identified as possible targets for *Cuscuta pentagona*, which grew far larger on tomato hosts deficient in SA or insensitive to JA (Runyon et al., 2010) and for *O. cumana*, growth of which was suppressed by application of BTH, a synthetic SA analogue, on sunflower (Sauerborn et al, 2002).

For plants, accumulation of SA is known to be involved in regulation of cell death via ROS production and therefore a strong protective measure against invading organisms when detected and essential for an adequate immune response against biotrophic pathogens (Huang et al., 2018; Klessig et al., 2018). Plant pathogens are thought to interact with hormone pathways in a variety of ways including reduction of SA accumulation by conversion to inactive derivatives, interrupting biosynthesis through attack of the pathway or by disrupting downstream signalling processes (Qi et al., 2018). Chorismate mutase (CM) and isochorismatase (ICM) are just two examples of putative effectors modulating immune suppression by attacking these pathways. In the parasitic nematode, *Hirschmanniella oryzae*, CM has been functionally characterised as being involved in reducing secondary metabolite content in host rice plants, with CM acting upon chorismate, the starting molecule in one pathway for the production of SA (Bauters et al., 2020). While there may be a large number of targets for immune suppression by virulence factors/effectors, studies by Mukhtar et al. (2011) suggested that ‘immune hubs’ may be a more likely target for evolutionary diverse

pathogens to converge onto. For the parasitic plant, *P. aegyptiaca*, one such putative immunity hub Pfd6 from *Arabidopsis* was found as a potential target, with overexpression of PFD6 genes leading to a significant increase in resistance to the parasite (Clarke et al., 2020).

In addition to proteinaceous effectors which have been more widely identified, RNA-seq studies have also shown the influence of effectors such as small RNAs secreted from parasite to host to increase virulence. Specifically in *Cuscuta,* a large number of very short sequence micro RNAs are accumulated in the parasite haustorium and are functionally uncharacterised, though thought to be involved in short interfering RNA production and subsequent gene silencing (Shahid et al, 2018).

4.1.5 Virulence gene discovery

Virulence associated genes are typically found in plant parasite or pathogen secretomes, that is, sets of proteins expressed and secreted by cells into the apoplastic space. While secretion from cells is a process that occurs in all plants and is a critical biological system for sessile organisms to interact with their environment, for parasites such as *Striga*, it is thought that delivery of secreted proteins into a host can often aid infection. Identification of putative secreted proteins in parasitic plants can be performed bioinformatically through analysis of molecular signatures (Alexandersson et al., 2013). However, these in-silico prediction pipelines cannot rely solely on homology to known 'effectors' due to the diverse array of proteins that are secreted across plant species. Instead, shared features such as a secretion signal peptide, high cysteine content, small molecular weight and absence of transmembrane domains, can be used to identify candidate proteins (Carreón-Anguiano et al., 2020).

In recent years the sequencing of more plant genomes has significantly increased the range of virulence-associated genes that have been characterised while open-source databases have allowed comparisons with well documented species such as *A. thaliana*. Selective pressures, the driving forces of natural selection, will differ between populations but act to create certain phenotypic traits more favourable for organism fitness in its particular environment; this can lead to the response of local adaptation (Rellstab et al., 2015). Local adaptation refers to the process by which genetic changes confer greater fitness in a local environment which may be related to adaptation to local or more prevalent hosts in that region. Genes underlying the heritable phenotypic variation, specifically here associated with virulence, are therefore of great interest. The expansion in availability of genomic data has also allowed a great emergence in adaptive landscape genomics, a field that combines population genomics, landscape ecology and spatial analytical techniques to gain a deeper understanding of the effect of different environmental factors on both adaptive and neutral genetic variation (Balkenhol et al., 2017).

Traditionally methods for identification of adaptive loci can be classified as either top-down or bottom-up approaches. Top-down methods including GWAS and QTL mapping focus on identifying specific phenotypic

measurements to relate to genotype data while bottom-up approaches such as landscape genomics utilise genomic information to correlate environmental factors to genotypes in a genotype-association analysis (GEA) (Rellstab et al., 2015). While this study shares many similarities with GWAS experiments in the large diverse population sense, QTL analyses are more different relying on controlled crosses of phenotypically different parents to pull out specific traits. A recent study by Qiu et al., (2022) to identify candidate virulence loci in *S. hermonthica* used an multi-faceted approach, pooled parasite DNA collected from either a susceptible (control) or resistant host was sequenced to expose those loci under selection for virulence. In landscape genomics, environmental variables, for which outlier adaptive loci can be associated, can be abiotic (such as temperature or rainfall) or biotic in nature (such as disease prevalence or vegetation type), however, the vast majority of studies in this field have thus far focused primarily on climatic variables (Dauphin, et al., 2023). Some exceptions for biotic factor-focused environment-genotype association landscape genomic studies include prevalence of devil facial tumour disease in the Tasmanian devil (Fraik et al., 2020) and survival of Pacific Ocean perch (Maselko et al, 2020).

The GEA method used to identify virulence-related genes in this chapter is through a form of redundancy analysis (RDA) which utilises constrained ordination to model linear relationships between genetic and phenotypic variation. This association can identify covarying adaptive loci which share an association with certain multivariate (more than one variable) predictors, essentially SNP loci covarying with virulence-association. This can be achieved through pooled sampling (Hoey et al., 2018) or on an individual level (Schweizer et al., 2016) making the technique applicable for large datasets and to elucidate candidate loci associated with different variables. One example of these studies is on the South American conifer *Araucaria araucana*, where over 1000 potentially adaptive loci based on temperature range were identified from a set of over 49,000 RAD-seq markers; this provided valuable information on gene flow for ecological restoration programs (Varas-Myrik et al., 2022).

Similarly, inference of the ability of wild tomato to colonise new environmental niches in Chile was made by Wei et al. (2022) where loci under selection for local adaptation were highlighted through RDA analysis and subsequent in-silico predictions were made to determine GO-term enrichments. With a focus on virulence, a similar technique has also been shown to identify candidate effector genes in the fungal leaf rust, *Puccina triticina* (Wu et al., 2017). Wu et al. found a set of 20 candidate effectors for the virulence gene Lr20 which showed significant counts of synonymous mutations having performed phenotype-genotype association with regression of virulence phenotype on genotype and applying a modified principal coordinate analysis to phenotype data. Following identification of these loci under selection, WZA can combine scores from closely located SNPs within a genic window to better identify the most significantly enriched gene functions which are under selection. This statistical technique has originated from methods for calculating F_{ST} and has been found to identify candidates undiscovered from a traditional RDA (Booker et al., 2023).

4.1.6 Aims and objectives

The focus of investigation within this chapter was to identify genes associated with variation in virulence among *S. asiatica* accessions. This combined a genotype-environment association (GEA), partial redundancy analysis (pRDA) approach with a weighted-Z score analysis, relying upon *S. asiatica* genotype data generated for individuals in chapter 3 and virulence phenotype data from chapter 2, utilising the virulence pattern of *S. asiatica* accessions across all hosts. Accession-level genotype data was not used for this analysis due to the unit of 'accession' not being particularly meaningful and accessions being more genetically variable than expected (see Chapter 3, section 3.4.2). However accession-level virulence phenotype data was used, despite discovery that accessions were not always representative of single selfing lineages for each accession as the sampling was completed ahead of any re-sequencing of *S. asiatica* individuals.

Identification of candidate virulence-associated loci through RDA genotype-environment association studies using WGS data for *S. asiatica* through the methods of:

I. Utilising neutral background loci, location, virulence phenotype and genic SNPs in a partial redundancy analysis (pRDA) to discover possible virulence-associated genes under selection after taking into account background structure.
II. Analysis of enriched gene functions and exploration into the most interesting features of the highest ranked genes and testing predictions of secretome associations and cell-wall related functions.

4.2 Materials and Methods

4.2.1 Plant materials, DNA extraction and genome re-sequencing

S. asiatica genotype data was generated from 25 individuals from across 9 accessions (Table 4.1). For the generation of the *S. asiatica* tissue for each individual, as previously described in section 3.2.1, Striga plants were grown in pots in a controlled environment chamber with a photon flux density of 500 µmol m^{-2} s^{-1} at plant height, a 12 h photoperiod, relative humidity of 60% and a day/night temperature of 27C/25C. Striga seeds of the Madagascan accessions were directly collected from the field, whereas those from Ethiopia, Tanzania and USA had been grown for at least one generation in the lab to procure a greater abundance of seeds. The youngest possible Striga leaf tissue was harvested from individual plants that emerged from pots and were immediately flash frozen in liquid nitrogen before genomic DNA extraction from approximately 400 mg of Striga tissue, using the Qiagen DNeasy Plant mini kit (QIAGEN, Hilden, Germany) following manufacturers instructions. Sequencing of 21 Madagascan individuals was performed by BGI Tech solutions (Shenzhen, China) using DNBseq™ NGS whole genome re-sequencing on paired end libraries with target 15X genome coverage using 150bp insert sizes, while the remaining 4 samples from Ethiopia, USA and Tanzania had been previously sequenced using Illumina and kindly provided by Dr. James Bradley, University of

Toronto, a former student. In total, 25 individuals across 9 accessions were used for this further analysis, as described in Table 4.1. We only included 25 individuals out of the 40 sequences generated in Chapter 2 for virulence-loci association because these were the individuals for which we also had corresponding phenotypic data.

Table 4.1 – *Striga asiatica* sampling site locations where seeds of the different accessions were collected. These seed batches were sown into pots and grown-on in a controlled environment chamber to produce tissue for genome resequencing. Virulence of each of these *S. asiatica* accessions was measured in chapter 2.

Accession code	Site No	Commune Location	Field coordinates (lat/lon)	No individuals sequenced
Amb1	1	Ambalamiadana, Madagscar	-18.4349, 46.0118	4
Amb2	2	Ambalamiadana, Madagascar	-18.4350, 46.0117	4
And	5	Androvasoa, Madagascar	-19.0246, 46.4377	4
Ant	7	Antsakarivo, Madagascar	-19.6815, 46.5889	4
Bel	8	Belanitra, Madagascar	-19.7198, 46.6125	5
Eth1	9	Ethiopia (Collection 1)	Unknown	1*
Eth2	10	Ethiopia (Collection 2)	Unknown	1*
Tan	11	Kyela, Tanzania	-9.35, 33.48	1*
USA	12	North Carolina, USA	34.3834,-78.9460	1*

*Raw reads from Illumina sequencing kindly provided by Dr. James Bradley, University of Toronto.

4.2.2 Sequence alignment, filtering and preparation of genotype data

Sequence data for the Madagascan accessions from BGI was delivered pre-trimmed after SOAPnuke processing and had been initially treated to remove reads containing over 28% adapter sequence, those with a quality score lower than 20 for over 50% of bases and any reads with 2% or higher 'N' bases. These reads were then mapped to the *S. asiatica* reference genome (Yoshida et al., 2019) using the Burrows-Wheeler Alignment (BWA-MEM 0.7.17), as were the previously sequenced reads of the USA, Ethiopian and Tanzanian accessions, which had been subsampled to ensure an average 15X coverage for all samples.

As described in more detail in section 3.2.2, GATK (Poplin et al., 2018) and Samtools (Danecek et al., 2021) were used for refinement of reads before GATK HaplotypeCaller (Poplin et al., 2018) generated a number of potential variant sites for each sample against the reference genome (Yoshida et al., 2019). Variant called files for the 25 individuals were combined and re-genotyped with GATK (Poplin et al., 2018) before selecting only biallelic SNP polymorphisms. Bcftools filtering (Danecek et al., 2021) was performed by mapping quality (FS>60.0, SOR>3, MQ<30, MQRankSum<-5.0, QD<2.0, ReadPosRankSum<-5.0, INFO/DP>2940), genotype quality & depth (FMT/DP<3, FMT/GQ<20) and missingness & minor allele frequency (F_MISSING>0.1, MAF<=0.015). Due to a high proportion of missing data (around 10% per individual), Beagle (Browning et al., 2021) was used to impute missing values across the dataset using default options. Complex and non-biallelic SNPs were also removed with vcftools.

SNPeff variant annotation tool (Cingolani et al., 2012a) was used to assign an annotation prediction for each SNP site, with reference to the *S. asiatica* genome (Yoshida et al., 2019). This categorised variant sites by location such as downstream of a genic region, upstream of a genic region or within an exon, as well as information on predicted function types such as a non-synonymous change or gain of a stop codon, this is presented in Figure 4.3. The purpose of using snpEff was to be able to categorise loci as either genic or non-genic, using the -ud 2000 flag to specify the interval length, up and down-stream of genes to be 2000 bp. SnpSift (Cingolani et al., 2012b) was then used to separate the loci into intergenic and genic sets of SNPs, classifying all sites other than intergenic and downstream as genic. Intergenic sites were then subsampled using bcftools vcfrandomsample leaving a dataset of ~250,000 SNPs to be used as a proxy for neutral population structure in RDA analysis. This was due to computational load trade-off, with likely little additional insight gained with a larger dataset.

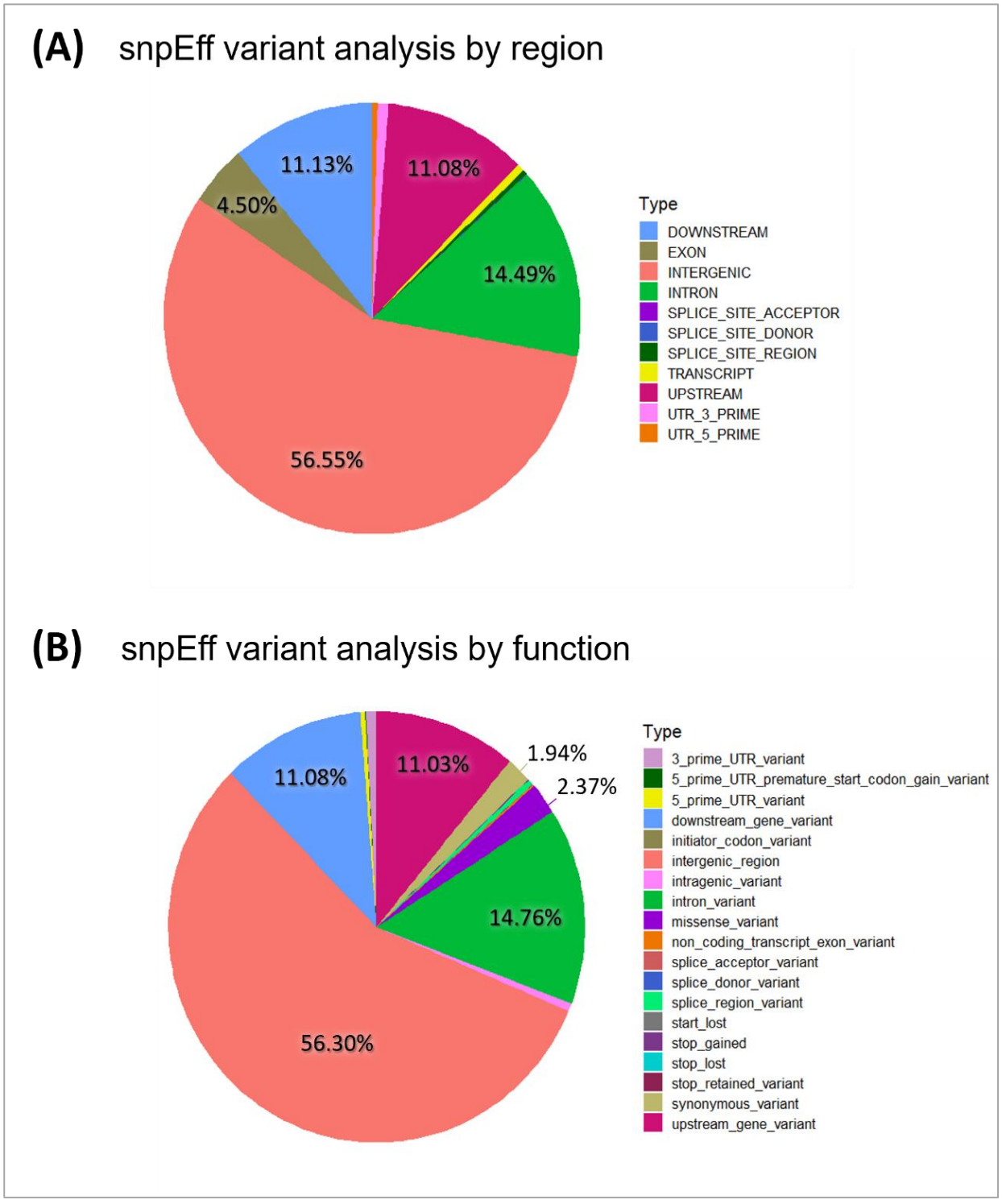


Figure 4.3 – Pie charts explaining the proportion of *S. asiatica* SNP variants by (A) location and by (B) putative function. Analysis was produced by snpEff using a combined dataset of 25 individuals against the *S. asiatica* reference genome (Yoshida et al., 2019).

For the genic SNP set, additional MAF filtering (MAF<=0.05) was performed with bcftools in order to remove rare, noisy variants which would not be significant in correlation with virulence. No additional MAF filtering was done on the intergenic dataset as these small changes would likely be involved in determining population structure. After filtering, a total of 405,047 genic SNPs remained. These genic and intergenic sets of SNPs formed then the basis for partial redundancy analysis (pRDA), as outlined in Figure 4.4, following the pipeline laid out by Capblancq and Forester et al. (2021) and following the tutorial available at https://github.com/Capblancq/RDA-landscape-genomics.

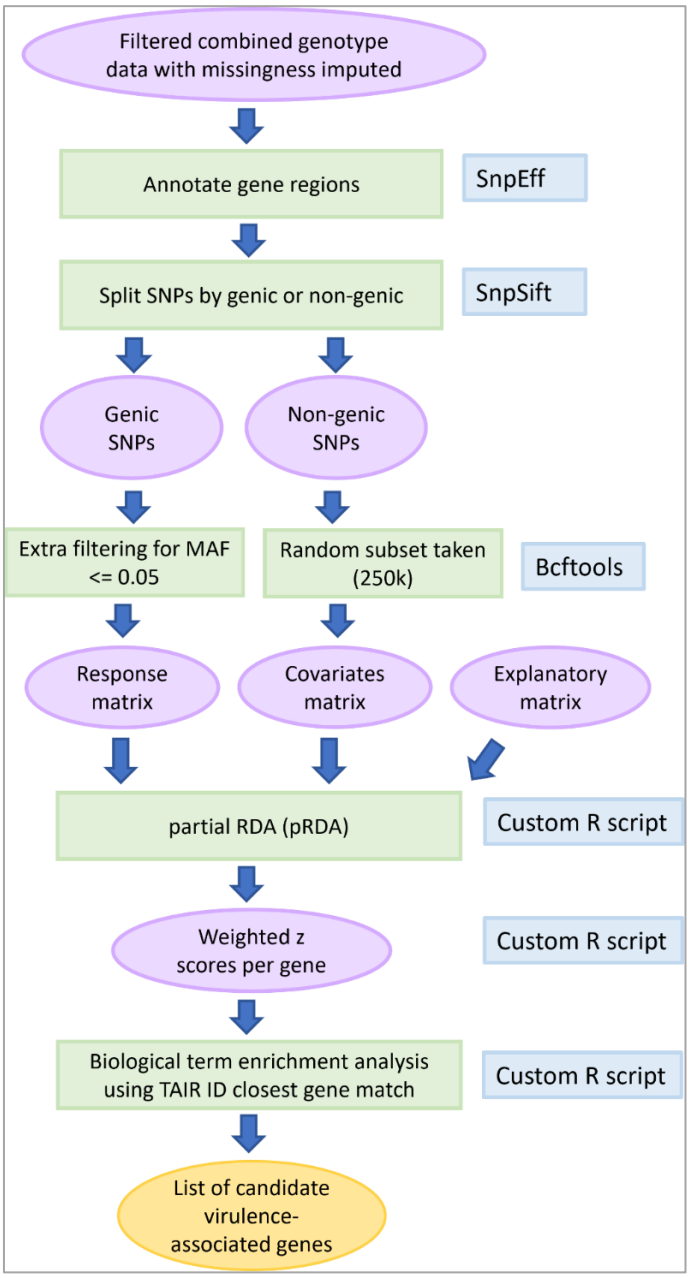
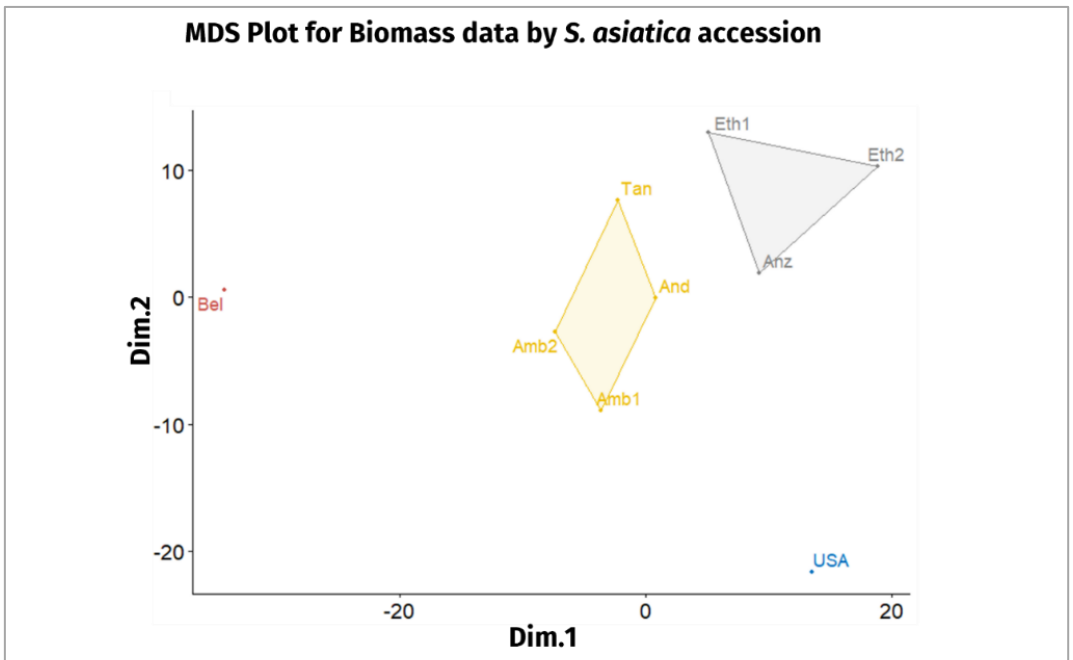


Figure 4.4 –Flowchart showing the bioinformatic pipeline following initial SNP discovery and filtering, for the detection of candidate virulence associated genes in *S. asiatica* using an integrated pRDA and WZA approach.

4.2.3 Building environmental and phenotypic datasets

Along with genotype data, virulence phenotype information was collated from chapter 2. In the former study, batches of *S. asiatica* seeds from the same accessions as described in Table 4.1 were used in virulence screens to assess the variation in virulence of Striga accessions across 9 different rice host varieties. The virulence measures of total number of Striga collected per host plant, cumulative length of Striga collected per host plant and total dry biomass of Striga per host plant were collected. Biomass was chosen as the most representative phenotypic measure of virulence for this current study. The virulence patterns of each Striga accession across all rice varieties was visualised through multidimensional scaling (MDS) plots using cmdscale in R v4.2.2 (R core team, 2022) package stats v.3.6.2.



	Amb1	Amb2	Anz	And	Bel	Eth2	Eth1	Tan	USA
Dim.1	-3.64	-7.36	9.21	0.84	-34.28	18.88	5.14	-2.29	13.50
Dim.2	-8.90	-2.74	1.88	-0.03	0.60	10.30	12.94	7.64	-21.68

Figure 4.5 - Multidimensional scaling (MDS) plot showing the variation in virulence patterns for each Striga accession used in this study, across a panel of different rice host varieties. Variation is explained across two dimensions and coordinates for each accession are given below the plot. Clustering of accessions into groups was performed by k means clustering, to partition accessions into the most similar groups.

Coordinates for each Striga accession datapoint across the first two dimensions and MDS plot are shown in Figure 4.5 where dimension 1 explained much of the variation between the mostly avirulent Striga accession from Belanitra, Madagascar and the other, more virulent accessions on our panel of hosts. Dimension 2, conversely, explained much of the difference between the USA accession and all other remaining accessions. The USA accession being derived from a collection of seeds which were sampled in an area of North Carolina, USA where Striga is non-native; this was also the accession used for the generation of the *S. asiatica* reference genome. Both Dimension 1 and Dimension 2 show meaningful variation in virulence across all *S. asiatica* accessions. These datapoint values from the MDS plot formed the explanatory variables matrix for the partial redundancy analysis (Figure 4.6).

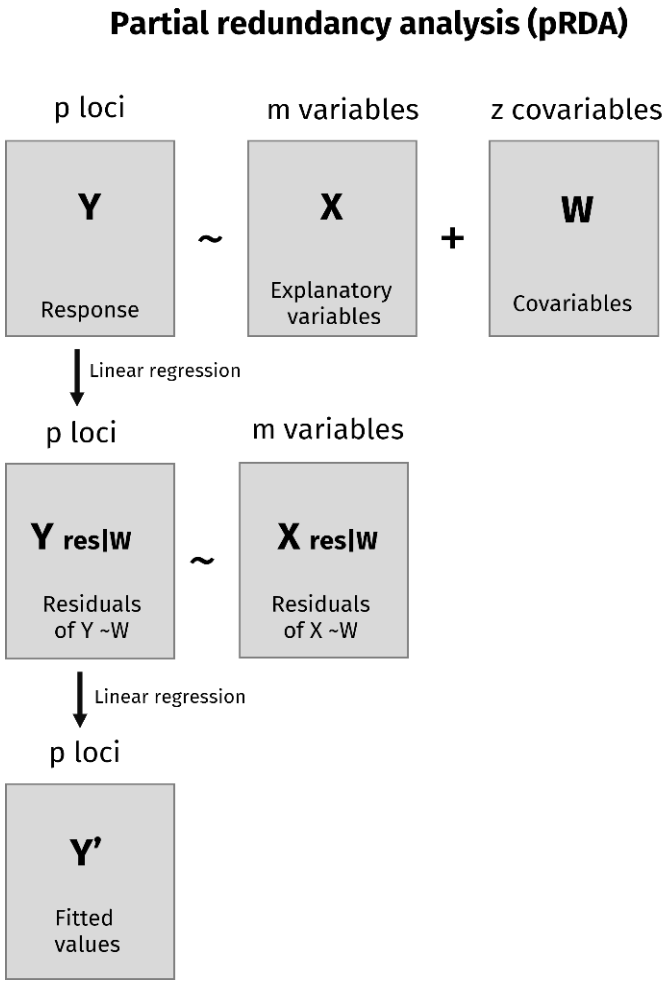


Figure 4.6 - A comparison of procedures for fitting a partial redundancy analysis (pRDA), adapted from Capblancq and Forester (2021). In a pRDA analysis, the Y genic loci are regressed against X explanatory virulence variables and using W covariables of latitude/longitude and PC axes of intergenic SNP loci, in order to account for population structure. From this analysis P values were calculated per locus to show whether relationships exist between virulence measures and genic loci, to facilitate the weighted-score analysis.

4.2.4 Genotype-environment associations with RDA

The aim of an RDA analysis is, using an extension of multiple linear regression, to test whether two variables from separate datasets behave independently. Benefits of RDA are the possibility of inclusion of multiple response variables as well as efficient analysis with large sets of SNPs (over 1 milllion). In addition, variance partitioning within the RDA can allow separation of neutral genetic structure from any virulence related genetic structure. In order to account for neutral population structure in this study, a principal component analysis (PCA) was performed on the subset of intergenic SNPs using R package VEGAN v2.6-4 (Oksanen et al., 2022). Scores from this PCA for the first few PC axes were tested for use as a proxy for neutral population structure. These scores were combined with latitude and longitude per Striga accession as covariates, along with virulence measures as explanatory variables.

Virulence measures and geographic location coordinates were duplicated for each individual within an accession. Variance partitioning was applied to different pRDA models to decompose the contribution of neutral population structure, geography and the virulence phenotype in explaining genic variation as well as to determine the number of neutral background PCs to include for the best explanation of background population structure. The proportion of variance explained (R^2) by each of the full, neutral, geography and virulence models was measured by a regression of response variables (genic loci) against the predictor variables: PCs of neutral intergenic loci, latitude & longitude and virulence phenotype. The models given below show how the predictor variables were split into explanatory and covariables (in brackets) in order to test the variance explained by each of the variable groups alone.

Full Model: Genic loci ~ Virulence phenotype + long. + lat. + Intergenic PCs

Virulence Model: Genic loci ~ Virulence phenotype + (long. + lat. + Intergenic PCs)

Pure structure Model: Genic loci ~ Intergenic PCs + (Virulence phenotype + long. + lat.)

Pure geog. Model: Genic loci ~ long. + lat. + (Virulence phenotype + Intergenic PCs)

Four PC axes were decided as the most suitable neutral background; this was determined by the lack of additional variance explained by the full model with addition of more than 4 PC axes.

Final model: Genic loci ~ Virulence phenotype + (long. + lat. + Intergenic PC1:4)

Following the procedure as outlined by Capblancq et al., (2018) which determines outlier loci from the most extreme values along a distribution of Mahalanobis distances, the distance between a locus position and the centre of an RDA space, using the 2 axes (K) of RDA1 and RDA2, P-values were assigned to loci. For this a redundancy analysis (RDA) was performed with R package VEGAN. The P-values were ascribed following

Bonferroni correction, to reduce the false-positive rate and an arbitrary p-value threshold was set to -log(p-value)>5 for 'outliers'. The practice of employing an arbitrary cut-off, typically around the top 10%, to evaluate statistical inference in redundancy analysis is widely observed (Gu et al., 2022). RDA biplot and Manhattan plot of loci were built using R package ggplot2 (Wickham, H., 2016), where the arbitrary p-value cut off was for demonstration purposes to be able to exhibit how those on the outer portion of the plots were theoretically most important in explaining variation in the explanatory variables, while those in the centre of the RDA space and bottom of the Manhattan plot were more likely to be loci without associations to virulence.

4.2.5 Implementation of weighted Z-score to gene windows

A window-based method for characterising outlier genes from these potentially important loci was adapted from Booker et al. (2023). Booker et al. identified that Weighted Z-score analysis, which averages scores across SNPs in set windows based on genes and characterises SNP informativeness based on allele frequency, outperformed or bettered the GEA, particularly in excluding a number of false positive loci. The aim for this analysis was to combine information from multiple, linked SNP sites and determine whether particular regions show associations with variation in virulence in some way. To create windows from SNP loci, P-values for each locus according to the RDA analysis were first converted to Z-scores by using the base R stats function qnorm.

Functional gene annotation of the *S. asiatica* genome (Yoshida et al, 2019) was downloaded from the datadryad repository https://datadryad.org/stash/dataset/doi:10.5061/dryad.53t3574 and merged with the genotype data and allele frequencies for SNPs in all 'genic regions' using the Bedtools (Quinlan, 2010) intersectBed command, before also merging with the file containing P-values and Z-scores for genic loci, in order to find and extract which SNPs from the outlier analysis intersected with gene coordinates. The information generated about genetic correlations with virulence phenotype and biallelic SNP sites was combined into a single weighted-Z score for a focal region: in this case, the genomic analysis windows were based on genes identified from the functional annotation provided by Yoshida et al. The formula for Stouffer's weighted-z score is detailed below where the weighted Z is calculated for each genomic region k for the n number of SNPs in a gene (Booker et al., 2023).

$$Z_{W,k} = \frac{\sum_{i=1}^{n} \overline{p}_i \overline{q}_i \overline{z}_i}{\sqrt{\sum_{i=1}^{n} (\overline{p}_i \overline{q}_i)^2}},$$

And where,

p_i = mean allele frequency across the population.

q_i = 1- p_i.

z_i = standard normal deviate from the one-sided p-value for each SNP.

