The genetics of host adaptation in the parasitic plant *Striga hermonthica*

A thesis submitted by

Oliver L. Pescott

In partial fulfilment of the requirement for admittance to the degree of Doctor of Philosophy at the University of Sheffield

The University Of Sheffield.

Department of Animal & Plant Sciences
University of Sheffield
Sheffield
S10 2TN
United Kingdom

April 2013

Acknowledgements

I would like to thank my two supervisors, Professors Julie Scholes and Roger Butlin, for accepting me onto this project, and for their encouragement, supervision, and ideas during the past four years. I also thank them both for reading all parts of this thesis and for making numerous suggestions for improvement and clarity. I am very grateful to Drs Tom van Mourik and Hamidou Traore, both for coordinating the sampling of *Striga* in Bukina Faso and Mali, and for sharing their data on field emergence. Dr Anna Santure provided essential guidance on the use of CRI-MAP at a time of need. I thank Dr Raj Whitlock for sharing the latest version of his AFLP scoring program, and for explaining the procedure. Maria-Elena Mannarelli was a regular sounding board for all problems AFLP-related, and therefore regularly kept my sanity in check; Andy Krupa provided advice in the initial stages of the project. I acknowledge the funding of the BBSRC.

In Lab C45, I would especially like to thank Maggi Killion for her assistance. My lab colleagues over the past four years were a frequent source of support and good cheer, particularly Dr Rodrigo Echegoyén-Nava, Dr Kay Guccione, Dr Mamadou Cissoko, Dr Arnaud Boisnard, Anne-Laure Hepp and Serina Aktar, as well as many others. Thank you also to my friend Dr Chris Galley, who provided excellent proof-reading at short notice.

I would like to thank my parents, partly for not asking too many questions, but mainly for always encouraging me to do what I loved. Finally, my lovely wife Fumiko provided me with boundless support and understanding, and I thank her with all my heart.

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

The obligately outbreeding root hemiparasite *Striga hermonthica* (Orobanchaceae) is a serious threat to subsistence agriculture in sub-Saharan Africa. Resistance to this parasite in its crop hosts, such as rice, sorghum and maize, is not common, and the evolution of host adaptations that are able to overcome new sources of resistance is an ever-present risk. Research into host adaptation in *S. hermonthica* has generally sought to correlate the genetic relationship between *Striga* individuals with host identity; however, such approaches must be supported by lab-based evidence of host adaptation, otherwise *ad hoc* field sampling may result in the confounding of host identity with isolation-by-distance. Additionally, genetic variation used to reconstruct relationships is unlikely to provide an insight into relationships at functional loci underlying host adaptation. In this thesis, I use a range of new approaches to investigate several different aspects of parasite adaptation in the *S. hermonthica*-*Sorghum bicolor* pathosystem.

Host adaptations, or pre-adaptations, are commonly revealed using tests for differential virulence between *Striga* populations and host genotypes; that is, by demonstrating population-level genotype-by-genotype interactions. Evidence for such interactions was found between three West African populations of *S. hermonthica* and five sorghum cultivars. These interactions were shown to be strongest at the parasite post-attachment life stage, and to depend on the parasite virulence metric used. Environmental influences on host-parasite interactions were strong and variable between years and sites.

Candidate genetic loci for virulence, responding to selection in a micro-evolutionary fashion, were identified by F_{ST} differentiation-based approaches ('outlier analyses') that aim to uncover associations between particular loci and environmental drivers, such as host identity. An AFLP outlier analysis was used on *Striga* plants parasitising nine sorghum cultivars in a field trial in Burkina Faso. Significant locus-specific differentiation was detected at 14 out of 1275 loci. However, predicted allele frequencies at these loci did not correlate with a field measure of *Striga* virulence across host-selected sub-populations. Simulation results suggested that the estimated levels of F_{ST} at outlying loci could mean that alleles underlying host adaptation exist at intermediate frequencies in populations.

A three-generation pedigree, created from a cross between *S. hermonthica* individuals from an East African and a West African population, enabled further insights into the genetics of adaptation. Individuals from a pseudo-backcrossed F_1 (BCF_1) generation, grown on two different sorghum hosts and in axenic culture, indicated significant host-related segregation distortion. Analyses of virulence in the BCF_1 also provided strong evidence for epistasis, and for an effect of maternal identity. A second outlier analysis of host adaptation, investigating the East African population used in the pedigree, indicated some correspondence between outlier loci and loci found to be differentially segregating between different hosts in the BCF_1 generation, and demonstrated the differing genomic extents of these phenomena. The results accumulated across these experiments provide evidence for a complex, polygenic basis to virulence in *S. hermonthica*.

Table of Contents

Chapter 1. General Introduction

1.1 The genus *Striga*: Parasitic weeds in Africa

Within the flowering plants, nutritional parasitism has arisen on at least 11 occasions in many different families (Barkman *et al.* 2007). One family, the Orobanchaceae, now consists almost entirely of parasitic plants; this is due to the breaking-up of the Scrophulariaceae *sensu lato*, and the reorganisation of its constituent genera based on several recent phylogenetic analyses (Park *et al.* 2008). A monophyletic Orobanchaceae suggests that the parasitic plant haustorium, the 'specialized organ of absorption' (Kuijt 1977) essential to the root parasitic lifestyle, evolved only once; variations on the root parasitic theme, e.g. facultative or obligatory parasitism and hemiparasitism or holoparasitism, are therefore derived characters which vary across the clade (Yoder *et al.* 2007). Hemiparasitism, the nutritional life-style of the genus of plant parasites with which this thesis is concerned, is the case where a parasitic plant contains chlorophyll, and can photosynthesise, but depends on its host for water and nutrients (Press & Graves 1995).

Hemiparasites of the Orobanchaceae are known to have important effects on ecosystems, often gaining for them the label of ecosystem engineers; these effects are mediated through a range of ecophysiological traits (reviewed in Phoenix & Press 2005). Agroecosystems constitute an extreme case, demonstrating the ability of root hemiparasites to exert large effects on their hosts over large areas of land. The reduction of host plant productivity resulting from nutritional parasitism can increase community diversity in (semi-)natural ecosystems (Phoenix & Press 2005); however, in farmers' fields, the most obvious and important outcome is the reduction of crop yields.

Within the Orobanchaceae, *Striga* is a genus of obligate root hemiparasites, which has been called 'the greatest biological constraint to food production in Africa' (Ejeta 2007). *Striga*, the witchweeds, is the most important genus of cereal crop weeds in the semi-arid tropics of sub-Saharan Africa (Parker & Riches 1993). *Striga* can cause a total loss of crop yield, but the loss from a typical infestation is thought to be around 50% (Ejeta 2007). The crop species primarily parasitised by *Striga* species include a large range of the staple food crops of African subsistence farmers; these include maize (*Zea mays*), sorghum (*Sorghum bicolor*; Fig. 1.1), pearl millet (*Pennisetum glaucum*), upland rice (*Oryza sativa*), and the grain legume, cowpea (*Vigna unguiculata*). Estimates from the 1990s put the level of infestation of *Striga* at around 50 million hectares, two-thirds of the total area of cereals and legumes in sub-Saharan Africa at that time (Musselman *et al.* 2001); the full range of estimates varies from 21 to 50 million hectares (Emechebe *et al.* 2004). However, the sources from which these figures were originally gathered are very dated, and it is generally considered that the spread of *Striga* has been accelerating in recent years (Ejeta 2007; Parker 2009). Estimates of the numbers of people affected by *Striga* are also quite out-of-date, but the most recent figure cited is 300 million people (Emechebe *et al.* 2004; Ejeta 2007). Information from farmers suggests a recent spread of *Striga*, linked to human population growth and the associated expansion of agriculture onto marginal lands (Parker 2009); these trends are themselves bound up with increased poverty and hunger, compounding the problem through time. This has led to the current desperate situation, recently characterised as a *Striga* 'pandemic' (Ejeta 2007).

Figure 1.1. *S. hermonthica* parasitising sorghum in an Ethiopian field. Picture courtesy of Prof. J.D. Scholes, University of Sheffield.

Of the 28 species of *Striga* currently recognised in Africa, out of around 40 worldwide, 9 are known to parasitise crop hosts to some extent (Mohamed *et al.* 2001). Of these, the bulk of lost crop yield is due to 3 species (Berner *et al.* 1995): *Striga asiatica* (L.) Kuntz; *Striga hermonthica* (Del.) Benth.; and *Striga gesnerioides* (Wald.) Vatke. Thus,

most African *Striga* species are not weedy, and can be found in natural and semi-natural habitats; however, some species have evolved, in tandem with their crop hosts, to become the serious threats to food production that they are today. For example, *S. hermonthica*, probably the single most serious parasitic weed in the world (Parker & Riches 1993), is only known from agroecosystems (Mohamed *et al.* 2001). *Striga asiatica* and *S. hermonthica* only infect monocotyledons of the Poaceae; *S. gesnerioides* can infect a range of dicotyledonous families, but its most serious impact on subsistence agriculture is due to its parasitism of cowpea (Fabaceae). These plant pathogens also differ in their geographical distributions: the main range of *S. asiatica* is southern and central Africa; *S. hermonthica* extends west to east across central Africa, and south to Tanzania (Fig. 1.2); *S. gesnerioides* has the largest distribution of any witchweed in Africa, it is widespread in west and east Africa, and extends down the eastern side of the continent as far as South Africa (Mohamed *et al.* 2001).

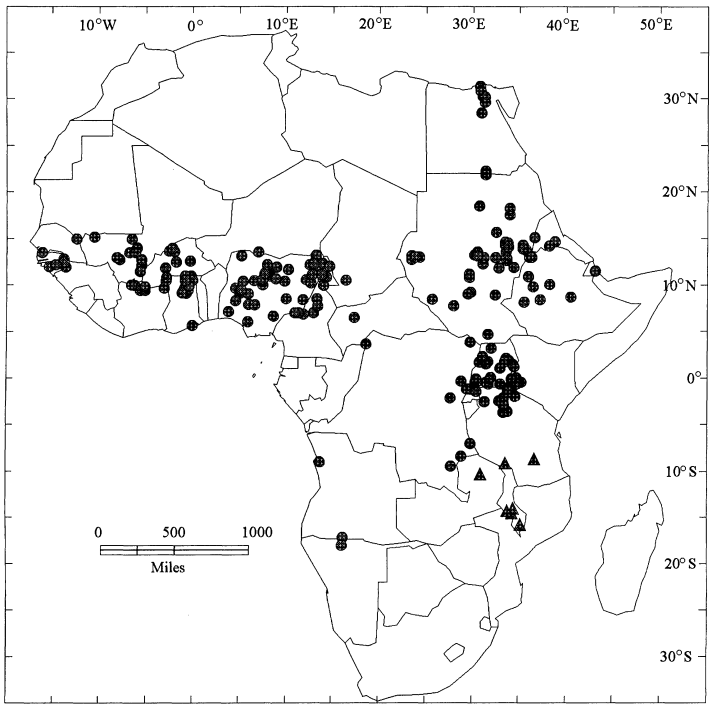

Figure 1.2. Distribution map of *S. hermonthica* (circles) in Africa, based on herbarium specimens (Mohamed *et al*. 2001). (Triangles represent the non-weedy *S. gracillima.*)

One of the main reasons why weedy *Striga* are so much of a problem for agriculture is their extreme fecundity. A single plant can produce from 50,000 to over 200,000 seeds depending on the species, with each tiny seed measuring approximately 0.3 × 0.2 mm (Parker & Riches 1993); these seeds may remain viable in the soil for up to 14 years (Parker & Riches 1993; Berner *et al.* 1995; Yoder & Musselman 2006). This means that seed banks can be vast, and may act as reservoirs of diversity. Genetic diversity is likely to equate to evolvability: genotypes capable of parasitising a range of host species and cultivars may be present in the seed bank. This diversity may be generated in several different ways, depending on the species. Entire genotypes may be moved between fields or regions through farming practices and seed exchange or sale (Berner *et al.* 1995). In addition, differences in mating system among *S. asiatica*, *S. hermonthica* and *S. gesnerioides* affect the levels of gene flow between plants. *Striga asiatica* and *S. gesnerioides* are primarily autogamous (Musselman *et al.* 1982), whilst *S. hermonthica* is obligately allogamous (Safa *et al.* 1984). Models that attempt to evaluate the relative risks of plant pathogen evolution state that autogamous and allogamous pathogen strategies present different levels of risk for crop plants (McDonald & Linde 2002): plants exhibiting autogamy (which is, however, often incomplete; for example, see Nickrent & Musselman 1979) may not be evolving rapidly, but if pathogenic genotypes are created through infrequent outcrossing events, or mutation, subsequent inbreeding can mean rapid increases in the number of parasites, all with the same pathogenic genotype. In addition, founder events are more likely, because a single plant is sufficient to establish a new population. On the other hand, allogamous plant populations are continually changing through the exchange and recombination of genetic material within and between populations; populations may therefore contain considerable

diversity, and the associated evolutionary potential (McDonald & Linde 2002; Mohamed *et al.* 2007).

1.2 The *Striga* life-cycle and host resistance

Striga seeds require exposure to a class of chemicals called strigolactones in order to germinate (Fig. 1.3a). Strigolactones had been known to induce germination in certain parasitic plants for around 25 years before it was discovered that this class of chemicals was also involved in plant shoot branching, and in inducing the branching of mycorrhizal fungi in the soil (Yoder & Scholes 2010). *Striga* exploits these essential host plant hormones in order to locate host plant roots (Mach 2010). After germination, the *Striga* radicle elongates and the tip of the radicle begins to differentiate into the specialised organ of parasitism, the haustorium (Kuijt 1977), that will be used to penetrate the host root. So-called 'haustorium inducing factors' (HIFs) are thought to be produced by the oxidation of host root cell-wall localised benzoquinones (Kim *et al.* 1998; Fig. 1.3b); the oxidation is carried out by host peroxidases activated by the release of hydrogen peroxide from the *Striga* radicle (Keyes *et al.* 2007; Scholes & Press 2008). *Striga* therefore actually elicits its own HIFs from the host root (Keyes *et al.* 2001). To date, only one HIF has been isolated, from the roots of sorghum, this is 2,4-dimethoxy-*p*-benzoquinone (DMBQ) (Chang & Lynn 1986); however, this compound may potentially be widespread in the plant kingdom, derived as it is from lignin and phenolic acids (Yoder & Scholes 2010). Host resistance can be linked to low production of germination stimulants or HIFs (Yoder & Scholes 2010). In these cases the resistance is known as 'pre-attachment' resistance (e.g. Jamil *et al.* 2011); that is, the resistance operates before the attachment of the haustorium to the host root.