A custom R script was used to generate these weighted z-scores for a total of 25,085 *S. asiatica* genic regions before genes were ranked by WZA score and the top 5% were considered as candidates. Weighted z-scores were also generated from the two RDA axis loadings to look at differences in contributions of particular gene functions; z scores for WZA were produced from the axis loadings for each locus using z = (loading - mean)/stdev.

4.2.6 Annotation and enrichment analyses

A secretome prediction excluding those proteins with a transmembrane spanning helix for the *S. asiatica* genome using SignalP version 5.0, was kindly provided by Dr James Bradley (University of Toronto) along with closest matching hits for Arabidopsis genes, as a widely annotated genome. Details of this secretome prediction pipeline in *S. hermonthica* are published in Qiu et al. (2022). The prediction of whether genes were in the secretome was annotated to the genic windows for weighted Z-scores. R package ggplot was used to show the distribution histogram of WZA scores, for genes in the secretome compared to WZA-scores overall. A subset of the top 5% of genes (n=1250) was made according to WZA-scores and these were sorted according to presence in the secretome then separate files were made for secreted and non-secreted gene products. A chi squared test was performed in R to look for any overrepresentation of the secretome within this top 5% of genes

Functional enrichment analyses of GO terms and Pfam domains for *S. asiatica* putative candidate virulence genes (both secretome and non-secretome) were carried out with online analysis tool, DAVID (https://david.ncifcrf.gov/) (Huang et al., 2009; Sherman et al., 2021) where the ID's of Arabidopsis analogues of the top 5% of genes by WZA score were used as the input, selecting 'TAIR_ID' as identifier, using secretome and non-secretome genes. Of the 1254 top 5% of genes, 794 also had an Arabidopsis analogue ID and of those 794 only 55 were also in the secretome; due to these low overall numbers of secreted gene-products, non-secretome genes were also investigated.

Two different backgrounds were used for enrichment analysis of the 55 top genes in the secretome: (1) the ID's for Arabidopsis analogues for all genes with a WZA score associated (including the remaining 95%); of these 25,085 genes, 15192 had an Arabidopsis analogue ID which could be used, and (2) all genes with a WZA associated and also in the secretome; of these genes, only 977 had an Arabidopsis analogue ID which could be used. All other DAVID steps were set as default. The output from DAVID provided fold enrichment for different biological process categories such as GO-term and INTERPRO with P-values for each. False discovery rate (FDR) correction was not applied. In addition to these analyses, enrichment analysis was performed on the top 5% of genes by WZA score which were not in the secretome (n=744) against a background of all genes with an Arabidopsis analogue ID and WZA score (n=15,192).

Functional enrichment analysis was also performed on the 2 RDA axes independently, to assess the different influences contributing to variation in virulence, again using DAVID. Of the 1254 top 5% of genes associated with RDA1, 809 also had an Arabidopsis analogue ID and of the 1254 top 5% of genes associated with RDA2, 752 also had an Arabidopsis analogue ID. The same background 15,192 Arabidopsis genes were used for both analyses. Plots for all enrichment analyses were made in R using package ggplot2.

4.3 Results

4.3.1 A genotype-environment association study of *S. asiatica* virulence using partial redundancy analysis (pRDA)

In order to run an RDA analysis of genotype and virulence with the most appropriate data, variance partitioning must first be enacted to look at the influence of any available geographic location data and in particular, any demographic history which could be elucidated with neutral genetic structure. This involved first, producing the most appropriate proxy for a neutral population structure, in this study the combination of several PC axes for 'intergenic' loci. A screeplot showing the amount of variation explained by the different principal component axes for the intergenic SNPs is shown in Figure 4.7.

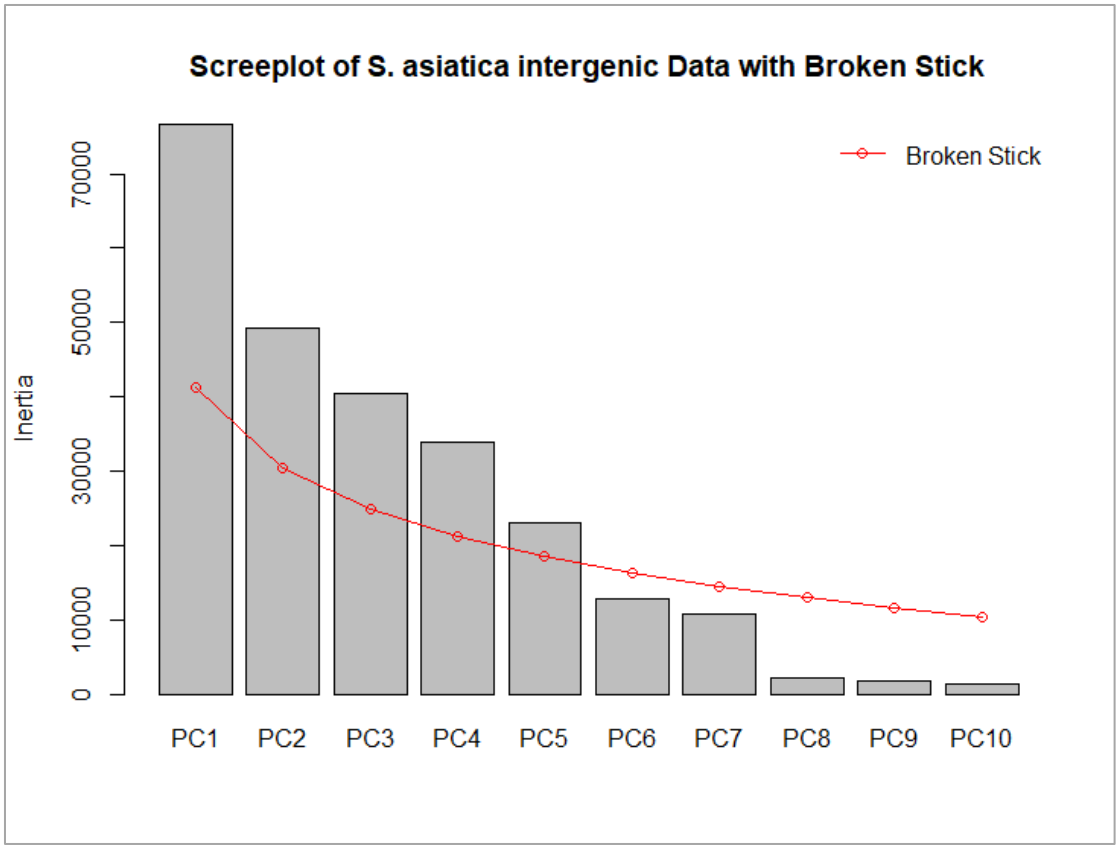


Figure 4.7 – Screeplot of the first 10 PC axes of the PCA performed on a subset of ~250,000 intergenic randomly selected *S. asiatica* loci, following SnpEff classification. The first four PCs from this plot were chosen to represent the neutral genetic background for RDA analysis of genotype-environmental correlations. Broken stick line shows the expected frequency distribution of data under the assumption of random and independent events.

Comparison of R^2 scores, emphasizing the model's goodness of fit, for the full RDA models when containing different numbers of PC axes showed no additional benefit to including more than four PC axes as the neutral background; the model R^2 being 0.101 with 4 PCs and 0.096 with 5 PCs in the model. Following this selection of a neutral background for population structure, the different models were able to be tested for each variable's contribution to explain variance.

Overall, the full model containing genic loci, virulence phenotype data, latitude and longitude and neutral population structure was found to explain nearly 95% of the variation (Table 4.2). The majority of variation was explained by the neutral genetic structure while geography (latitude and longitude) and virulence measures each contributed to roughly 7% of the variation in virulence seen. The final model shown in section 4.2.3 and comparison of all models can be seen in Table 4.2. The proportion of confounded explainable virulence, was 39%, suggesting a degree of collinearity amongst variables and therefore effects that could not be attributed to any one component.

Table 4.2 – Table showing the influence of virulence, neutral genetic structure and geography on genetic variation decomposed with pRDA (partial redundancy analysis). The proportion of explainable variance represents the total constrained variation explained by the full model.

Partial RDA models	Inertia (variance)	R^2	p (>F)	Proportion of explainable variance
Full model Genic loci ~ Virulence + long+lat + neutral PCs	231237	0.945	0.001	1.000
virulence model Genic loci ~ Virulence + (long+lat + neutral PCs)	16243	0.066	0.001	0.070
neutral structure model Genic loci ~ neutral PCs + (Virulence + long+lat)	106663	0.436	0.001	0.461
geography model Genic loci ~ long+lat+(Virulence + neutral PCs)	18078	0.074	0.001	0.078
Confounded virulence/neutral/geography	90253			0.390
Total unexplained	13495			
Total inertia	244732			

Correlation between the effects of the different explanatory variables and covariables was visualised by graphic display using a correlogram matrix (Figure 4.8). The plot graphically showed the collinearity between variables and suggested the highest correlation with latitude and longitude, but very little collinearity between the two most important explanatory variables of Dim.1 and Dim.2, representing the virulence measures, thus both measures were retained. However, eigenvalues of the constrained axes seen in Figure 4.9 suggested a much greater proportion of variance on RDA1 than on RDA2. The correlogram matrix identified many stronger correlations between Dim.2 and the covariables, than for Dim.1.

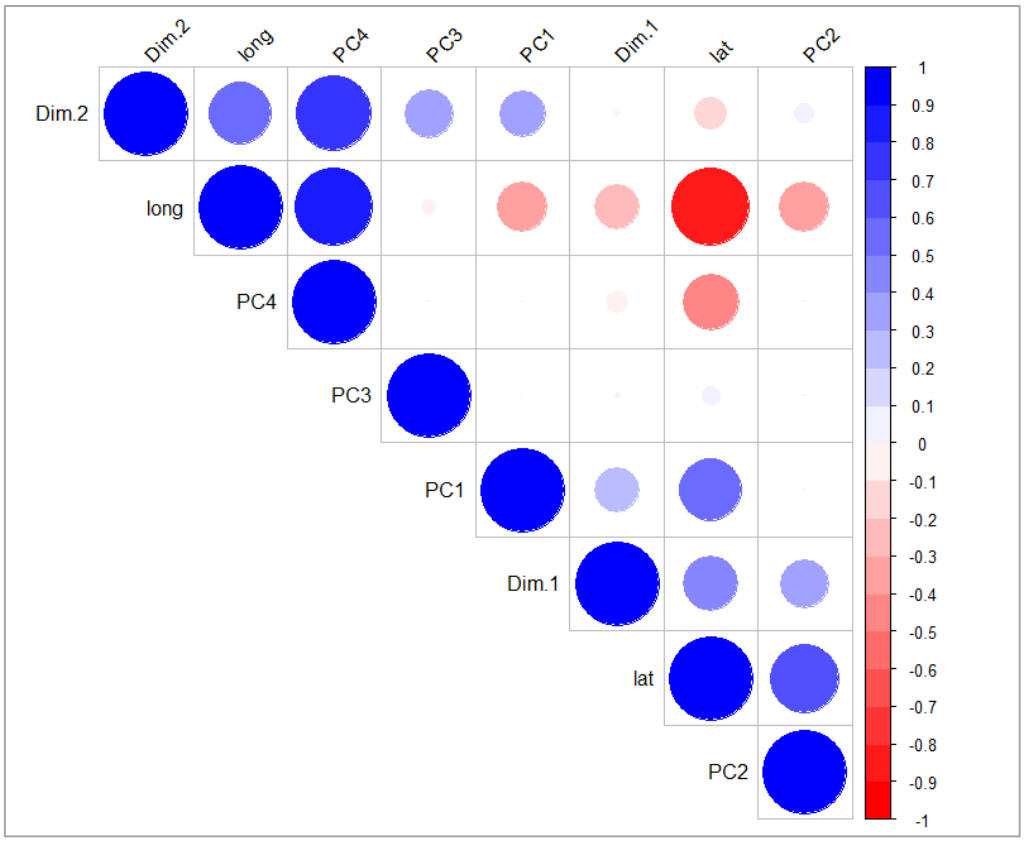


Figure 4.8 – Correlogram matrix showing a graphical display of the degree of correlation among the different variables of Dim.1 and Dim.2 (virulence measures from MDS plot), PC1, PC2, PC3 and PC4 (principal coordinates from the neutral background and lat and long geographic coordinates. Among the majority of these variables and covariables there is a relatively low level of collinearity. The two explanatory variables of Dim.1 and Dim.2 share different levels and patterns of correlation across the covariables. Colour and size are proportional to the correlation coefficients.

A genome scan of *S. asiatica* loci was performed using partial redundancy analysis and p-values were assigned to each of the 405,047 genic SNPs to indicate the level of influence on variation in virulence and highlight those SNPs which were more strongly correlated to be virulence than expected, having allowed for background population (neutral SNPs) and geographic structure (latitude and longitude). P-values for each locus were visualised with a Manhattan plot and RDA space plot, where RDA1 is correlated with Dim.1 and Dim.2 and RDA2 is only influenced by Dim.1 (Figure 4.9).

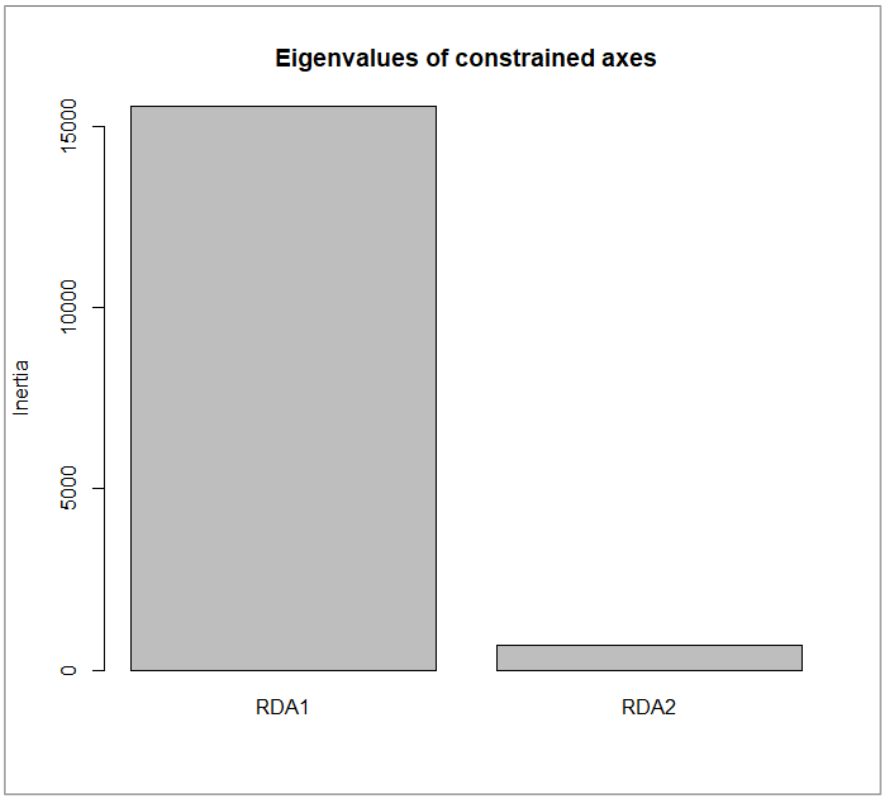


Figure 4.9 – Histogram screeplot showing the inertia (variance) across the two constrained axes of RDA1 and RDA2 from an RDA analysis, explained by the two explanatory variables of virulence, Dim.1 and Dim.2. Much more of the variation in virulence of *S. asiatica* appears to be explained by RDA1 than RDA2.

While both figures highlighted a dense central cluster of less significant p-values, the Manhattan plot, which plots loci in order along scaffolds, showed very large clusters of low -log10(p) values indicating that those loci were inherited together. However, the relatively low significance -log10(p) values suggested that these would not be of great influence in determining variation in virulence and were just an artifact from selfing. For these loci it is more likely that they represent any unique alleles in an accession which have been implicated in association with virulence. This could be due to large differences in virulence seen at the accession level. To circumvent this overestimation of loci, it is possible to set a higher cut off to show only the strongest outliers. Several loci with exceptionally high p-values were identified from this analysis and most interestingly, a total of 18 loci out of the top 50 SNPs by P-values were in a similar position all along scaffold219 (Fig. 4.10A and Fig. 4.11). Six genes on this scaffold219 were associated with these SNPs and descriptions of these genes are given in Table 4.3. In addition to focusing on the strongest outliers, the scores for SNPs were averaged across genes using the WZA approach which reduced the noise of many of the spurious associations to virulence. The RDA space plot (Fig. 4.10B) also highlighted that variation in virulence associated with both Dim.1 and Dim.2 was seen across RDA1, while comparatively, RDA2 was mainly influenced by the variation in virulence seen across Dim.1 (Fig. 4.5).

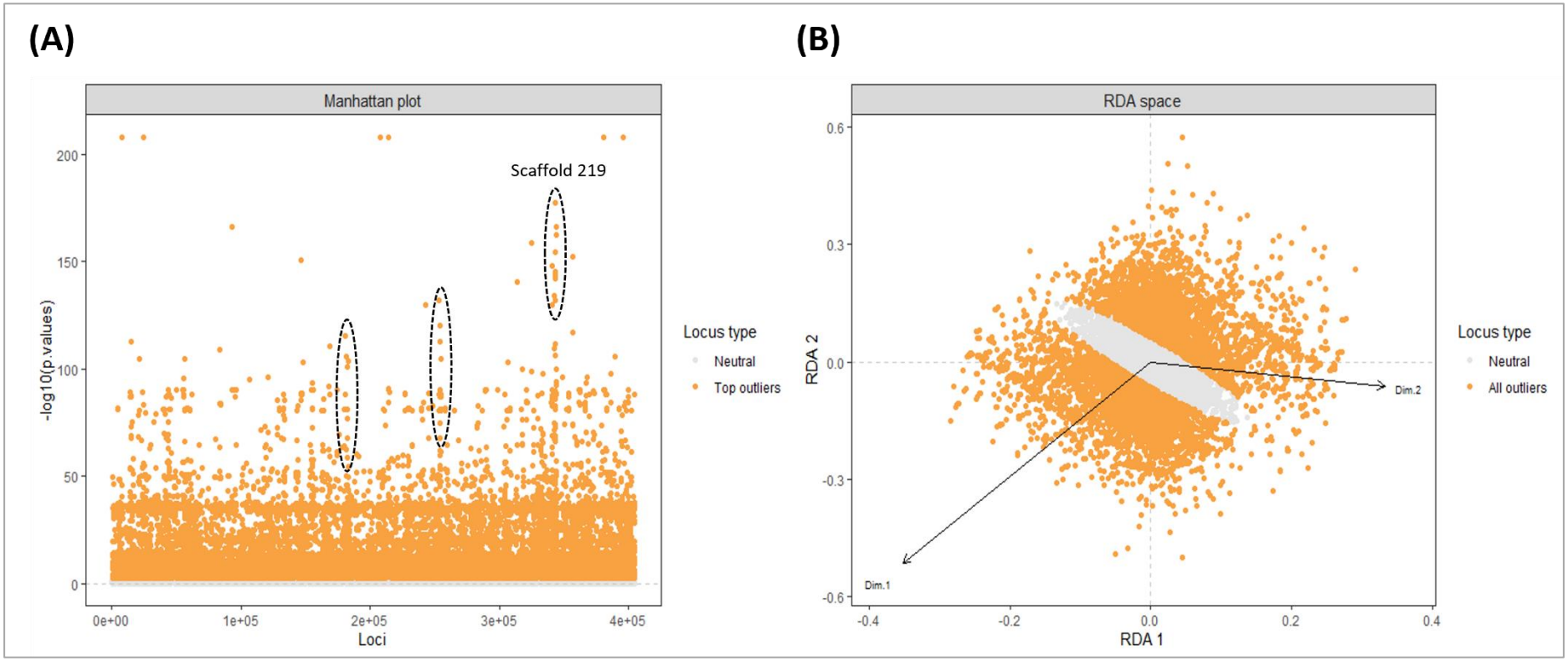


Figure 4.10 - Results of the genotype-environment association using partial redundancy analysis (pRDA) to showing the linkage between *S. asiatica* loci and environmental variables. (A) shows a Manhattan plot of the data demonstrating the distribution of -log10(p-values) for each locus ordered randomly. Despite the generally low-level linkage, several very high y-axis peaks in p values are seen for loci in the same scaffold location and have been circled, one such genic region is scaffold 219. (B) shows this data in an RDA space, with an ellipse of neutral loci in the centre of the plot and some outlier loci on the edges. The loci more associated with directions of the Dim.1 and Dim.2 arrows may explain variation in virulence for those particular virulence measures.

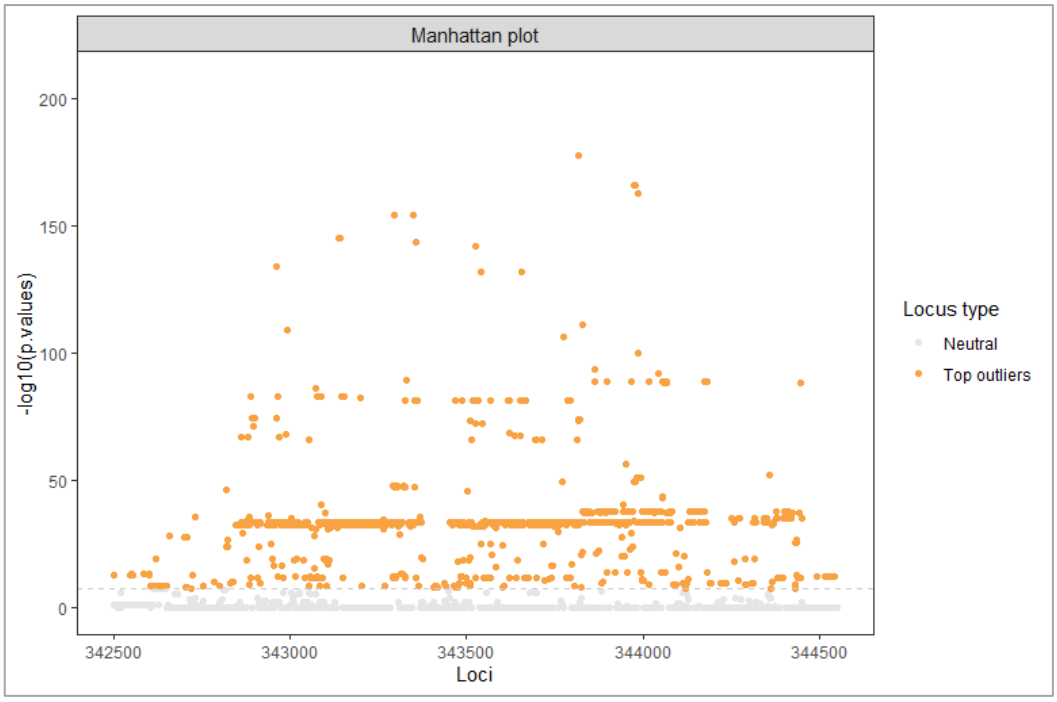


Figure 4.11 - Manhattan plot showing the distribution of -log10(p-values) for each locus ordered across contigs of the *S. asiatica* genome focusing on the region of scaffold219 which showed a large number of outlier loci with scores over 50.

Table 4.3 – Pfam descriptions of the *S. asiatica* genes from scaffold 219 with positions of several SNP loci with exceptionally high P-values according to pRDA analysis. These loci showed a strong peak on the pRDA Manhattan plot suggesting they have a strong association with variation in virulence.

Gene ID	P-value	Pfam domain description
SGA_v2.0_scaffold219G42539	4.25E-155	WW domain; FF domain
SGA_v2.0_scaffold219G42541	4.25E-155 1.31E-144	Amino acid permease 6; Transmembrane amino acid transporter protein
SGA_v2.0_scaffold219G42523	5.72E-146	Transcription initiation factor IIA: gamma subunit: helical domain
SGA_v2.0_scaffold219G42541	4.25E-155 1.31E-144	Amino acid permease 6; Transmembrane amino acid transporter protein
SGA_v2.0_scaffold219G42548	5.53E-143	NAD(P)-binding Rossmann-like domain; Squalene epoxidase
SGA_v2.0_scaffold219G42557	1.36E-132 1.36E-132	Pectinesterase

4.3.2 Weighted Z-score (WZA) analysis identified significant over-representation in genes within the *S. asiatica* secretome compared to the whole genome.

To better understand the influences of genotype on variation in virulence, a weighted Z analysis (WZA) was applied to the data to combine the information from multiple SNP sites into genic windows; this approach being particularly useful for when sample population sizes are relatively small and statistically noisy (Booker et al., 2023). SNP loci were combined into windows using Z-scores converted from the P-values generated by the pRDA. Weighted z-scores were then applied to the 25,085 gene-based windows calculated with Stouffer's weighted-z score given in section 4.2.4. *S. asiatica* gene annotations, including details of occurrence in the secretome were merged with WZA score data. Distribution of WZA scores across all genic windows was plotted using ggplot2 with a mean WZA score of 7.86 and WZA scores for the top 250 genes (upper 1% of genes) ranging between 180.16 and 77.43 while scores for the 1250 top 5% of genes ranged from and WZA scores for the top 250 genes (upper 1% of genes) ranging between 136.74 and 48.34. Of the 25,085 genes, 1240 were identified as being in the secretome (Fig. 4.12 and Fig. 4.13). Due to the expected enrichment of the secretome for virulence factors, it was tested whether the secreted genes were overrepresented in the top 5% of all genes a Chi-square test was performed (Supplementary Figure 4.1).

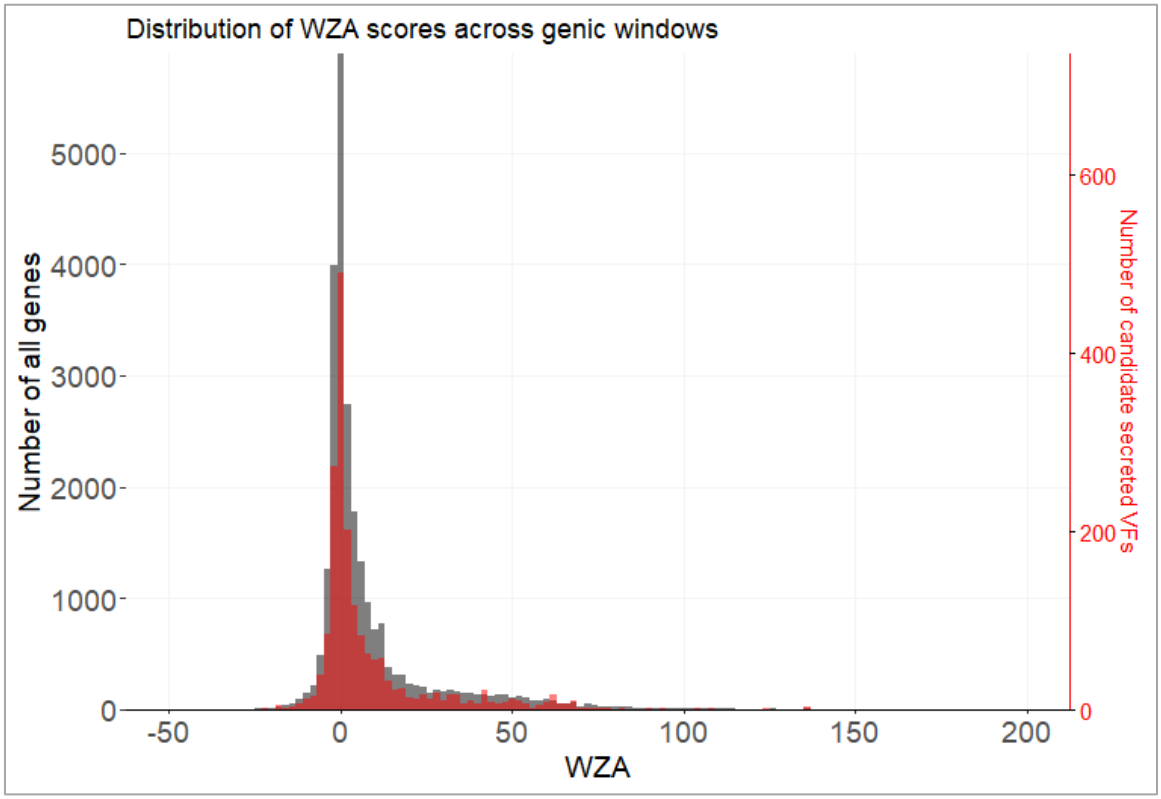


Figure 4.12 – Distribution plot of WZA scores for *S. asiatica* genes with higher WZA scores indicating a greater influence of that gene on variation in virulence, based on P-values generated from pRDA. Plots for all 25,085 genes (illustrated as black bars) are overlaid with the 1,240 secretome only genes (illustrated as red bars). The left and right y-axes show different scales for all genes and secreted genes, respectively.

This found a significant relationship between being in the top 5% and being in the secretome, X^2 (1, N=25085) = 4475.4, p=2.2e-16. A Chi-square test for overrepresentation of secretome genes was also repeated for the top 1% of all genes which found no significant relationship, X^2 (1, N=25085) = 0.70233, p=0.402. The first 100 genes by WZA score are given in Supplementary table 4.1.

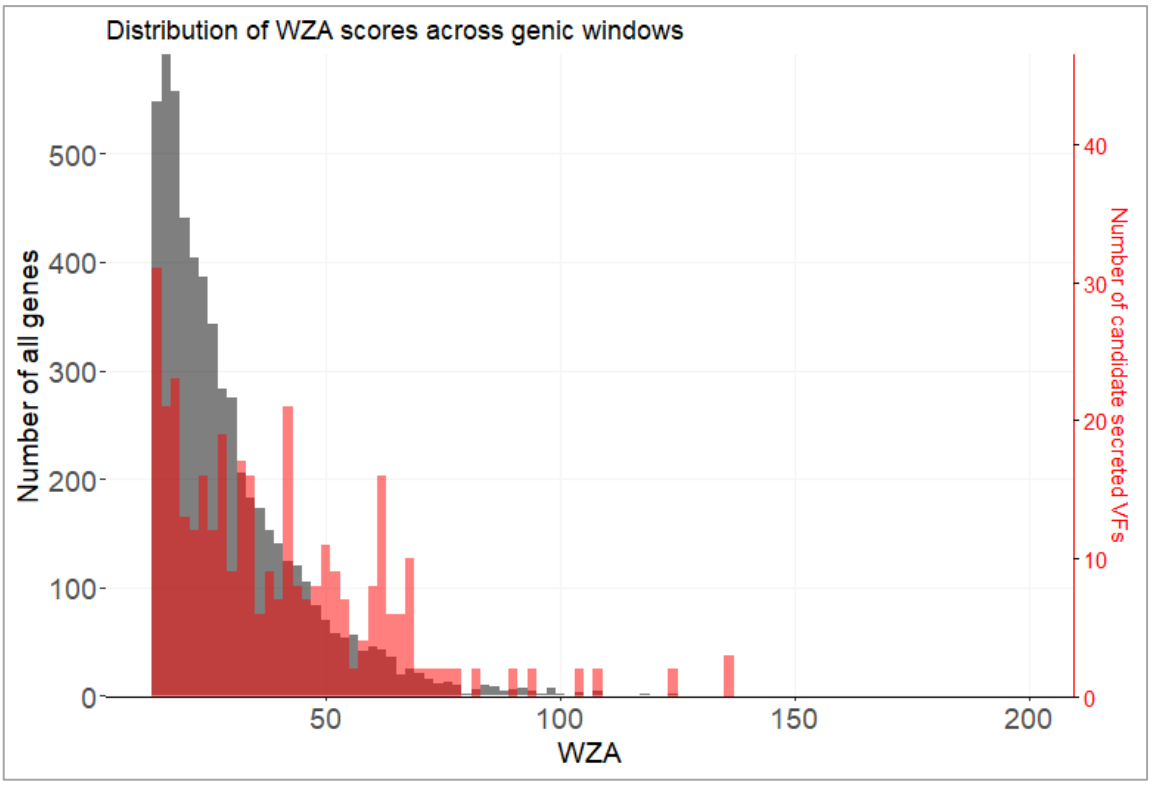


Figure 4.13 – Modified view of Figure 4.11, showing the distribution plot for WZA scores over 50 for *S. asiatica* genes. Higher WZA scores indicating a greater influence of that gene on variation in virulence, based on P-values generated from pRDA. Plots for all genes (illustrated as black bars) are overlaid with the secretome only genes (illustrated as red bars). The left and right y-axes show different scales for all genes and secreted genes, respectively.