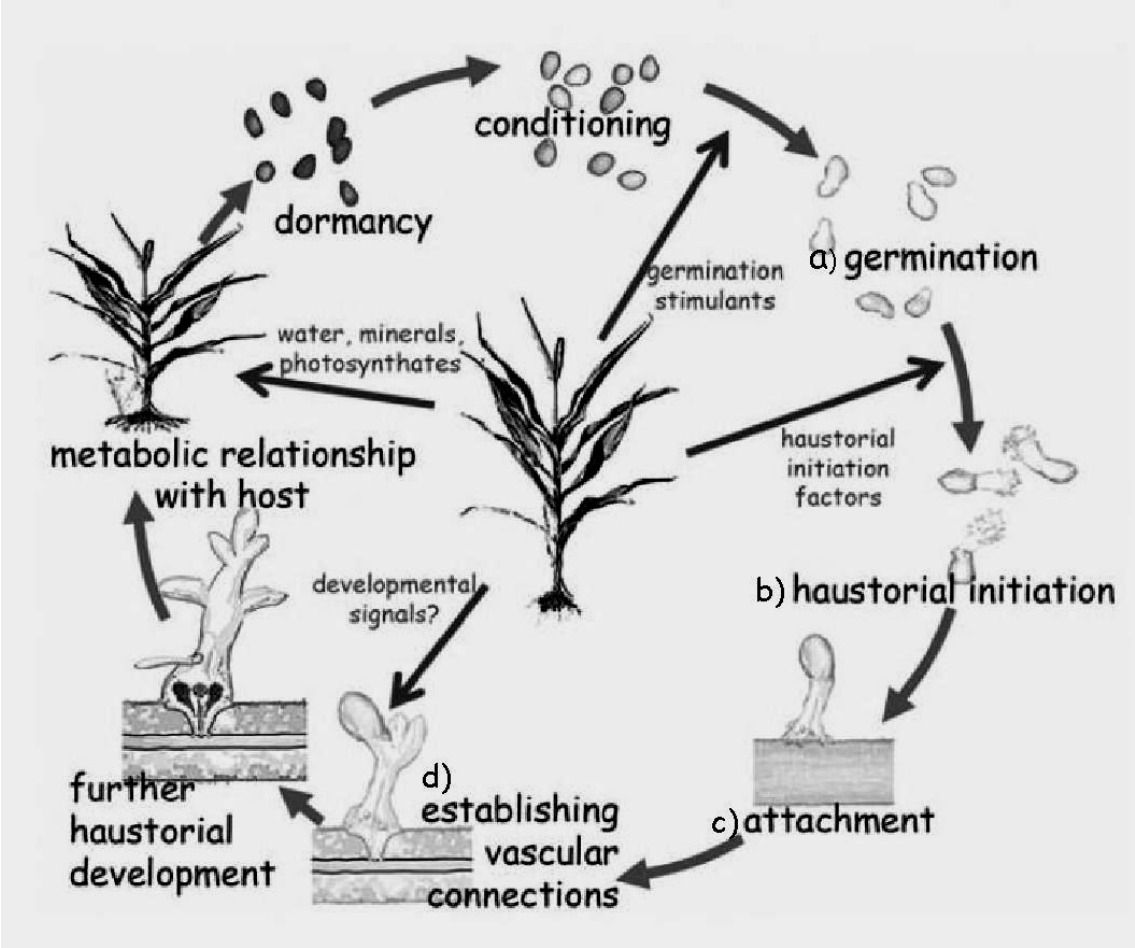

Figure 1.3. The *Striga* life-cycle. Adapted from Rich & Ejeta (2007).

As it develops, the haustorium becomes covered in sticky hairs that assist in anchoring it to the host root (Reiss & Bailey 1998; Fig. 1.3c). Following *Striga* attachment, the parasite penetrates the host root cortex and endodermis, and connections between host and parasite xylems are established (Fig. 1.3d). As an obligate parasite, this step is essential for the survival of *Striga*, and allows the transfer of host-derived nutrients and water (Press & Graves 1995). Clearly, many plant-plant molecular interactions must take place between the attachment of the parasite and the successful establishment of xylem-xylem continuity. Much less is known about the molecular basis of host

resistance acting at this stage, which is normally known as 'post-attachment' resistance (e.g. Cissoko *et al.* 2011). However, various host post-attachment resistance phenotypes have been described through microscopic examination of incompatible interactions: for example, the blocking of parasite growth in the host root cortex, at the endodermis, and before, or after, connection to the host vasculature have all been reported, either for *Striga*, or for the closely related holoparasitic genus *Orobanche* (Yoshida & Shirasu 2009; Yoder & Scholes 2010). These observations suggest that a variety of molecular responses may underlie post-attachment resistance, even within a single host species; for example, for various sorghum cultivars, endodermal thickening, pericycle lignification and silica crystal deposition have all been observed in post-attachment resistance to *S. hermonthica* (El Hiweris 1987).

The most recent advance in understanding the molecular basis of host post-attachment resistance to *Striga* has been for the *S. gesnerioides*-cowpea interaction. Work on the existence of 'races' of *S. gesnerioides*, with particular patterns of virulence against sets of cowpea cultivars (Lane *et al.* 1993; Botanga & Timko 2006), recently led to the identification of the first sequenced resistance gene (or 'R-gene') against a parasitic plant, and confirmation that gene-for-gene resistance can occur in these interactions (Li & Timko 2009). The R-gene encoded protein contained the 'coiled-coil nucleotide-binding-site leucine-rich-repeat' (CC-NBS-LRR) pattern of amino acid structural motifs, commonly found in proteins involved in plant resistance to microbial pathogens (Jones & Dangl 2006) and insect herbivory (de Meaux & Mitchell-Olds 2003). When the coding gene (*RSG3-301*) in the previously resistant cowpea cultivar B301 was silenced, susceptibility to the *S. gesnerioides* race 'SG3' was created; however, the cultivar remained resistant to other races of the parasite, thus supporting a gene-for-gene

interaction hypothesis (Li & Timko 2009). As noted above (section 1.1), *S. gesnerioides* is strongly autogamous (Musselman *et al.* 1982), and the formation of parasite races with particular reactions to different host species or cultivars (often called host specificity; Vasudeva Rao & Musselman 1987) may be more likely to occur for inbreeding species, because of the almost clonal nature of successful genotypes, where most genetic variation is between populations (McDonald & Linde 2002). It seems likely, therefore, that this breakthrough in understanding one particular interaction between *S. gesnerioides* and cowpea may not automatically provide an insight into the interactions between the obligately outcrossing *S. hermonthica* and its hosts; indeed, because of its mating system, *S. hermonthica* may present a higher risk to agroecosystems and sustainable host resistance (McDonald & Linde 2002). The genetics of *Striga* species, and their interactions with host genotypes, are discussed further below (section 1.4).

1.3 Strategies for controlling *Striga*

Many control strategies for weedy *Striga* species have been developed or proposed. The problem is often that these strategies are unaffordable for subsistence farmers, who make up approximately 75–80% of farmers in sub-Saharan Africa (Gressel *et al.* 2004); for example, nitrogen fertilisation, soil fumigation, and chemical herbicides all have some degree of efficacy for controlling *Striga*, but are beyond the reach of the average resource-limited farmer in sub-Saharan Africa (Gressel *et al.* 2004). Simpler and cheaper solutions more often used by subsistence farmers include hand-pulling, sowing rotation crops, trap-cropping, mixed cropping and altering the time or method of sowing

(Parker & Riches 1993). These techniques aim to reduce levels of infestation through depletion of the *Striga* seed bank and by preventing the setting of seed. By reducing field infestations these techniques could also improve farmers' yields. Another well-established method for reducing parasitism lies with crop breeding (e.g. Haussmann, Hess, Koyama, *et al.* 2000). Desirable crop genotypes for the reduction of parasitism can be characterised as either tolerant or resistant: tolerant genotypes are those showing smaller reductions in yield than susceptible genotypes under the same level of *Striga* infestation, whilst resistant genotypes show less *Striga* attachment and higher yields than susceptible genotypes under the same conditions (Haussmann, Hess, Welz, *et al.* 2000).

As yet, no crop cultivar, or wild relative, with full resistance (i.e. immunity) to any *Striga* species has been found. However, tolerant cultivars of maize, sorghum and rice have been identified (Scholes *et al.* 2007; Kaewchumnong & Price 2008); and novel types of post-attachment resistance have been described in rice (Gurney *et al.* 2006), sorghum (Mohamed *et al.* 2003), and in a wild relative of maize (Gurney *et al.* 2003). Progress has been made in breeding complex traits underlying broad-spectrum resistance in sorghum into farmer-preferred or locally-adapted cultivars; these cultivars, with high levels of resistance and high yields, are now starting to have a positive impact in several African countries (Ejeta *et al.* 2007; Kapran *et al.* 2007).

Resistant cultivars have the potential to both deplete the soil seed bank and prevent *Striga* reproduction; tolerant cultivars can also deplete the seed bank, but other methods must be used in conjunction to prevent seed-setting (Rodenburg *et al.* 2005). It is generally recognised that any successful *Striga* control programme must be one which integrates a range of methods that are appropriate for the local conditions, both

environmental and socio-cultural (Ransom *et al.* 2007). This approach extends to the design and eventual deployment of improved cultivars: tolerant cultivars can increase the *Striga* seed bank without supplementary control; and where plants do have some degree of resistance, this is often variable across environments (Haussmann, Hess, Welz, *et al.* 2000), meaning that the breakdown of host resistance could occur if new, more virulent, genotypes of *Striga* are selected for (e.g. Lane *et al.* 1994). Therefore, it is of great importance for the durability of resistant germplasm, even if the resistance is broad-spectrum, that measures are taken to ensure that virulent genotypes of *Striga* are not selected for in the field by improved cultivars (McDonald & Linde 2002). Given the frequency of crop seed contamination by *Striga* seed that has been found in surveys of African markets (20–40%; Berner *et al.* 1994), and other potential routes of genotype flow (Berner *et al.* 1995), it is possible that resistance breakdown in one area could spread. Because of this, the requirement for an improved understanding of the relationship between *Striga* genetic variation and host resistance has been noted frequently in the parasitic plant literature (Vasudeva Rao & Musselman 1987; Parker & Riches 1993; Haussmann, Hess, Welz, *et al.* 2000; Mohamed *et al.* 2007; Scholes & Press 2008). Even so, relatively few studies on *Striga* genetic diversity have been carried out, and the potential for large increases in knowledge exists, with broad implications for cultivar testing and deployment, and the understanding of the molecular and population genetic bases of *Striga* virulence (Mohamed *et al.* 2007).

1.4 *Striga*-host interactions: studies and conceptual approaches

It has long been noted that there is variation in host specificity within *Striga* species (see

Vasudeva Rao & Musselman 1987 for a review of studies up to that date). It has generally been the practice to divide the observed patterns of adaptation found for a species of *Striga* into two types: inter-species host specificity and intra-species host specificity. Within *S. hermonthica* it has been suggested that there will be populations that are specific to certain crop species (inter-species), or to a certain subset of genotypes (e.g. cultivars or landraces) within a crop species (intra-species) (Vasudeva Rao & Musselman 1987; Parker & Riches 1993; Mohamed *et al.* 2007).

For *S. hermonthica*, the most cited example of host specificity is the inter-species specificity reported from parts of Africa and India for sorghum and pearl millet (Parker & Riches 1993; Ejeta 2007). Vasudeva Rao & Musselman (1987) cite four 'cross-infectivity' studies investigating the degree of *S. hermonthica* population specificity for either sorghum or pearl millet. Three of these studies (Wilson-Jones 1955; Parker & Reid 1979; Hosmani & Parker 1980) found, through either cross-inoculation of hosts, or cross-stimulation of germination using host root exudates, that host preferences could be demonstrated for the populations of *S. hermonthica* tested. The fourth study (Ramaiah 1984) found less specificity in cross-inoculation tests, and suggested that this was due to *Striga* populations representing 'intermediate forms'. Field-based virulence trials, where the same set of crop hosts are grown in different locations, have also suggested broad geographic patterns in *S. hermonthica* inter-species specificity (Vasudeva Rao & Musselman 1987). However, given the age of these studies, it is unclear how many of these patterns of virulence are extant; plant pathosystems are continually evolving, and human agency in moving crop and parasite seed around the landscape is a significant factor in *Striga* spread and evolution, as noted above.

Ejeta (2007) gives a brief and partial report on the current state of one example of inter-

species specificity in the field. Interestingly, he reports that for *S. hermonthica* in Sudan, where inter-species specificity for sorghum and pearl millet was previously found (Wilson-Jones 1955), changes in the regions growing pearl millet in the 1980s resulted in the gradual appearance of *S. hermonthica* virulent on pearl millet, where previously there had been specificity for sorghum (Ejeta 2007). This phenomenon suggests that *Striga* host specificity can be based on regional separation of crop cultivation, which can break down when these patterns are disrupted, suggesting that *S. hermonthica* adaptations to host species may change on a scale of years rather than tens of years (see also Parker & Riches 1993 pp. 6–7). However, in the western Eritrea region of Ethiopia, sorghum and pearl millet are grown in the same areas, but *Striga* is reportedly not currently parasitic on pearl millet (Ejeta 2007); this may be due to the suggested lag phase for *S. hermonthica* populations switching between sorghum and pearl millet (Parker & Riches 1993), suggesting that adapting to both species simultaneously may also take time (although evidence suggests that this is possible, e.g. Estep *et al.* 2011).