4.3.3 The two explanatory RDA axes contain different genes associated with virulence.

To identify whether genes putatively associated with virulence are present across both RDA1 and RDA2, (despite the low variance explained by RDA2 as seen in Figure 4.9) or whether contributions to the different axes can be distinguished. WZA was performed on RDA1 and RDA2 independently using z-scores generated from RDA space coordinates. The 1250 genes in the top 5% by WZA score were filtered based on presence in the secretome and secreted gene products were compared between the two RDA axes to see any differences in gene function (Figure 4.14). Among the 5 highest scoring secretome genes (out of 72 genes) associated with RDA1 were a number of enzymes such as pectin acetylesterase, cytochrome P450, trehalase, peroxidase and alpha galactosidase and for RDA2 (out of 60 genes) were glycosyl hydrolase, peptidase, protein phosphatase and peroxidase.

(A) Across all genes

GeneID	Pfam description	A.thaliana description	GO terms
SGA_v2.0_scaffold112G31570	Pectinacetylesterase	Pectinacetylesterase family protein	GO:0016787
SGA_v2.0_scaffold3G01798	Glycosyl hydrolases family 31; Galactose mutarotase-like	Glycosyl hydrolases family 31 protein	GO:0004553, GO:0005975
SGA_v2.0_scaffold59G20672	Ribonuclease T2 family; 3-dehydroquinate synthase II	No_hit	GO:0003723, GO:0003856, GO:0009073, GO:0016491, GO:0033897, GO:0055114
SGA_v2.0_scaffold10G05336	Pectinacetylesterase; Pectinacetylesterase	Pectinacetylesterase family protein	GO:0016787
SGA_v2.0_scaffold337G49189	Glycosyl hydrolase family 10; Carbohydrate binding domain	Glycosyl hydrolase family 10 protein	GO:0004553, GO:0005975, GO:0016798

(B) Genes from RDA1

GeneID	Pfam description	A.thaliana description	GO terms
SGA_v2.0_scaffold10G05336	Pectinacetylesterase; Pectinacetylesterase	Pectinacetylesterase family protein	GO:0016787
SGA_v2.0_scaffold13G06642	Cytochrome P450; Cytochrome P450; Cytochrome P450	cytochrome P450; family 83; subfamily B; polypeptide 1	GO:0005506,GO:0016705,GO:0020037, GO:0055114
SGA_v2.0_scaffold198G41058	Trehalase	trehalase 1	GO:0004555,GO:0005991
SGA_v2.0_scaffold292G47377	Peroxidase	Peroxidase superfamily protein	GO:0004601,GO:0006979,GO:0020037, GO:0055114
SGA_v2.0_scaffold100G29574	Alpha galactosidase C-terminal beta sandwich domain; Alpha galactosidase A; Alpha galactosidase A	Melibiase family protein	GO:0004553,GO:0005975

(C) Genes from RDA2

GeneID	Pfam description	A.thaliana description	GO terms
SGA_v2.0_scaffold19G08766	Glycosyl hydrolases family 35; Beta-sandwich domain in beta galactosidase; Galactose binding lectin domain	beta galactosidase 9	GO:0030246
SGA_v2.0_scaffold145G36077	Peptidase inhibitor I9; PA domain; Subtilase family; Fibronectin type-III domain	Subtilisin-like serine endopeptidase family protein	GO:0004252,GO:0006508
SGA_v2.0_scaffold19G08720	X8 domain; Glycosyl hydrolases family 17	O-Glycosyl hydrolases family 17 protein	GO:0004553,GO:0005975
SGA_v2.0_scaffold1G00532	Protein phosphatase 2C; Protein kinase domain	Protein phosphatase 2C family protein	GO:0003824,GO:0004672,GO:0005524, GO:0006468
SGA_v2.0_scaffold60G20919	Peroxidase	Peroxidase superfamily protein	GO:0004601,GO:0006979,GO:0020037, GO:0055114

Figure 4.14 – A comparison of the number of genes in the top 5% of candidates for association with variation in virulence according to WZA scores, and the number of those which are in the secretome. Tables show the details of the top 5 (by WZA score) *S. asiatica* genes, also in the secretome which have GO terms associated.

(Figure 4.14 continued) (A) shows which are the most significant genes overall, (B) shows the top secreted genes associated with explanatory virulence variable RDA1 from genotype-environment association analysis and (C) shows the top secreted genes associated with explanatory virulence variable RDA2 from genotype-environment association analysis.

Table 4.4 – Descriptions of the 10 secreted genes found in the top 5% of virulence-associated *S. asiatica* genes candidates, which were present in both RDA1 and RDA2 constrained genotype-environment analysis, for which Pfam or *A. thaliana* descriptions were available. An 11th gene, SGA_v2.0_scaffold219G42540, did not have an annotation.

GeneID	Pfam description	*A.thaliana* description	GO terms
SGA_v2.0_scaffold337G49183	Cysteine-rich secretory protein family	CAP (Cysteine-rich secretory proteins; Antigen 5; and Pathogenesis-related 1 protein) superfamily protein	NA
SGA_v2.0_scaffold145G36077	Peptidase inhibitor I9; PA domain; Subtilase family; Fibronectin type-III domain	Subtilisin-like serine endopeptidase family protein	GO:0004252, GO:0006508
SGA_v2.0_scaffold79G25525	Probable lipid transfer	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	NA
SGA_v2.0_scaffold163G37997	Leucine rich repeat; Leucine rich repeat; Protein kinase domain; Leucine rich repeat N-terminal domain	Leucine-rich receptor-like protein kinase family protein	GO:0004672, GO:0005515, GO:0005524, GO:0006468
SGA_v2.0_scaffold145G36043	Probable lipid transfer	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	NA
SGA_v2.0_scaffold121G33051	Pectinacetylesterase	Pectinacetylesterase family protein	GO:0016787
SGA_v2.0_scaffold49G18164	WRKY DNA -binding domain; WRKY DNA -binding domain	WRKY DNA-binding protein 2	GO:0003700, GO:0006355, GO:0043565
SGA_v2.0_scaffold13G06566	PAC2 family	clast3-related	NA
SGA_v2.0_scaffold54G19682	Zinc carboxypeptidase; Zinc carboxypeptidase	carboxypeptidase D; putative	GO:0004181, GO:0006508, GO:0008270
SGA_v2.0_scaffold478G52225	Glycerophosphoryl diester phosphodiesterase family	SHV3-like 1	GO:0006629,GO:0008081

Overall, 11 of these secreted genes overlapping in the top 5% were identified in both RDA1 and RDA2 gene sets. A table of these genes with Pfam and *A. thaliana* descriptions for ten of these are shown in Table 4.4.

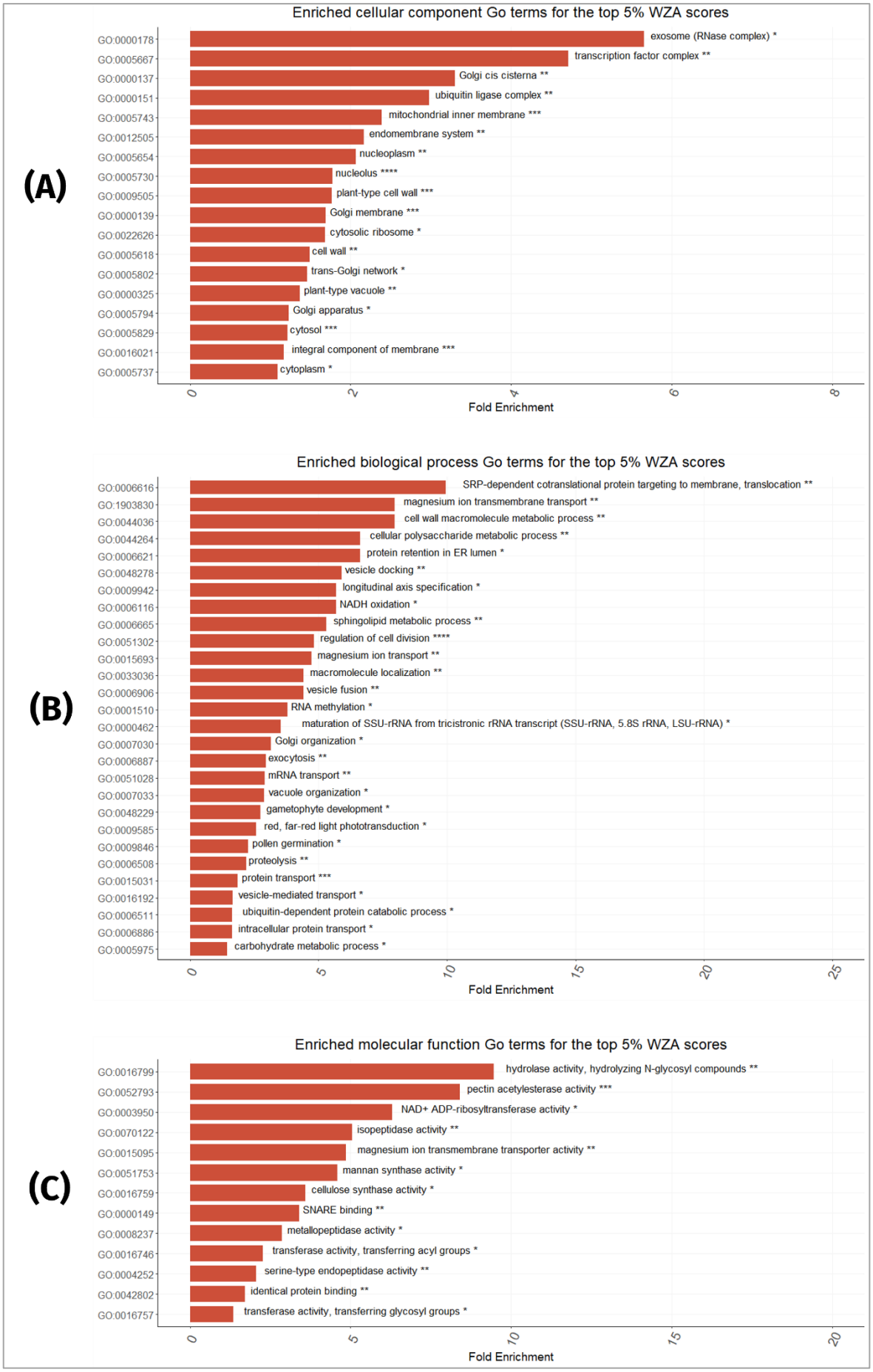
4.3.4 Enriched GO-terms in the top 5% of genes associated with variation in virulence across the *S. asiatica* genome.

Enrichment analysis was performed using online functional annotation tool, DAVID to measure the enrichment of GO-terms associated with the top 5% of genes by WZA score, regardless of whether genes were in the secretome or not. This list of genes was compared to a background of all genes in the *S. asiatica* genome (Fig. 4.15).

These enriched terms are shown in Figure 4.15, where the cellular component GO-terms were most related to terms, such as exosome (GO:0000178) and transcription factor complex (GO:0005667). The highest biological process enrichment was for co-translational protein translocation (GO:0006616) and highest molecular function go term enrichment was for hydrolase activity (GO:0016799). INTERPRO terms for the top genes highlighted strong enrichment in pectin acetylesterase (IPRO004963).

Another analysis was performed to separate out functions associated with secretome and non-secretome genes. First, fold enrichment of functional terms associated with the top 5% of genes by WZA score which were also in the secretome was assessed against a background of all secretome gene (Fig 4.16). Secretome genes were used as background because it was identified that there was a significant difference in expected representation of secretome genes in the top 5% of all genes. Therefore, several terms were already expected to be overrepresented in the top 5% of secretome genes. The enrichment analysis performed using all secretome genes as a background showed, as expected, many fewer enriched terms, but all were statistically significant with up to 10-fold enrichment for both pectin acetylesterase activity (GO:0052793) and Leucine-rich repeat activity (IPR003591) (Fig. 4.16).

Enrichment analysis of biological functions for non-secretome genes potentially associated with virulence (Figure 4.17) found many genes with enzymatic products. The highest biological process enrichment was for SRP-dependent co-translational protein targeting to membrane (GO:0006616) and highest molecular function go term enrichment was for hydrolase activity (GO:0016799). A comparison of cysteine content and protein length for genes from the top 5% of WZA scores between secreted and non-secreted gene products is shown in Table 4.5.


(A)
Enriched cellular component Go terms for the top 5% WZA scores
GO:0000178 exosome (RNase complex) *
GO:0005667 transcription factor complex **
GO:0000137 Golgi cis cisterna **
GO:0000151 ubiquitin ligase complex **
GO:0005743 mitochondrial inner membrane ***
GO:0012505 endomembrane system **
GO:0005654 nucleoplasm **
GO:0005730 nucleolus ****
GO:0009505 plant-type cell wall ***
GO:0000139 Golgi membrane ***
GO:0022626 cytosolic ribosome *
GO:0005618 cell wall **
GO:0005802 trans-Golgi network *
GO:0000325 plant-type vacuole **
GO:0005794 Golgi apparatus *
GO:0005829 cytosol ***
GO:0016021 integral component of membrane ***
GO:0005737 cytoplasm *
0 2 4 6 8
Fold Enrichment
(B)
Enriched biological process Go terms for the top 5% WZA scores
GO:0006616 SRP-dependent cotranslational protein targeting to membrane, translocation **
GO:1903830 magnesium ion transmembrane transport **
GO:0044036 cell wall macromolecule metabolic process **
GO:0044264 cellular polysaccharide metabolic process **
GO:0006621 protein retention in ER lumen *
GO:0048278 vesicle docking **
GO:0009942 longitudinal axis specification *
GO:0006116 NADH oxidation *
GO:0006665 sphingolipid metabolic process **
GO:0051302 regulation of cell division ****
GO:0015693 magnesium ion transport **
GO:0033036 macromolecule localization **
GO:0006906 vesicle fusion **
GO:0001510 RNA methylation *
GO:0000462 maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) *
GO:0007030 Golgi organization *
GO:0006887 exocytosis **
GO:0051028 mRNA transport **
GO:0007033 vacuole organization *
GO:0048229 gametophyte development *
GO:0009585 red, far-red light phototransduction *
GO:0009846 pollen germination *
GO:0006508 proteolysis **
GO:0015031 protein transport ***
GO:0016192 vesicle-mediated transport *
GO:0006511 ubiquitin-dependent protein catabolic process *
GO:0006886 intracellular protein transport *
GO:0005975 carbohydrate metabolic process *
0 5 10 15 20 25
Fold Enrichment
(C)
Enriched molecular function Go terms for the top 5% WZA scores
GO:0016799 hydrolase activity, hydrolyzing N-glycosyl compounds **
GO:0052793 pectin acetylesterase activity ***
GO:0003950 NAD+ ADP-ribosyltransferase activity *
GO:0070122 isopeptidase activity **
GO:0015095 magnesium ion transmembrane transporter activity **
GO:0051753 mannan synthase activity *
GO:0016759 cellulose synthase activity *
GO:0000149 SNARE binding **
GO:0008237 metallopeptidase activity *
GO:0016746 transferase activity, transferring acyl groups *
GO:0004252 serine-type endopeptidase activity **
GO:0042802 identical protein binding **
GO:0016757 transferase activity, transferring glycosyl groups *
0 5 10 15 20
Fold Enrichment

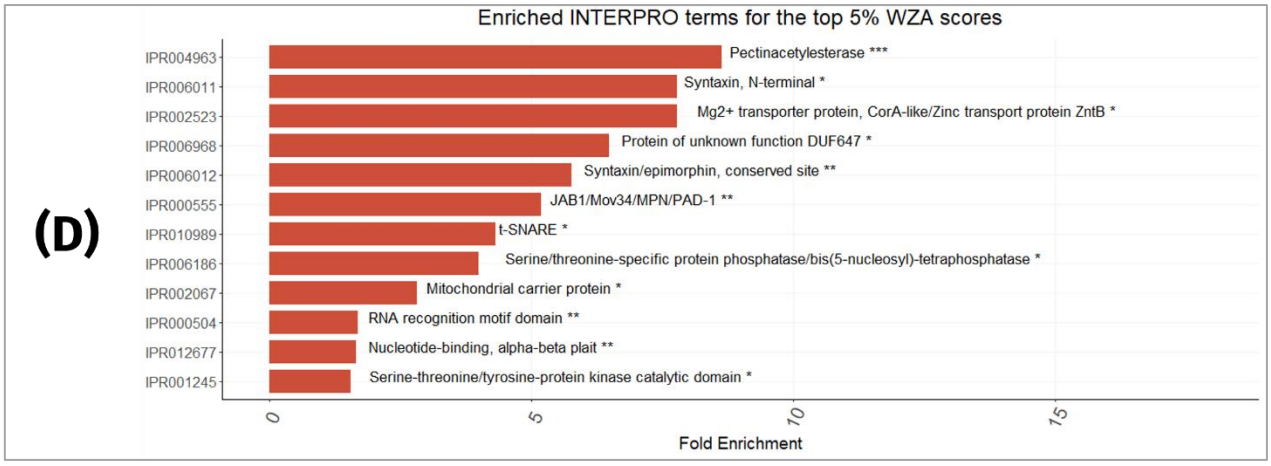


Figure 4.15 – Enriched GO and INTERPRO terms associated with variation in virulence using genes in the top 5% of WZA scores against a background of all *S. asiatica* genes for (A) cellular components (B) biological processes (C) molecular function and (D) INTERPRO terms. Significance of fold enrichment is indicated by stars - **** P<0.001, *** P<0.01, ** P<0.05, *P<0.1.

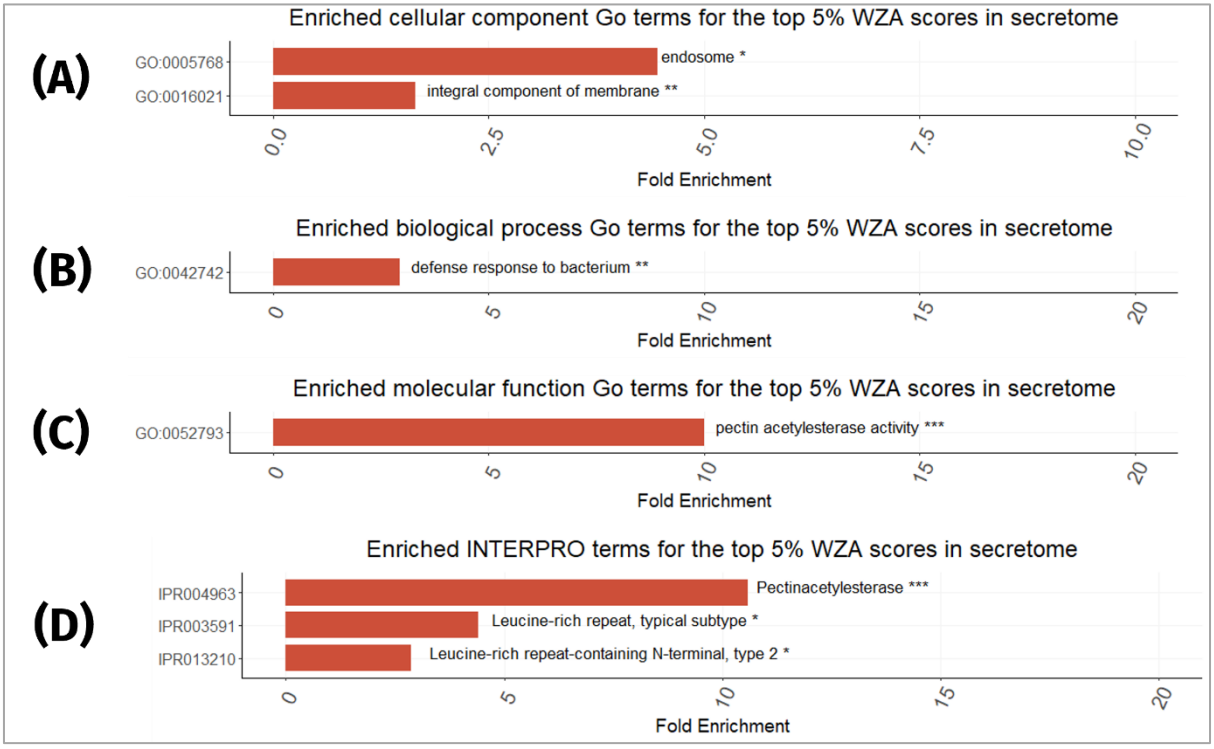
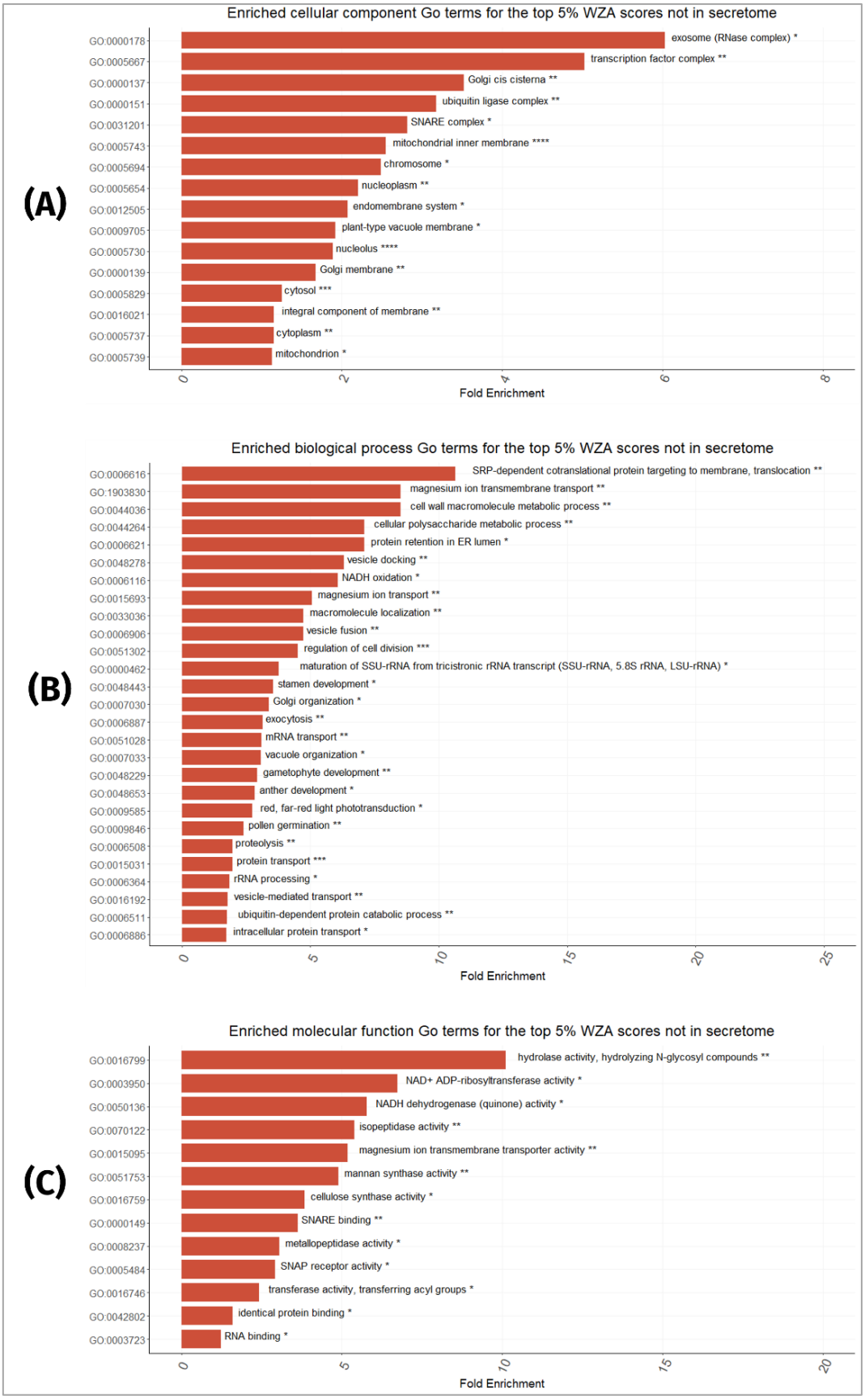


Figure 4.16 – Enriched GO and INTERPRO terms associated with variation in virulence using genes in the top 5% of WZA scores that were also in the *S. asiatica* secretome, against a background of all secretome *S. asiatica* genes for (A) cellular components (B) biological processes (C) molecular function and (D) INTERPRO terms. Significance of fold enrichment is indicated by stars - **** P<0.001, *** P<0.01, ** P<0.05, *P<0.1.

Table 4.5 – descriptions of cysteine content and length for gene products from the top 1250 candidate virulence-associated genes in the *S. asiatica* genome.

		Secretome	Non secretome
	Number of proteins	67	1183
Protein length (aa)	Mean	505.00	492.63
	Min	102.00	100.00
	Max	1301.00	4191.00
	SE	35.95	11.59
Cysteine content (%)	Mean	2.12	1.87
	Min	0.11	0.00
	Max	7.14	7.88
	SE	0.18	0.03


(A)
Enriched cellular component Go terms for the top 5% WZA scores not in secretome
GO:0000178 exosome (RNase complex) *
GO:0005667 transcription factor complex **
GO:0000137 Golgi cis cisterna **
GO:0000151 ubiquitin ligase complex **
GO:0031201 SNARE complex *
GO:0005743 mitochondrial inner membrane ****
GO:0005694 chromosome *
GO:0005654 nucleoplasm **
GO:0012505 endomembrane system *
GO:0009705 plant-type vacuole membrane *
GO:0005730 nucleolus ****
GO:0000139 Golgi membrane **
GO:0005829 cytosol ***
GO:0016021 integral component of membrane **
GO:0005737 cytoplasm **
GO:0005739 mitochondrion *
0 2 4 6 8
Fold Enrichment
(B)
Enriched biological process Go terms for the top 5% WZA scores not in secretome
GO:0006616 SRP-dependent cotranslational protein targeting to membrane, translocation **
GO:1903830 magnesium ion transmembrane transport **
GO:0044036 cell wall macromolecule metabolic process **
GO:0044264 cellular polysaccharide metabolic process **
GO:0006621 protein retention in ER lumen *
GO:0048278 vesicle docking **
GO:0006116 NADH oxidation *
GO:0015693 magnesium ion transport **
GO:0033036 macromolecule localization **
GO:0006906 vesicle fusion **
GO:0051302 regulation of cell division ***
GO:0000462 maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) *
GO:0048443 stamen development *
GO:0007030 Golgi organization *
GO:0006887 exocytosis **
GO:0051028 mRNA transport **
GO:0007033 vacuole organization *
GO:0048229 gametophyte development **
GO:0048653 anther development *
GO:0009585 red, far-red light phototransduction *
GO:0009846 pollen germination **
GO:0006508 proteolysis **
GO:0015031 protein transport ***
GO:0006364 rRNA processing *
GO:0016192 vesicle-mediated transport **
GO:0006511 ubiquitin-dependent protein catabolic process **
GO:0006886 intracellular protein transport *
0 5 10 15 20 25
Fold Enrichment
(C)
Enriched molecular function Go terms for the top 5% WZA scores not in secretome
GO:0016799 hydrolase activity, hydrolyzing N-glycosyl compounds **
GO:0003950 NAD+ ADP-ribosyltransferase activity *
GO:0050136 NADH dehydrogenase (quinone) activity *
GO:0070122 isopeptidase activity **
GO:0015095 magnesium ion transmembrane transporter activity **
GO:0051753 mannan synthase activity **
GO:0016759 cellulose synthase activity *
GO:0000149 SNARE binding **
GO:0008237 metallopeptidase activity *
GO:0005484 SNAP receptor activity *
GO:0016746 transferase activity, transferring acyl groups *
GO:0042802 identical protein binding *
GO:0003723 RNA binding *
0 5 10 15 20
Fold Enrichment

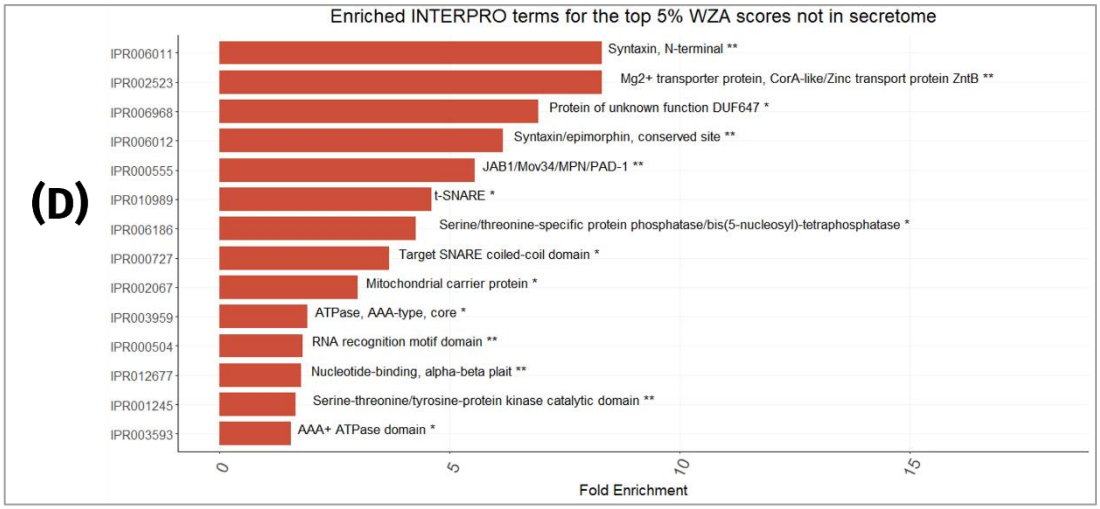


Figure 4.17 – Enriched GO and INTERPRO terms associated with variation in virulence using genes in the top 5% of WZA scores that were not in the *S. asiatica* secretome, against a background of all *S. asiatica* genes for (A) cellular components (B) biological processes (C) molecular function and (D) INTERPRO terms. Significance of fold enrichment is indicated by stars - **** P<0.001, *** P<0.01, ** P<0.05, *P<0.1.

4.3.5 Three putative virulence-associated genes were identified as also being present in *S. hermonthica*.

Data from Qiu et al. (2022) were retrieved from a data repository and top gene candidates from the study, both secretome and non-secretome were compared to the top 5% of genes identified in the current study through closest matching *A. thaliana* accession ID comparison. A total of 794 genes from the current study were matched to the top 823 genes from the Qui et al. study. Three of these putative virulence-associated genes were identified in both studies (Table 4.6). Pfam descriptions suggest these homologs possess (1) peptidase inhibition activity (SGA_v2.0_scaffold145G36077 /SHERM_26448), (2) nucleic acid deaminase activity (SGA_v2.0_scaffold5G02803/ SHERM_15413) and (3) helicase activity (SGA_v2.0_scaffold171G38937/ SHERM_27708). Despite only 3 identical matches to *A. thaliana* genes between the top sets of *S. hermonthica* and *S. asiatica* virulence candidates, many functional classes were seen in both studies including glycosyl hydrolases, kinases and proteases as well as proteins relating to transcription and translation processes.

Table 4.6 – Descriptions of the three putative virulence factors present in the top 5% of WZA scores across all *S. asiatica* genes which were also found as potential virulence candidates in a study of *S. hermonthica* virulence (Qui et al., 2022). Gene ID for both *S. asiatica* and *S. hermonthica* are given.

Gene ID	SGA_v2.0_scaffold145G36077/ SHERM_26448	SGA_v2.0_scaffold5G02803/ SHERM_15413	SGA_v2.0_scaffold171G38937/ SHERM_27708
Protein Length (aa)	823	1257	1203
Cysteine percent (%)	1.58	2.39	1.83
In secretome?	YES	NO	NO
WZA score	90.934	79.111	84.620
Nº variants in promoter region	7	2	0
Nº synonymous variants	8	11	7
Nº missense variants	3	10	3
Nº other coding region variants	20	12	11
Nº intron variants	21	43	51
Pfam terms	PF05922; PF02225; PF00082; PF17766	PF14432; PF01535; PF01535; PF01535; PF03129; PF13041; PF13041	PF04408; PF00271; PF00270; PF07717; PF00035
Pfam description	Peptidase inhibitor I9; PA domain; Subtilase family; Fibronectin type-III domain	DYW family of nucleic acid deaminases; PPR repeat; PPR repeat; PPR repeat; Anticodon binding domain; PPR repeat family; PPR repeat family	Helicase associated domain (HA2); Helicase conserved C-terminal domain; DEAD/DEAH box helicase; Oligonucleotide/oligosaccharide-binding (OB)-fold; Double-stranded RNA binding motif
A. thaliana acc.	AT1G20160	AT1G29880	AT1G48650
A. thaliana description	Subtilisin-like serine endopeptidase family protein	glycyl-tRNA synthetase / glycine--tRNA ligase	DEA(D/H)-box RNA helicase family protein
GO terms	GO:0004252, GO:0006508	GO:0008270	GO:0003676, GO:0004386, GO:0005524

4.4 Discussion

4.4.1 A combined RDA and WZA approach to genotype-environment association has revealed *S. asiatica* genes associated with virulence.

In this study the candidate virulence-associated genes identified corroborate previous findings on the classes of genes known in other species, particularly S hermonthica (Qiu et al., 2022). Typical virulence-associated factors are likely to fall into one of several groups including (but not limited to) (1) degradation of the physical cell wall barrier (2) suppression of host immunity or (3) nutrient acquisition (Nemri et al., 2014). Functional annotations of the genes found in this study suggested that cell wall modifying enzymes in particular were significant in influencing variation in *S. asiatica* virulence.

The method of GEA for use with *S. asiatica* virulence is a novel undertaking and was able to identify a large number of putative virulence-associated candidate genes, many of which have been identified in other studies. Despite the large number of weakly linked loci being present in GEA analysis, likely as a result of selfing, the weighted-Z score approach was able to elucidate the true positive associations.

One major challenge in utilising and adapting this method was ensuring that candidate virulence loci elucidated in the highly selfing *S. asiatica*, were truly associated with virulence. While GEA using redundancy analysis seeks to identify genetic variants that are associated with specific phenotypic traits or environmental variables, the association must be strong enough to detect outliers. As mentioned in section 4.2, accession-level genotype data was not used for this analysis due to the unit of 'accession' not being particularly meaningful and accessions being more genetically variable than expected, so instead individual-level genotypes were used, despite phenotype information being collected at accession level and encompassing an overall virulence measure, averaged across all hosts according to MDS scores (section 4.2.3). The reason being for the use of these MDS scores for phenotypic measures was that differences in variation of virulence for *S. asiatica* accession patterns were visibly stronger for virulence across hosts than variation in virulence seen on each host independently. This was due to the greater effect of rice host variety on virulence, with less influence of *S. asiatica* accession.

4.4.2 Both secreted and non-secreted gene products can be involved in *S. asiatica* virulence.

The top candidates by WZA score for virulence association (both secretome and non-secretome genes) were analysed in this study. While many virulence factors may be secreted to best target the host, some virulence factors may be instead involved with regulation of their expression or processing within the parasite cell (Evan et al., 2021). A comparison of the cysteine content and amino acid length of gene products identified that many of the non-secreted proteins were larger, although the mean length was similar between secreted and non-secreted gene products and cysteine content was generally lower for non-secreted proteins; cysteine

content being known to enhance effector stability in the apoplastic space for many pathogens (Kamoun, 2006). The mean number of amino acids in the putative secreted proteins was around 500, a larger cut off than has been used in many studies. For example, in *Ceratocystis cacaofunesta*, secreted 'effectors' were classified as those small proteins with fewer than 200 amino acids and a cystine content of over 4% (Molano et al., 2018), whereas fewer than 300 amino acids with at least 3% cysteine content was suggested based on comparison of several fungal secretomes by de Queiroz and Santana (2020). It was first observed in the parasitic interaction between *Cuscuta pentagona* and *Cucurbita maxima* that information was able to freely pass between host and parasite (Roney et al., 2006) where questions were raised regarding how parasites were able to manipulate their hosts, not only by utilising nutrients but also by acquiring host-derived transcripts along with secretion of their own virulence factors. Studies by Yang et al. (2015) and Reiss and Bailey (1998) have suggested that many of these gene products are secreted where their impact will be most felt such as the middle lamella as seen in *S. gesnerioides* attacking cowpea.

In terms of functional classes of candidate virulence genes, several differences were seen between secretome and non-secretome based genes. Enriched terms for secretome-based proteins included many enzymes such as pectin acetylesterases, glycosyl hydrolases and multi-Copper oxidases, many of which have also been implicated in the growth and differentiation of haustoria (Ashapkin et al., 2023). In addition, a number of genes associated with leucine rich repeats were identified through homology to *A. thaliana* gene accession IDs (including AT2G26330, AT5G46330, AT3G20820 and AT5G61240).

The leucine-rich repeat domain is known to be a platform for pathogen recognition, associated with detection of pathogen products of avirulence such as PAMPs (Padmanabhan et al., 2009). Recently, a putative virulence factor from *S. gesnerioides* was suggested to act as a 'decoy effector', interfering with the defence response pathway by disrupting parasite recognition (Su et al., 2020). Su et al. identified the *S. gesnerioides* leucine-rich repeat receptor kinase, SHR4z, which was shown to have significant similarity to somatic embryogenesis receptor kinase (SERK) in the conserved domain region. Several other SERK homologs are known, such as SISERK3A in tomato which binds to a flagellin-sensing receptor (FLS2) and on detection of flagellin, activates a PTI response (Albert et al., 2020). However, SHR4z was by Su et al. to bind to a ubiquitin ligase cowpea-interacting protein (VuPOB1) which led to silencing of its HR response mechanism. This is most definitely a possible mode of action for the leucine-rich repeats identified in the *S. asiatica* genome and the Su et al. study gives an excellent pipeline for future analysis of these gene products using transcriptomic analysis to examine upregulation over the infection timeline.

In comparison, enriched terms for non-secreted gene products included magnesium ion transport and SRP-dependent co-translational translocation. Both of these terms are interesting as they will no doubt be important for all biological processes but also have specific implications for virulence. Metal ions including

magnesium and copper often serve key roles as cofactors in enzymes which are crucial for cell wall degradation in host invasion by a pathogen (Porcheron et al., 2013) while co-translational processes may not be directly responsible for parasite invasion but may facilitate the activity through translocation (Wassenaar and Gaastra, 2001).

4.4.3 Two major cell wall degrading enzymes were implicated in virulence-association in *S. asiatica* - pectin acetylesterases and glycosyl hydrolases

With respect to the secretome, pectin acetylesterases were one group of well-known PCWDE found to have highly enriched molecular function GO terms (GO:0052793) in *S. asiatica* (Fig. 4.16). Pectins are known to be an incredibly diverse array of galacturonic acid-rich polysaccharides and are by far the most complex components of the plant cell wall (Qiu et al., 2021), exhibiting extensive structural variation from being constantly remodeled to allow for morphological changes which take place during plant growth and in response to stress (Wu et al., 2018). These changes can include cell wall stiffening, softening and loosening involving different combinations of expansins, pectin methylesterases, pectin lyases, pectin acetylesterases, xyloglucan endotransglucosylases/hydrolases etc. As well as in this current research, pectin acetylesterases have also been identified in a number of transcriptomic studies as key differentially expressed genes including in *Phytophthora parasitica* infection of Lupin (Blackman et al., 2015), root knot nematode (*Pratylenchus penetrans*) transient expression in *Nicotiana benthamiana* and S. *asiatica* and *S. hermonthica* on sorghum (Anyolo et al., 2015) and rice (Qiu et al., 2022). An interesting functional study of the pectin acetylesterase, PAE5 in the pathogenic oomycete, *Peronophythora litchii* was performed by Kong et al. (2019). Knockout of the pectin acetylesterase genes from *P. litchii* followed by virulence assays on lychee showed a decrease in infectivity; the study proposed the mechanism by which pectin acetylesterase enables virulence is through the modification of O-acetyl pectin in the cell wall (Kong et al., 2019). Pectin degradation products from the cell wall breakdown have also been shown to act as damage-associated molecular patterns (DAMPs) and found to initiate strong immune response in the *Botryris cinerea* infection of *Arabidopsis* (Benedetti et al., 2015; Shin et al., 2021) and possibly why studies have found that pectin acetylesterases can be known to increase stress-related transcription and improve defence against biotic stresses (Kloth et al., 2019).

While Pectin acetylesterases are involved in degradation of one component of the cell wall, glycoside hydrolases are involved in degradation of another, cellulose. In this study glycoside hydrolase activity was found to have strong fold enrichment (Fig. 4.15C) and these genes were frequently found in the top candidates by WZA score (Supplementary table 4.1). However, this functional group was not seen when using a background of only secretome genes, suggesting that the class of gene may purely be secreted and not necessarily also involved in viruelnce function. However, given that the enzyme products are commonly identified virulence factors it is likely that they could be identified in further studies. These enzymes are

responsible for hydrolysis of the glycosidic bond between carbohydrate molecules and their activity in pathogenesis has been recognised as being during mid-late stage infection as well as playing an unknown role in immunity (Rafiei et al., 2021). Glycosyl hydrolases have routinely been associated with pathogenicity; in the phytopathogenic fungus *Alternatia alternara*, endo-1,4-β-glucanase expression at high pH levels led to increased virulence on fruit development (Eshel et al., 2002), while in the soil-borne oomycete *Phytophtora sojae*, a cellulase family gene, PsGH7a was highly upregulated during infection of soybean and is highly conserved among families (Tan et al., 2020).

Postulated by Blackman et al. (2015), another function of these enzymes has been thought to involve neutralisation of host basal defence by disruption of callose deposition in the cell walls. However, the story is complicated by the fact that these enzymes are also able to act as elicitors of immune response in the host. Recently a rice gene has somewhat elucidated one effect of these enzymes in *Arabidopsis*, suggesting that upregulation of these glycosyl-hydrolase cell wall modifying genes was associated with increased defence through increased ROS-enabled programmed cell death which greatly hindered biotrophic attacks (Kim et al., 2022). With regards to parasitic plants, comparative studies of *S. hermonthica*, *T. versicolor* and *P. aegyptiaca* have found that glycosyl hydrolases play key roles in cell wall modification (Yang et al., 2015). The study by Yang et al. identified enriched cellular component terms highly associated with cell wall and extracellular localisation, as was seen in this thesis for *S. asiatica* (Section 4.3.4, Fig. 4.15). In addition, they were able to show through transcriptomic analysis of different life stages of the parasites, that glycosyl hydrolases were particularly enriched gene ontology terms among haustorial genes.

4.4.4 The two independent RDA axes corresponded to different virulence-related genes.

From analysis of the two separate RDA axes, generally candidate SNPs found on RDA axis 1 represented an association between both Dim.1 and Dim.2 virulence measures with variation in virulence while RDA axis 2 represented only associations with Dim.1 with virulence. While RDA1 and RDA2 explained different patterns of virulence by *S. asiatica* accession, RDA1 was found to explain more variation. Focusing of virulence-associated candidate loci into WZA gene windows for each RDA axis independently identified differences between them. RDA1 was more associated with a number of enzymes such as pectin acetylesterase, cytochrome P450, trehalase, peroxidase and alpha galactosidase while the top gene-products for RDA2 were glycosyl hydrolase, peptidase, protein phosphatase and peroxidase. For this study, it was not possible to detect outlier loci associated with a particular *S. asiatica* accession however, for future studies, the inclusion of more wide-ranging virulence profile in the phenotyping (including more *S. asiatica* accessions and rice varieties) may lead to stronger differences in genes utilised for that virulence.

4.5 Conclusion

Despite genetic limitations brought up with analysis methods due to the use of the highly inbreeding nature of *S. asiatica* as well as phenotypic data which was not necessarily derived from selfing lineages, this study identified a large number of virulence genes from the *S. asiatica* genome. Several of these gene products were found to be cell-wall degrading enzymes including pectin acetylesterase, which has been shown to influence virulence in a number of studies. The work done here can be utilised in future work for functional validation of candidates to ascertain mechanisms of action for candidate virulence-associated gene products and help determine how they can best be utilised in development of new resistant crop lines.

Chapter 5

General Discussion

In the first experimental chapter of this thesis, it was confirmed that both rice host variety and genotype of Striga accession have a significant impact on the virulence profile of *S. asiatica*, though host variety had the most significant effect in determining Striga virulence as well as a significant interaction effect. The variation in virulence seen for the numerous geographically separate Striga accessions grown on a range of different rice hosts is suggestive of a degree of host adaptation in *S. asiatica* and coevolution of host and parasite genotypes. This research was unique in the utilisation of a more diverse panel of hosts and wider range of geographically separate Striga accession than ever studied previously. While some earlier studies have shown variation in virulence of *S. asiatica* across hosts such as the NERICA rice varieties (Rodenburg et al., 2015), quality protein maize (Nayakurwa et al., 2018) and IITA maize lines (Gasura et al., 2019), others have identified variation in virulence dependent on parasite populations (Botanga et al., 2002).

However, until now a comprehensive study profiling virulence across a wide range of hosts and *S. asiatica* accessions has not been achieved, despite large morphological variability seen in specimens, supposedly indicative of differentially infecting 'ecotypes' noted decades prior (Cochrane and Press, 1997). However, this new evidence of variation in virulence builds a clearer picture of *S. asiatica* and while more extensive panels of hosts and wider ranges of Striga accessions would be beneficial to explore, this work can immediately be utilised for use in appropriate selection of rice varieties which show less susceptibility to Striga infection, particularly in Madagascar, a major focus of this study.

In chapter 3 of this thesis the population structure of *S. asiatica* was identified to reflect that of a highly inbreeding species, with low genetic diversity between individuals within sampled accession sites and greater diversity between accessions. The chapter also indicated long lineages between countries, providing a base for the evolutionary history of *S. asiatica*. This may be indicative of little evolutionary change between accessions in different countries or it may suggest that inadequate sampling had been performed leading to 'missing links'. The selfing nature of the species has been highlighted previously in population based genetic diversity analyses (Gethi et al., 2005; Botanga et al., 2002; Werth et al., 1984). However, these studies have shown mixed results concerning the extent of population structuring within *S. asiatica,* with Botanga et al. (2002) detecting much stronger population structuring than Gethi et al., (2004) identified in Kenya. As a highly autogamous species, it was hypothesised that a significant population structure would exist, in a similar way to *S. gesnerioides* where populations could be genetically identified and those that differ in neutral variants

would also differ in virulence phenotype, which would be independent of environment. Within this thesis' research some of the accessions studied represented multiple selfing lineages while other accessions contained a single lineage. While this could be suggestive of experimental sampling of too small an area for a 'population', these mixed selfing lineages may represent more recently colonised areas in Madagascar, as was seen in Kenyan populations (Gethi et al., 2004). The occurrence of multiple introductions of S. asiatica into Madagascar over the last century could explain the observed mixing phenomenon. It is possible that multiple accessions of Striga belonging to a single selfing lineage resulted from multiple introductions to the same fields, potentially influenced by factors like seed or tool sharing. This could also explain why certain accessions are distributed across multiple selfing lineages, as a batch of seeds arriving in Madagascar via a ship or similar means may have inadvertently been divided and sent to multiple farms. An alternative to this theory could also be that some inadvertent mixing of accession names happened during a part of the analysis leading to erroneous results which suggest mixed selfing lineages but may actually be mislabelled.

While Madagascan populations in this study did emphasize inbreeding, the use of the multiple individuals coming from a lab-grown Ethiopian inbred line confused this picture. The Ethiopian individuals had very low genetic diversity in comparison to individuals within the field-collected Madagascan accessions and as such were not able to be used for calculations of nucleotide diversity (π) or inbreeding coefficient (F_{is}). However, Ethiopian individuals also exhibited higher than expected heterozygosity, possibly related to potential outcrossing under lab conditions. While the extent of outcrossing of these Ethiopian accessions with accessions from a spatially and genetically distant population is unknown this crossing in the lab may have occurred as pollination bags were not used in one instance. Although this study found a clear neutral population structure, particularly on a wider spatial scale, future work would benefit from inclusion of field-collected accessions across an even wider range of areas with more individuals sequenced within these areas; These field-collected samples would provide a much more reliable picture of natural *S. asiatica* populations. Nevertheless, despite this limitation, the study managed to yield valuable insights. Firstly, the identification of outcrossing in the Ethiopian samples through genetic comparison was particularly significant. It not only demonstrates the relative ease of crossing in *S. asiatica* but also emphasises the inadvisability of employing these same batches of samples for subsequent research. Furthermore, the study revealed the existence of mixed selfing lineages within the field sites, highlighting the necessity to generate selfing lines for future analyses. The SNP genotyping of individuals in this work may also provide a base for selection of a genotyping panel to be used for screening of fields to map the spread of Striga with ever threatening climate changes.

Major findings from Chapter 4 include the novel population genomics approach for identification of *S. asiatica* virulence-associated genes. This involved separation of neutral loci from loci under selection based on the variation in virulence seen in chapter 2. This in-silico prediction suggested strong involvement of cell wall

degrading enzymes such as pectin acetylesterases and glycosyl hydrolases as well as Leucine-rich repeats being highly enriched terms participating in the *S. asiatica* virulence process.

Several of these genes would be important candidates that could be functionally validated to examine their participation in Striga infection in the future. However, in this analysis a large number of candidate genes were proposed to be associated with virulence and therefore follow-up population genetic approaches may also be required to narrow these candidates down to a more manageable number as well as confirm their importance. Alternative approaches for assessment of *S. asiatica* candidates could include the comparison of SNP variants from *S. asiatica* pooled genomic DNA grown on either susceptible or resistant hosts in order to determine loci which are under selection for increased virulence when grown on a resistant host. This is in a similar experiment as Qiu et al. (2022) who identified candidates from bulked *S. hermonthica* DNA samples grown on NERICA-7 or NERICA-17 hosts. In addition, RNA-seq and transcriptomics can be a useful method to identify differentially expressed genes in the parasite at different stages of infection as has been evidenced by Qiu et al. (2022) with *S. hermonthica* on rice but also the plant parasitic nematode *Bursaphelenchus mucronatus* on *Pinus thunbergia* (Zhou et al., 2016) and the bacterial wilt-causing *Ralstonia solanacearum* on potato (de Pedro-Jove et al., 2021).

For functional validation, it may be necessary to generate a successful transformation protocol for *S. asiatica* to be able to manipulate any identified candidates. However, this could also be achieved through transient expression vehicles such as Agrobacterium-mediated *Nicotiana benthamiana* expression of selected candidates to explore localisation and cellular effects. Agroinfiltration is a particularly useful technique for expression of avirulence genes which can be co-expressed with the prospective host R gene to test resistance (Ma et al., 2012). Agrobacterium-mediated transformation has been previously performed on sunflower to test the effects of silencing of holoparasitic plant *O. cumana* expansin genes (Jiang et al., 2021); the study found a good silencing efficiency and subsequently confirmed the from effect the *O. cumana* genes have on parasitism. For breeders of potential host crops, further work on these potential effectors and the extent of how conserved they are across the parasitic species would prove highly beneficial for determining how resistance can be effectively used in breeding programmes.

Nevertheless, several flaws in the experimental design have been apparent, not least the lack of a large number of true selfing lines for determination of virulence. While this study relied upon the available virulence data from chapter 2 and genetic data from chapter 3, this did not provide a particularly clear picture of differences in virulence. Because of this, the statistical approach of the partial redundancy analysis was not able to detect outlier virulence-associated loci for each accession based on Striga parasitism on a range of hosts with differing susceptibilities and instead had to rely upon identifying overall differences in virulence patterns, using the MDS plot. What should be taken away from this study is that selfing lines, should be used

for assessment of phenotyping such as this. This may be best achieved through collection of *S. asiatica* seed from smaller sampling sites to avoid multiple lineages being collected as one, then several individuals within batches can be genotyped and those batches which contain the same selfing lineage may be combined.

Overall, this thesis investigated novel research themes within the *S. asiatica* virulence genre, identifying variation in virulence and determining putative effector genes associated with pathogenicity by utilising the recent *S. asiatica* reference genome as a most valuable tool.

Supplementary materials

Supplementary Table 2.1 – Permutation analysis of variance (ANOVA) showing the effects of rice host variety and Striga accession on Striga **dry weight (biomass)** as a measure of virulence using 9 *S. asiatica* accessions grown on 9 rice hosts as combination.

Sources of variation	df	Mean sq	F-value	p-value
Rice variety	8	4178.1504	53.5994	0.0002
Striga accession	8	1378.7161	17.76869	0.0002
Rice variety:Striga accession	64	220.2772	2.8258	0.0002
Residuals	511	8.829		

Supplementary Table 2.2 – Permutation analysis of variance (ANOVA) showing the effects of rice host variety and Striga accession on Striga **dry weight (biomass)** as a measure of virulence using 8 *S. asiatica* accessions, excluding the avirulent Bel (site 8), grown on 9 rice hosts as combination.

Sources of variation	df	Mean sq	F-value	p-value
Rice variety	8	4334.9287	56.9303	0.0002
Striga accession	7	340.2071	4.4679	0.0002
Rice variety:Striga accession	56	201.6981	2.6489	0.0002
Residuals	436	8.726		

Supplementary Table 2.3 – Permutation analysis of variance (ANOVA) showing the effects of rice host variety and Striga accession **cumulative length** of Striga attachments as a measure of virulence using 9 *S. asiatica* accessions grown on 9 rice hosts as combination.

Sources of variation	df	Mean sq	F-value	p-value
Rice variety	8	49345.898	73.4587	0.0002
Striga accession	8	24958.325	37.1542	0.0002
Rice variety:Striga accession	64	2773.679	4.1290	0.0002
Residuals	511	25.918		

Supplementary Table 2.4 – Permutation analysis of variance (ANOVA) showing the effects of rice host variety and Striga accession on **cumulative length** of Striga attachments as a measure of virulence using 8 *S. asiatica* accessions, excluding the avirulent Bel (site 8), grown on 9 rice hosts as combination.

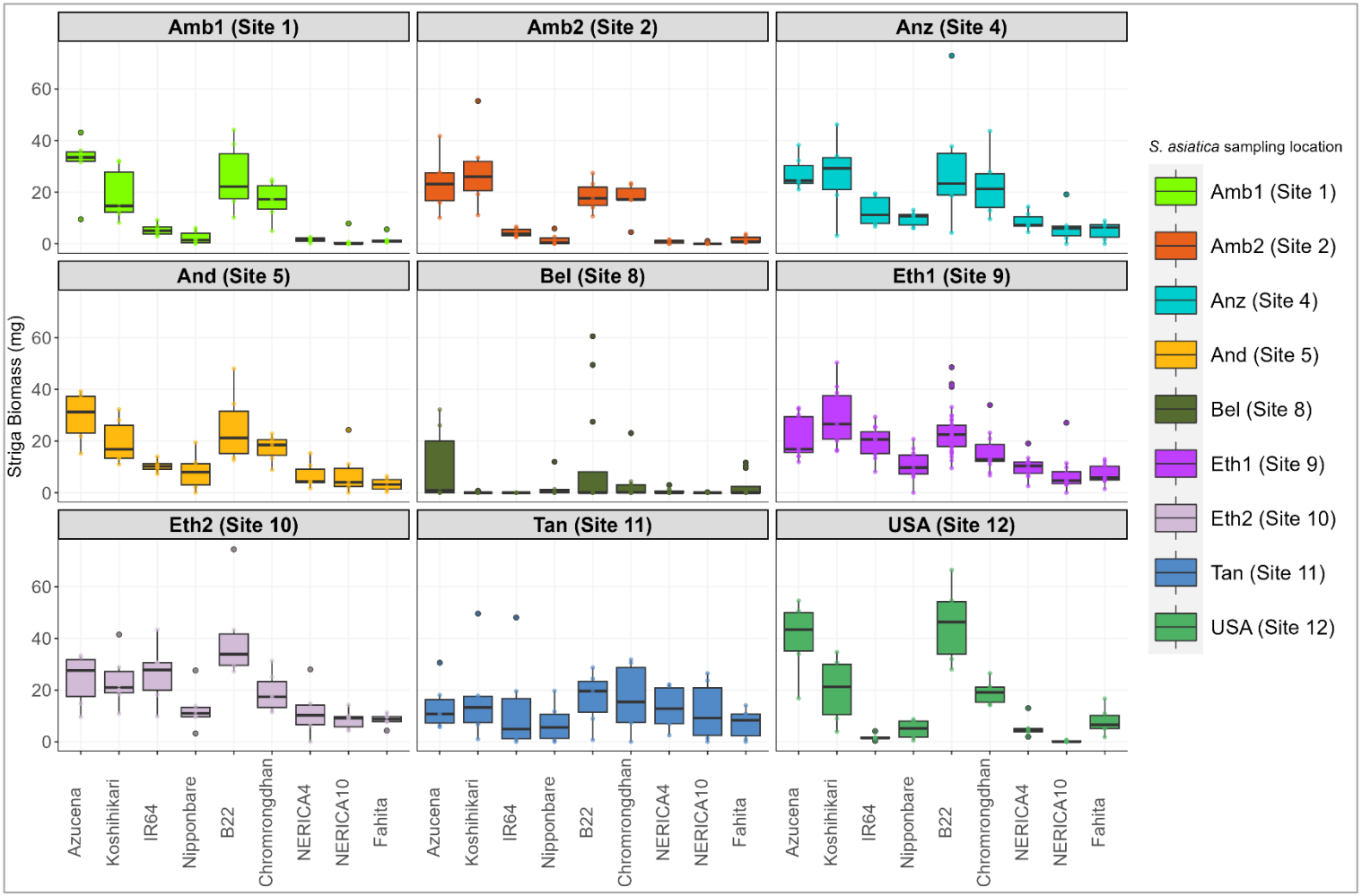
Sources of variation	df	Mean sq	F-value	p-value
Rice variety	8	54090.73	74.4133	0.0002
Striga accession	7	14224.39	19.5687	0.0002
Rice variety:Striga accession	56	2622.41	3.6077	0.0002
Residuals	436	29.96		

Supplementary Table 2.5 – Permutation analysis of variance (ANOVA) showing the effects of rice host variety and Striga accession on **total number** of Striga attachments as a measure of virulence using 9 *S. asiatica* accessions grown on 9 rice hosts as combination.