Intra-species specificity is defined here as the presence of reproducible, differential cross-reactions between *Striga* populations and a set of genotypes within a host species. Evidence for this within *Striga* species has in the past been considered equivocal (Vasudeva Rao & Musselman 1987). However, recent evidence for populations of *S. gesnerioides* with specific differential reactions across cultivars of cowpea is very clear, as discussed above (section 1.2; Botanga & Timko 2006; Li & Timko 2009). Vasudeva Rao & Musselman (1987) equivocated on the existence of *Striga* intra-species specificity for three reasons: they pointed out that the observed differential reactions in the field could also be due to: "(1) *Striga* intensity [i.e. infestation] differences, (2) instability of the resistance of the host, or (3) the several soil and environmental factors

which affect (weaken/strengthen) the resistance of the host cultivars" (Vasudeva Rao & Musselman 1987). The last two of these can be thought of as environmental components of the *Striga*-host interaction; that is, to use the terminology of analysis of variance, the main effect of the environment and its interaction effects. Vasudeva Rao & Musselman (1987) also suggested that certain types of host resistance mechanism, such as the mechanical barrier of lignification in sorghum cultivar N13 (Maiti *et al.* 1984), would be unlikely to be overcome by variation in *Striga* virulence mechanisms, suggesting that variation in parasite virulence between cultivars was more likely to be due to variation in the expression of host resistance. However, as Parker & Riches (1993 pp. 53–54) point out, a variety of mechanisms may actually underlay the resistance of a variety such as N13, and more research is required before it can be concluded that variation in the field resistance of particular cultivars is not at least partially due to parasite genetic variability. Clearly, the observations of Vasudeva Rao & Musselman (1987) must be taken into account when assessing *Striga* intra-species specificity in the field (Haussmann, Hess, Welz, *et al.* 2000): *Striga* population-host genotype interactions are most convincingly examined in controlled, lab-based experiments, where the environmental component is stable and moderate, and where host and parasite main effects and their interactions can be examined. Of course, environmental effects are a real part of any natural host-parasite system (Wolinska & King 2009), and their impact should also be examined in well-designed field trials or common garden experiments (Nuismer & Gandon 2008). Although work on the genetics of host resistance has progressed through the use of lab-based pot and root-observation chamber studies (e.g. Gurney *et al.* 2003, 2006), few studies have used these techniques to evaluate the genetic component of the host specificity of *Striga* populations. One exception to this is

the work of Lane and colleagues on the *S. gesnerioides*-cowpea interaction (Lane *et al.* 1993, 1996). Their work demonstrated the existence of differential host-parasite resistance reactions under laboratory conditions; specifically, they reported a post-attachment necrotic response of cowpea cultivar 58-57 to a population of *S. gesnerioides* from Burkina Faso, but not to a population from Mali (Lane *et al.* 1993). Subsequent pot studies and molecular marker work confirmed the existence of at least 7 races of *S. gesnerioides* on cowpea in West Africa (Botanga & Timko 2006); work on the basis of resistance to these races in cowpea led to the discovery of gene-for-gene resistance in this pathosystem (Li & Timko 2009; Timko *et al.* 2012). Work of this type, combining lab-based assessments of virulence with population genetics, can therefore help to direct workers to host-parasite interactions that stand the best chance of yielding interesting information concerning the genetic basis of differential virulence.

At this point, it is necessary to express a partial dissatisfaction with the conceptual framework of intra- and inter-species host specificity that has so far dominated research on *Striga* population-host genotype interactions. The word 'specificity' suggests a focused virulence, implying a previous degree of exposure and host adaptation between parasite and host populations (Parker & Riches 1993). The word also suggests restriction of the phenomenon to the parasite populations and host species or genotypes studied in any particular system, when, in fact, differential virulence between parasite and host populations could sometimes be due to pre-adaptive genetic diversity. Whilst the term may be suitable for *Striga*-host associations in the strongly inbreeding *S. gesnerioides*, where, at the level of host species, differentiated parasite lineages exist, even sympatrically (Parker & Riches 1993; Mohamed *et al.* 2001), and, possibly, for the inbreeding *S. asiatica* (e.g. Botanga *et al.* 2002), this conceptual framework seems

unsuitable for the outbreeding *S. hermonthica*, where host preference appears to be a matter of degree than of absolutes (i.e. quantitative rather than qualitative). Whilst this point has been made before (Parker & Riches 1993 p. 198), it has also been ignored (e.g. Mohamed *et al.* 2007). Therefore, throughout this thesis I will express the phenomenon of differential virulence between *S. hermonthica* populations and host genotypes as a population-level genotype-by-genotype interaction (G × G); this terminology is increasingly used in epidemiological and ecological research of all kinds (e.g. Lambrechts *et al.* 2006; Nuismer & Gandon 2008; Wolinska & King 2009; Rowntree *et al.* 2011). It also carries less connotation of historical interaction between species, and admits, by its neutrality, of a greater potential for dynamism and change in species interactions, that are a part of all interactions between species in the natural world (Thompson 1999; Barrett *et al.* 2008; Burdon & Thrall 2009), but perhaps particularly for outbreeding pathogens in agroecosystems (McDonald & Linde 2002).

Additionally, in this thesis I use the word 'population', rather than the word 'race', to describe *S. hermonthica* seed collections with particular population-level reactions against sets of host-species genotypes. Although research on the *S. gesnerioides*-cowpea interaction has used the word 'race' for parasite groups exhibiting differential virulence between host genotypes (Lane *et al.* 1993; Botanga & Timko 2005, 2006), the word is usually associated with plant-microbe pathosystems, and often implies a particular, well-defined genetic interaction, involving R-genes, for example (Agrios 2005). Whilst this has actually proven to be the case for certain interactions between the strongly inbreeding *S. gesnerioides* and genotypes of its cowpea host (Li & Timko 2009), the current lack of experimental evidence for qualitative, all-or-nothing, host specificity for *S. hermonthica* suggests that this terminology may be less appropriate in this instance.

Certainly, the widespread finding of quantitative G × G population-level interactions between *S. hermonthica* populations and host genotypes (Kim *et al.* 1994; Cissoko *et al.* 2011; Huang *et al.* 2012), and the discovery of multiple quantitative trait loci (QTL) for *S. hermonthica* resistance in rice (Gurney *et al.* 2006; Swarbrick *et al.* 2009) and tolerance in sorghum (Kaewchumnong & Price 2008), suggest that host adaptation at this level is a matter of degree (i.e. resistance and virulence are quantitative traits). Unfortunately, this conclusion is somewhat complicated by the fact that most research on *S. hermonthica* has quantified virulence (or host resistance) at the population, or, more accurately, the seed-batch level. That is, counts are usually made of the number of *Striga* plants attached to a host root system (Gurney *et al.* 2003, 2006; Swarbrick *et al.* 2009; Cissoko *et al.* 2011), or emerging from pots or in the field (e.g. Kim *et al.* 1994). Whilst this is a natural approach in the sense that host plants are typically attacked by multiple parasites, in terms of the genetic basis of such a measure, it will be seen that it is not an individual-level quantitative trait, but rather a frequency measurement of a qualitative trait (i.e. a count of the individual parasites that are able to successfully parasitise the host). Therefore, both of the preceding points concerning the evidence for the quantitative bases of virulence and resistance in the *S. hermonthica*-host pathosystem, could also represent situations where populations consist of multiple *Striga* genotypes, differing at a small number of R-gene loci; in this case, QTL for resistance could theoretically represent qualitative resistance genes responding to different *Striga* virulence genotypes in the genepool. Clearly, in a pathosystem where the interaction of practical importance involves infection of a host by multiple genotypes, and the extent of damage to the host is a decreasing function of the amount of co-infection (Fig. 1.4; Gurney *et al.* 1999), a population-level approach is likely to

yield the most useful information concerning interactions of agroecological importance. However, for genetic inference, the subtle difference between a true quantitative trait and the population-level frequency of a qualitative trait should not be forgotten.

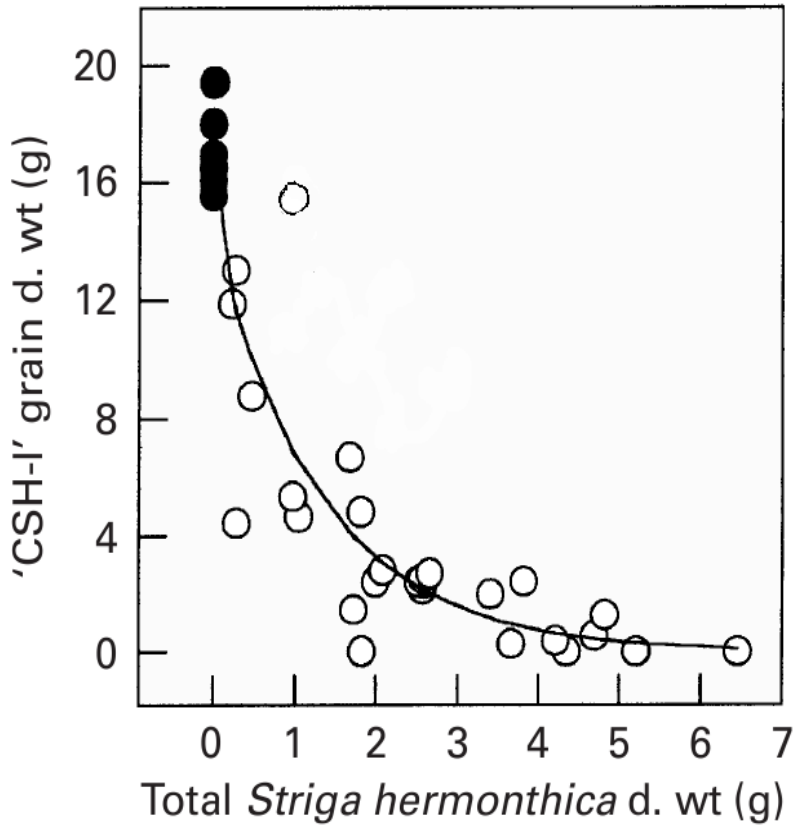

Figure 1.4. The relationship between the infection density of *S. hermonthica* and the grain yield of the sorghum cultivar 'CSH-I' at 100 days after planting. Closed circles indicate the grain yield in the absence of *Striga* infection; open circles represent plants infected at different initial densities of *Striga* seed. Adapted from Gurney *et al.* (1999); d. wt = dry weight.

Related to this point is the use of the term 'virulence'; this term is often singled out for the variety of meanings that have been attached to it over time (Jarosz & Davelos 1995; Barrett *et al.* 2009). Its traditional meaning in plant pathology refers to the ability of a pathogen genotype to infect a particular host genotype, often relating to the gene-for-gene relationship, with the classic example being the flax/flax-rust system (Flor 1956);

therefore, traditionally, a highly virulent pathogen genotype was one that could infect a wide range of host genotypes (Jarosz & Davelos 1995). However, in the ecological and animal pathology literature, 'virulence' has been used to refer to the extent of a pathogen's impacts on its host (Jarosz & Davelos 1995; Barrett *et al.* 2009). The plant pathology world has partly adopted this second definition (Jarosz & Davelos 1995; Sacristán & García-Arenal 2008; Barrett *et al.* 2009), adding to the potential for confusion. In the case of *S. hermonthica*, the term virulence has essentially been used very casually, without any rigorous attempt to relate the word to one or other of the established definitions (e.g. Gurney *et al.* 2003; Cissoko *et al.* 2011; Huang *et al.* 2012). Indeed, given the fact that *S. hermonthica* virulence is usually measured as the frequency of an individual-level qualitative trait (i.e. binary infectivity), and that increasing amounts of attached parasite increasingly reduces host fitness (Fig. 1.4; Gurney *et al.* 1999), the word, and the feature of *S. hermonthica* pathosystems that it seeks to represent, can be thought of as containing both established definitions. In this thesis, I continue with the use of the word 'virulence' to indicate a quantitative (i.e. frequency) measure of the population-level occurrence of successful parasitism, which scales with the extent of a parasite population's impact on a host.

1.5 Prior work on the genetics of *Striga*-host interactions

Some data concerning genetic variation in *Striga* species are available. For the strongly inbreeding species *S. asiatica*, the study of Botanga *et al*. (2002) provided the first information on the relationship between genetic variation and the host adaptations of *Striga* populations. Botanga *et al.* (2002) followed the approach of Lane *et al*. (1993) in

using pot studies to minimise environmental variation, ensuring that genetic variation in *S. asiatica* could be confidently associated with observed differential virulence. Other studies that have sampled *Striga* species from hosts in the field, with the apparent aim of investigating host adaptation, have tended to assume that the host association recorded in the field represents the 'adapted' host of that parasite deme (e.g. Estep *et al.* 2011; Welsh & Mohamed 2011), without using complementary lab investigations to prove the existence of significant host-parasite G × G interactions (i.e. the ability of different *Striga* populations to preferentially parasitise different host genotypes).

Botanga *et al*. (2002) demonstrated the existence of population-level G × G interactions at both the host species and host genotype level for *S. asiatica* populations in Benin: none of the *S. asiatica* populations collected from maize or wild grasses (*Panicum* spp. and *Rottboellia* spp.) were observed to parasitise a susceptible sorghum cultivar (CK60B) in laboratory pot studies, neither would some of the populations collected from wild grasses parasitise maize. Host genotype-by-parasite population interactions were also shown in the lab by the cross-inoculation of the different maize hosts sampled in the field with *Striga* seed collected from these same hosts: the existence of interactions was shown by the fact that maize hosts from two of the six field sites were differentially resistant to the *S. asiatica* populations (Botanga *et al*., 2002). Significantly, the genetic analyses used in Botanga *et al.* (2002) revealed some evidence for among-population genetic variation in *S. asiatica* correlating with host preference. Based on the clustering observed in some of the resulting molecular marker-based dendrograms, the authors proposed that host selection could be the main force driving the observed inter-population genetic differentiation within *S. asiatica* in Benin. For example, one dendrogram showed three *S. asiatica* populations separating into two

'major classes' correlating with host species (wild grasses or maize, Fig. 1.5; Botanga *et al*. 2002); however, the bootstrapped branch supports were very low in some cases (Fig. 1.5; Botanga *et al.* 2002). In addition, Botanga *et al.* (2002) also found a strong effect (R^2 = 0.61) of geographic distance on inter-population genetic distance. Clearly, as for most host-pathogen interactions, much more evidence and knowledge of causal mechanisms will be needed before firm conclusions, or even broad generalisations, can be made regarding host-parasite interactions that have been formed by host selection in the field (Barrett *et al.* 2008; Burdon & Thrall 2009).

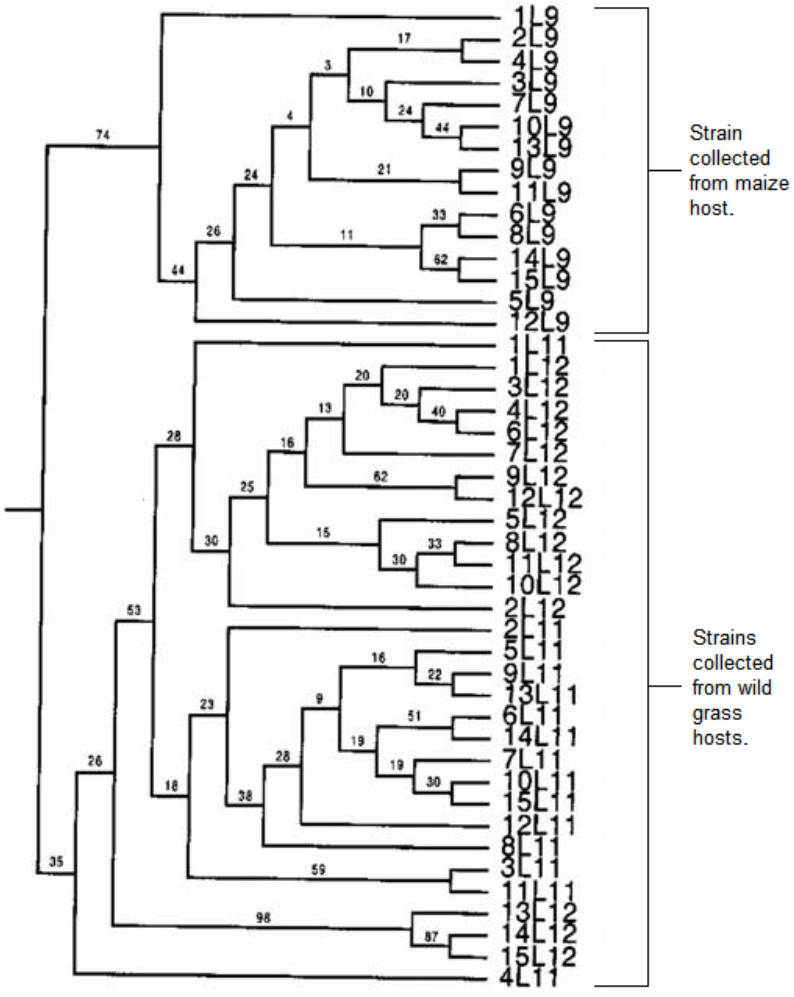

Figure 1.5. Dendrogram showing the relationships between individuals from three 'strains' of *S. asiatica* (L9, L11 & L12; adapted from Botanga *et al*. 2002). The dendrogram was constructed using the 'unweighted pair-group method using averages' (UPGMA) clustering algorithm with 117 amplified fragment length polymorphism (AFLP) markers.

Studies of genetic variation within *S. hermonthica* also exist, and attempts to relate this variation to host-parasite G × G interactions have been made. Olivier *et al*. (1998) used allozymes to investigate populations of *S. hermonthica* growing on different host species across Africa. Olivier *et al.* (1998) suggested that high levels of genetic divergence between *S. hermonthica* plants parasitising particular host species would indicate an ancestral host specialisation, whilst low divergence would be more indicative of recent adaptation. Presumably, however, even an ancestral specialisation for host species specificity within *S. hermonthica* would not rule out more recent, flexible adaptations to host genotypes within species. Olivier *et al*. (1998) found little evidence of genetic divergence between the populations of *S. hermonthica* sampled from the different host species investigated. They concluded that even adaptation to different host species may be recently evolved in *S. hermonthica*, although no cross-infectivity analysis was presented to prove that the populations investigated were truly preferentially adapted to the different host species sampled; that is, G × G interactions showing differential virulence on the different host species were not demonstrated (Olivier *et al.* 1998). If we accept the two allozyme loci investigated by Olivier *et al*. (1998) as representative of inter-population divergence relevant for host-preference (which is, admittedly, unlikely for a very small number of markers if there is gene flow between the putative host-specialised populations), then we could conclude that: (1) *S. hermonthica* has recently adapted to host species (and, by extension, host genotypes), or, (2), if there was actually low host-specialisation amongst populations, then *S. hermonthica* has enough standing genetic variation to parasitise a range of host species and genotypes. In either case, the outlook for crop breeders would not be bright (cf. the conclusions of McDonald & Linde 2002 on the risks presented by outbreeding pathogen

species).

Gethi *et al*. (2005) suggest "that it may not be necessary to develop [host] cultivar and resistance breeding or selection programmes targeting particular regions", based on a finding of low *S. hermonthica* among-population variation in Kenya; however, this is a heavily assumption-laden prediction, given that, like Olivier *et al*. (1998), Gethi and colleagues did not test for parasite-by-host species or genotype interactions, and that high genetic differentiation may not necessarily correlate with specificity (Olivier *et al*., 1998). Low among-population genetic divergence may be due to most genetic variation being within-population, however, even a low among-population component to genetic variation could create different population-level host-parasite G × G interactions if that variation was at functional sites for parasite virulence (Huyse *et al.* 2005; Barrett *et al.* 2008).

Other studies of genetic diversity in *S. hermonthica* have also concluded that divergence due to isolation-by-distance is higher than that due to host species (Bharathalakshmi *et al.* 1990; Musselman *et al.* 1991; Kuiper *et al.* 1996), or that neither was significant (Gethi *et al*. 2005); this would not be unexpected for neutral markers, even if there is strong differentiation for functional (i.e. host-adapted) genetic variation. This pattern seems likely to be at least partially due to the obligately outcrossing life history of *S. hermonthica*: populations growing on host species and genotypes locally may freely interbreed, whilst the fecundity of the species, and the large seed bank, help to ensure that there are always likely to be genotypes retaining the ability to parasitise different hosts; isolation-by-distance will then be primarily responsible for population structure at neutral markers in the absence of high migration. It is possible that understanding genetic variability in *S. hermonthica* may help to target appropriate sources of host

resistance against particular populations; however, as Olivier *et al*. (1998) suggest, high adaptability to new hosts in *S. hermonthica* may correspond to a low probability of finding high levels of host resistance in cultivars, because the time period for a strong coevolutionary response has been too short and sexual recombination can create large numbers of different genotypes, especially if host resistance is quantitative (McDonald & Linde 2002; Barrett *et al.* 2008). This has been the case so far (Hearne 2009), and crop breeding programmes now focus on attempting to 'pyramid' resistance genes in order to create broad-spectrum, quantitative resistance to *Striga* (Ejeta *et al.* 2007).

Further evidence for the existence of high adaptability of *S. hermonthica* has been provided by studies examining within-population selection by different host genotypes in the laboratory; that is, the individual basis of population-level G × G interactions has been investigated. For example, Koyama (2000a) demonstrated genetic divergence between subsets of parasites from one population of *S. hermonthica* selected for by five different sorghum cultivars. Huang *et al.* (2012) came to a similar conclusion using population genetic outlier analyses for the detection of loci under selection (see Chapter 3 for further information); Huang and colleagues also demonstrated population-level G × G interactions between three rice cultivars and the *S. hermonthica* population used in the outlier experiment, suggesting that variation at a small number of loci can drive virulence differences between populations of *S. hermonthica*. Such results indicate that standing diversity in populations of *S. hermonthica* relevant for host adaptation is likely to respond quickly to changes in selection pressure. These observations provide evidence for similar mechanisms leading to population-level G × G interactions in the field, albeit, interactions underlain by high adaptability, and thus potentially leading to relatively rapid changes from very low to very high virulence in *Striga* populations

repeatedly presented with the same, previously resistant host (cf. Ejeta 2007). We should also be aware that what appears to be the gradual adaptation of existing populations may also be due to an influx of new genotypes via the routes noted above (e.g. seed exchange, wind or livestock movement; Berner *et al.* 1994).

1.6 The *S. hermonthica*-host pathosystem in the wider context of plant pathology

Whilst parasitic plants have tended to receive less attention than microbial pathogens (e.g. Musselman *et al.* 2001), there is a rich theoretical literature on parasite ecology in general (see the recent book-length reviews of Poulin 2008 and Schmid-Hempel 2011), and this work may guide our expectations of the patterns of host adaptation that are likely to be found in particular pathosystems, such as between the outcrossing *S. hermonthica* and its gramineous hosts.

Sexual reproduction is thought to be common amongst parasites, although there appears to have been no systematic attempt to assess its relative level of occurrence compared to non-parasitic organisms. Despite the abundance of work on the parasite-driven maintenance of sex in hosts, comparatively little work has been performed on the phenomenon of sex in parasites themselves (Clay & Kover 1996; Thompson 2005). Recent modelling efforts have suggested that sex should increase virulence in parasites (Galvani 2003); the normal contention being that sex should be important for staying ahead of a host's responses to infection (Thompson 2005). Thompson (2005) points out that there is considerable scope for further evaluation of the particular circumstances that favour the evolution and maintenance of mixed mating strategies in parasites; this is particularly relevant for plant pathogens, where a number of the most economically

important plant pathogens exhibit mixed mating (McDonald & Linde 2002). The plant pathogen risk evaluation framework put forward by McDonald & Linde (2002) suggests that a mixed reproductive system is likely to lead to the highest risk of host resistance breakdown. This is due to the fact that occasional outbreeding can produce new multi-locus genotypes, which, if successful, can subsequently spread asexually across large areas. However, sex without an asexual phase can also be a successful strategy for parasites, as demonstrated by the current *S. hermonthica* 'pandemic' (Ejeta 2007). Whilst it is important to understand the conditions that favour the evolution and maintenance of sex in parasites, the patterns of adaptation within a parasite species that confront us across a landscape over short time scales are perhaps more important for applied questions in the area of pathology (Burdon & Thrall 2009): this brings us to the subject of local adaptation, which has been another domain of intense theoretical and empirical investigation for the ecology and genetics of host–parasite interactions (Thompson 2005; Poulin 2008).

Parasite local adaptation is the situation where local parasite genotypes are more virulent on sympatric hosts than on allopatric ones; it is essentially a type of G × G interaction based on variable adaptation across space and time (Thompson 2005). Research in this area grew from the recognition that merely focusing on interactions occurring in a single deme would not capture the range of evolutionary possibilities for particular host–parasite interactions, which could be extremely important for evolution at the landscape scale (Burdon & Thrall 2009); it is increasingly appreciated that variation in interactions, for example the strength of selection, across the landscape can drive different ecological and evolutionary patterns and processes (Thompson 2005; Tack *et al.* 2012). The interactions between the numerous phenomena that can influence

local adaptation, such as the population genetics, life history, traits and spatial structure of a pathogen, have not been well studied in many plant pathosystems (Barrett *et al.* 2008; Tack *et al.* 2012), and the links between these variables are an area where considerable advances in knowledge could be made (Burdon & Thrall 2009). Whilst a number of broad inferences concerning the influences of various pathogen life histories and traits on pathogen genetic structure have been drawn (Barrett *et al.* 2008), the links between pathogen genetic structure and local adaptation are not necessarily straightforward (Gandon & Nuismer 2009). Huyse *et al.* (2005) suggested that very small parasite effective population sizes (N_e) would reduce the likelihood of local adaptation. Gandon & Nuismer (2009) confirmed this, but also found that local adaptation was actually highest for parasite populations that were still quite small (10–100 individuals), as long as parasite migration exceeded that of the host and all interactions were occurring within one habitat type. *Striga hermonthica* populations are likely to be much larger than 10–100 individuals, due to their extremely high fecundity and the size and longevity of their seed bank (Van Delft *et al.* 1997; Yoder & Musselman 2006), and the fact that in agroecosystems host plants are unlikely to be rare.

Models of local adaptation are normally based on variations on the theme of major gene (i.e. qualitative) resistance (for example, gene-for-gene or matching allele models of infection; Clay & Kover 1996), and there has been little work on quantitative resistance in this context (Lannou 2012; but see Zhan *et al.* 2002). Intuitively, if the degree of virulence depends on a variable number of locus-specific interactions, we might expect considerable variation in the virulence of pathogen populations across the landscape, which does in fact seem to be the case for many plant pathosystems (Laine *et al.* 2011;

Tack *et al.* 2012). A recent individual-based model has indicated that quantitative resistance, based on a set of correlated host traits, may be a strong constraint on parasite virulence (Gilman *et al.* 2012). Earlier modelling efforts suggested that quantitative resistance may be more likely to select for higher parasite virulence than qualitative resistance (Gandon & Michalakis 2000), although the speed at which this evolution occurs is likely to be dependent on the genetic architecture of the virulence and resistance traits involved (Gilman *et al.* 2012). Indeed, it has usually been thought that the evolution of virulence against quantitative resistance is much slower than against qualitative resistance (van der Plank 1968; Frankel & Soulé 1981), and empirical evidence exists to support this (Zhan *et al.* 2002).

Qualitative and quantitative resistance are likely to be two ends of a spectrum, and interactions based on the presence of both a number of R-genes and a degree of quantitative resistance are plausible (Thompson & Burdon 1992; Burdon *et al.* 1996); indeed, the prevalence of gene-for-gene relationships in the plant pathogen literature may merely be due to subtle biases in the types of pathosystems that have been investigated (Thompson & Burdon 1992). Even in pathosystems where reproducible, qualitative virulence among pathogen races had been previously demonstrated, for example the rice brown planthopper *Nilaparvata lugens* (Homoptera: Delphacidae), subsequent studies found the basis of virulence to be polygenic (Thompson & Burdon 1992). For *S. hermonthica*, the presence of some R-genes in the genome is almost certain: they have been found for *S. gesnerioides* (Li & Timko 2009), and they are known from across the plant kingdom, suggesting an early evolutionary origin and on-going utility against a wide variety of pathogens (Meyers *et al.* 2005). Despite these uncertainties concerning the bases of virulence/resistance interactions in any particular

pathosystem, a number of general factors that are likely to be associated with the genetic structure and effective population size of parasite populations have been identified (Table 1.1; Huyse *et al.* 2005; Barrett *et al.* 2008).

Table 1.1. Host and pathogen life history and demographic variables likely to be associated with the effective population size of parasites. Adapted from Barrett *et al.* (2008).