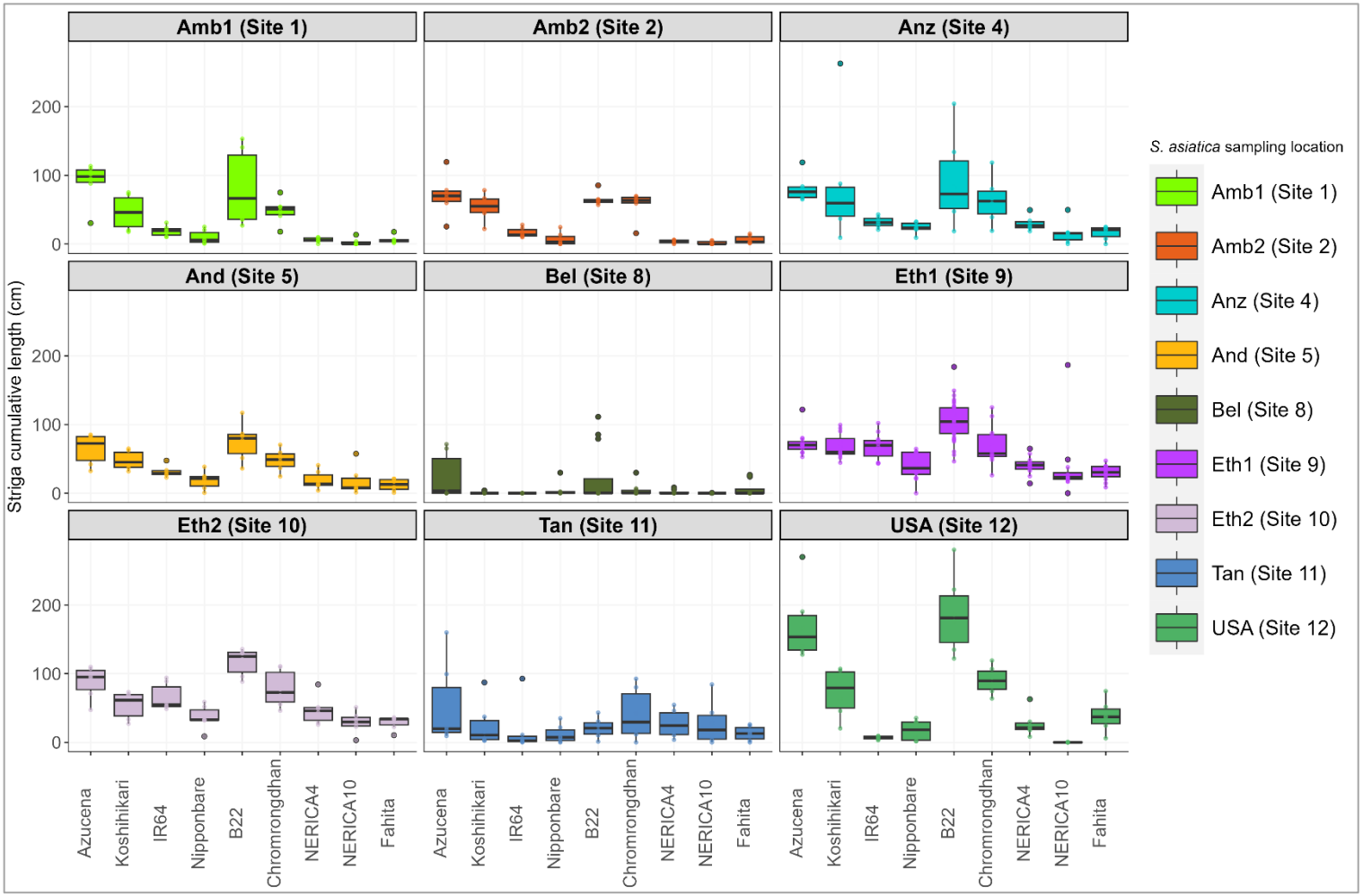
Sources of variation	df	Mean sq	F-value	p-value
Rice variety	8	171.8350	70.1718	0.0002
Striga accession	8	201.9846	82.4839	0.0002
Rice variety:Striga accession	64	7.0052	2.8607	0.0002
Residuals	511	1.5649		

Supplementary Table 2.6 – Permutation analysis of variance (ANOVA) showing the effects of rice host variety and Striga accession on **total number** of Striga attachments as a measure of virulence using 8 *S. asiatica* accessions, excluding the avirulent Bel (site 8), grown on 9 rice hosts as combination

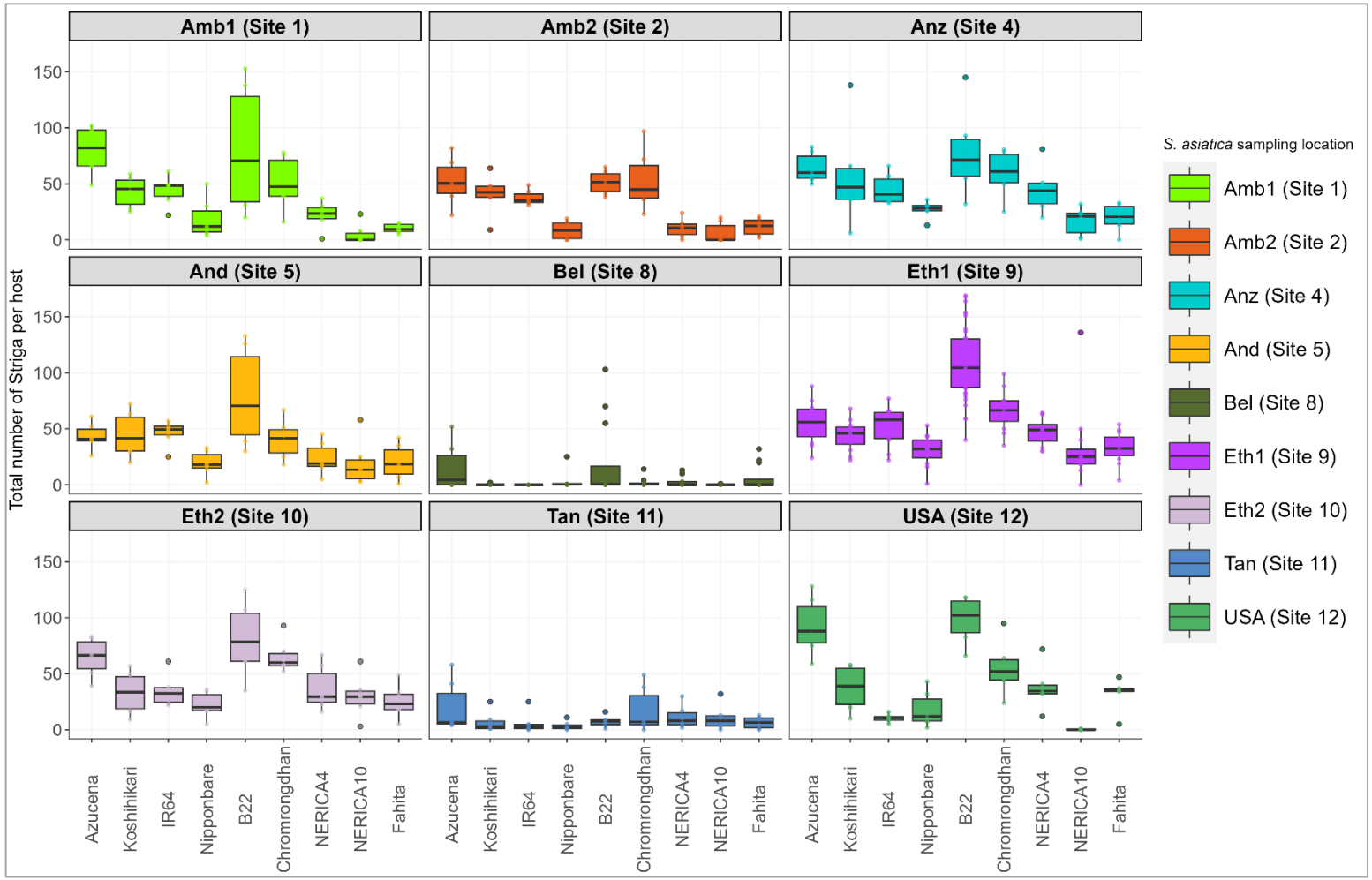
Sources of variation	df	Mean sq	F-value	p-value
Rice variety	8	179.0788	75.8242	0.0002
Striga accession	7	92.6881	39.2453	0.0002
Rice variety:Striga accession	56	6.0800	2.5743	0.0002
Residuals	436	1.5368		



Supplementary Figure 2.1 - Comparison of *S. asiatica* virulence profiles on a panel of rice varieties using Striga dry biomass collected from the host root systems as a measure of virulence. *S. asiatica* sampling sites are given in the figure legend. Sites 1,24,5 and 8 are from Madagascar, sites 9 and 10 are from Ethiopia and site 11 and site 12 are from Tanzania and USA, respectively. The plot shows a wide variation in Striga virulence by Striga accession.



Supplementary Figure 2.2 - Comparison of *S. asiatica* virulence profiles on a panel of rice varieties using cumulative length of Striga collected from the host root systems as a measure of virulence. *S. asiatica* sampling sites are given in the figure legend. Sites 1,24,5 and 8 are from Madagascar, sites 9 and 10 are from Ethiopia and site 11 and site 12 are from Tanzania and USA, respectively. The plot shows a wide variation in Striga virulence by Striga accession.



Supplementary Figure 2.3 - Comparison of *S. asiatica* virulence profiles on a panel of rice varieties using total number of Striga collected from the host root systems as a measure of virulence. *S. asiatica* sampling sites are given in the figure legend. Sites 1,24,5 and 8 are from Madagascar, sites 9 and 10 are from Ethiopia and site 11 and site 12 are from Tanzania and USA, respectively. The plot shows a wide variation in Striga virulence by host variety.

X^2	Top 5%	Remaining 95%
Secretome	1183	22662
Non-secretome	67	1173

X^2	Top 1%	Remaining 99%
Secretome	241	23604
Non secretome	9	1231

Supplementary Figure 4.1 –chi squared tables for total numbers of secretome and non secretome-based genes in the top 1% and top 5% of virulence-associated genes by WZA score.

Supplementary Table 4.1 – Details of *S. asiatica* virulence gene candidates identified in this study.

***S. asiatica* gene ID**	**In secretome?**	**Protein length (aa)**	**Cysteine %**	**Pfam ID**	**Pfam description**	***A.thaliana* accession ID**	**A.thaliana description**	**WZA score**	**GO_terms**
SGA_v2.0_scaffold112G31602	NO	145	1.38	no_Pfam	no_Pfam	No_hit	No_hit	180.16	NA
SGA_v2.0_scaffold11G05607	NO	304	1.64	no_Pfam	no_Pfam	No_hit	No_hit	176.57	NA
SGA_v2.0_scaffold10G05394	NO	184	2.17	no_Pfam	no_Pfam	No_hit	No_hit	163.86	NA
SGA_v2.0_scaffold24G10339	NO	293	3.41	no_Pfam	no_Pfam	No_hit	No_hit	156.52	NA
SGA_v2.0_scaffold45G17315	NO	601	1.5	no_Pfam	no_Pfam	No_hit	No_hit	149.07	NA
SGA_v2.0_scaffold244G44358	NO	959	0.73	PF03989; PF03989; PF03989; PF03989; PF03989; PF03989; PF00521	DNA gyrase C-terminal domain: beta-propeller; DNA gyrase/topoisomerase IV: subunit A	AT3G10690	DNA GYRASE A	148.70	GO:0003677, GO:0003916, GO:0003918, GO:0005524, GO:0005694, GO:0006265
SGA_v2.0_scaffold279G46783	NO	870	2.18	PF13966; PF00566	zinc-binding in reverse transcriptase; Rab-GTPase-TBC domain	No_hit	No_hit	148.68	NA
SGA_v2.0_scaffold281G46892	NO	346	2.02	no_Pfam	no_Pfam	No_hit	No_hit	146.42	NA
SGA_v2.0_scaffold197G41027	NO	112	1.79	no_Pfam	no_Pfam	No_hit	No_hit	146.11	NA
SGA_v2.0_scaffold112G31574	NO	149	4.03	no_Pfam	no_Pfam	No_hit	No_hit	141.42	NA
SGA_v2.0_scaffold367G50060	NO	222	2.25	no_Pfam	no_Pfam	No_hit	No_hit	140.66	NA
SGA_v2.0_scaffold258G45747	NO	493	3.25	no_Pfam	no_Pfam	AT3G19895	RING/U-box superfamily protein	138.33	NA
SGA_v2.0_scaffold115G32164	NO	536	0.93	PF03949; PF03949; PF00390	Malic enzyme: NAD binding domain; Malic enzyme: NAD binding domain; Malic enzyme: N-terminal domain	AT4G00570	NAD-dependent malic enzyme 2	137.63	GO:0004471, GO:0051287, GO:0055114
SGA_v2.0_scaffold292G47381	YES	212	0.94	PF13462	Thioredoxin	AT1G76020	Thioredoxin superfamily protein	136.74	NA
SGA_v2.0_scaffold24G10342	NO	125	0	no_Pfam	no_Pfam	No_hit	No_hit	136.26	NA
SGA_v2.0_scaffold112G31570	YES	309	3.24	PF03283	Pectinacetylesterase	AT4G19410	Pectinacetylesterase family protein	135.60	GO:0016787
SGA_v2.0_scaffold93G28345	NO	110	4.55	no_Pfam	no_Pfam	No_hit	No_hit	135.50	NA
SGA_v2.0_scaffold177G39442	NO	272	0.74	no_Pfam	no_Pfam	No_hit	No_hit	132.67	NA
SGA_v2.0_scaffold272G46375	NO	1418	1.62	PF00271; PF07717; PF00270; PF04408	Helicase conserved C-terminal domain; Oligonucleotide/oligosaccharide-binding (OB)-fold; DEAD/DEAH box helicase; Helicase associated domain (HA2)	AT1G58060	RNA helicase family protein	129.04	GO:0003676, GO:0004386, GO:0005524

***S. asiatica* gene ID**	**In secretome?**	**Protein length (aa)**	**Cysteine %**	**Pfam ID**	**Pfam description**	***A.thaliana* accession ID**	**A.thaliana description**	**WZA score**	**GO_terms**
SGA_v2.0_scaffold13G06749	NO	461	0.65	PF06027; PF06027; PF06027	Solute carrier family 35; Solute carrier family 35; Solute carrier family 35	No_hit	No_hit	128.82	GO:0006810, GO:0016021
SGA_v2.0_scaffold213G42229	NO	222	1.35	no_Pfam	no_Pfam	No_hit	No_hit	126.81	NA
SGA_v2.0_scaffold213G42230	NO	108	2.78	no_Pfam	no_Pfam	No_hit	No_hit	126.81	NA
SGA_v2.0_scaffold367G50065	NO	201	5.97	no_Pfam	no_Pfam	No_hit	No_hit	126.77	NA
SGA_v2.0_scaffold3G01798	YES	927	0.65	PF01055; PF13802	Glycosyl hydrolases family 31; Galactose mutarotase-like	AT5G63840	Glycosyl hydrolases family 31 protein	123.15	GO:0004553, GO:0005975
SGA_v2.0_scaffold196G40952	NO	307	2.28	PF01267	F-actin capping protein alpha subunit	AT3G05520	Subunits of heterodimeric actin filament capping protein Capz superfamily	122.11	GO:0008290, GO:0051016
SGA_v2.0_scaffold145G36046	NO	116	2.59	no_Pfam	no_Pfam	No_hit	No_hit	121.53	NA
SGA_v2.0_scaffold22G09893	NO	246	1.22	PF07993	Male sterility protein	AT4G33790	Jojoba acyl CoA reductase-related male sterility protein	119.84	NA
SGA_v2.0_scaffold171G38855	NO	3858	1.53	PF06025; PF00632; PF00627; PF06012; PF14377; PF14377; PF14377	Domain of Unknown Function (DUF913); HECT-domain (ubiquitin-transferase); UBA/TS-N domain; Domain of Unknown Function (DUF908); Ubiquitin binding region;	AT1G55860	ubiquitin-protein ligase 1	117.99	GO:0004842
SGA_v2.0_scaffold11G05551	NO	107	3.74	no_Pfam	no_Pfam	No_hit	No_hit	117.34	NA
SGA_v2.0_scaffold112G31605	NO	140	1.43	no_Pfam	no_Pfam	No_hit	No_hit	115.59	NA
SGA_v2.0_scaffold308G48072	NO	192	2.08	no_Pfam	no_Pfam	No_hit	No_hit	114.95	NA
SGA_v2.0_scaffold11G05742	NO	195	1.03	no_Pfam	no_Pfam	No_hit	No_hit	114.78	NA
SGA_v2.0_scaffold337G49187	NO	451	1.11	PF00331	Glycosyl hydrolase family 10	AT4G33840	Glycosyl hydrolase family 10 protein	114.41	GO:0004553, GO:0005975
SGA_v2.0_scaffold112G31569	NO	117	4.27	no_Pfam	no_Pfam	No_hit	No_hit	114.40	NA
SGA_v2.0_scaffold107G30768	NO	307	2.93	no_Pfam	no_Pfam	No_hit	No_hit	114.08	NA
SGA_v2.0_scaffold424G51476	NO	213	0.47	no_Pfam	no_Pfam	No_hit	No_hit	113.72	NA
SGA_v2.0_scaffold81G25999	NO	695	2.01	PF00514; PF00514; PF00651	Armadillo/beta-catenin-like repeat; Armadillo/beta-catenin-like repeat; BTB/POZ domain	AT5G19330	ARM repeat protein interacting with ABF2	112.79	GO:0005515
SGA_v2.0_scaffold101G29757	NO	455	1.54	no_Pfam	no_Pfam	AT3G22990	ARM repeat superfamily protein	111.96	NA
SGA_v2.0_scaffold466G52020	NO	447	2.68	no_Pfam	no_Pfam	No_hit	No_hit	111.38	NA
SGA_v2.0_scaffold272G46372	NO	770	1.43	PF08145; PF00400; PF00400; PF00400	BOP1NT (NUC169) domain; WD domain: G-beta repeat	AT2G40360	Transducin/WD40 repeat-like superfamily protein	110.79	GO:0005515, GO:0006364

***S. asiatica* gene ID**	**In secretome?**	**Protein length (aa)**	**Cysteine %**	**Pfam ID**	**Pfam description**	***A.thaliana* accession ID**	**A.thaliana description**	**WZA score**	**GO_terms**
SGA_v2.0_scaffold27G11473	NO	311	1.29	PF03171; PF14226	2OG-Fe(II) oxygenase superfamily; non-haem dioxygenase in morphine synthesis N-terminal	AT4G21690	gibberellin 3-oxidase 3	110.59	GO:0016491, GO:0055114
SGA_v2.0_scaffold275G46549	NO	463	1.94	PF07821; PF00128	Alpha-amylase C-terminal beta-sheet domain; Alpha amylase: catalytic domain	AT1G76130	alpha-amylase-like 2	109.89	GO:0003824, GO:0004556, GO:0005509, GO:0005975
SGA_v2.0_scaffold40G15422	NO	787	1.02	PF02737; PF00725; PF00378	3-hydroxyacyl-CoA dehydrogenase: NAD binding domain; 3-hydroxyacyl-CoA dehydrogenase: C-terminal domain; Enoyl-CoA hydratase/isomerase	AT4G29010	Enoyl-CoA hydratase/isomerase family	109.80	GO:0003824, GO:0003857, GO:0006631, GO:0008152, GO:0016491, GO:0055114
SGA_v2.0_scaffold59G20672	YES	627	2.87	PF00445; PF01959	Ribonuclease T2 family; 3-dehydroquinate synthase II	No_hit	No_hit	108.63	GO:0003723, GO:0003856, GO:0009073, GO:0016491, GO:0033897, GO:0055114
SGA_v2.0_scaffold219G42591	NO	942	2.55	PF00201; PF00201	UDP-glucoronosyl and UDP-glucosyl transferase; UDP-glucoronosyl and UDP-glucosyl transferase	No_hit	No_hit	108.50	GO:0008152, GO:0016758
SGA_v2.0_scaffold69G23353	NO	899	2	PF13691	tRNase Z endonuclease	AT1G52160	tRNAse Z3	107.95	GO:0008033
SGA_v2.0_scaffold112G31572	NO	606	1.65	PF00069; PF13499; PF13499	Protein kinase domain; EF-hand domain pair; EF-hand domain pair	AT2G38910	calcium-dependent protein kinase 20	107.90	GO:0004672, GO:0005509, GO:0005524, GO:0006468
SGA_v2.0_scaffold22G09943	NO	137	3.65	no_Pfam	no_Pfam	No_hit	No_hit	107.73	NA
SGA_v2.0_scaffold357G49813	NO	174	0.57	no_Pfam	no_Pfam	No_hit	No_hit	107.62	NA
SGA_v2.0_scaffold22G09942	NO	571	2.28	PF08392; PF08541	FAE1/Type III polyketide synthase-like protein; 3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C terminal	AT2G26640	3-ketoacyl-CoA synthase 11	107.55	GO:0006633, GO:0016020, GO:0016747
SGA_v2.0_scaffold145G36062	NO	390	1.79	PF00297	Ribosomal protein L3	AT1G43170	ribosomal protein 1	107.30	GO:0003735, GO:0005622, GO:0005840, GO:0006412
SGA_v2.0_scaffold274G46506	NO	1895	0.9	PF00575; PF00575; PF00575; PF00575; PF05843	S1 RNA binding domain; Suppressor of forked protein (Suf)	AT3G11964	RNA binding;RNA binding	106.97	GO:0003676, GO:0005634, GO:0006397

***S. asiatica* gene ID**	**In secretome?**	**Protein length (aa)**	**Cysteine %**	**Pfam ID**	**Pfam description**	***A.thaliana* accession ID**	**A.thaliana description**	**WZA score**	**GO_terms**
SGA_v2.0_scaffold258G45746	NO	684	1.75	PF03167; PF00450; PF00450	Uracil DNA glycosylase superfamily; Serine carboxypeptidase; Serine carboxypeptidase	No_hit	No_hit	106.64	GO:0004185, GO:0006508
SGA_v2.0_scaffold22G09895	NO	243	2.06	PF03015; PF07993	Male sterility protein; Male sterility protein	AT4G33790	Jojoba acyl CoA reductase-related male sterility protein	106.44	NA
SGA_v2.0_scaffold112G31603	NO	406	1.48	PF00149	Calcineurin-like phosphoesterase	AT5G57140	purple acid phosphatase 28	106.40	GO:0016787
SGA_v2.0_scaffold165G38254	NO	294	2.04	PF09785	Prp31 C terminal domain	No_hit	No_hit	105.86	NA
SGA_v2.0_scaffold132G34513	NO	573	1.75	PF17849; PF13638; PF17216; PF00773	Dis3-like cold-shock domain 2 (CSD2); PIN domain; Rrp44-like cold shock domain; RNB domain	AT2G17510	ribonuclease II family protein	105.51	NA
SGA_v2.0_scaffold80G25734	NO	111	0.9	no_Pfam	no_Pfam	No_hit	No_hit	105.37	NA
SGA_v2.0_scaffold106G30732	NO	1141	2.54	PF16589; PF11799; PF00817	BRCT domain: a BRCA1 C-terminus domain; impB/mucB/samB family C-terminal domain; impB/mucB/samB family	AT5G44750	DNA-directed DNA polymerases	105.29	GO:0003684, GO:0006281
SGA_v2.0_scaffold79G25538	NO	287	2.44	no_Pfam	no_Pfam	No_hit	No_hit	104.88	NA
SGA_v2.0_scaffold145G36061	NO	291	1.37	PF10075; PF10075	CSN8/PSMD8/EIF3K family; CSN8/PSMD8/EIF3K family	AT4G14110	COP9 signalosome; subunit CSN8	104.75	NA
SGA_v2.0_scaffold131G34491	NO	785	1.66	PF00749; PF03950; PF00043	tRNA synthetases class I (E and Q): catalytic domain; tRNA synthetases class I (E and Q): anti-codon binding domain; Glutathione S-transferase: C-terminal domain	AT5G26710	Glutamyl/glutaminyl-tRNA synthetase; class Ic	104.48	GO:0000166, GO:0004812, GO:0005524, GO:0005737, GO:0006418, GO:0016876, GO:0043039
SGA_v2.0_scaffold132G34546	NO	186	3.23	PF13966	zinc-binding in reverse transcriptase	AT2G02650	Ribonuclease H-like superfamily protein	104.44	NA
SGA_v2.0_scaffold198G41064	NO	177	3.39	no_Pfam	no_Pfam	No_hit	No_hit	104.41	NA
SGA_v2.0_scaffold62G21483	NO	945	0.95	PF07766	LETM1-like protein	AT1G65540	LETM1-like protein	103.97	NA
SGA_v2.0_scaffold182G39813	NO	499	2	no_Pfam	no_Pfam	No_hit	No_hit	103.94	NA
SGA_v2.0_scaffold10G05336	YES	351	4.56	PF03283; PF03283	Pectinacetylesterase; Pectinacetylesterase	AT5G23870	Pectinacetylesterase family protein	103.89	GO:0016787
SGA_v2.0_scaffold4G02415	NO	1041	1.06	PF10408; PF04564	Ubiquitin elongating factor core; U-box domain	AT5G15400	U-box domain-containing protein	103.59	GO:0000151, GO:0004842, GO:0006511, GO:0016567, GO:0034450

***S. asiatica* gene ID**	**In secretome?**	**Protein length (aa)**	**Cysteine %**	**Pfam ID**	**Pfam description**	***A.thaliana* accession ID**	**A.thaliana description**	**WZA score**	**GO_terms**
SGA_v2.0_scaffold165G38273	NO	616	1.79	PF00133; PF08264; PF09334	tRNA synthetases class I (I: L: M and V); Anticodon-binding domain of tRNA; tRNA synthetases class I (M)	AT3G55400	methionyl-tRNA synthetase / methionine--tRNA ligase / MetRS (cpMetRS)	102.99	GO:0000166, GO:0004812, GO:0005524, GO:0006418
SGA_v2.0_scaffold196G40885	NO	157	0.64	no_Pfam	no_Pfam	No_hit	No_hit	102.44	NA
SGA_v2.0_scaffold25G11140	NO	1586	3.03	PF07496; PF00856; PF17907	CW-type Zinc Finger; SET domain; AWS domain	AT1G77300	histone methyltransferases(H3-K4 specific);histone methyltransferases(H3-K36 specific)	101.85	GO:0005515, GO:0008270
SGA_v2.0_scaffold93G28346	NO	134	0	no_Pfam	no_Pfam	No_hit	No_hit	101.53	NA
SGA_v2.0_scaffold42G15981	NO	235	3.83	no_Pfam	no_Pfam	No_hit	No_hit	100.99	NA
SGA_v2.0_scaffold90G27914	NO	2555	3.05	PF02259	FAT domain	AT3G48190	ataxia-telangiectasia mutated	100.57	GO:0005515
SGA_v2.0_scaffold244G44338	NO	715	1.96	PF03364; PF03364	Polyketide cyclase / dehydrase and lipid transport; Polyketide cyclase / dehydrase and lipid transport	AT5G08720	Unknown protein	100.24	NA
SGA_v2.0_scaffold75G24800	NO	392	1.02	PF00113; PF03952	Enolase: C-terminal TIM barrel domain; Enolase: N-terminal domain	AT2G36530	Enolase	99.85	NA
SGA_v2.0_scaffold191G40427	NO	745	1.34	PF09742	Dyggve-Melchior-Clausen syndrome protein	AT1G04200	Unknown protein	99.35	NA
SGA_v2.0_scaffold11G05586	NO	308	1.3	no_Pfam	no_Pfam	No_hit	No_hit	99.31	NA
SGA_v2.0_scaffold198G41085	NO	2576	0.66	PF00005; PF00005; PF00664; PF00664; PF00664; PF00664; PF00005; PF00005	ABC transporter transmembrane region; ABC transporter	No_hit	No_hit	99.29	GO:0005524, GO:0006810, GO:0016021, GO:0016887, GO:0042626, GO:0055085
SGA_v2.0_scaffold117G32377	NO	326	2.76	no_Pfam	no_Pfam	No_hit	No_hit	99.19	NA
SGA_v2.0_scaffold11G05672	NO	1376	1.82	PF08030; PF08030; PF01794; PF08022; PF08022	Ferric reductase NAD binding domain; Ferric reductase NAD binding domain; Ferric reductase like transmembrane component; FAD-binding domain; FAD-binding domain	No_hit	No_hit	99.00	GO:0016491, GO:0055114

***S. asiatica* gene ID**	**In secretome?**	**Protein length (aa)**	**Cysteine %**	**Pfam ID**	**Pfam description**	***A.thaliana* accession ID**	**A.thaliana description**	**WZA score**	**GO_terms**
SGA_v2.0_scaffold198G41113	NO	422	1.18	PF00764; PF00764	Arginosuccinate synthase;	AT4G24830	arginosuccinate synthase family	98.79	GO:0004055, GO:0005524, GO:0006526
SGA_v2.0_scaffold196G40963	NO	1057	1.51	PF17846; PF17846; PF17846; PF03159; PF00098	Xrn1 helical domain; XRN 5'-3' exonuclease N-terminus; Zinc knuckle	AT1G54490	exoribonuclease 4	98.43	GO:0003676, GO:0004527, GO:0008270
SGA_v2.0_scaffold343G49460	NO	1126	1.42	PF05664	Plant family of unknown function (DUF810)	AT4G11670	Protein of unknown function (DUF810)	98.23	NA
SGA_v2.0_scaffold4G02372	NO	115	2.61	no_Pfam	no_Pfam	No_hit	No_hit	98.17	NA
SGA_v2.0_scaffold40G15404	NO	1002	2.2	PF14381; PF07714	Ethylene-responsive protein kinase Le-CTR1; Protein tyrosine kinase	AT5G03730	Protein kinase superfamily protein	97.99	GO:0004672, GO:0006468
SGA_v2.0_scaffold5G02773	NO	774	2.2	PF00515; PF13181; PF14559; PF12895; PF07719	Tetratricopeptide repeat; Anaphase-promoting complex: cyclosome: subunit 3; Tetratricopeptide repeat	AT2G20000	CDC27 family protein	97.33	GO:0005515
SGA_v2.0_scaffold280G46868	NO	131	2.29	no_Pfam	no_Pfam	No_hit	No_hit	96.63	NA
SGA_v2.0_scaffold5G02833	NO	786	1.91	PF00069	Protein kinase domain	No_hit	No_hit	96.61	GO:0004672, GO:0005524, GO:0006468
SGA_v2.0_scaffold4G02383	NO	612	3.59	PF02548; PF01535; PF01535; PF01535; PF01535	Ketopantoate hydroxymethyltransferase; PPR repeat	AT3G02330	Pentatricopeptide repeat (PPR) superfamily protein	96.58	GO:0003864, GO:0015940
SGA_v2.0_scaffold163G38075	NO	345	1.45	no_Pfam	no_Pfam	No_hit	No_hit	96.13	NA
SGA_v2.0_scaffold11G05624	NO	352	0.85	no_Pfam	no_Pfam	No_hit	No_hit	95.96	NA
SGA_v2.0_scaffold36G14479	NO	1654	2.24	PF00067	Cytochrome P450	No_hit	No_hit	95.87	GO:0005506, GO:0016705, GO:0020037, GO:0055114
SGA_v2.0_scaffold10G05124	NO	269	2.23	PF12697	Alpha/beta hydrolase family	AT3G50440	methyl esterase 10	95.59	NA
SGA_v2.0_scaffold152G36839	NO	176	4.55	no_Pfam	no_Pfam	No_hit	No_hit	95.56	NA
SGA_v2.0_scaffold24G10341	NO	194	2.58	no_Pfam	no_Pfam	No_hit	No_hit	95.45	NA
SGA_v2.0_scaffold81G26040	NO	218	3.67	no_Pfam	no_Pfam	No_hit	No_hit	95.35	NA
SGA_v2.0_scaffold308G48076	NO	357	1.12	PF14802	TMEM192 family	AT2G06005	FRIGIDA interacting protein 1	95.31	NA
SGA_v2.0_scaffold62G21434	NO	962	1.46	PF04100	Vps53-like: N-terminal	AT1G50500	Membrane trafficking VPS53 family protein	95.30	NA

References

Agrawal, G. K., Jwa, N.-S., Lebrun, M.-H., Job, D., & Rakwal, R. (2010). Plant secretome: Unlocking secrets of the secreted proteins. *Proteomics*, *10*(4), 799–827. https://doi.org/10.1002/pmic.200900514

Aguilar, R., Ashworth, L., Galetto, L., & Aizen, M. A. (2006). Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. *Ecology Letters*, *9*(8), 968–980. https://doi.org/10.1111/j.1461-0248.2006.00927.x

Ai, B., Kang, M., & Huang, H. (2014). Assessment of Genetic Diversity in Seed Plants Based on a Uniform π Criterion. *Molecules*, *19*(12), 20113–20127. https://doi.org/10.3390/molecules191220113

Albenne, C., Canut, H., & Jamet, E. (2013). Plant cell wall proteomics: the leadership of Arabidopsis thaliana. *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00111

Albert, I., Böhm, H., Albert, M., Feiler, C. E., Imkampe, J., Wallmeroth, N., Brancato, C., Raaymakers, T. M., Oome, S., Zhang, H., Krol, E., Grefen, C., Gust, A. A., Chai, J., Hedrich, R., van den Ackerveken, G., & Nürnberger, T. (2015). An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered immunity. *Nature Plants*, *1*(10), 15140. https://doi.org/10.1038/nplants.2015.140

Albert, I., Hua, C., Nürnberger, T., Pruitt, R. N., & Zhang, L. (2020). Surface Sensor Systems in Plant Immunity. *Plant Physiology*, *182*(4), 1582–1596. https://doi.org/10.1104/pp.19.01299

Alexandersson, E., Ashfaq, A., Resjö, S., & Andreasson, E. (2013). Plant secretome proteomics. *Frontiers in Plant Science*, *4*. https://www.frontiersin.org/articles/10.3389/fpls.2013.00009

Amini, M., Saify Nabiabad, H., & Deljou, A. (2018). The role of cuscutain-propeptide inhibitor in haustoria parasitism and enhanced resistance to dodder in transgenic alfalfa expressing this propeptide. *Plant Biotechnology Reports*, *12*(3), 165–173. https://doi.org/10.1007/s11816-018-0481-z

Amusan, I. O., Rich, P. J., Menkir, A., Housley, T., & Ejeta, G. (2008). Resistance to Striga hermonthica in a maize inbred line derived from Zea diploperennis. *New Phytologist*, *178*(1), 157–166. https://doi.org/https://doi.org/10.1111/j.1469-8137.2007.02355.x

Anteyi, W. O., & Rasche, F. (2020). Role and In Vivo Localization of Fusarium oxysporum f. sp. Strigae and Bacillus subtilis in an Integrated Striga hermonthica Biocontrol System. *PhytoFrontiers™*, *1*(1), 51–61. https://doi.org/10.1094/PHYTOFR-08-20-0011-R

Anyolo, A. L. (2015). *Determination of significantly differentially expressed candidate effectors of Striga at stages 3 and 4.1 of infection in the host and their functions*.