Life history/demographic variable	**Increase effective population size**	**Decrease effective population size**
Host exploitation	Generalist species; multiple hosts	Specialised; single host
Pathogen reproduction	Sexual	Clonal or inbreeding
Pathogen dispersal	Long-distance	Local
Environmental variability	Stable environment/host population	Frequent extinction/recolonisation
Host longevity	Perennial/long-lived host	Annual/ephemeral host
Host population size/structure	Large, well-connected	Small, fragmented
Epidemiology	Endemic	Epidemic ('boom-and-bust')

Huyse *et al.* (2005) present a similar list of factors likely to influence the genetic structure of parasite populations. Whilst effective population size and genetic structure are not necessarily related, under certain conditions, for example low migration and small population sizes, low effective population size is likely to lead to increased genetic structure through genetic drift; however, for large parasite populations, with stable, long-lived hosts, but low parasite dispersal between populations, it is possible that a degree of significant among-population genetic structure could be present even though individual effective population sizes are large. This is perhaps the most likely situation for *S. hermonthica*, at least in the absence of large-scale genotype-flow from human activities (Berner *et al.* 1994). If hosts vary across the landscape due to the use of local crop landraces, or particular farming practices and local environments, then genetic structure could be present at loci involved in host adaptation (cf. Bierne *et al.* 2013). However given the fact that *S. hermonthica* populations have been found to be

highly heterozygous (Bharathalakshmi *et al.* 1990; Olivier *et al.* 1998; Gethi *et al.* 2005; Yoshida *et al.* 2010; Estep *et al.* 2011; Welsh & Mohamed 2011; Huang *et al.* 2012), it seems likely that populations will retain considerable genetic variation; therefore, even if some locally adapted interactions slowly evolve against quantitative host resistance, it is likely that populations will still contain genetic variation for infecting unencountered host genotypes (i.e. populations will be pre-adapted to some degree). Indeed, as noted, the finding of population-level G × G interactions between host genotypes is not uncommon in lab experiments and field trials for *S. hermonthica* (e.g. Kim *et al.* 1994; Cissoko *et al.* 2011), therefore the molecular basis of virulence between *S. hermonthica* and its hosts seems likely to be polygenic and highly adaptable (Olivier *et al.* 1998).

1.7 Thesis aims

To summarise, the current evidence suggests that populations of *S. hermonthica* are genetically diverse, not strongly differentiated by host (species or genotypes), and capable of rapid evolution in response to new hosts. Characterising populations with differential virulence reactions against host species or genotypes in different regions should still be useful; this information could be used by farmers and seed-merchants to ensure a measure of local resistance or tolerance against *S. hermonthica*. This approach would need to be supported by pest management strategies designed to reduce the chances of increased virulence evolving. Additionally, the potential effects of the environment on virulence can be large, and should be taken into account (Kaltz & Shykoff 1998; Wolinska & King 2009). Characterising the genetic structure of meta-populations at different scales may also be of use for predicting the evolution of

virulence (e.g. Estep *et al.* 2011), although it should be remembered that broad scale genetic structure may have no bearing on patterns of differential virulence if this is dependent on a small number of loci and/or their interactions (Huyse *et al.* 2005). A useful step for African agriculture in the long-term would be to characterise those loci responsible for virulence within *Striga* species. Knowledge of the causal mechanisms of virulence, and of their variability, could offer new approaches for parasite control, and underscore older ones (Hearne 2009). Understanding the nature of the genes responsible for successful parasitism could also assist with the on-going search for resistance genes in host and non-host germplasm.

Therefore, the broad aims of this thesis are:

1. To investigate the magnitude of population-level G × G interactions for populations of *S. hermonthica* and genotypes of its sorghum host, and to assess the impact of the environment on them (Chapter 2).

2. To extend the within-population approach to detecting candidate virulence loci underlying population-level G × G interactions to field populations of *S. hermonthica* and its sorghum genotype hosts (Chapter 3).

3. To develop and test new approaches for investigating the genetic basis of interactions between individual *S. hermonthica* genotypes and different host genotypes, and to compare this approach to the outlier analysis approach used in Chapter 3 and in Huang *et al.* (2012) (Chapters 4 and 5).

Chapter 2. Population-level G × G interactions in a host-parasite interaction depend on parasite life-stage but vary in their ability to predict outcomes in the field

2.1 Introduction

In some host-parasite systems the level of resistance of host genotypes, or the virulence of parasite genotypes, may be the main determinant of the interactions between species (e.g. Grech *et al.* 2006), but it is also not unusual to discover interactions between host and parasite genotypes (Wolinska & King 2009; Tack *et al.* 2012). Here, the word 'interaction' designates the situation where parasite fitness depends on the host genotype under consideration (or, conversely, relative host resistance depends on the parasite genotype). This is usually demonstrated as a significant statistical interaction, and should be evaluated in an experimental set-up that enables the 'genotype-by-genotype' (G × G) interaction component to be separated from confounding environmental effects; for example in a common garden experiment, or a lab assay (Tack *et al.* 2012). Although a strong G × G interaction may change the relative virulence rankings of parasite genotypes between hosts, this is not necessary for an interaction to be significant (in the statistical sense, at least), and it may simply be the case that the relative differences in virulence between parasites are increased or decreased, without a change in parasite rank virulence (Wolinska & King 2009). G × G interactions can, in theory, be investigated at either the individual or population level, although differences in interpretation may apply. At the individual level, different clones of two species could be used to investigate the genotype-dependence of an interaction such as parasitism (Whitham *et al.* 2003). Alternatively, for obligately outbreeding species such as *S.*

hermonthica, different populations can be used to investigate how a change of deme affects the interactions with a set of host genotypes.

G × G interactions are often of interest to researchers, although the reason for the interest varies between areas of research. For example, workers in the field of community genetics have stressed the importance of G × G interactions in determining ecological outcomes at the community or ecosystem level (Neuhauser *et al.* 2003; Whitham *et al.* 2006). The aim in this area is to link evolutionary and ecological phenomena more tightly, by discovering the precise genetic loci and phenomena that influence larger scale ecological patterns and processes (Hersch-Green *et al.* 2011). Alternatively, taxonomic studies of parasites have attempted to better define host races by demonstrating the existence of G × G interactions: a change in parasite virulence rank between hosts may suggest the presence of genetic variation worth formally recognising with a taxonomic rank, especially if the parasite populations demonstrating host preferences have morphological correlates and can be distinguished in the field (Thorogood *et al.* 2009). Cryptic (morphologically indistinguishable) host races may also be of interest to researchers if the parasite is of economic importance (Jerome & Ford 2002), or if researchers are interested in more fundamental biological questions, such as speciation (Huyse *et al.* 2005). G × G interactions can also be of practical importance in crop breeding programs seeking to utilise resistant crop germplasm: if a parasite is geographically widespread and genetically variable (like *Striga hermonthica*; Mohamed *et al.* 2007) then assessing the stability of crop host resistance against many different parasite populations is likely to help ensure that host resistance is not going to be overcome in the areas in which a crop variety is going to be deployed (Haussmann, Hess, Welz, *et al.* 2000; Cissoko *et al.* 2011). This may highlight regions that require

monitoring for resistance breakdown, potentially increasing the efficiency of other parasite control methods that could then be targeted to those areas. Finally, the local adaptation of parasites to their hosts, a much discussed topic in the area of eco-evolutionary dynamics (Gandon 2002; Tack *et al.* 2012), is essentially a form of G × G interaction, where the ability of a parasite genotype to infect a host genotype depends on their degree of prior exposure to each other.

Within these several different contexts, host-parasite G × G interactions involving parasitic plants have been investigated using pot, root observation chamber ('rhizotron') and common garden (or field trial) experiments. An obvious shortcoming of such experiments is the absence of information on the effects of environmental variation on the outcomes observed (Wolinska & King 2009). The only situation in which the effect of the environment on the host-parasite interaction might not be immediately relevant is where basic cellular or molecular knowledge of the interaction is desired; for example, studies of quantitative trait loci (QTL) may simply aim to identify loci involved in parasite virulence or host resistance (Gurney *et al.* 2006; Swarbrick *et al.* 2009); however, eventually, a fuller understanding of the interactions of QTL with the environment may also be desired (Anderson *et al.* 2011, 2013). In all of the other situations described above, where G × G interactions are of interest, the desire to extrapolate lab or common garden findings to the wider environment makes any environmental effects on the observed G × G interactions an important part of the system under study (i.e. G × G × E interactions should be considered; Wolinska & King 2009; Tack *et al.* 2012). In the area of host-parasite local adaptation, a field-based experimental design using reciprocal common garden experiments has been shown to separate out the different components of variance in species' interactions across

genotypes and environments (Nuismer & Gandon 2008).

For parasitic plants, there are very few examples of pathosystems where an effort to quantify the impact of environmental variability on G × G interactions has been made (but see Omanya *et al.* 2000). This is perhaps not surprising, given that Nuismer & Gandon (2008) reported only one example of such an experimental strategy being used for any type of interaction between two species. Moreover, for parasitic weeds, such as *Striga*, it is often impossible to translocate parasites between field sites due to quarantine measures (Prof. J.D. Scholes, pers. comm.); removing existing 'home site' infestations may also be costly (Haussmann, Hess, Welz, *et al.* 2000). Therefore, the comparison of lab experiments with experimental field trials, which are restricted to the home sites from which the parasites originate, is the only option available in the study of the agriculturally important *S. hermonthica*/sorghum host-parasite system. Whilst this approach provides an insight into how G × G interactions may change between a controlled environment and the field, it unfortunately precludes experimentation across a range of environments, which would provide further information on the size of any G × G × E component (Gandon & Nuismer, 2008; Wolinska & King, 2009). However, conducting field trials across multiple years may be able to provide information on the effects of a varying environment on G × G interactions.

There is considerable evidence that environmental variation will play an important part in host-parasitic plant interactions. For the *S. hermonthica*/sorghum interaction, both nitrogen and phosphorous deficiency have been found to promote strigolactone production in sorghum (Yoneyama *et al.* 2007), potentially leading to greater parasite infection. Field-based research programs have also found soil nitrogen status to have a complex relationship with host resistance to *Striga* (Press *et al.* 1999; Schulz *et al.*

2003). In general, the importance of environmental variability for *Striga* virulence has long been appreciated (Haussmann, Hess, Welz, *et al.* 2000).

The 'multi-step' nature of parasite infection is another topic of importance when considering G × G interactions. *Striga* requires a chemical signal from a host plant in order to germinate, a signal for haustorium differentiation, and compatible molecular interactions in order to successfully penetrate host tissue (Yoder & Scholes 2010). This suggests at least three plant-plant interactions, each, potentially, with its own genetic basis. Recent theoretical work has suggested that the existence of two-step infection processes could have strong effects on host-parasite coevolution and related patterns of gene frequencies and selection (Fenton *et al.* 2012). However, most work on parasitic plant G × G interactions has either not separated out the 'foraging' stage from host attachment and penetration (Mutikainen *et al.* 2000; Koskela *et al.* 2000; Rowntree *et al.* 2011), or, has focused solely on the post-attachment part of the host-parasite interaction by using pre-germinated parasite seed (Cissoko *et al.* 2011; Huang *et al.* 2012).

One exception to this general pattern is the work of Thorogood *et al.* (2009). These authors examined both germination and attachment in the interaction between two intra-specific taxa within *Orobanche minor s.l.* and the host plants *Trifolium repens* and *Daucus carota* ssp. *gummifer*. Thorogood *et al.* (2009) showed the presence of host-parasite interactions at both the pre- and post-attachment stages; that is, the relative virulence of the two parasite taxa depended on the host and on the stage of interaction. However, the two taxa investigated, *O. minor* ssp. *minor* (synonym var. *minor*; Stace 2010) and *O. minor* ssp. *maritima*, have long been separated taxonomically on the basis of morphological difference and host preference (e.g. Druce 1930), and have been found

to be genetically differentiated (Thorogood *et al.* 2008); therefore this particular instance of a host-parasite G × G interaction may actually represent a pair of taxa that are well on the way to full host-specialisation and reproductive isolation.

Another recent study has investigated both pre- and post-attachment resistance for the interaction between rice and *S. hermonthica* (Jamil *et al.* 2011); however, in this instance, different parasite populations were used in the different parts of the study. Unfortunately, this strategy forfeits the opportunity to comment on the strength of the genetic basis of any observed similarities or differences in the virulence rankings of the two stages of this host-parasite interaction. Even though the rankings of Jamil *et al.* (2011) were broadly similar between the two stages of infection, we cannot be sure that this means that the two stages of virulence are actually similar for a single population of the parasite: the use of different populations of the parasite gives us no information on the strength of any parasite genotype life-stage × host interaction; that is, parasite life-stage and parasite population were confounded in the experimental design.

The work of Omanya, Haussmann and colleagues (Omanya *et al.* 2000; Haussmann *et al.* 2001, 2004) may be the most comprehensive assessment of pre- and post-attachment resistance, and of the effects of the environment on these phenomena, for any parasitic plant-host interaction to date. They studied the interaction between *S. hermonthica* and two recombinant inbred populations of sorghum in the lab and the field. Their controlled experiments consisted of both germination assays and pot studies, and correlations between these results and metrics of *Striga* field emergence were calculated. Their main conclusion was that pot studies were poor predictors of the field performance of sorghum lines, but that germination assays showed more promise, although even the predictive ability of this assay type varied considerably between sites

(Omanya *et al.* 2000).

The need to explicitly consider separate G × G interactions at different stages of parasite infection may depend on the focus of a study. For example, for the early stages of a community genetic research program, such as that of Rowntree *et al.* (2011), the fact that an overall G × G interaction was found between hemiparasitic *Rhinanthus* species and populations and a set of barley (*Hordeum vulgare*) cultivars is simply presented as evidence that the genetics of interacting species are likely to have indirect effects on ecosystem function, because *Rhinanthus* species have been considered to be 'ecosystem engineers' (Rowntree *et al.* 2011). However, if this research progresses to the stage where coevolutionary patterns of gene frequencies are investigated in natural populations, or genomic signatures of selection are sought (e.g. Mhedbi-Hajri *et al.* 2011), then a better knowledge of the stages at which genetic interactions are occurring would be desirable. The distinction between pre- and post-attachment G × G interactions might similarly be of importance for work on host-parasite local adaptation (Lively 1999; Agrawal & Lively 2003; Fenton *et al.* 2012).