Araya, A., Keesstra, S. D., & Stroosnijder, L. (2010). A new agro-climatic classification for crop suitability zoning in northern semi-arid Ethiopia. *Agricultural and Forest Meteorology*, *150*(7), 1057–1064. https://doi.org/10.1016/j.agrformet.2010.04.003

Ashapkin, V. v., Kutueva, L. I., Aleksandrushkina, N. I., Vanyushin, B. F., Teofanova, D. R., & Zagorchev, L. I. (2023). Genomic and Epigenomic Mechanisms of the Interaction between Parasitic and Host Plants. *International Journal of Molecular Sciences*, *24*(3), 2647. https://doi.org/10.3390/ijms24032647

Atera, E. A., Itoh, K., & Onyango, J. C. (2011). Evaluation of ecologies and severity of Striga weed on rice in sub-Saharan Africa. *Agriculture and Biology Journal of North America*, *2*, 752–760.

Atera, E., Itoh, K., Aazuma, T., & Ishii, T. (2012). Farmers' perspectives on the biotic constraint of Striga hermonthica and its control in western Kenya. *Weed Biology and Management*, *12*(1), 53–62. https://doi.org/10.1111/j.1445-6664.2012.00435.x

Ayongwa, G. C., Stomph, T. J., Hoevers, R., Ngoumou, T. N., & Kuyper, T. W. (2010). Striga infestation in northern Cameroon: Magnitude, dynamics and implications for management. *NJAS: Wageningen Journal of Life Sciences*, *57*(2), 159–165. https://doi.org/10.1016/j.njas.2010.04.003

Badu-Apraku, B., Obisesan, O., Abiodun, A., & Obeng-Bio, E. (2021). Genetic Gains from Selection for Drought Tolerance during Three Breeding Periods in Extra-Early Maturing Maize Hybrids under Drought and Rainfed Environments. *Agronomy*, *11*(5), 831. https://doi.org/10.3390/agronomy11050831

Badu-Apraku, B., Adewale, S., Paterne, A., Gedil, M., & Asiedu, R. (2020). Identification of QTLs Controlling Resistance/Tolerance to Striga hermonthica in an Extra-Early Maturing Yellow Maize Population. *Agronomy*, *10*(8), 1168. https://doi.org/10.3390/agronomy10081168

Baird, Wm. V., & Riopel, J. L. (1984). Experimental studies of haustorium initiation and early development in Agalinis purpurea (L.) RAF. (Scrophulariaceae). *American Journal of Botany*, *71*(6), 803–814. https://doi.org/10.1002/j.1537-2197.1984.tb14145.x

Baird, Wm. V., & Riopel, J. L. (1985). Surface Characteristics of Root and Haustorial Hairs of Parasitic Scrophulariaceae. *Botanical Gazette*, *146*(1), 63–69. http://www.jstor.org.sheffield.idm.oclc.org/stable/2474479

Balint-Kurti, P. (2019). The plant hypersensitive response: concepts, control and consequences. *Mol Plant Pathol*, *20*(8), 1163–1178. https://doi.org/10.1111/mpp.12821

Balkenhol, N., Dudaniec, R. Y., Krutovsky, K. v., Johnson, J. S., Cairns, D. M., Segelbacher, G., Selkoe, K. A., von der Heyden, S., Wang, I. J., Selmoni, O., & Joost, S. (2017). *Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations* (pp. 261–322). https://doi.org/10.1007/13836_2017_2

Barkman, T., Klooster, M., Gaddis, K., Franzone, B., Lavigne, S., Manickam, S., Vessabutr, S., Sasirat, S., & Davis, C. (2017). Reading between the vines: Hosts as islands for extreme holoparasitic plants. *American Journal of Botany*, *104*. https://doi.org/10.3732/ajb.1700117

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauters, L., Kyndt, T., de Meyer, T., Morreel, K., Boerjan, W., Lefevere, H., & Gheysen, G. (2020). Chorismate mutase and isochorismatase, two potential effectors of the migratory nematode Hirschmanniella oryzae, increase host susceptibility by manipulating secondary metabolite content of rice. *Molecular Plant Pathology*, *21*(12), 1634–1646. https://doi.org/10.1111/mpp.13003

Beardon, E. (2018). *Discovery and analysis of novel resistance genes in rice to the parasitic weed Striga hermonthica*.

Bebawi, F. F., Eplee, R. E., Harris, C. E., & Norris, R. S. (1984). Longevity of Witchweed (Striga asiatica) Seed. *Weed Science*, *32*(4), 494–497. http://www.jstor.org/stable/4044010

Bekunda, M., Sanginga, N., & Woomer, P. L. (2010). Chapter Four - Restoring Soil Fertility in Sub-Sahara Africa. *Advances in Agronomy*, *108*, 183–236. https://doi.org/10.1016/S0065-2113(10)08004-1

Bellincampi, D., Cervone, F., & Lionetti, V. (2014). Plant cell wall dynamics and wall-related susceptibility in plant-pathogen interactions. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00228

Bellis, E. S., McLaughlin, C. M., dePamphilis, C. W., & Lasky, J. R. (2021). The geography of parasite local adaptation to host communities. *Ecography*, *44*(8), 1205–1217. https://doi.org/10.1111/ecog.05730

Bellis, E. S., von Münchow, C. S., Odero, C. O., Kronberger, A., Kelly, E., Xia, T., Huang, X., Wicke, S., Runo, S. M., dePamphilis, C. W., & Lasky, J. R. (2022). Genomic signatures of host-specific selection in a parasitic plant. *BioRxiv*, 2022.02.01.478712. https://doi.org/10.1101/2022.02.01.478712

Benedetti, M., Pontiggia, D., Raggi, S., Cheng, Z., Scaloni, F., Ferrari, S., Ausubel, F. M., Cervone, F., & de Lorenzo, G. (2015). Plant immunity triggered by engineered in vivo release of oligogalacturonides, damage-associated molecular patterns. *Proceedings of the National Academy of Sciences*, *112*(17), 5533–5538. https://doi.org/10.1073/pnas.1504154112

Berner, D. K., Kling, J. G., & Singh, B. B. (1995). Striga Research and Control: A Perspective from Africa. *Plant Disease*, *79*(7), 652–660. https://doi.org/10.1094/PD-79-0652

Bharathalakshmi, & Jayachandra, R. (1979). Physiological variations in Striga asiatica. *2nd Int. Symp. Parasitic Weeds*, 132–143.

Bharathalakshmi, Werth, C. R., & Musselman, L. J. (1990). A study of genetic diversity among host-specific populations of the witchweed Striga hermonthica (Scrophulariaceae) in Africa. *Plant Systematics and Evolution*, *172*(1/4), 1–12.

Blackman, L. M., Cullerne, D. P., Torreña, P., Taylor, J., & Hardham, A. R. (2015). RNA-Seq Analysis of the Expression of Genes Encoding Cell Wall Degrading Enzymes during Infection of Lupin (Lupinus angustifolius) by Phytophthora parasitica. *PLOS ONE*, *10*(9), e0136899-. https://doi.org/10.1371/journal.pone.0136899

Bleischwitz, M., Albert, M., Fuchsbauer, H.-L., & Kaldenhoff, R. (2010). Significance of Cuscutain, a cysteine protease from Cuscuta reflexa, in host-parasite interactions. *BMC Plant Biology*, *10*(1), 227. https://doi.org/10.1186/1471-2229-10-227

Block, A., Li, G., Fu, Z. Q., & Alfano, J. R. (2008). Phytopathogen type III effector weaponry and their plant targets. *Current Opinion in Plant Biology*, *11*(4), 396–403. https://doi.org/10.1016/j.pbi.2008.06.007

Booker, T. R., Yeaman, S., Whiting, J. R., & Whitlock, M. C. (2023). The WZA: A window-based method for characterizing genotype-environment association. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13768

Botanga, C., Kling, J., K. Berner, D., & Timko, M. (2002). Genetic variability of Striga asiatica (L.) Kuntz based on AFLP analysis and host-parasite interaction. In *Euphytica* (Vol. 128). https://doi.org/10.1023/A:1021200213599

Botanga, C., & Timko, M. (2006). Phenetic relationships among different races of Striga gesnerioides (Willd.) Vatke from West Africa. *Genome / National Research Council Canada = Génome / Conseil National de Recherches Canada*, *49*, 1351–1365. https://doi.org/10.1139/g06-086

Boukteb, A., Sakaguchi, S., Ichihashi, Y., Kharrat, M., Nagano, A., Shirasu, K., & Bouhadida, M. (2021). Analysis of Genetic Diversity and Population Structure of Orobanche foetida Populations From Tunisia Using RADseq. *Front Plant Sci*, *12*(618245). https://doi.org/10.3389/fpls.2021.618245

Bouwmeester, H., Sinha, N., & Scholes, J. (2021). Parasitic plants: physiology, development, signaling, and ecosystem interactions. *Plant Physiology*, *185*(4), 1267–1269. https://doi.org/10.1093/plphys/kiab055

Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, *108*(10), 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005

Buckingham, L. J., & Ashby, B. (2022). Coevolutionary theory of hosts and parasites. *Journal of Evolutionary Biology*, *35*(2), 205–224. /https://doi.org/10.1111/jeb.13981

Burgarella, C., & Glémin, S. (2017). Population Genetics and Genome Evolution of Selfing Species. In *eLS* (pp. 1–8). https://doi.org/https://doi.org/10.1002/9780470015902.a0026804

Butler, L. G. (1994). Chemical Communication Between the Parasitic Weed Striga and Its Crop Host. In *Allelopathy* (Vol. 582, pp. 158–168). American Chemical Society. https://doi.org/10.1021/bk-1995-0582.ch012

Caffall, K. H., & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research*, *344*(14), 1879–1900. https://doi.org/10.1016/j.carres.2009.05.021

Calderan-Rodrigues, M. J., Guimarães Fonseca, J., de Moraes, F. E., Vaz Setem, L., Carmanhanis Begossi, A., & Labate, C. A. (2019). Plant Cell Wall Proteomics: A Focus on Monocot Species, Brachypodium distachyon, Saccharum spp. and Oryza sativa. *International Journal of Molecular Sciences*, *20*(8), 1975. https://doi.org/10.3390/ijms20081975

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298–2309. https://doi.org/10.1111/2041-210X.13722

Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J., & Nickerson, D. A. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research*, *15*(11), 1553–1565. https://doi.org/10.1101/gr.4326505

Carreón-Anguiano, K. G., Islas-Flores, I., Vega-Arreguín, J., Sáenz-Carbonell, L., & Canto-Canché, B. (2020). EffHunter: A Tool for Prediction of Effector Protein Candidates in Fungal Proteomic Databases. *Biomolecules*, *10*(5), 712. https://doi.org/10.3390/biom10050712

Chabaud, M., Auriac, M.-C., Boniface, M.-C., Delgrange, S., Folletti, T., Jardinaud, M.-F., Legendre, A., Pérez-Vich, B., Pouvreau, J.-B., Velasco, L., Delavault, P., & Muños, S. (2022). Wild Helianthus species: A reservoir of resistance genes for sustainable pyramidal resistance to broomrape in sunflower. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.1038684

Charlesworth, B., Charlesworth, D., & Barton, N. H. (2003). The Effects of Genetic and Geographic Structure on Neutral Variation. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 99–125. https://doi.org/10.1146/annurev.ecolsys.34.011802.132359

Chauvin, N. D., Mulangu, F., & Porto, G. (2012). *Food Production and Consumption Trends in Sub-Saharan Africa: Prospects for the Transformation of the Agricultural Sector* (No. 2012–011; African Human Development Report).

Chen, Y., Liu, Z., & Halterman, D. A. (2012). Correction: Molecular Determinants of Resistance Activation and Suppression by Phytophthora infestans Effector IPI-O. *PLoS Pathogens*, *8*(9). https://doi.org/10.1371/annotation/75775518-f06e-4148-a639-31cfc6972b2e

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., Zhang, X., Wang, J., Yang, H., Fang, L., & Chen, Q. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience*, *7*(1). https://doi.org/10.1093/gigascience/gix120

Chikoye, D., Fontem, L. A., & Menkir, A. (2011). Seed coating herbicide tolerant maize hybrids with imazapyr for Striga hermonthica (Del.) Benth control in the West African savanna. *Ournal of Food, Agriculture & Environment*, *9*(1), 416–421.

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F. M., Ponce, M. R., Micol, J. L., & Solano, R. (2007). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature*, *448*(7154), 666–671. https://doi.org/10.1038/nature06006

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012a). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012b). Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics*, *3*. https://doi.org/10.3389/fgene.2012.00035

Cissoko, M., Boisnard, A., Rodenburg, J., Press, M. C., & Scholes, J. D. (2011). New Rice for Africa (NERICA) cultivars exhibit different levels of post-attachment resistance against the parasitic weeds Striga hermonthica and Striga asiatica. *New Phytologist*, *192*(4), 952–963. https://doi.org/10.1111/j.1469-8137.2011.03846.x

Clarke, C. R., Park, S.-Y., Tuosto, R., Jia, X., Yoder, A., van Mullekom, J., & Westwood, J. (2020). Multiple immunity-related genes control susceptibility of *Arabidopsis thaliana* to the parasitic weed *Phelipanche aegyptiaca*. *PeerJ*, *8*, e9268. https://doi.org/10.7717/peerj.9268

Cochrane, V., & Press, M. C. (1997). Geographical distribution and aspects of the ecology of the hemiparasitic angiosperm Striga asiatica (L.) Kuntze: a herbarium study. *Journal of Tropical Ecology*, *13*, 371–380.

Collemare, J., O'Connell, R., & Lebrun, M.-H. (2019). Nonproteinaceous effectors: the terra incognita of plant–fungal interactions. *New Phytologist*, *223*(2), 590–596. https://doi.org/10.1111/nph.15785

Cressler CE, McLeod DV, Rozins C, van den Hoogen J, & Day T. (2016). The adaptive evolution of virulence: a review of theoretical predictions and empirical tests. *Parasitology*, *143*(7), 915–930. https://doi.org/10.1017/S003118201500092X

Csurhes, S., Markula, A., & Zhou, Y. (2016). *Invasive plant risk assessment: Witchweeds Striga species*.

Cui, S., Wakatake, T., Hashimoto, K., Saucet, S. B., Toyooka, K., Yoshida, S., & Shirasu, K. (2016). Haustorial Hairs Are Specialized Root Hairs That Support Parasitism in the Facultative Parasitic Plant *Phtheirospermum japonicum*. *Plant Physiology*, *170*(3), 1492–1503. https://doi.org/10.1104/pp.15.01786

Cui, S., Wada, S., Tobimatsu, Y., Takeda, Y., Saucet, S. B., Takano, T., Umezawa, T., Shirasu, K., & Yoshida, S. (2018). Host lignin composition affects haustorium induction in the parasitic plants Phtheirospermum japonicum and Striga hermonthica. *New Phytologist*, *218*(2), 710–723. https://doi.org/10.1111/nph.15033

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Dauphin, B., Rellstab, C., Wüest, R. O., Karger, D. N., Holderegger, R., Gugerli, F., & Manel, S. (2023). Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution*, *38*(3), 261–274. https://doi.org/https://doi.org/10.1016/j.tree.2022.10.010

David, O. A., Obiakara, M. C., Fabolude, G. O., Akomolafe, G. F., & Ajiboye, M. D. (2022). Habitat suitability and dispersal of invasive *Striga* species under climate change in Africa. *African Journal of Ecology*, *60*(4), 1143–1154. https://doi.org/10.1111/aje.13064

Dayou, O., Kibet, W., Ojola, P., Gangashetty, P. I., Wicke, S., & Runo, S. (2021). Two-tier witchweed (Striga hermonthica) resistance in wild pearl millet (Pennisetum glaucum) 29Aw. *Weed Science*, *69*(3), 300–306. https://doi.org/DOI: 10.1017/wsc.2021.12

de Pedro-Jové, R., Puigvert, M., Sebastià, P., Macho, A. P., Monteiro, J. S., Coll, N. S., Setúbal, J. C., & Valls, M. (2021). Dynamic expression of Ralstonia solanacearum virulence factors and metabolism-controlling genes during plant infection. *BMC Genomics*, *22*(1), 170. https://doi.org/10.1186/s12864-021-07457-w

Dell'Oste, V., Spyrakis, F., & Prandi, C. (2021). Strigolactones, from Plants to Human Health: Achievements and Challenges. *Molecules*, *26*(15), 4579. https://doi.org/10.3390/molecules26154579

Deshpande, S. P., Mohamed, A., & Thomas Hash Jr., C. (2013). Molecular Breeding for Striga Resistance in Sorghum. In *Translational Genomics for Crop Breeding* (pp. 77–93). https://doi.org/10.1002/9781118728475.ch6

Dick, C. W., & Patterson, B. D. (2007). Against all odds: Explaining high host specificity in dispersal-prone parasites. *International Journal for Parasitology*, *37*(8), 871–876. https://doi.org/10.1016/j.ijpara.2007.02.004

Dieringer, G., & Werth, C. R. (1994). Population Genetic Structure of Agalinis strictifolia (Scrophulariaceae). *Bulletin of the Torrey Botanical Club*, *121*(2), 148–153. https://doi.org/10.2307/2997166

Dow, M., Newman, M.-A., & von Roepenack, E. (2000). The Induction and Modulation of Plant Defense Responses by Bacterial Lipopolysaccharides. *Annual Review of Phytopathology*, *38*(1), 241–261. https://doi.org/10.1146/annurev.phyto.38.1.241

Duriez, P., Vautrin, S., Auriac, M.-C., Bazerque, J., Boniface, M.-C., Callot, C., Carrère, S., Cauet, S., Chabaud, M., Gentou, F., Lopez-Sendon, M., Paris, C., Pegot-Espagnet, P., Rousseaux, J.-C., Pérez-Vich, B., Velasco, L., Bergès, H., Piquemal, J., & Muños, S. (2019). A receptor-like kinase enhances sunflower resistance to Orobanche cumana. *Nature Plants*, *5*(12), 1211–1215. https://doi.org/10.1038/s41477-019-0556-z

Ejeta, G. (2007). The Striga scourge in Africa: A growing pandemic. In *Integrating New Technologies for Striga Control* (pp. 3–16). WORLD SCIENTIFIC. https://doi.org/10.1142/9789812771506_0001

Ekawa, M., & Aoki, K. (2017). Phloem-Conducting Cells in Haustoria of the Root-Parasitic Plant Phelipanche aegyptiaca Retain Nuclei and Are Not Mature Sieve Elements. *Plants*, *6*(4), 60. https://doi.org/10.3390/plants6040060

Emechebe, A. M., Ellis-Jones, J., Schulz, S., Chikoye, D., Douthewaite, B., Kureh, I., Tarawali, G., Hussaini, M. A., Kormawa, P., & Sanni, A. (2004). Farmer's perception of the Striga problem and its control in Northern Nigeria. *Experimental Agriculture*, *40*(2), 215–232. https://doi.org/DOI: 10.1017/S0014479703001601

Eplee, R. (1983). Progress on control of Striga asiatica in the United States. In V. Kumble (Ed.), *2nd International Striga Workshop*. IDRC/ICRISAT.

Eplee, R. E. (1992). Witchweed (Striga asiatica): an overview of management strategies in the USA. *Crop Protection*, *11*(1), 3–7. https://doi.org/https://doi.org/10.1016/0261-2194(92)90071-C

Eshel, D., Miyara, I., Ailing, T., Dinoor, A., & Prusky, D. (2002). pH Regulates Endoglucanase Expression and Virulence of *Alternaria alternata* in Persimmon Fruit. *Molecular Plant-Microbe Interactions®*, *15*(8), 774–779. https://doi.org/10.1094/MPMI.2002.15.8.774

Fatima, U., & Senthil-Kumar, M. (2015). Plant and pathogen nutrient acquisition strategies. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00750

Felix, G., Regenass, M., & Boller, T. (1993). Specific perception of subnanomolar concentrations of chitin fragments by tomato cells: induction of extracellular alkalinization, changes in protein phosphorylation, and establishment of a refractory state. *The Plant Journal*, *4*(2), 307–316. https://doi.org/10.1046/j.1365-313X.1993.04020307.x

Fernández-Aparicio, M., Pérez-de-Luque, A., Prats, E., & Rubiales, D. (2008). Variability of interactions between barrel medic (Medicago truncatula) genotypes and Orobanche species. *Annals of Applied Biology*, *153*(1), 117–126. https://doi.org/10.1111/j.1744-7348.2008.00241.x

Ferrari, S. (2013). Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development. *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00049

Fishman, L., & Wyatt, R. (1999). Pollinator-Mediated Competition, Reproductive Character Displacement, and the Evolution of Selfing in Arenaria uniflora (Caryophyllaceae). *Evolution*, *53*(6), 1723. https://doi.org/10.2307/2640435

Flor, H. H. (1971). Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology*, *9*(1), 275–296. https://doi.org/10.1146/annurev.py.09.090171.001423

Flury, P., Klauser, D., Schulze, B., Boller, T., & Bartels, S. (2013). The Anticipation of Danger: Microbe-Associated Molecular Pattern Perception Enhances AtPep-Triggered Oxidative Burst . *Plant Physiology*, *161*(4), 2023–2035. https://doi.org/10.1104/pp.113.216077

Foll, M., Fischer, M. C., Heckel, G., & Excoffier, L. (2010). Estimating population structure from AFLP amplification intensity. *Molecular Ecology*, *19*(21), 4638–4647. https://doi.org/10.1111/j.1365-294X.2010.04820.x

Fortuna, M. A., Albaladejo, R. G., Fernández, L., Aparicio, A., & Bascompte, J. (2009). Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences*, *106*(45), 19044–19049. https://doi.org/10.1073/pnas.0907704106

Fraik, A. K., Margres, M. J., Epstein, B., Barbosa, S., Jones, M., Hendricks, S., Schönfeld, B., Stahlke, A. R., Veillet, A., Hamede, R., McCallum, H., Lopez-Contreras, E., Kallinen, S. J., Hohenlohe, P. A., Kelley, J. L., & Storfer, A. (2020). Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (Sarcophilus harrisii) populations. *Evolution*, *74*(7), 1392–1408. https://doi.org/10.1111/evo.14023

García-Garrido JM, Lendzemo V, Castellanos-Morales V, Steinkellner S, & Vierheilig H. (2009). Strigolactones, signals for parasitic plants and arbuscular mycorrhizal fungi. *Mycorrhiza*, *19*(7), 449–459. https://doi.org/10.1007/s00572-009-0265-y

García-Dorado, A. (2012). Understanding and Predicting the Fitness Decline of Shrunk Populations: Inbreeding, Purging, Mutation, and Standard Selection. *Genetics*, *190*, 1461–1476. https://doi.org/10.1534/genetics.111.135541

Gasura, E., Setimela, P., Mabasa, S., Rwafa, R., Kageler, S., & Nyakurwa, C. (2019). Response of IITA maize inbred lines bred for Striga hermonthica resistance to Striga asiatica and associated resistance mechanisms in southern Africa. *Euphytica*, *215*(10), 151. https://doi.org/10.1007/s10681-019-2467-5

Gautier, C., Fournet, S., Piriou, C., Renault, L., Yvin, J.-C., Nguema-Ona, E., Grenier, E., & Montarry, J. (2020). Plant–parasite coevolution: A weak signature of local adaptation between Peruvian Globodera pallida populations and wild potatoes. *Ecology and Evolution*, *10*(9), 4156–4163. https://doi.org/https://doi.org/10.1002/ece3.6248

Gbèhounou, G., Pieterse, A. H., & Verkleij, J. A. C. (2003). Longevity of Striga seeds reconsidered: results of a field study on purple witchweed (Striga hermonthica) in Bénin. *Weed Science*, *51*(6), 940–946. https://doi.org/DOI: 10.1614/ws03-0221r1

Gethi, J. G., Smith, M. E., Mitchell, S. E., & Kresovich, S. (2005). Genetic diversity of Striga hermonthica and Striga asiatica populations in Kenya. *Weed Research*, *45*(1), 64–73.

Gethi, J. G., & Smith, M. E. (2004). Genetic Responses of Single Crosses of Maize to Striga hermonthica (Del.) Benth. and Striga asiatica (L.) Kuntze. *Crop Science*, *44*(6), 2068–2077. https://doi.org/https://doi.org/10.2135/cropsci2004.2068

Gimenez-Ibanez, S., Boter, M., Fernández-Barbero, G., Chini, A., Rathjen, J. P., & Solano, R. (2014). The Bacterial Effector HopX1 Targets JAZ Transcriptional Repressors to Activate Jasmonate Signaling and Promote Infection in Arabidopsis. *PLoS Biology*, *12*(2), e1001792. https://doi.org/10.1371/journal.pbio.1001792

Gimenez-Ibanez, S., Hann, D. R., Ntoukakis, V., Petutschnig, E., Lipka, V., & Rathjen, J. P. (2009). AvrPtoB Targets the LysM Receptor Kinase CERK1 to Promote Bacterial Virulence on Plants. *Current Biology*, *19*(5), 423–429. https://doi.org/10.1016/j.cub.2009.01.054

Gobena, D., Shimels, M., Rich, P. J., Ruyter-Spira, C., Bouwmeester, H., Kanuganti, S., Mengiste, T., & Ejeta, G. (2017). Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance. *Proceedings of the National Academy of Sciences*, *114*(17), 4471–4476. https://doi.org/10.1073/pnas.1618965114

González-Verdejo, C. I., Barandiaran, X., Moreno, M. T., Cubero, J. I., & di Pietro, A. (2006). A peroxidase gene expressed during early developmental stages of the parasitic plant Orobanche ramosa. *Journal of Experimental Botany*, *57*(1), 185–192. https://doi.org/10.1093/jxb/erj024

Goyet, V., Wada, S., Cui, S., Wakatake, T., Shirasu, K., Montiel, G., Simier, P., & Yoshida, S. (2019). Haustorium Inducing Factors for Parasitic Orobanchaceae. *Frontiers in Plant Science*, *10*. https://www.frontiersin.org/articles/10.3389/fpls.2019.01056

Gu, X., Chen, I.-G., Harding, S. A., Nyamdari, B., Ortega, M. A., Clermont, K., Westwood, J. H., & Tsai, C.-J. (2021). Plasma membrane phylloquinone biosynthesis in nonphotosynthetic parasitic plants. *Plant Physiology*, *185*(4), 1443–1456. https://doi.org/10.1093/plphys/kiab031

Gu, F., Yung, Y.-F., Cheung, M. W.-L., Joo, B.-K. (Brian), & Nimon, K. (2022). Statistical Inference in Redundancy Analysis: A Direct Covariance Structure Modeling Approach. *Multivariate Behavioral Research*, 1–17. https://doi.org/10.1080/00273171.2022.2141675

Gurney, A. L., Press, M. C., & Scholes, J. D. (1999). Infection Time and Density Influence the Response of Sorghum to the Parasitic Angiosperm Striga hermonthica. *The New Phytologist*, *143*(3), 573–580. http://www.jstor.org/stable/2588330

Gurney AL, Slate J, Press MC, & Scholes JD. (2006). A novel form of resistance in rice to the angiosperm parasite Striga hermonthica. *New Phytol.*, *169*(1), 199–208. https://doi.org/10.1111/j.1469-8137.2005.01560.x

Hadley, W. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag.

Hailu, G., Niassy, S., Zeyaur, K. R., Ochatum, N., & Subramanian, S. (2018). Maize–Legume Intercropping and Push–Pull for Management of Fall Armyworm, Stemborers, and Striga in Uganda. *Agronomy Journal*, *110*(6), 2513–2522. https://doi.org/10.2134/agronj2018.02.0110

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*, 618–620.

Harholt, J., Suttangkakul, A., & Vibe Scheller, H. (2010). Biosynthesis of Pectin. *Plant Physiology*, *153*(2), 384–395. https://doi.org/10.1104/pp.110.156588

Harris, K. (2018). The randomness that shapes our DNA. *ELife*, *7*. https://doi.org/10.7554/eLife.41491

Haussmann, B. I. G., Hess, D. E., Omanya, G. O., Folkertsma, R. T., Reddy, B. V. S., Kayentao, M., Welz, H. G., & Geiger, H. H. (2004). Genomic regions influencing resistance to the parasitic weed Striga hermonthica in two recombinant inbred populations of sorghum. *Theoretical and Applied Genetics*, *109*(5), 1005–1016. https://doi.org/10.1007/s00122-004-1706-9

He, X., Wang, C., Wang, H., Li, L., & Wang, C. (2020). The Function of MAPK Cascades in Response to Various Stresses in Horticultural Plants. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00952

Hegenauer, V., Fürst, U., Kaiser, B., Smoker, M., Zipfel, C., Felix, G., Stahl, M., & Albert, M. (2016). Detection of the plant parasite Cuscuta reflexa by a tomato cell surface receptor. *Science*, *353*(6298), 478 LP – 481.

Hegenauer, V., Slaby, P., Körner, M., Bruckmüller, J.-A., Burggraf, R., Albert, I., Kaiser, B., Löffelhardt, B., Droste-Borel, I., Sklenar, J., Menke, F. L. H., Maček, B., Ranjan, A., Sinha, N., Nürnberger, T., Felix, G., Krause, K., Stahl, M., & Albert, M. (2020). The tomato receptor CuRe1 senses a cell wall protein to identify Cuscuta as a pathogen. *Nature Communications*, *11*(1), 5299. https://doi.org/10.1038/s41467-020-19147-4

Heide-Jørgensen, H. S. (2013). Introduction: The Parasitic Syndrome in Higher Plants. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.), *Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies* (pp. 1–18). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38146-1_1

Hemetsberger, C., Herrberger, C., Zechmann, B., Hillmer, M., & Doehlemann, G. (2012). The Ustilago maydis Effector Pep1 Suppresses Plant Immunity by Inhibition of Host Peroxidase Activity. *PLOS Pathogens*, *8*(5), e1002684-. https://doi.org/10.1371/journal.ppat.1002684

Hijmans, R. J., Karney, C., Williams, E., & Vennes, C. (2022). *R Package "Geosphere"* (1.5-18).

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoey, J. A., & Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evolutionary Applications*, *11*(9), 1732–1747. https://doi.org/10.1111/eva.12676

Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, *21*(6), 797–807. https://doi.org/10.1007/s10980-005-5245-9

Honaas, L. A., Wafula, E. K., Yang, Z., Der, J. P., Wickett, N. J., Altman, N. S., Taylor, C. G., Yoder, J. I., Timko, M. P., Westwood, J. H., & dePamphilis, C. W. (2013). Functional genomics of a generalist parasitic plant: Laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression. *BMC Plant Biology*, *13*(1), 9. https://doi.org/10.1186/1471-2229-13-9

Hosmani, M. M. (1978). *Striga (a noxious root parasitic weed)*. India Royal Printers.