For research into *S. hermonthica*, or other agriculturally important parasitic weeds, information on G × G interactions at different infection stages is likely to be of considerable importance (Haussmann, Hess, Welz, *et al.* 2000). If G × G interactions are more common at the post-attachment stage of infection (Cissoko *et al.* 2011; Huang *et al.* 2012) than at the pre-attachment stage, then one interpretation might be that post-attachment virulence evolves more quickly between populations than pre-attachment virulence (or 'foraging' efficiency). This could indicate that, between pre- and post-attachment resistance, pre-attachment is likely to be the more robust strategy for breeding crops with durable *Striga* resistance (Kim *et al.* 1994; Omanya *et al.* 2004).

Here, I investigate the existence of population-level G × G interactions between three populations of *S. hermonthica* and five sorghum cultivars. Specifically, I test the following hypotheses: (1) Is there evidence for G × G interactions at the pre-attachment (germination stimulation) stage? (2) Is there evidence for G × G interactions at the post-attachment (host root penetration) stage? (3) Does the choice of virulence metric (where different measures may represent different aspects of parasite fitness) affect conclusions regarding post-attachment G × G interactions? And (4), do any of the lab-based estimates of virulence for each *Striga* population/sorghum host combination consistently predict the outcomes of a set of multi-year West African field trials of the same *Striga* and sorghum plant material? The investigation of (4) provides an insight into the presence of G × G × E interactions in this host-parasite system.

2.2 Materials and Methods

2.2.1 Plant materials

The *S. hermonthica* seed populations used in this chapter were collected from three agricultural research stations in West Africa; these sites are part of the International Sorghum and Millet Collaborative Research Support Program (INTSORMIL CRSP; http://intsormil.org). As a part of the INTSORMIL program, multi-year field trials have been in progress at these research stations, the primary aim of which has been to characterise *Striga*-resistant and *Striga*-tolerant varieties of sorghum (Dr T. van Mourik, pers. comm.) One of these stations was at Kouare, Burkina Faso (11°95'N:00°30'E), the other two were in Mali, at Sotuba (12°66'N:07°91'W) and Samanko (12°52'N:08°07'W), both on the outskirts of the capital Bamako. The geographical distance between the two Mali sites is approximately 21 km, whereas the Burkina Faso site is approximately 906

km from the two Mali sites (measured from Bamako, Mali to Kouare, Burkina Faso).

The INTSORMIL field trials at these sites evaluated fifteen sorghum cultivars; five of these cultivars were chosen as the subject of the research presented in this chapter. These five were selected because they represented a variety of host responses to *S. hermonthica* across the West African sites. As judged by *Striga* emergence across the three field trial sites, they were either relatively resistant, relatively susceptible, or showed differing responses across sites (Dr T. van Mourik, pers. Comm.; Chapter 3, Table 3.2). The five were: Brhan; CSM 388 (hereafter called CSM); Malisor 92-1 (hereafter called Malisor); Mota Galmi (hereafter called MG); and SRN 39 (hereafter called SRN). Seeds of all sorghum cultivars were obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Mali. Due to low germination of the seed of some of the sorghum cultivars on receipt, seed stocks were bulked-up by growing plants in pots in a controlled environment growth room before use in lab experiments; inflorescences were bagged before anther dehiscence to prevent crossing taking place between cultivars.

2.2.2 Host root exudate germination stimulant assays

Five replicates were used for each sorghum cultivar. Sorghum seeds were germinated in rock-wool blocks in a controlled environment growth room under the following conditions: photon flux density: 500 μmol m^{-2} s^{-1} at plant height; 12 h photoperiod; relative humidity 60%. The seedlings were transferred after five days to a hydroponic system in the same growth room. This system consisted of individual sorghum plants growing in 50 mL plastic Falcon tubes (Sarstedt, UK), with the tubes wrapped in aluminium foil. The plants sat in sponge bungs; the sponge around the plant was covered with Parafilm (Pechiney, USA) to reduce evaporation. The tubes were filled

with 40% Long Ashton nutrient solution (Hewitt 1966) with 2 mM ammonium nitrate, topped-up as required. After 10 days the nutrient solution was replaced with distilled water. Twelve hours were allowed for the collection of root exudates, then the water from each replicate was collected and frozen at -20°C until use. The sorghum root systems were collected, dried at 45°C and weighed. *S. hermonthica* seeds were conditioned at 27°C for 12 days on moistened glass-fibre filter paper (GF/A, Whatman) in 9 cm Petri dishes sealed with Parafilm. Germination assays were performed in 48-well plates (Nunc, Themo Scientific, USA) by cutting 10 mm discs of filter paper containing between 50 and 200 conditioned seeds with a metal borer, and suspending the seeds in 2 mL of root exudate in a well of the 48-well plate. Two technical replicates were performed for each biological replicate.

In order to account for differences in the production of root exudate due to differences in root biomass, the root exudate solutions were normalised to the lowest root dry weight. However, this resulted in the exudates being too dilute to compare (no *Striga* seeds germinated). Attempts were also made to quantify the phenol content of the exudates using the Folin Ciocalteu method (Singleton *et al.* 1999) as a proxy for root exudation: this revealed no differences between replicates, either within or between sorghum cultivars. Therefore, root exudate dilutions (1×; 0.5×; 0.2×; 0.1×; 0.01×) were made from the neat root exudates with sterile, distilled water; the dilutions were intended to increase the range over which differences between cultivars might be observed. Rather than normalising samples to remove the influence of root biomass, the effect of root dry weight was accounted for by its inclusion in all statistical models.

The 48-well plates were incubated at 27°C for 24 h, whereupon germination (number of germinated seeds/total number of seeds) was assessed for each sample using a

stereomicroscope. Negative (sterile de-ionised water) and positive (the germination stimulant GR-24, 0.1 mg L^{-1}) controls were used. The germination counts for each technical replicate were added together to give the final data for each biological replicate, as is appropriate for proportion data (Crawley 2005).

2.2.3 Post-attachment resistance assays

Sorghum seeds were germinated between strips of kitchen-towel, supporting by blocks of wet horticultural rockwool (Aquaculture, Sheffield, UK). After five days the sorghum plants were transferred to individual root observation chambers (rhizotrons). A rhizotron consisted of a square, perspex tissue culture plate of 300 mm side and 30 mm depth. Rhizotrons were packed with moist vermiculite, except for a rockwool block at the base to aid drainage. A mesh (100 µm polyester multi; Plastic Group, Birkenhead, UK) was placed over the vermiculite, onto which individual sorghum plants were placed; an opening at the top of the rhizotron allowed for shoot growth. Individual sorghum root systems were infected with 12.5 mg of conditioned, pre-germinated *Striga* seed 12 days after sowing. *S. hermonthica* seed was conditioned as described in section 2.2.2. *S. hermonthica* seeds were artificially germinated (using a 0.1 mg L^{-1} solution of the artificial germination stimulant GR-24) in order to promote synchronous attachment of *Striga* to the sorghum roots (Gurney *et al.* 2006). The *S. hermonthica* seeds were carefully aligned against the sorghum roots using a fine paintbrush.

After the rhizotron lids were closed, the rhizotrons were wrapped in aluminium foil to prevent light from reaching the sorghum root systems. A diagram of a rhizotron and photos of the experimental set-up are given in Figures 2.1 and 2.2 respectively. Each rhizotron was watered automatically four times a day with approximately 100 mL of 40% Long Ashton solution, with 2 mM ammonium nitrate (Hewitt 1966). Rhizotrons

were placed in a controlled environment growth room under the conditions described above (section 2.2.2).

Six replicate rhizotrons for each sorghum cultivar/*S. hermonthica* population combination were used. The surviving *Striga* parasites were harvested from their host root systems at 30 days post-inoculation under a stereomicroscope. The harvested parasites were placed on filter paper in 9 cm Petri dishes and digitally photographed. Counts and length measurements of parasites were made from the digital images using ImageJ v. 1.45 (http://rsb.info.nih.gov/ij/). Total parasite dry weight was calculated by drying the Petri dishes containing the parasites at 45°C for two days, and then taking the difference between the weight of the filter paper plus the parasite, and the filter paper minus the parasite.

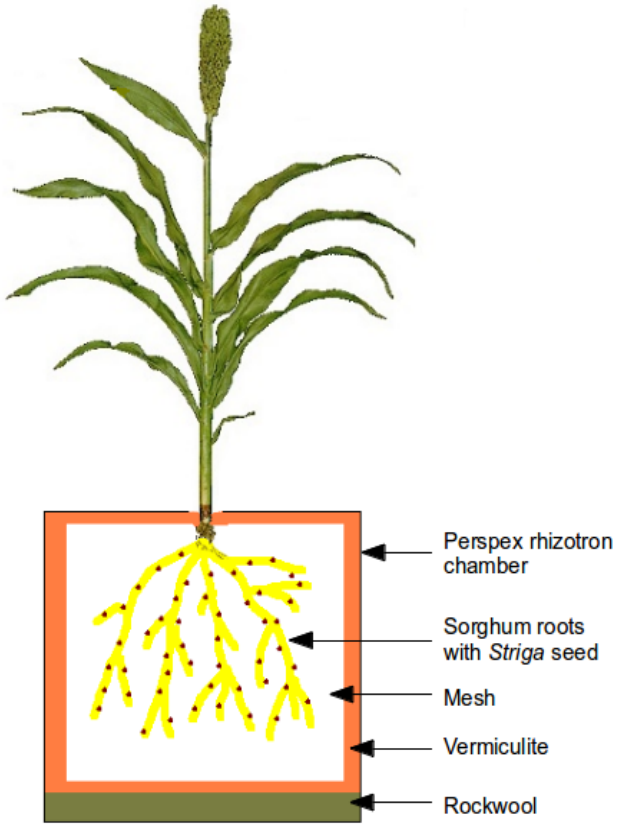

Figure 2.1. The rhizotron setup used in the post-attachment sorghum resistance assays. Once the chamber lid is affixed, the rhizotron is wrapped in aluminium foil to keep the root system in the dark.

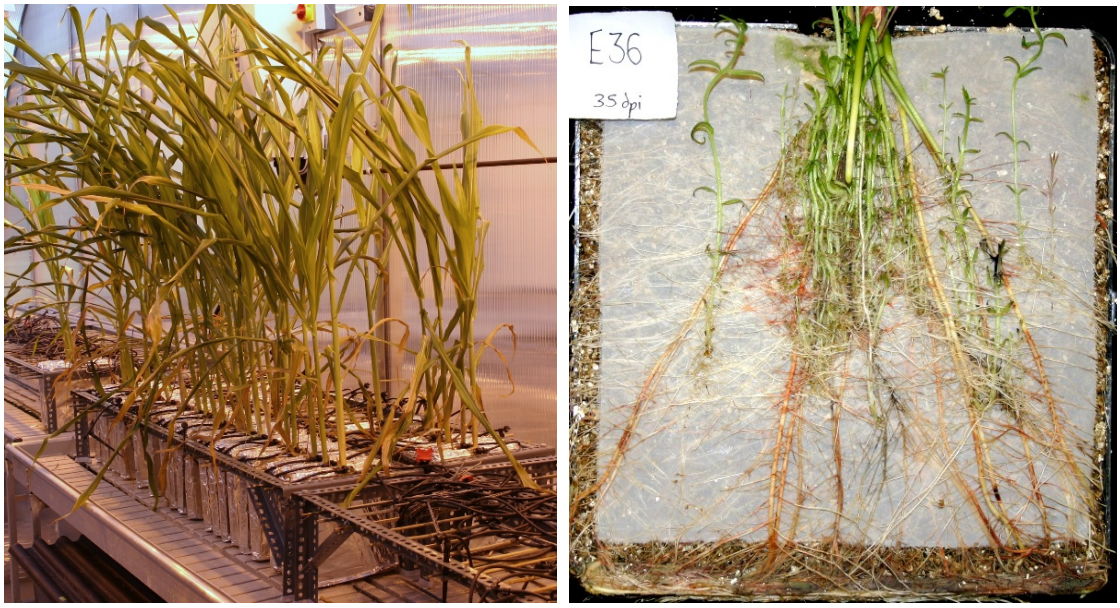

Figure 2.2. A typical rhizotron set-up in a growth room (left), and, an individual rhizotron (the susceptible sorghum cultivar E36 infected with *S. hermonthica*) opened up to reveal the infected root system at 35 days post-infection (right).

2.2.4 Field trials of sorghum cultivar resistance to S. hermonthica

The field trials were carried out by staff of the West African research stations listed above (section 2.2.1). All of the information that immediately follows concerning the field trials was communicated to me by Drs Tom van Mourik (ICRISAT, Mali) and Hamidou Traore (INERA, Burkina Faso).

The field trials were conducted between 2008 and 2011 using randomized complete block designs with either three or four replicate plots, depending on year and site. Each replicate at a site consisted of a row of sub-plots, one sub-plot for each of the 15 sorghum cultivars evaluated; a sub-plot either measured 4.0 m by 1.5 m (Samanko) or 3.0 m by 1.6 m (Sotuba & Kouare). An example of the experimental field layout is given in Figure 2.3. The sorghum resistance trait measured was the average number of *S. hermonthica* parasites emerged above ground at 90 d after host planting, per host plant, per sub-plot (hereafter, 'emerged *Striga* at 90 d host plant^{-1} sub-plot^{-1}'). Due to different initial planting schemes, and to the initial differential survival of sorghum plants post-planting, differing final numbers of sorghum plants were assessed for *Striga* emergence per sub-plot, both between and within sites and years (hence the normalisation to the number of host plants in the resistance metric used). The overall mean number of sorghum plants (± s.d.) assessed per sub-plot was 12.26 (± 6.62; range = 5-34). Further information about the field trials is summarised in Table 2.1. The parasite seeds used for artifical infestation at a field site were from the same batch as the seeds received from a site at the University of Sheffield for use in the pre-attachment and post-attachment resistance assays described above.

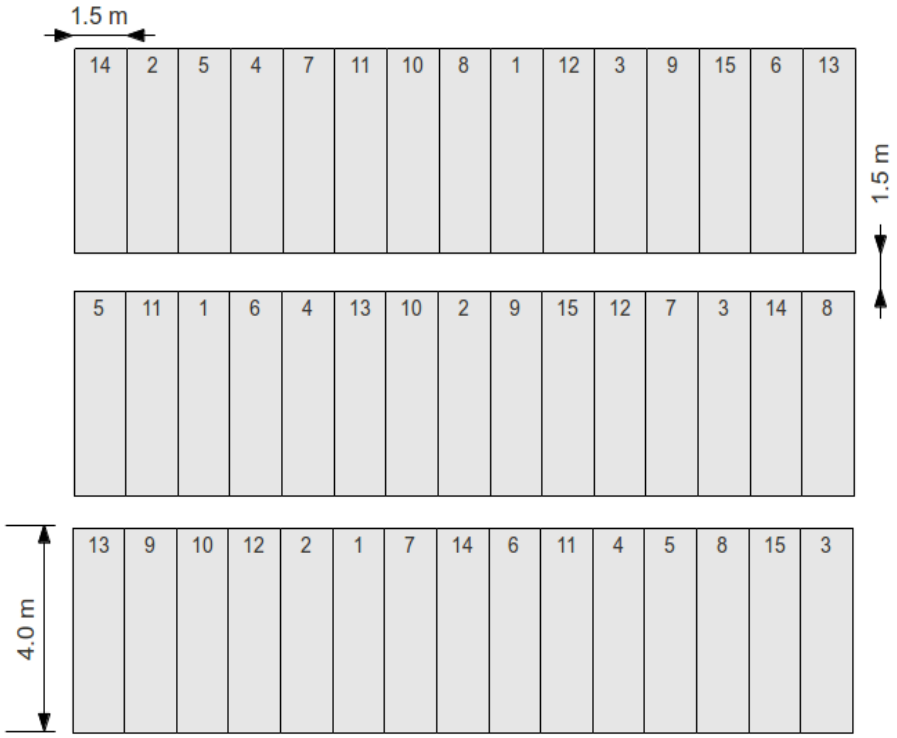

Figure 2.3. An example of the experimental design used in the sorghum cultivar resistance field trials. This example consists of three replicate plots, each consisting of 15 sub-plots. The sub-plot measurements shown here are those used at Samanko; sub-plots measured 3.0 m by 1.6 m at Sotuba and Kouare. Each of the 15 sorghum cultivars (here designated by a number between 1 and 15) was randomly assigned to a sub-plot within a replicate plot. A sub-plot contained between 4 and 34 sorghum plants; see Table 2.1 for further details on the specific layouts, and for further information on the numbers of sorghum plants assessed in a sub-plot at each site.

Table 2.1. Infomation relating to the West African sorghum *S. hermonthica* resistance field trials conducted in Burkina Faso and Mali between 2008 and 2011.

Site	Geographic coordinates	Years trial conducted (and number of replicates)	Final number of sorghum plants per sub-plot assessed: mean ± s.d. (range)	Sub-plot size (m)	Artifical *Striga* infestation method	Amount of *Striga* used in artificial infestation	NPK fertiliser used?[1]
Kouare, Burkina Faso	11°95'N:00°30'E	2008 (3) 2009 (-)[2] 2010 (3)	14.50 ± 5.89 (5-28)	3.0 by 1.6	Infested planting holes	3000 germinable seeds per planting hole	Y
Samanko, Mali	12°52'N:08°07'W	2008 (3) 2009 (4) 2010 (4)	8.51 ± 0.81 (5-9)	4.0 by 1.5	Infested planting ridges	1 g of 70% germinable *S. hermonthica* seed per ridge[3]	N
Sotuba, Mali	12°66'N:07°91'W	2009 (3) 2010 (3) 2011 (3)	15.43 ± 8.62 (5-34)	3.0 by 1.6	Infested planting holes	3000 germinable seeds per planting hole	Y

1 Field applications followed the recommendations given in Berner *et al.* (1997).

2 Whilst average *Striga* emergence data were received from Kouare for 2009, the raw data could not be retrieved, hence their exclusion from the analysis presented in this chapter.

3 There was one (soil) ridge per sub-plot. If one *Striga* seed weighs approximately 5×10^{-6} g (Berner *et al.* 1997), this is around 140,000 seeds (assuming 70% germination). For a 4.0 m length planting ridge, this is 3500 seeds for every 10 cm, i.e. comparable to the number of seeds used per planting hole at the other sites.

2.2.5 Statistical analyses

2.2.5.1 Host root exudate germination stimulant assays

To determine the effects of the sorghum cultivar root exudates on the germination of the three *S. hermonthica* seed populations, the proportion of seeds germinated was analysed in a binomial generalised linear model (GLM). The response was a two column vector of 'successes' and 'failures' (Crawley 2007); that is, counts of the numbers of seeds germinated and ungerminated for each combination of host, *Striga* population, and root exudate dilution factor. Using a two column vector utilises information on sample size to weight the model predictions, information which is lost when using a simple proportion as a response variable (Crawley 2007). The germination dataset was found to be overdispersed; that is, there was more variability than could be accounted for by the binomial distribution (Warton & Hui 2011). Overdispersion was assessed by comparing the ratio of the residual deviance (deviance is a likelihood-based measure of goodness of fit, which can be compared to sums of squares) to its degrees of freedom. One recommended solution when dealing with overdispersed proportion data is to fit a quasi-binomial model; these models use an empirical scaling factor to account for overdispersion (Crawley 2007; Warton & Hui 2011).

The standard 'link' function for a binomial GLM is the log of the odds (the 'logit'; Crawley 2007); however, in the current experiment the response of seed germination to an increasing concentration of root exudate was asymptotic, and exhibited a form similar to the well-known Michaelis-Menten equation (Soetaert & Herman 2009). The Michaelis-Menten equation can be linearised by taking the inverse of x and y (the Lineweaver-Burk linearisation); therefore, the inverse link function was used, and the inverse of the dilution factor was also taken before modelling. This approach was found to minimise the residual deviances of the models compared to alternative approaches.

The inclusion of covariates and their interactions in models ('model selection') was assessed by likelihood ratio tests using the *F* distribution, as is recommended for overdispersed models (Crawley 2007). Main effects were included in models if any of their interactions were significant, even if they themselves were not significant (Venables & Ripley 2002). Model checking was carried out by inspecting plots of residuals against fitted model values for heteroscedasticity (unequal group variances) and non-normality of residuals (Warton & Hui 2010).

2.2.5.2 Post-attachment resistance assays

Striga attachment data from post-attachment resistance assays were analysed differently depending on the response metric. Counts of *Striga* attachment were analysed in a quasi-Poisson GLM due to overdispersion (Crawley 2007). Median length of the parasites and total dry weight of attachments were log-transformed to stabilise the variances, and were subsequently analysed in linear models. Likelihood ratio tests to check the significance of terms in the model and model checking (for heteroscedasticity and non-normality of residuals) were carried out in all cases. Tukey's Honestly Significant Difference (HSD) was used for *post hoc* pairwise group comparisons.

2.2.5.3 Field trials of sorghum cultivar resistance

The metric 'emerged *Striga* at 90 d host plant^{-1} sub-plot^{-1}' (section 2.2.4) was calculated in a spreadsheet program. Plots of sub-group (that is, data from a particular combination of year, host and *Striga* field site) means and variances showed that heteroscedasticity was present in this dataset, therefore the square root of the *Striga* emergence metric was taken to stabilise the variances. The transformed *Striga* emergence metric was analysed in a linear model. The alternative was to analyse counts of emerged *Striga* at 90 d sub-plot^{-1}, and to use the number of sorghum host plants per sub-plot as an 'offset term' in a

Poisson or negative binomial model (Gelman & Hill 2006), so that the count data can be modelled as rates. However, lower Akaike's Information Criterion (AIC; a commonly used measure of the relative goodness of fit of competing models, Bolker 2008) values indicated that the use of a linear model, and the square root transformation, were both justified over the alternative Poisson or negative binomial approach. Model selection was performed using likelihood ratio tests (for nested models; Bolker 2008) and AIC comparisons (for nested or non-nested models; Burnham *et al.* 2011). Tukey's HSD was used for *post hoc* comparisons between sorghum cultivars within sites and years.

2.2.5.4 Using lab-assay data to predict field outcomes

In order to evaluate the ability of the data on pre- and post-attachment resistance for each *Striga* population/sorghum host combination to predict sorghum cultivar resistance in the field trials, the percentage of seeds germinated for the undiluted root exudates (1×), as predicted by the minimum adequate models (i.e. the most parsimonious model for the data; Crawley 2007) of the germination data, and the mean values of the three post-attachment estimates of *Striga* virulence from the rhizotron experiments (number of attached *Striga*; total *Striga* dry-weight; and median length of attached *Striga*), were used as covariates in the minimum adequate models of the field trial *Striga* emergence data.

Because the expression of post-attachment virulence depends on *Striga* having first germinated in response to host root exudates, estimates of combined pre- and post-attachment virulence were created. These additional virulence covariates were created by multiplying each post-attachment metric by the pre-attachment germination metric for every *Striga* population/sorghum host combination (i.e. downweighting the post-attachment metrics by the percentage of *Striga* seeds germinated). These new covariates

were tested in the same way as the separate pre- and post-attachment resistance estimates, by using them as covariates in the minimum adequate models of the field trial *Striga* emergence data.

The use of these different metrics in place of the categorical variable 'sorghum host cultivar' in the models of the field data enabled an evaluation of the ability of these lab-derived resistance metrics to explain the field resistance of the sorghum cultivars. AIC and R^2 values were used to compare the explanatory powers of the different host resistance (or, conversely, parasite virulence) pre- and post-attachment metrics.

All analyses were performed using the statistical programming language R v. 2.12.2 (R Core Team 2012).

2.3 Results

2.3.1 Host root exudate germination stimulant assays

2.3.1.1 Baseline differences in *S. hermonthica* population germination

The mean germination of the three *Striga* seed populations with the GR-24 artificial germination stimulant control were (means ± binomial standard errors): Kouare, Burkina Faso: 79.84% ± 2.69; Samanko, Mali: 74.18% ± 2.68; Sotuba, Mali: 82.58% ± 3.03 (n = 7 for all populations). There were no significant differences in baseline germination between the populations (quasi-binomial GLM: $F = 2.55$, $p = 0.106$, d.f. = 2), indicating that *Striga* germination responses to the host root exudates could be directly compared between the three sites without normalisation to the amount of germinable seed per population.

2.3.1.2 The effects of host cultivar and dilution on *Striga* germination

Initial analyses of *Striga* germination were undertaken within populations (Table 2.2). Figures 2.4a–e, 2.5a–e and 2.6a–e show the percentage germination responses to each host cultivar root exudate for the *Striga* populations at Kouare, Samanko and Sotuba respectively. The overlaid regression lines represent the predicted germination values from the minimum adequate models (Table 2.2; Crawley 2007) for each *S. hermonthica* population/host combination (Figs 2.4a–e, 2.5a–e and 2.6a–e). Figure 2.7a–c, displays the same minimum adequate model predictions for each *Striga* population's response to the five different sorghum cultivar root exudates without the raw data for comparison. The statistics for the terms in the minimum adequate models of the germination responses of the *Striga* populations are given in Table 2.2.

For each *Striga* population, significant effects of host (sorghum cultivar), root exudate dilution factor and host root weight were found (Table 2.2). Figures 2.4a–e, 2.5a–e and 2.6a–e clearly indicate that, within *Striga* populations, clear differences were found in the percentage germination response between different cultivar root exudates, and that, around the 1 in 5 to 1 in 10 point, dilution of the host root exudates significantly reduced percentage germination. A comparison of the effects of host cultivar on *Striga* percentage germination between the three *Striga* populations suggests that, in each case, the germination responses to the different cultivars were very similar between *Striga* populations (Fig. 2.7a–c). Both the order in which the host root exudates stimulated germination, and the absolute levels of percentage germination, are comparable between populations (Fig. 2.7a–c), with the possible exception of the relatively lower percentage germination in the Sotuba *Striga* population in response to root exudates from the cultivars SRN and Brhan (Fig. 2.7c). This suggests that the *Striga* population × sorghum

host interaction detected affected the sizes of the differences in germination sensitivity to the different hosts between *Striga* populations, but did not change the order (in terms of the percentage germination) in which populations were stimulated by the sorghum cultivar root exudates.

The models also indicate that host root dry weight had a significant effect on percentage germination for all three *Striga* populations, although this direct effect of host root dry weight had the smallest effect of the significant model terms in all three *Striga* populations (Table 2.2). Significant interaction terms between host cultivar and dilution factor were found for all three *Striga* populations, indicating that particular combinations of cultivar root exudate and dilution level had specific effects on *Striga* percentage germination. This is particularly clear in the Samanko and Sotuba populations, where the models clearly show that the shape of the curve relating *Striga* percentage germination to dilution factor depends strongly on cultivar (Fig. 2.7b and c). The relationship between host cultivar and dilution factor was also influenced by variation in the root dry weight of particular replicates for the Kouare and Samanko populations, although for Sotuba the effect of root dry weight did not interact with cultivar identity, suggesting that variation in this factor was similar across cultivars in the Sotuba population experiments (Table 2.2).