Hu, L., Wang, J., Yang, C., Islam, F., Bouwmeester, H., Munos, S., & Zhou, W. (2020). The Effect of Virulence and Resistance Mechanisms on the Interactions between Parasitic Plants and Their Hosts. *Int J Mol Sci.*, *21*(23). https://doi.org/10.3390/ijms21239013

Huang, K., Whitlock, R., Press, M. C., & Scholes, J. D. (2012). Variation for host range within and among populations of the parasitic plant Striga hermonthica. *Heredity*, *108*(2), 96–104. https://doi.org/10.1038/hdy.2011.52

Huang, R., Chu, Q.-H., Lu, G.-H., & Wang, Y.-Q. (2019). Comparative studies on population genetic structure of two closely related selfing and outcrossing Zingiber species in Hainan Island. *Scientific Reports*, *9*(1), 17997. https://doi.org/10.1038/s41598-019-54526-y

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, X., Zhu, G., Liu, Q., Chen, L., Li, Y., & Hou, B. (2018). Modulation of Plant Salicylic Acid-Associated Immune Responses via Glycosylation of Dihydroxybenzoic Acids. *Plant Physiology*, *176*(4), 3103–3119. https://doi.org/10.1104/pp.17.01530

Hussaini, M., Kling, J., BERNER , D., & Ikie, F. (2003). Evaluation of integrated Striga hermonthica control technologies under farmer management. *Experimental Agriculture*, *39*, 99–108. https://doi.org/10.1017/S0014479702001084

Hyten, D. L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R. L., Costa, J. M., Specht, J. E., Shoemaker, R. C., & Cregan, P. B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences*, *103*(45), 16666–16671. https://doi.org/10.1073/pnas.0604379103

Jamil, M., Charnikhova, T., Houshyani, B., van Ast, A., & Bouwmeester, H. J. (2012). Genetic variation in strigolactone production and tillering in rice and its effect on Striga hermonthica infection. *Planta*, *235*(3), 473–484. https://doi.org/10.1007/s00425-011-1520-y

Jamil, M., Charnikhova, T., Verstappen, F., & Bouwmeester, H. (2010). Carotenoid inhibitors reduce strigolactone production and Striga hermonthica infection in rice. *Archives of Biochemistry and Biophysics*, *504*(1), 123–131. https://doi.org/10.1016/j.abb.2010.08.005

Jamil, M., Kountche, B. A., & Al-Babili, S. (2021). Current progress in Striga management. *Plant Physiology*, *185*(4), 1339–1352. https://doi.org/10.1093/plphys/kiab040

Jamil, M., Rodenburg, J., Charnikhova, T., & Bouwmeester, H. J. (2011). Pre-attachment Striga hermonthica resistance of New Rice for Africa (NERICA) cultivars based on low strigolactone production. *New Phytologist*, *192*(4), 964–975. https://doi.org/10.1111/j.1469-8137.2011.03850.x

Jiang, Z., Zhao, Q., Bai, R., Yu, R., Diao, P., Yan, T., Duan, H., Ma, X., Zhou, Z., Fan, Y., & Wuriyanghan, H. (2021). Host sunflower-induced silencing of parasitism-related genes confers resistance to invading *Orobanche cumana*. *Plant Physiology*, *185*(2), 424–440. https://doi.org/10.1093/plphys/kiaa018

Jhu, M.-Y., Farhi, M., Wang, L., Zumstein, K., & Sinha, N. R. (2022). Investigating Host and Parasitic Plant Interaction by Tissue-Specific Gene Analyses on Tomato and Cuscuta campestris Interface at Three Haustorial Developmental Stages. *Frontiers in Plant Science*, *12*. https://www.frontiersin.org/articles/10.3389/fpls.2021.764843

Jhu, M.-Y., Ichihashi, Y., Farhi, M., Wong, C., & Sinha, N. R. (2021). Lateral organ boundaries domain 25 functions as a key regulator of haustorium development in dodders. *Plant Physiology*, *186*(4), 2093–2110. https://doi.org/10.1093/plphys/kiab231

Jiang, Z., Zhao, Q., Bai, R., Yu, R., Diao, P., Yan, T., Duan, H., Ma, X., Zhou, Z., Fan, Y., & Wuriyanghan, H. (2021). Host sunflower-induced silencing of parasitism-related genes confers resistance to invading *Orobanche cumana*. *Plant Physiology*, *185*(2), 424–440. https://doi.org/10.1093/plphys/kiaa018

Joel, K., Runo, S., & Muchugi, A. (2018). Genetic diversity and virulence of Striga hermonthica from Kenya and Uganda on selected sorghum varieties. *Nusantara Bioscience*, *10*(2), 111–120. https://doi.org/10.13057/nusbiosci/n100208

Joel, D. M., Hershenhorn, J., Eizenberg, H., Aly, R., Ejeta, G., Rich, P. J., Ransom, J. K., Sauerborn, J., & Rubiales, D. (2007). Biology and Management of Weedy Root Parasites. In *Horticultural Reviews* (pp. 267–349). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470168011.ch4

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Jordal, B. H., Beaver, R. A., & Kirkendall, L. R. (2001). Breaking taboos in the tropics: incest promotes colonization by wood-boring beetles. *Global Ecology and Biogeography*, *10*(4), 345–357. https://doi.org/10.1046/j.1466-822X.2001.00242.x

Kaewchumnong, K., & Price, A. H. (2008). A study on the susceptibility of rice cultivars to Striga hermonthica and mapping of Striga tolerance quantitative trait loci in rice. *New Phytologist*, *180*(1), 206–216. https://doi.org/10.1111/j.1469-8137.2008.02568.x

Kaloshian, I. (2004). Gene-for-gene disease resistance: bridging insect pest and pathogen defence. *Journal of Chemical Ecology*, *30*(12), 2419–2438. https://doi.org/10.1007/s10886-004-7943-1

Kamara, A. Y., Menkir, A., Chikoye, D., Solomon, R., Tofa, A. I., & Omoigui, L. O. (2020). Seed Dressing Maize with Imazapyr to Control Striga hermonthica in Farmers' Fields in the Savannas of Nigeria. *Agriculture*, *10*(3), 83. https://doi.org/10.3390/agriculture10030083

Kamaya, Y., Nakatsubo, F., Higuchi, T., & Iwahara, S. (1981). Degradation of d,l-syringaresinol, a β-β' linked lignin model compound, by Fusarium solani M-13-1. *Archives of Microbiology*, *129*(4), 305–309. https://doi.org/10.1007/BF00414702

Kamoun, S. (2006). A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. *Annual Review of Phytopathology*, *44*(1), 41–60. https://doi.org/10.1146/annurev.phyto.44.070505.143436

Kanampiu, F., Diallo, A., & Karaya, H. (2007). Herbicide-seed coating technology: A unique approach for Striga control in maize. In K. Z. Ahmed (Ed.), *African Crop Science Conference Proceedings* (Vol. 8, pp. 1095–1098).

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241.

Kebede, M., & Ayana, B. (2018). Economically Important Parasitic Weeds and Their Management Practices in Crops. *Journal of Environment and Earth Science*, *8*(12), 104–115.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Keller, H., Boyer, L., & Abad, P. (2016). Disease susceptibility in the Zig-Zag model of host-microbe interactions: only a consequence of immune suppression? *Molecular Plant Pathology*, *17*(4), 475–479. https://doi.org/10.1111/mpp.12371

Kim, C.-Y., Park, J.-Y., Choi, G., Kim, S., Vo, K. T. X., Jeon, J.-S., Kang, S., & Lee, Y.-H. (2022). A rice gene encoding glycosyl hydrolase plays contrasting roles in immunity depending on the type of pathogens. *Molecular Plant Pathology*, *23*(3), 400–416. https://doi.org/10.1111/mpp.13167

King, L. J. (1966). *Weeds of the World. Biology and Control.* Interscience.

Klessig, D. F., Choi, H. W., & Dempsey, D. A. (2018). Systemic Acquired Resistance and Salicylic Acid: Past, Present, and Future. *Molecular Plant-Microbe Interactions®*, *31*(9), 871–888. https://doi.org/10.1094/MPMI-03-18-0067-CR

Kloth, K. J., Abreu, I. N., Delhomme, N., Petřík, I., Villard, C., Ström, C., Amini, F., Novák, O., Moritz, T., & Albrectsen, B. R. (2019). Pectin acetylesterase9 Affects the Transcriptome and Metabolome and Delays Aphid Feeding. *Plant Physiology*, *181*(4), 1704–1720. https://doi.org/10.1104/pp.19.00635

Koh, S., André, A., Edwards, H., Ehrhardt, D., & Somerville, S. (2005). Arabidopsis thaliana subcellular responses to compatible Erysiphe cichoracearum infections. *The Plant Journal*, *44*(3), 516–529. https://doi.org/10.1111/j.1365-313X.2005.02545.x

Kokla, A., & Melnyk, C. W. (2018). Developing a thief: Haustoria formation in parasitic plants. *Developmental Biology*, *442*(1), 53–59. https://doi.org/10.1016/j.ydbio.2018.06.013

Koskela, Salonen, & Mutikainen. (2000). Local adaptation of a holoparasitic plant, Cuscuta europaea: variation among populations. *Journal of Evolutionary Biology*, *13*(5), 749–755. https://doi.org/https://doi.org/10.1046/j.1420-9101.2000.00226.x

Koskela, T., Salonen, V., & Mutikainen, P. (2001). Interaction of a host plant and its holoparasite: effects of previous selection by the parasite. *Journal of Evolutionary Biology*, *14*(6), 910–917. https://doi.org/10.1046/j.1420-9101.2001.00352.x

Kover, P., & Schaal BA. (2002). Genetic variation for disease resistance and tolerance among Arabidopsis thaliana accessions. *Proc Natl Acad Sci USA*, *99*(17), 11270–11274. https://doi.org/10.1073/pnas.102288999

Koyama, M. (2000). Molecular markers for studying pathogen variability: implications for breeding for resistance to Striga hermonthica. In B. Haussmann, D. Hess, M. Koyama, L. Grivet, H. Rattunde, & H. Geiger (Eds.), *Breeding for Striga Resistance in Cereals* (pp. 227–254). Margraf Verlag.

Kujit, J. (1969). The Biology of Parasitic Flowering Plants. In *Science* (Issue 3935). University of California Press.

Kurotani, K., Wakatake, T., Ichihashi, Y., Okayasu, K., Sawai, Y., Ogawa, S., Cui, S., Suzuki, T., Shirasu, K., & Notaguchi, M. (2020). Host-parasite tissue adhesion by a secreted type of β-1,4-glucanase in the parasitic plant Phtheirospermum japonicum. *Communications Biology*, *3*(1), 407. https://doi.org/10.1038/s42003-020-01143-5

Kushanov, F. N., Turaev, O. S., Ernazarova, D. K., Gapparov, B. M., Oripova, B. B., Kudratova, M. K., Rafieva, F. U., Khalikov, K. K., Erjigitov, D. Sh., Khidirov, M. T., Kholova, M. D., Khusenov, N. N., Amanboyeva, R. S., Saha, S., Yu, J. Z., & Abdurakhmonov, I. Y. (2021). Genetic Diversity, QTL Mapping, and Marker-Assisted Selection

Technology in Cotton (Gossypium spp.). *Frontiers in Plant Science*, *12*. https://www.frontiersin.org/articles/10.3389/fpls.2021.779386

Lamb, C., & Dixon, R. A. (1997). The oxidative burst in plant disease resistance. *Annual Review of Plant Physiology and Plant Molecular Biology*, *48*(1), 251–275. https://doi.org/10.1146/annurev.arplant.48.1.251

le Corre, W., Reibel, C., & Gibot-Leclerc, S. (2014). Development of microsatellite markers in the branched broomrape Phelipanche ramosa L. (Pomel) and evidence for host-associated genetic divergence. *Int J Mol Sci*, *15*(1), 994–1002. https://doi.org/10.3390/ijms15010994

Li, J., & Timko, M. P. (2009). Gene-for-Gene Resistance in Striga-Cowpea Associations. *Science*, *325*(5944), 1094 LP – 1094.

Lionetti, V., Cervone, F., & Bellincampi, D. (2012). Methyl esterification of pectin plays a role during plant–pathogen interactions and affects plant resistance to diseases. *Journal of Plant Physiology*, *169*(16), 1623–1630. https://doi.org/10.1016/j.jplph.2012.05.006

Lionetti, V., Fabri, E., de Caroli, M., Hansen, A. R., Willats, W. G. T., Piro, G., & Bellincampi, D. (2017). Three Pectin Methylesterase Inhibitors Protect Cell Wall Integrity for Arabidopsis Immunity to Botrytis . *Plant Physiology*, *173*(3), 1844–1863. https://doi.org/10.1104/pp.16.01185

Little, T. J., Watt, K., & Ebert, D. (2006). Parasite-Host Specificity: Experimental Studies on the Basis of Parasite Adaptation. *Evolution*, *60*(1), 31–38. http://www.jstor.org/stable/4095259

Liu, Z., Zhang, Z., Faris, J. D., Oliver, R. P., Syme, R., McDonald, M. C., McDonald, B. A., Solomon, P. S., Lu, S., Shelver, W. L., Xu, S., & Friesen, T. L. (2012). The Cysteine Rich Necrotrophic Effector SnTox1 Produced by Stagonospora nodorum Triggers Susceptibility of Wheat Lines Harboring Snn1. *PLoS Pathogens*, *8*(1), e1002467. https://doi.org/10.1371/journal.ppat.1002467

Lloyd, D. G. (1979). Some Reproductive Factors Affecting the Selection of Self-Fertilization in Plants. *The American Naturalist*, *113*(1), 67–79. https://doi.org/10.1086/283365

Lopez, L., Bellis, E. S., Wafula, E., Hearne, S. J., Honaas, L., Ralph, P. E., Timko, M. P., Unachukwu, N., dePamphilis, C. W., & Lasky, J. R. (2019). Transcriptomics of Host-Specific Interactions in Natural Populations of the Parasitic Plant Purple Witchweed (Striga hermonthica). *Weed Science*, *67*(4), 397–411. https://doi.org/10.1017/wsc.2019.20

Loveless, M. D., & Hamrick, J. L. (1984). Ecological Determinants of Genetic Structure in Plant Populations. *Annual Review of Ecology and Systematics*, *15*(1), 65–95. https://doi.org/10.1146/annurev.es.15.110184.000433

Luo, D., Ganesh, S., Koolaard, J., & Luo, M. D. (2018). *Package 'predictmeans.'*

Ma, L.-J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W.-B., Woloshuk, C., Xie, X., Xu, J.-R., Antoniw, J., Baker, S. E., Bluhm, B. H., … Rep, M. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature*, *464*, 367.

Ma, Y., Zhao, Y., & Berkowitz, G. A. (2017). Intracellular Ca2+ is important for flagellin-triggered defense in Arabidopsis and involves inositol polyphosphate signaling. *Journal of Experimental Botany*, *68*(13), 3617–3628. https://doi.org/10.1093/jxb/erx176

Ma, L., Lukasik, E., Gawehns, F., & Takken, F. L. W. (2012). *The Use of Agroinfiltration for Transient Expression of Plant Resistance and Fungal Effector Proteins in Nicotiana benthamiana Leaves* (pp. 61–74). https://doi.org/10.1007/978-1-61779-501-5_4

Madany, M. M. Y., Obaid, W. A., Hozien, W., AbdElgawad, H., Hamed, B. A., & Saleh, A. M. (2020). Salicylic acid confers resistance against broomrape in tomato through modulation of C and N metabolism. *Plant Physiology and Biochemistry*, *147*, 322–335https://doi.org/10.1016/j.plaphy.2019.12.028

Makaza, W., En-nahli, Y., & Amri, M. (2023). Harnessing plant resistance against Striga spp. parasitism in major cereal crops for enhanced crop production and food security in Sub-Saharan Africa: a review. *Food Security*. https://doi.org/10.1007/s12571-023-01345-9

Mallu, T., Irafasha, G., Mutinda, S., Owuor, E., Githri, S., Odeny, D., & Runo S. (2022). Mechanisms of pre-attachment Striga resistance in sorghum through genome-wide association studies. *Mol Genet Genomics*, *297*(3), 751–762. https://doi.org/10.1007/s00438-022-01882-6

Mandumbu, R., Mutengwa, C., Mabasa, S., & Mwenje, E. (2019). Challenges to the exploitation of host plant resistance for Striga management in cereals and legumes by farmers in sub-Saharan Africa: a review. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, *69*(1), 82–88. https://doi.org/10.1080/09064710.2018.1494302

Mandumbu, R., Mutengwa, C., Mabasa, S., & Mwenje, E. (2016). Existence of different physiological "strains" of Striga asiatica (L.) Kuntze on sorghum species [Sorghum bicolor (L.) Moench and Sorghum arundinaceum (Desv.) Stapf] in Zimbabwe. *Research on Crops*, *17*, 468–478. https://doi.org/10.5958/2348-7542.2016.00077.2

Mandumbu, R., Mutengwa, C., Mabasa, S., & Mwenje, E. (2019). Challenges to the exploitation of host plant resistance for Striga management in cereals and legumes by farmers in sub-Saharan Africa: a review. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, *69*(1), 82–88. https://doi.org/10.1080/09064710.2018.1494302

Marley, P. S., Ahmed, S. M., Shebayan, J. A. Y., & Lagoke, S. T. O. (1999). Isolation of Fusarium oxysporum with Potential for Biocontrol of the Witchweed (Striga hermonthica) in the Nigerian Savanna. *Biocontrol Science and Technology*, *9*(2), 159–163. https://doi.org/10.1080/09583159929749

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Maselko, J., Andrews, K. R., & Hohenlohe, P. A. (2020). Long-lived marine species may be resilient to environmental variability through a temporal portfolio effect. *Ecology and Evolution*, *10*(13), 6435–6448. https://doi.org/10.1002/ece3.6378

Mbuvi, D. A., Masiga, C. W., Kuria, E., Masanga, J., Wamalwa, M., Mohamed, A., Odeny, D. A., Hamza, N., Timko, M. P., & Runo, S. (2017). Novel sources of witchweed (Striga) resistance from wild sorghum accessions. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00116

McDonald BA, & Linde C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol*, *40*, 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., … Wilson, J. F. (2008). Runs of Homozygosity in

European Populations. *The American Journal of Human Genetics*, *83*(3), 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Melnyk, C. W., Schuster, C., Leyser, O., & Meyerowitz, E. M. (2015). A Developmental Framework for Graft Formation and Vascular Reconnection in Arabidopsis thaliana. *Current Biology*, *25*(10), 1306–1318. https://doi.org/10.1016/j.cub.2015.03.032

Menkir, A., Chikoye, D., & Lum, F. (2010). Incorporating an herbicide resistance gene into tropical maize with inherent polygenic resistance to control Striga hermonthica (Del.) Benth. *Plant Breeding*, *129*(4), 385–392. https://doi.org/10.1111/j.1439-0523.2009.01687.x

Midega, C. A. O., Khan, Z. R., Amudavi, D. M., Pittchar, J., & Pickett, J. A. (2010). Integrated management of Striga hermonthica and cereal stemborers in finger millet (Eleusine coracana (L.) Gaertn.) through intercropping with Desmodium intortum. *International Journal of Pest Management*, *56*(2), 145–151. https://doi.org/10.1080/09670870903248843

Miller, G., Shulaev, V., & Mittler, R. (2008). Reactive oxygen signaling and abiotic stress. *Physiologia Plantarum*, *133*(3), 481–489. https://doi.org/10.1111/j.1399-3054.2008.01090.x

Mishra, S., Upadhyay, S., & Shukla, R. K. (2017). The Role of Strigolactones and Their Potential Cross-talk under Hostile Ecological Conditions in Plants. *Frontiers in Physiology*, *7*. https://www.frontiersin.org/articles/10.3389/fphys.2016.00691

Mitsumasu, K., Seto, Y., & Yoshida, S. (2015). Apoplastic interactions between plants and plant root intruders. *Frontiers in Plant Science*, *6*, 617. https://doi.org/10.3389/fpls.2015.00617

Mohamed, A., Ellicott, A., Housley, T. L., & Ejeta, G. (2003). Hypersensitive Response to Striga Infection in Sorghum. *Crop Science*, *43*(4), 1320–1324. https://doi.org/10.2135/cropsci2003.1320

Mohamed, A., Ejeta, Butler, & Housley. (1998). Moisture content and dormancy in Striga asiatica seeds. *Weed Research*, *38*(4), 257–265. https://doi.org/10.1046/j.1365-3180.1998.00102.x

Mohamed, A., Housley, T., & Ejeta, G. (2010). Inheritance of hyper sensitive response to Striga parasitism in sorghum [Sorghum bicolor (L.) Moench]. *African Journal of Agricultural Research*, *5*(19), 2720–2729.

Mohamed, K. I., Musselman, L. J., & Riches, C. R. (2001). The Genus Striga (Scrophulariaceae) in Africa. *Annals of the Missouri Botanical Garden*, *88*(1), 60–103. https://doi.org/10.2307/2666132

Molano, E. P. L., Cabrera, O. G., Jose, J., do Nascimento, L. C., Carazzolle, M. F., Teixeira, P. J. P. L., Alvarez, J. C., Tiburcio, R. A., Tokimatu Filho, P. M., de Lima, G. M. A., Guido, R. V. C., Corrêa, T. L. R., Leme, A. F. P., Mieczkowski, P., & Pereira, G. A. G. (2018). Ceratocystis cacaofunesta genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). *BMC Genomics*, *19*(1), 58. https://doi.org/10.1186/s12864-018-4440-4

Molinero-Ruiz, L., García-Carneros, A. B., Collado-Romero, M., Raranciuc, S., Domínguez, J., & Melero-Vara, J. M. (2014). Pathogenic and molecular diversity in highly virulent populations of the parasitic weed Orobanche cumana (sunflower broomrape) from Europe. *Weed Research*, *54*(1), 87–96. https://doi.org/10.1111/wre.12056

Mukhtar, M. S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., Pevzner, S. J., Donovan, S. E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M. M., Gebreab, F., Gutierrez, B. J., Tam, S., … Dangl, J. L. (2011). Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science*, *333*(6042), 596–601. https://doi.org/10.1126/science.1203659

Murage, A. W., Obare, G., Chianu, J., Amudavi, D. M., Pickett, J., & Khan, Z. R. (2011). Duration analysis of technology adoption effects of dissemination pathways: A case of 'push–pull' technology for control of Striga weeds and stemborers in Western Kenya. *Crop Protection*, *30*(5), 531–538. https://doi.org/https://doi.org/10.1016/j.cropro.2010.11.009

Murashima, K., Kosugi, A., & Doi, R. H. (2003). Synergistic Effects of Cellulosomal Xylanase and Cellulases from Clostridium cellulovorans on Plant Cell Wall Degradation. *Journal of Bacteriology*, *185*(5), 1518–1524. https://doi.org/10.1128/JB.185.5.1518-1524.2003

Musselman, L. J. (1980). The Biology of Striga, Orobanche, and other Root-Parasitic Weeds. *Annual Review of Phytopathology*, *18*(1), 463–489. https://doi.org/10.1146/annurev.py.18.090180.002335

Musselman LJ, Bhrathalakshmi, Safa SB, Knepper DA, & Mohamed KI. (1988). Recent research on the biology of Striga asiatica, S. gesnerioides and S. hermonthica. In Kim SK (Ed.), *Combating Striga in Africa: proceedings of the international workshop* (pp. 31–41). International Institute of Tropical Agriculture.

Mutikainen, P. K., & Koskela, T. (2002). Population structure of a parasitic plant and its perennial host. *Heredity*, *89*, 318–324.

Mutinda, S. M., Masanga, J., Mutuku, J. M., Runo, S., & Alakonya, A. (2018). KSTP 94, an Open-pollinated Maize Variety Has Postattachment Resistance to Purple Witchweed (*Striga hermonthica*). *Weed Science*, *66*(4), 525–529. https://doi.org/10.1017/wsc.2018.24

Mutuku, J. M., Cui, S., Hori, C., Takeda, Y., Tobimatsu, Y., Nakabayashi, R., Mori, T., Saito, K., Demura, T., Umezawa, T., Yoshida, S., & Shirasu, K. (2019). The Structural Integrity of Lignin Is Crucial for Resistance against Striga hermonthica Parasitism in Rice . *Plant Physiology*, *179*(4), 1796–1809. https://doi.org/10.1104/pp.18.01133

Mutuku, J. M., Yoshida, S., Shimizu, T., Ichihashi, Y., Wakatake, T., Takahashi, A., Seo, M., & Shirasu, K. (2015). The WRKY45-Dependent Signaling Pathway Is Required For Resistance against *Striga hermonthica* Parasitism. *Plant Physiology*, *168*(3), 1152–1163. https://doi.org/10.1104/pp.114.256404

Nambafu, G., Onwonga, Karuku, G., Ariga, E., Vanlauwe, B., & Röing de Nowina, K. (2014). Knowledge, Attitude and Practices Used in the Control of Striga in Maize by Smallholder Farmers of Western Kenya. *Journal of Agricultural Science and Technology B*, *B4*, 237–248. https://doi.org/10.17265/2161-6264/2014.03B.010

Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G. J., Jones, D. A., Kamoun, S., Ellis, J. G., & Dodds, P. N. (2014). The genome sequence and effector complement of the flax rust pathogen Melampsora lini. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00098

Neumann, U., Vian, B., Weber, H. C., & Sallé, G. (1999). Interface between haustoria of parasitic members of the Scrophulariaceae and their hosts: A histochemical and immunocytochemical approach. *Protoplasma*, *207*(1), 84–97. https://doi.org/10.1007/BF01294716

Ngou, B. P. M., Ding, P., & Jones, J. D. G. (2022). Thirty years of resistance: Zig-zag through the plant immune system. *The Plant Cell*, *34*(5), 1447–1478. https://doi.org/10.1093/plcell/koac041

Nickrent, D. L., & Musselman, L. J. (1979). Autogamy in the American Strain of Witchweed, Striga Asiatica (Scrophulariaceae). *Brittonia*, *31*(2), 253–256. https://doi.org/10.2307/2806182

Nothnagel, E. A., McNeil, M., Albersheim, P., & Dell, A. (1983). Host-Pathogen Interactions. *Plant Physiology*, *71*(4), 916–926. https://doi.org/10.1104/pp.71.4.916

Nyakurwa, C. S., Gasura, E., Setimela, P. S., Mabasa, S., Rugare, J. T., & Mutsvanga, S. (2018). Reaction of New Quality Protein Maize Genotypes to Striga asiatica. *Crop Science*, *58*(3), 1201–1218. https://doi.org/10.2135/cropsci2017.10.0639

Nzioki, H. S., Oyosi, F., Morris, C. E., Kaya, E., Pilgeram, A. L., Baker, C. S., & Sands, D. C. (2016). Striga Biocontrol on a Toothpick: A Readily Deployable and Inexpensive Method for Smallholder Farmers. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/articles/10.3389/fpls.2016.01121

OECD, & FAO. (2016). *OECD-FAO Agricultural Outlook 2016-2025*. OECD Publishing. https://doi.org/10.1787/agr_outlook-2016-en

Ohlson, E. W., & Timko, M. P. (2020). Race Structure of Cowpea Witchweed (Striga gesnerioides) in West Africa and Its Implications for Striga Resistance Breeding of Cowpea. *Weed Science*, *68*(2), 125–133. https://doi.org/10.1017/wsc.2020.3

Okonkwo, S. N. C., & Raghavan, V. (1982). Studies on the Germination of Seeds of the Root Parasites, Alectra vogelii and Striga gesnerioides. I. Anatomical Changes in the Embryos. *American Journal of Botany*, *69*(10), 1636. https://doi.org/10.2307/2442918

Oksanen J, Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara R, Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., de Caceres, M., Durand, S., … Weedon, J. (2022). *Vegan: Community Ecology Package.* (R package version 2.6-4).

Okubamichael, D. Y., Griffiths, M. E., & Ward, D. (2016). Host specificity in parasitic plants—perspectives from mistletoes. *AoB PLANTS*, *8*, plw069. https://doi.org/10.1093/aobpla/plw069

Oula, D. A., Nyongesah, J. M., Odhiambo, G., & Wagai, S. (2020). The effectiveness of local strains of Fusarium oxysporium f. Sp. Strigae to control Striga hermonthica on local maize in western Kenya. *Food Science & Nutrition*, *8*(8), 4352–4360. https://doi.org/10.1002/fsn3.1732

Padmanabhan, M., Cournoyer, P., & Dinesh-Kumar, S. P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cellular Microbiology*, *11*(2), 191–198. https://doi.org/10.1111/j.1462-5822.2008.01260.x

Palmer, A., Liu, Y., Adkins, S., Zhang, X., Wu, I.-L., Chen, X., & Lynn, D. (2008). The molecular language of semagenesis. *Plant Signaling & Behavior*, *3*, 560–561. https://doi.org/10.4161/psb.3.8.5703

Parker, C. (2009). Observations on the current status of Orobanche and Striga problems worldwide. *Pest Management Science*, *65*(5), 453–459. https://doi.org/10.1002/ps.1713

Parker, C., & Riches, C. R. (1993). *Parasitic weeds of the world*. Wallingford : CAB International, c1993.

Parker, C. (2012). Parasitic Weeds: A World Challenge. *Weed Science*, *60*(2), 269–276. https://doi.org/DOI: 10.1614/WS-D-11-00068.1

Parker, C., & Polniaszek, T. I. (1990). Parasitism of cowpea by Striga gesnerioides: variation in virulence and discovery of a new source of host resistance. *Annals of Applied Biology*, *116*(2), 305–311. https://doi.org/https://doi.org/10.1111/j.1744-7348.1990.tb06611.x

Pekkala, N., Knmott, K., Kotiaho, J., Nissinen, K., & Puurtinen, M. (2014). The effect of inbreeding rate on fitness, inbreeding depression and heterosis over a range of inbreeding coefficients. *Evol Appl*, *7*(9), 1107–1119. https://doi.org/10.1111/eva.12145

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Pfunye, A., Rwafa, R., Mabasa, S., & Gasura, E. (2021). Genome-Wide Association Studies for Striga asiatica Resistance in Tropical Maize. *International Journal of Genomics*, *2021*, 1–8. https://doi.org/10.1155/2021/9979146

Piazza, S., Campa, M., Pompili, V., Costa, L. D., Salvagnin, U., Nekrasov, V., Zipfel, C., & Malnoy, M. (2021). The Arabidopsis pattern recognition receptor EFR enhances fire blight resistance in apple. *Horticulture Research*, *8*(1), 204. https://doi.org/10.1038/s41438-021-00639-3

Pineda-Martos, R., Pujadas-Salvà, A. J., Fernández-Martínez, J. M., Stoyanov, K., Velasco, L., & Pérez-Vich, B. (2014). The Genetic Structure of Wild Orobanche cumana Wallr. (Orobanchaceae) Populations in Eastern Bulgaria Reflects Introgressions from Weedy Populations. *The Scientific World Journal*, *2014*, 1–15. https://doi.org/10.1155/2014/150432

Plantwise. (2016). *FACTSHEET FOR FARMERS: Striga asiatica*. https://factsheetadmin.plantwise.org/Uploads/PDFs/20167800563.pdf

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178. https://doi.org/10.1101/201178

Porcheron, G., Garénaux, A., Proulx, J., Sabri, M., & Dozois, C. M. (2013). Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence. *Frontiers in Cellular and Infection Microbiology*, *3*. https://doi.org/10.3389/fcimb.2013.00090

Press, M. C., & Stewart, G. R. (1987). Growth and Photosynthesis in Sorghum bicolor Infected with Striga hermonthica. *Annals of Botany*, *60*(6), 657–662. https://doi.org/10.1093/oxfordjournals.aob.a087496

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, & Sham PC. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*.

Qi, G., Chen, J., Chang, M., Chen, H., Hall, K., Korin, J., Liu, F., Wang, D., & Fu, Z. Q. (2018). Pandemonium Breaks Out: Disruption of Salicylic Acid-Mediated Defense by Plant Pathogens. *Molecular Plant*, *11*(12), 1427–1439. https://doi.org/10.1016/j.molp.2018.10.002

Qiu, S., Bradley, J. M., Zhang, P., Chaudhuri, R., Blaxter, M., Butlin, R. K., & Scholes, J. D. (2022). Genome-enabled discovery of candidate virulence loci in Striga hermonthica, a devastating parasite of African cereal crops. *New Phytologist*, *236*(2), 622–638. https://doi.org/10.1111/nph.18305

Qiu, D., Xu, S., Wang, Y., Zhou, M., & Hong, L. (2021). Primary Cell Wall Modifying Proteins Regulate Wall Mechanics to Steer Plant Morphogenesis. *Frontiers in Plant Science*, *12*. https://www.frontiersin.org/articles/10.3389/fpls.2021.751372

Queiroz, C. B. de, & Santana, M. F. (2020). Prediction of the secretomes of endophytic and nonendophytic fungi reveals similarities in host plant infection and colonization strategies. *Mycologia*, *112*(3), 491–503. https://doi.org/10.1080/00275514.2020.1716566

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (4.1.2). R Foundation for Statistical Computing.

Radford, T. (2003, September). Perfect maize, in three simple steps. *The Guardian*.

Rafiei, V., Vélëz, H., & Tzelepis, G. (2021). The Role of Glycoside Hydrolases in Phytopathogenic Fungi and Oomycetes Virulence. *International Journal of Molecular Sciences*, *22*(17), 9359. https://doi.org/10.3390/ijms22179359

Raffoux, X., Bourge, M., Dumas, F., Martin, O. C., & Falque, M. (2018). Role of Cis, Trans, and Inbreeding Effects on Meiotic Recombination in Saccharomyces cerevisiae. *Genetics*, *210*(4), 1213–1226. https://doi.org/10.1534/genetics.118.301644

Ralandison, T., Arimoto, Y., Kono, H., Sakurai, T., & Takahashi, K. (2018). Rice flows across regions in Madagascar . *African Journal of Agricultural and Resource Economics*, *13*(1), 78–105.

Ramaiah, K., & Werder, J. (1987). Screening pearl millet for resistance to Striga hermonthica in West Africa. In A. Patancheru (Ed.), *Proceedings of the International Pearl Millet Workshop, International Crops Research Inst. for the Semi-Arid Tropics* (pp. 314–315). ICRISAT.

Ramaiah, K., & Werder, J. (1987). Screening pearl millet for resistance to Striga hermonthica in West Africa. In A. Patancheru (Ed.), *Proceedings of the International Pearl Millet Workshop, International Crops Research Inst. for the Semi-Arid Tropics* (pp. 314–315). ICRISAT.

Rampitsch, C., Day, J., Subramaniam, R., & Walkowiak, S. (2013). Comparative secretome analysis of Fusarium graminearum and two of its non-pathogenic mutants upon deoxynivalenol induction in vitro. *PROTEOMICS*, *13*(12–13), 1913–1921. https://doi.org/10.1002/pmic.201200446

Randrianjafizanaka, M. T., Autfray, P., Andrianaivo, A. P., Ramonta, I. R., & Rodenburg, J. (2018). Combined effects of cover crops, mulch, zero-tillage and resistant varieties on Striga asiatica (L.) Kuntze in rice-maize rotation systems. *Agriculture, Ecosystems & Environment*, *256*, 23–33. https://doi.org/10.1016/j.agee.2017.12.005

Rank, C., Rasmussen, L. S., Jensen, S. R., Pierce, S., Press, M. C., & Scholes, J. D. (2004). Cytotoxic constituents of Alectra and Striga species. *Weed Research*, *44*(4), 265–270. https://doi.org/10.1111/j.1365-3180.2004.00398.x

Rao, T. B., Chopperla, R., Prathi, N. B., Balakrishnan, M., Prakasam, V., Laha, G. S., Balachandran, S. M., & Mangrauthia, S. K. (2020). A Comprehensive Gene Expression Profile of Pectin Degradation Enzymes Reveals the Molecular Events during Cell Wall Degradation and Pathogenesis of Rice Sheath Blight Pathogen Rhizoctonia solani AG1-IA. *Journal of Fungi*, *6*(2), 71. https://doi.org/10.3390/jof6020071

Read, A. F. (1994). The evolution of virulence. *Trends in Microbiology*, *2*(3), 73–76. https://doi.org/10.1016/0966-842X(94)90537-1

Reiss, G. C., & Bailey, J. A. (1998). Striga gesnerioides Parasitising Cowpea: Development of Infection Structures and Mechanisms of Penetration. *Annals of Botany*, *81*(3), 431–440. https://doi.org/10.1006/anbo.1997.0577

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rich, P. J., & Ejeta, G. (2008). Towards effective resistance to Striga in African maize. *Plant Signaling & Behavior*, *3*(9), 618–621. https://doi.org/10.4161/psb.3.9.5750

Riopel, J., & Timko, M. (1995). Haustorial initiation and differentiation. In M. Press & J. Graves (Eds.), *Parasitic Plants* (pp. 39–79). Chapman & Hall.

Riopel, J. L., & Musselman, L. J. (1979). Experimental Initiation of Haustoria in Agalinis purpurea (Scrophulariaceae). *American Journal of Botany*, *66*(5), 570. https://doi.org/10.2307/2442506

Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, *67*(2), 175–185. https://doi.org/DOI: 10.1017/S0016672300033620

Rodenburg, J., Cissoko, M., Kayongo, N., Dieng, I., Bisikwa, J., Irakiza, R., Masoka, I., Midega, C. A. O., & Scholes, J. D. (2017). Genetic variation and host-parasite specificity of Striga resistance and tolerance in rice: the need for predictive breeding. *The New Phytologist*, *214*(3), 1267–1280. https://doi.org/10.1111/nph.14451

Rodenburg, J., Demont, M., Zwart, S. J., & Bastiaans, L. (2016). Parasitic weed incidence and related economic losses in rice in Africa. *Agriculture, Ecosystems & Environment*, *235*, 306–317. https://doi.org/10.1016/j.agee.2016.10.020

Rodenburg, J., Cissoko, M., Kayeke, J., Dieng, I., Khan, Z. R., Midega, C. A. O., Onyuka, E. A., & Scholes, J. D. (2015). Do NERICA rice cultivars express resistance to Striga hermonthica (Del.) Benth. and Striga asiatica (L.) Kuntze under field conditions? *Field Crops Research*, *170*, 83–94. https://doi.org/10.1016/j.fcr.2014.10.010

Rodenburg, J., Riches, C. R., & Kayeke, J. M. (2010). Addressing current and future problems of parasitic weeds in rice. *Crop Protection*, *29*(3), 210–221. https://doi.org/10.1016/j.cropro.2009.10.015

Rodenburg, J., Bastiaans, L., Kropff, M. J., & van Ast, A. (2006a). Effects of host plant genotype and seedbank density on Striga reproduction. *Weed Research*, *46*(3), 251–263. https://doi.org/10.1111/j.1365-3180.2006.00506.x

Rodenburg, J., Bastiaans, L., & Kropff, M. J. (2006b). Characterization of host tolerance to Striga hermonthica. *Euphytica*, *147*(3), 353–365. https://doi.org/10.1007/s10681-005-9030-2

Roney, J. K., Khatibi, P. A., & Westwood, J. H. (2007). Cross-Species Translocation of mRNA from Host Plants into the Parasitic Plant Dodder. *Plant Physiology*, *143*(2), 1037–1043. https://doi.org/10.1104/pp.106.088369

Rousset. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, *13*(1), 58–62. https://doi.org/https://doi.org/10.1046/j.1420-9101.2000.00137.x

Runo, S., & Kuria, E. K. (2018). Habits of a highly successful cereal killer, Striga. *PLOS Pathogens*, *14*(1), e1006731.

Runo, S. (2019). Modern Breeding Approaches for Durable Resistance Against the Parasitic Plant Striga. *Afrika Focus*, *32*(2), 109–115. https://doi.org/10.1163/2031356X-03202009

Runo, S., Macharia, S., Alakonya, A., Machuka, J., Sinha, N., & Scholes, J. (2012). Striga parasitizes transgenic hairy roots of Zea mays and provides a tool for studying plant-plant interactions. *Plant Methods*, *8*(1), 20. https://doi.org/10.1186/1746-4811-8-20

Runyon, J. B., Mescher, M. C., Felton, G. W., & de Moraes, C. M. (2010). Parasitism by Cuscuta pentagona sequentially induces JA and SA defence pathways in tomato. *Plant, Cell & Environment*, *33*(2), 290–303. https://doi.org/10.1111/j.1365-3040.2009.02082.x

Sacristán, S., Goss, E. M., & Eves-van den Akker, S. (2021). How Do Pathogens Evolve Novel Virulence Activities? *Molecular Plant-Microbe Interactions®*, *34*(6), 576–586. https://doi.org/10.1094/MPMI-09-20-0258-IA

Safa, S. B., Jones, B. M. G., & Musselman, L. J. (1984). Mechanisms Favouring Outbreeding in Striga hermonthica [Scrophulariaceae]. *The New Phytologist*, *96*(2), 299–305. http://www.jstor.org/stable/2432407

Samejima, H., & Sugimoto, Y. (2022). Phenotypic Diversity in Pre- and Post-Attachment Resistance to Striga hermonthica in a Core Collection of Rice Germplasms. *Plants*, *12*(1), 19. https://doi.org/10.3390/plants12010019

Sauerborn, J., Buschmann, H., Ghiasi, K. G., & Kogel, K.-H. (2002). Benzothiadiazole Activates Resistance in Sunflower (Helianthus annuus) to the Root-Parasitic Weed Orobanche cuman. *Phytopathology®*, *92*(1), 59–64. https://doi.org/10.1094/PHYTO.2002.92.1.59

Savary, S., Ficke, A., Aubertot, J.-N., & Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Security*, *4*(4), 519–537. https://doi.org/10.1007/s12571-012-0200-5

Savelkoul, P. H. M., Aarts, H. J. M., de Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J. L. W., Schouls, L., & Lenstra, J. A. (1999). Amplified-Fragment Length Polymorphism Analysis: the State of an Art. *Journal of Clinical Microbiology*, *37*(10), 3083–3091. https://doi.org/10.1128/JCM.37.10.3083-3091.1999

Sawadogo, P., Ouedraogo, T., Batieno, B. J., Dieni, Z., Sawadogo, N., Congo, A., Zida, F., Kabore, B., Karidiatou, G., Hamadou, Z., Leandre, P., Tignegre, S., & Mahamadou, S. (2020). *Genetic diversity among Striga gesnerioides (Willd.) Vatkerevealed by microsatellite markers*. *3*, 266–276.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Scholes, J. D., & Press, M. C. (2008). Striga infestation of cereal crops - an unsolved problem in resource limited agriculture. *Current Opinion in Plant Biology*, *11*(2), 180–186. https://doi.org/10.1016/j.pbi.2008.02.004

Schulte RD, Makus C, Hasert B, Michiels NK, & Schulenburg H. (2011). Host-parasite local adaptation after experimental coevolution of Caenorhabditis elegans and its microparasite Bacillus thuringiensis. *Proc Biol Sci.*, *278*(1719), 2832–2839. https://doi.org/10.1098/rspb.2011.0019

Scott, D., Scholes, J. D., Randrianjafizanaka, M. T., Randriamampianina, J. A., Autfray, P., & Freckleton, R. P. (2020). Mapping the drivers of parasitic weed abundance at a national scale: a new approach applied to Striga asiatica in the mid-west of Madagascar. *Weed Research*, *60*(5), 323–333. https://doi.org/10.1111/wre.12436

Scott, D., & Freckleton, R. P. (2022). Crop diversification and parasitic weed abundance: a global meta-analysis. *Scientific Reports*, *12*(1), 19413. https://doi.org/10.1038/s41598-022-24047-2

Shahid, S., Kim, G., Johnson, N. R., Wafula, E., Wang, F., Coruh, C., Bernal-Galeano, V., Phifer, T., dePamphilis, C. W., Westwood, J. H., & Axtell, M. J. (2018). MicroRNAs from the parasitic plant Cuscuta campestris target host messenger RNAs. *Nature*, *553*(7686), 82–85. https://doi.org/10.1038/nature25027

Shaibu, A. S., Badu-Apraku, B., & Ayo-Vaughan, M. A. (2021). Enhancing Drought Tolerance and Striga hermonthica Resistance in Maize Using Newly Derived Inbred Lines from the Wild Maize Relative, Zea diploperennis. *Agronomy*, *11*(1), 177. https://doi.org/10.3390/agronomy11010177

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, *50*(W1), W216–W221. https://doi.org/10.1093/nar/gkac194

Shin, Y., Chane, A., Jung, M., & Lee, Y. (2021). Recent Advances in Understanding the Roles of Pectin as an Active Participant in Plant Signaling Networks. *Plants*, *10*(8), 1712. https://doi.org/10.3390/plants10081712

Sibhatu, B. (2016). Review on Striga Weed Management. *Int. J. Life. Sci. Scienti. Res.*, *2*(2), 110–120.

Smith, C. E., Ruttledge, T., Zeng, Z., O'Malley, R. C., & Lynn, D. G. (1996). A Mechanism for Inducing Plant Development: The Genesis of a Specific Inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(14), 6986–6991. http://www.jstor.org/stable/39522

Smith, C. E., Ruttledge, T., Zeng, Z., O'Malley, R. C., & Lynn, D. G. (1996). A Mechanism for Inducing Plant Development: The Genesis of a Specific Inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(14), 6986–6991. http://www.jstor.org/stable/39522

Sonah, H., Deshmukh, R. K., & Bélanger, R. R. (2016). Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00126

Spallek, T., Mutuku, M., & Shirasu, K. (2013). The genus Striga: a witch profile. *Molecular Plant Pathology*, *14*(9), 861–869. https://doi.org/10.1111/mpp.12058

Stebbins, G. L. (1957). Self fertilization and population variability in the higher plants. *The American Naturalist*, *91*(861), 337–354.

Steinkellner, S., Lendzemo, V., Langer, I., Schweiger, P., Khaosaad, T., Toussaint, J.-P., & Vierheilig, H. (2007). Flavonoids and Strigolactones in Root Exudates as Signals in Symbiotic and Pathogenic Plant-Fungus Interactions. *Molecules*, *12*(7), 1290–1306. https://doi.org/10.3390/12071290

Stewart, C. N., & Via, L. E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques*, *14*(5), 748–750.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M. Jr. (1949). Adjustment during Army Life. In *The American Soldier* (Vol. 1). Princeton University Press.

Su, C., Liu, H., Wafula, E. K., Honaas, L., de Pamphilis, C. W., & Timko, M. P. (2020). SHR4z, a novel decoy effector from the haustorium of the parasitic weed Striga gesnerioides, suppresses host plant immunity. *New Phytologist*, *226*(3), 891–908. https://doi.org/10.1111/nph.16351

Sun, W., Cao, Y., Jansen Labby, K., Bittel, P., Boller, T., & Bent, A. F. (2012). Probing the Arabidopsis Flagellin Receptor: FLS2-FLS2 Association and the Contributions of Specific Domains to Signaling Function . *The Plant Cell*, *24*(3), 1096–1113. https://doi.org/10.1105/tpc.112.095919

Tan, H. (1987). Strategies in rubber tree breeding. In A. J. Abbott & R. K. Atkin (Eds.), *Improving vegetatively propagated crops* (pp. 27–62). Royaume Uni, Academic Press.

Taurino, M., Abelenda, J. A., Río-Alvarez, I., Navarro, C., Vicedo, B., Farmaki, T., Jiménez, P., García-Agustín, P., López-Solanilla, E., Prat, S., Rojo, E., Sánchez-Serrano, J. J., & Sanmartín, M. (2014). Jasmonate-dependent modifications of the pectin matrix during potato development function as a defense mechanism targeted by Dickeya dadantii virulence factors. *The Plant Journal*, *77*(3), 418–429. https://doi.org/10.1111/tpj.12393

Teixeira, S., Foerster, K., & Bernasconi, G. (2009). Evidence for inbreeding depression and post-pollination selection against inbreeding in the dioecious plant Silene latifolia. *Heredity*, *102*(2), 101–112. https://doi.org/10.1038/hdy.2008.86

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences*, *118*(10). https://doi.org/10.1073/pnas.2015096118

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. https://doi.org/10.1038/35048692

The World Bank. (2017). *World Development Indicators: Agricultural Inputs*. World Development Indicators.

Tian, M., Benedetti, B., & Kamoun, S. (2005). A Second Kazal-Like Protease Inhibitor from Phytophthora infestans Inhibits and Interacts with the Apoplastic Pathogenesis-Related Protease P69B of Tomato. *Plant Physiology*, *138*(3), 1785–1793. https://doi.org/10.1104/pp.105.061226

Timko, M. P., & Scholes, J. D. (2013). Host Reaction to Attack by Root Parasitic Plants. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.), *Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies* (pp. 115–141). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38146-1_7

Timko, M. P., Gowda, B. S., Ouedraogo, J., & Ousmane, B. (2007). molecular markers for analysis of resistance to Striga gesnerioides in cowpea. In *Integrating New Technologies for Striga Control* (pp. 115–128). WORLD SCIENTIFIC. https://doi.org/9789812771506_0009

Tompkins, Daniel. M., & Clayton, Dale. H. (1999). Host resources govern the specificity of swiftlet lice: size matters. *Journal of Animal Ecology*, *68*(3), 489–500. https://doi.org/10.1046/j.1365-2656.1999.00297.x

Tsuchiya, Y., Yoshimura, M., Sato, Y., Kuwata, K., Toh, S., Holbrook-Smith, D., Zhang, H., McCourt, P., Itami, K., Kinoshita, T., & Hagihara, S. (2015). Probing strigolactone receptors in Striga hermonthica with fluorescence. *Science*, *349*(6250), 864–868. https://doi.org/10.1126/science.aab3831

Tsuda, K., & Somssich, I. E. (2015). Transcriptional networks in plant immunity. *New Phytologist*, *206*(3), 932–947. https://doi.org/https://doi.org/10.1111/nph.13286

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 59. https://doi.org/10.1038/s43586-021-00056-9

Unachukwu, N. N., Menkir, A., Rabbi, I. Y., Oluoch, M., Muranaka, S., Elzein, A., Odhiambo, G., Farombi, E. O., & Gedil, M. (2017). Genetic diversity and population structure of Striga hermonthica populations from Kenya and Nigeria. *Weed Research*, *57*(5), 293–302. https://doi.org/10.1111/wre.12260

Underwood, W. (2012). The Plant Cell Wall: A Dynamic Barrier Against Pathogen Invasion. *Frontiers in Plant Science*, *3*. https://doi.org/10.3389/fpls.2012.00085

United Nations. (2022). *World Population Prospects 2022: Summary of results*.

USDA (United States Department of Agriculture). (1957). *Witchweed (Striga asiatica)—a new parasitic plant in the United States. USDA Research Service special publication 10.*

van Delft, G.-J., Graves, J. D., Fitter, A. H., & Pruiksma, M. a. (1997). Spatial distribution and population dynamics of Striga hermonthica seeds in naturally infested farm soils. *Plant and Soil*, *195*(1), 1–15. https://doi.org/10.1023/A:1004214015281

Varas-Myrik, A., Sepúlveda-Espinoza, F., Fajardo, A., Alarcón, D., Toro-Núñez, Ó., Castro-Nallar, E., & Hasbún, R. (2022). Predicting climate change-related genetic offset for the endangered southern South American conifer Araucaria araucana. *Forest Ecology and Management*, *504*, 119856. https://doi.org/10.1016/j.foreco.2021.119856

Velasco, L., Pérez-Vich, B., Jan, C. C., & Fernández-Martínez, J. M. (2007). Inheritance of resistance to broomrape (Orobanche cumana Wallr.) race F in a sunflower line derived from wild sunflower species. *Plant Breeding*, *126*(1), 67–71. https://doi.org/10.1111/j.1439-0523.2006.01278.x

Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic diversity. *Ecology Letters*, *8*(7), 767–781. https://doi.org/https://doi.org/10.1111/j.1461-0248.2005.00775.x

Vogel, A., Schwacke, R., Denton, A. K., Usadel, B., Hollmann, J., Fischer, K., Bolger, A., Schmidt, M. H.-W., Bolger, M. E., Gundlach, H., Mayer, K. F. X., Weiss-Schneeweiss, H., Temsch, E. M., & Krause, K. (2018). Footprints of parasitism in the genome of the parasitic flowering plant Cuscuta campestris. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04344-z

Voragen, A. G. J., Coenen, G.-J., Verhoef, R. P., & Schols, H. A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Structural Chemistry*, *20*(2), 263–275. https://doi.org/10.1007/s11224-009-9442-z

Wada, S., Cui, S., & Yoshida, S. (2019). Reactive Oxygen Species (ROS) Generation Is Indispensable for Haustorium Formation of the Root Parasitic Plant Striga hermonthica. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00328

Wan, J., He, M., Hou, Q., Zou, L., Yang, Y., Wei, Y., & Chen, X. (2021). Cell wall associated immunity in plants. *Stress Biology*, *1*(1), 3. https://doi.org/10.1007/s44154-021-00003-4

Wang, Y., Yao, R., Du, X., Guo, L., Chen, L., Xie, D., & Smith, S. M. (2021a). Molecular basis for high ligand sensitivity and selectivity of strigolactone receptors in *Striga*. *Plant Physiology*, *185*(4), 1411–1428. https://doi.org/10.1093/plphys/kiaa048

Wang, Y., Li, X., Fan, B., Zhu, C., & Chen, Z. (2021b). Regulation and Function of Defense-Related Callose Deposition in Plants. *International Journal of Molecular Sciences*, *22*(5), 2393. https://doi.org/10.3390/ijms22052393

Wassenaar, T. M., & Gaastra, W. (2001). Bacterial virulence: can we draw the line? *FEMS Microbiology Letters*, *201*(1), 1–7. https://doi.org/10.1111/j.1574-6968.2001.tb10724.x

Waweru, D. N., Kuria, E. K., Bradley, J. M., Scholes, J. D., & Runo, S. (2019). Tissue culture protocols for the obligate parasitic plant Striga hermonthica and implications for host-parasite co-cultivation. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *138*(2), 247–256. https://doi.org/10.1007/s11240-019-01621-7

Wei, K., Silva-Arias, G. A., & Tellier, A. (2023). Selective sweeps linked to the colonization of novel habitats and climatic changes in a wild tomato species. *New Phytologist*, *237*(5), 1908–1921. https://doi.org/10.1111/nph.18634

Welham, S., Cullis, B., Gogel, B., Gilmour, A., & Thompson, R. (2004). Prediction in linear mixed models. *Australian & New Zealand Journal of Statistics*, *46*(3), 325–347. https://doi.org/10.1111/j.1467-842X.2004.00334.x

Welsh, A. B., & Mohamed, K. I. (2011). Genetic Diversity of Striga hermonthica Populations in Ethiopia: Evaluating the Role of Geography and Host Specificity in Shaping Population Structure. *International Journal of Plant Sciences*, *172*(6), 773–782. https://doi.org/10.1086/660104

Werth, C. R., Riopel, J. L., & Gillespie, N. W. (1984). Genetic Uniformity in an Introduced Population of Witchweed (Striga asiatica) in the United States. *Weed Science*, *32*(5), 645–648. http://www.jstor.org.sheffield.idm.oclc.org/stable/4043983

Westwood, J. H., Yoder, J. I., Timko, M. P., & DePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends in Plant Science*, *15*(4), 227–235. https://doi.org/10.1016/j.tplants.2010.01.004

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://doi.org/https://ggplot2.tidyverse.org

Wilson, J. P., Hess, D. E., Hanna, W., Kumar, K., & Gupta, S. C. (2004). Pennisetum glaucum subsp. monodii accessions with Striga resistance in West Africa. *Crop Protection*, *23*. https://doi.org/10.1016/j.cropro.2004.01.006

Woldemariam, G. W. , Damot, G. A. , & Dereje, A. Z. (2022). Nitrogen fertilizer and cattle manure for Striga (Striga hermonthica) management and enhancement of sorghum productivity in northwest Ethiopia. *Journal of Plant Nutrition*, *45*(2), 232–245. https://doi.org/10.1080/01904167.2021.1952219

Woolhouse, M. E. J., Taylor, L. H., & Haydon, D. T. (2001). Population Biology of Multihost Pathogens. *Science*, *292*(5519), 1109–1112. http://www.jstor.org.sheffield.idm.oclc.org/stable/3083648

Wright, S., Kalisz, S., & Slotte, T. (2013). Evolutionary consequences of self-fertilization in plant. *Proc Biol Sci*, *280*(1760). https://doi.org/10.1098/rspb.2013.0133

Wright S. (1978). *Evolution and the genetics of populations. Variability within and among natural populations* (Vol. 4). University of Chicago Press.

Wu, J. Q., Sakthikumar, S., Dong, C., Zhang, P., Cuomo, C. A., & Park, R. F. (2017). Comparative Genomics Integrated with Association Analysis Identifies Candidate Effector Genes Corresponding to Lr20 in Phenotype-Paired Puccinia triticina Isolates from Australia. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00148

Wu, H.-C., Bulgakov, V. P., & Jinn, T.-L. (2018). Pectin Methylesterases: Cell Wall Remodeling Proteins Are Required for Plant Response to Heat Stress. *Frontiers in Plant Science*, *9*. https://www.frontiersin.org/articles/10.3389/fpls.2018.01612

Yacoubou, A.-M., Zoumarou Wallis, N., Menkir, A., Zinsou, V. A., Onzo, A., Garcia-Oliveira, A. L., Meseka, S., Wende, M., Gedil, M., & Agre, P. (2021). Breeding maize (Zea mays) for Striga resistance: Past, current and prospects in sub-saharan africa. *Plant Breeding*, *140*(2), 195–210. https://doi.org/10.1111/pbr.12896

Yang, Z., Wafula, E. K., Honaas, L. A., Zhang, H., Das, M., Fernandez-Aparicio, M., Huang, K., Bandaranayake, P. C. G., Wu, B., Der, J. P., Clarke, C. R., Ralph, P. E., Landherr, L., Altman, N. S., Timko, M. P., Yoder, J. I., Westwood, J. H., & dePamphilis, C. W. (2015). Comparative Transcriptome Analyses Reveal Core Parasitism Genes and Suggest Gene Duplication and Repurposing as Sources of Structural Novelty. *Molecular Biology and Evolution*, *32*(3), 767–790. https://doi.org/10.1093/molbev/msu343

Yohannes, T., Abraha, T., Kiambi, D. K., Folkertsma, R., Hash Charles, J., Ngugi, K., Mutitu, E., Abraha, N., Weldetsion, M., Mugoya, C., Masiga, C., & de Villiers, S. (2015). Marker-assisted introgression improves Striga resistance in an Eritrean Farmer-Preferred Sorghum Variety. *Field Crops Research*, *173*. https://doi.org/10.1016/j.fcr.2014.12.008

Yoneyama, K., Arakawa, R., Ishimoto, K., Kim, H. il, Kisugi, T., Xie, X., Nomura, T., Kanampiu, F., Yokota, T., Ezawa, T., & Yoneyama, K. (2015). Difference in Striga-susceptibility is reflected in strigolactone secretion

profile, but not in compatibility and host preference in arbuscular mycorrhizal symbiosis in two maize cultivars. *New Phytologist*, *206*(3), 983–989. https://doi.org/10.1111/nph.13375

Yonli, D., Traore, H., Sereme, P., Hess, D. E., & Sankara, P. (2010). Pathogenicity of Fusarium isolates to Striga hermonthica in Burkina Faso. *Pakistan Journal of Biological Sciences*, *13*(5), 201–208. https://doi.org/10.3923/pjbs.2010.201.208

Yoshida, S., & Shirasu, K. (2009). Multiple Layers of Incompatibility to the Parasitic Witch Weed, Striga Hermonthica. *The New Phytologist*, *183*(1), 180–189. http://www.jstor.org.sheffield.idm.oclc.org/stable/40302015

Yoshida, S., Kim, S., Wafula, E. K., Tanskanen, J., Kim, Y.-M., Honaas, L., Yang, Z., Spallek, T., Conn, C. E., Ichihashi, Y., Cheong, K., Cui, S., Der, J. P., Gundlach, H., Jiao, Y., Hori, C., Ishida, J. K., Kasahara, H., Kiba, T., … Shirasu, K. (2019). Genome Sequence of Striga asiatica Provides Insight into the Evolution of Plant Parasitism. *Current Biology*, *29*(18), 3041-3052.e4. https://doi.org/10.1016/j.cub.2019.07.086

Yoshida, S., Cui, S., Ichihashi, Y., & Shirasu, K. (2016). The Haustorium, a Specialized Invasive Organ in Parasitic Plants. *Annual Review of Plant Biology*, *67*(1), 643–667. https://doi.org/10.1146/annurev-arplant-043015-111702

Yuan, M., Ngou, B. P. M., Ding, P., & Xin, X.-F. (2021). PTI-ETI crosstalk: an integrative view of plant immunity. *Current Opinion in Plant Biology*, *62*, 102030. https://doi.org/10.1016/j.pbi.2021.102030

Yule, K. M., Koop, J. A. H., Alexandre, N. M., Johnston, L. R., & Whiteman, N. K. (2016). Population structure of a vector-borne plant parasite. *Molecular Ecology*, *25*(14), 3332–3343. https://doi.org/10.1111/mec.13693

Zhang, X., Berkowitz, O., Teixeira da Silva, J., Zhang, M., Ma, G., Whelan, J., & Duan, J. (2015). RNA-Seq analysis identifies key genes associated with haustorial development in the root hemiparasite Santalum album. *Frontiers in Plant Science*, *6*. https://www.frontiersin.org/articles/10.3389/fpls.2015.00661

Zhang, M., Suren, H., & Holliday, J. A. (2019). Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. *Genome Biology and Evolution*, *11*(8), 2256–2272. https://doi.org/10.1093/gbe/evz151

Zhang, J., & Zhou, J.-M. (2010). Plant Immunity Triggered by Microbial Molecular Signatures. *Molecular Plant*, *3*(5), 783–793. https://doi.org/10.1093/mp/ssq035

Zhong, R., Cui, D., & Ye, Z.-H. (2019). Secondary cell wall biosynthesis. *New Phytologist*, *221*(4), 1703–1723. https://doi.org/10.1111/nph.15537

Zhou, L., Chen, F., Pan, H., Ye, J., Dong, X., Li, C., & Lin, F. (2016). Identifying Virulence-Associated Genes Using Transcriptomic and Proteomic Association Analyses of the Plant Parasitic Nematode Bursaphelenchus mucronatus. *International Journal of Molecular Sciences*, *17*(9), 1492. https://doi.org/10.3390/ijms17091492