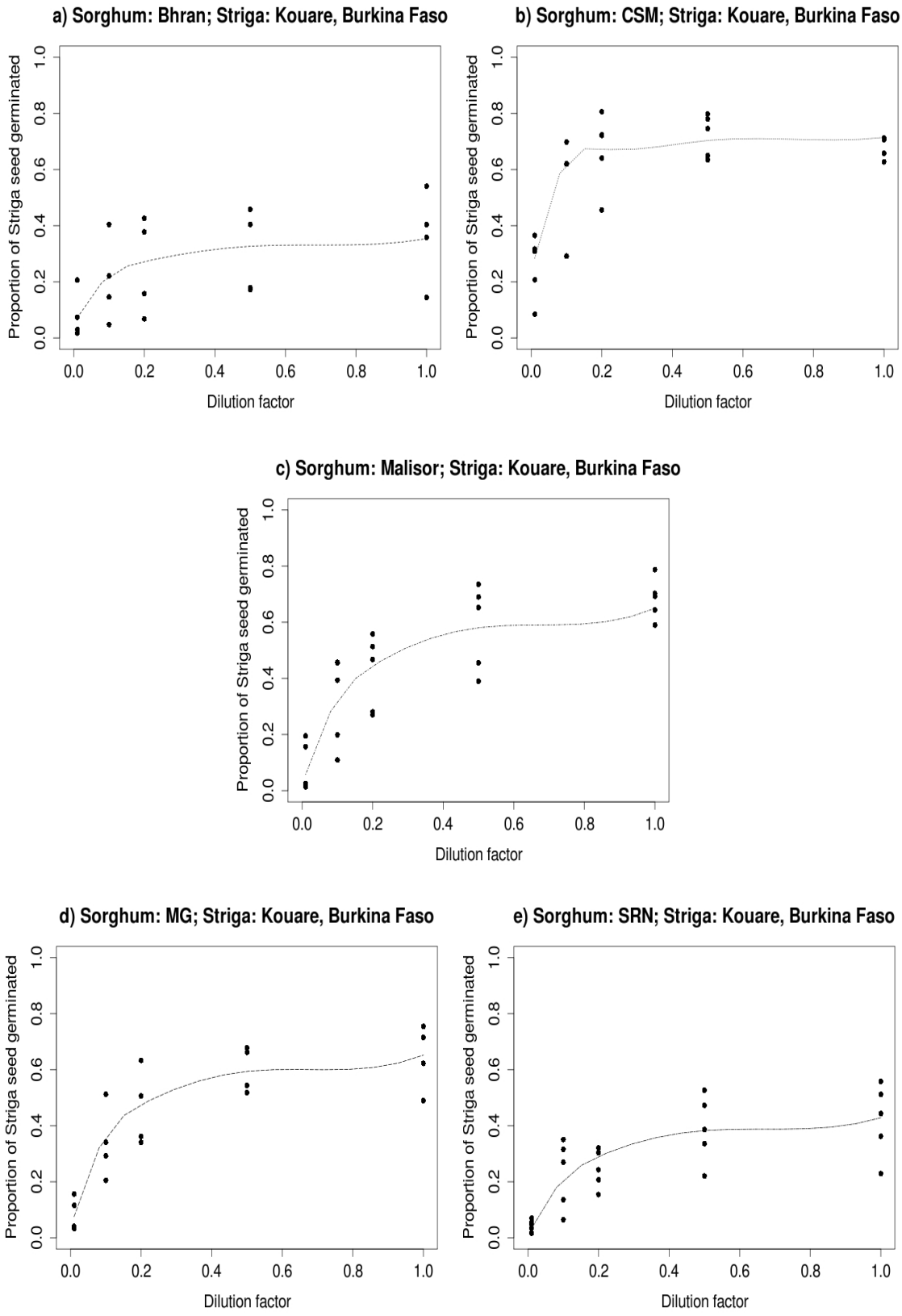

Figure 2.4a–e. The germination of *S. hermonthica* seed from Kouare, Burkina Faso, by root exudates of five sorghum cultivars. The broken lines plot the smoothed model predictions of the minimum adequate model of the response of the *Striga* seed to each cultivar's exudate.

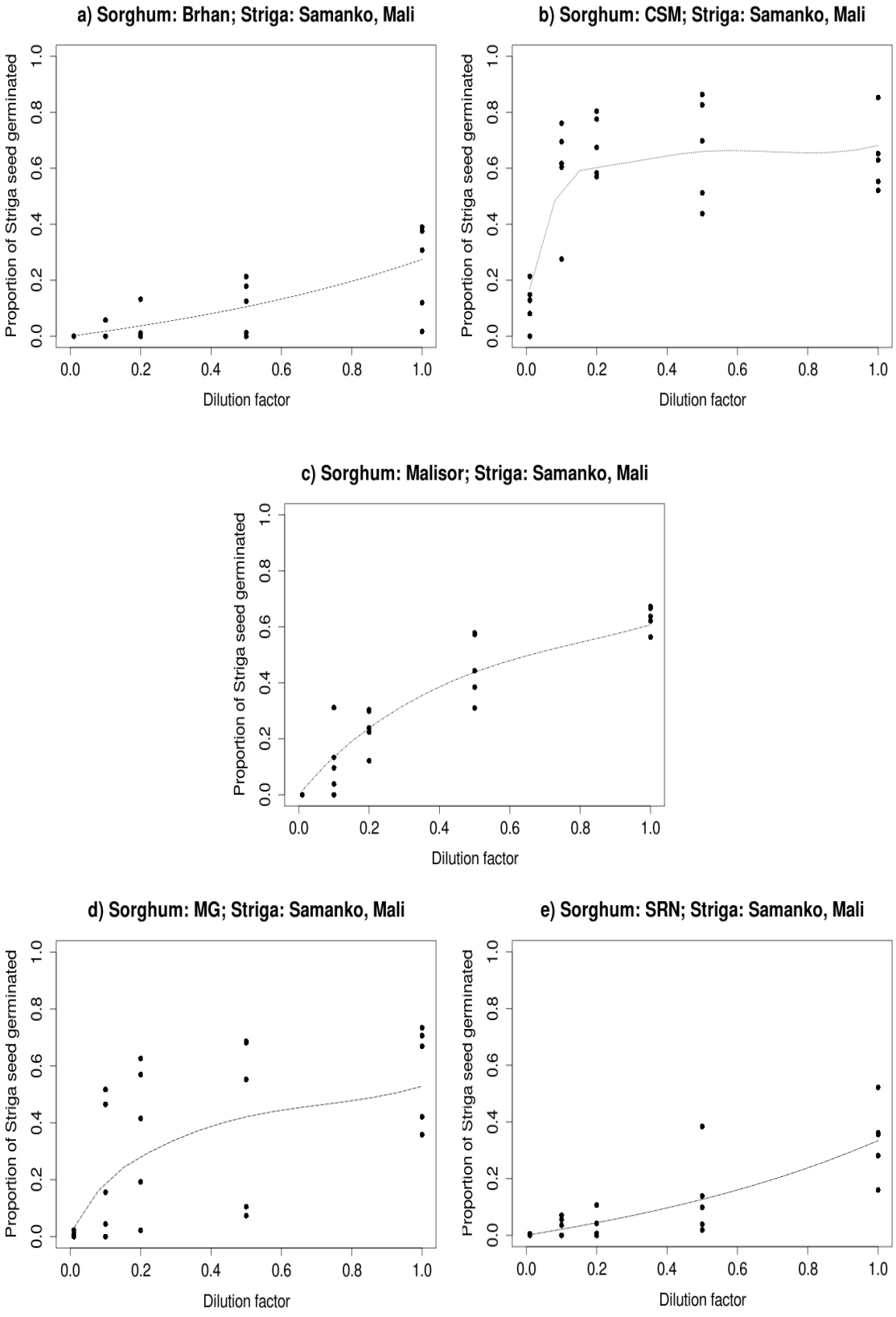

Figure 2.5a–e. The germination of *S. hermonthica* seed from Samanko, Mali, by root exudates of five sorghum cultivars. The broken lines plot the smoothed model predictions of the minimum adequate model of the response of the *Striga* seed to each cultivar's exudate.

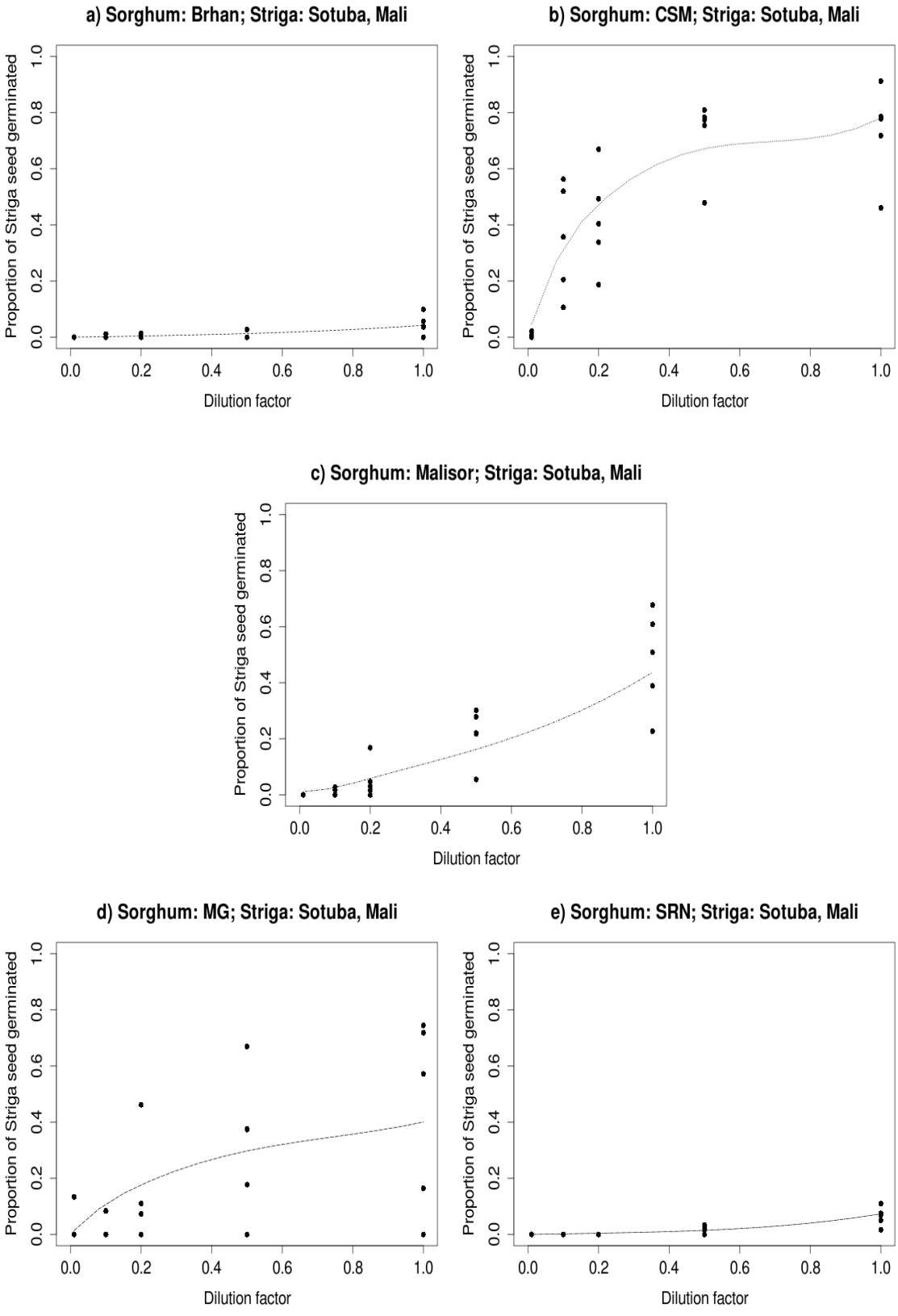

Figure 2.6a–e. The germination of *S. hermonthica* seed from Sotuba, Mali, by root exudates of five sorghum cultivars. The broken lines plot the smoothed model predictions of the minimum adequate model of the response of the *Striga* seed to each cultivar's exudate.

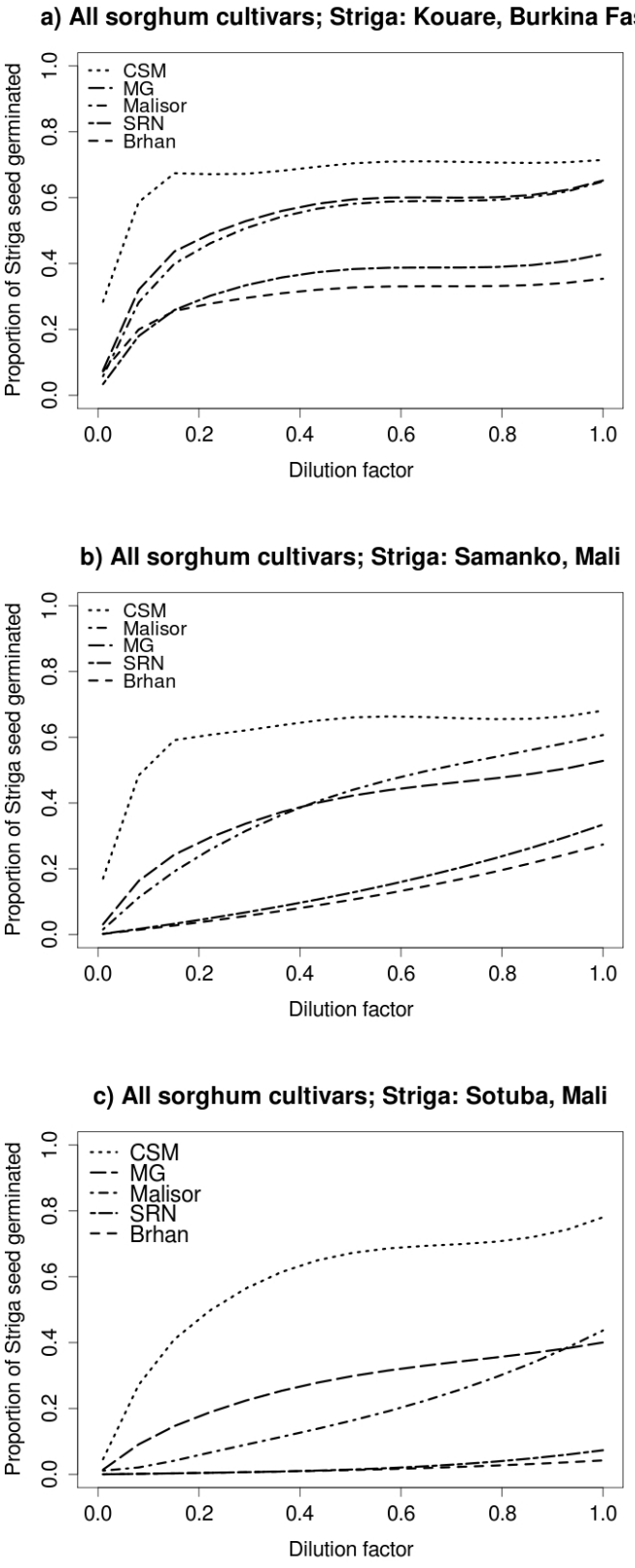

Figure 2.7a–c. For each *Striga* population, the graphs show the predictions of the minimum adequate models relating *Striga* seed germination to dilution factor for the root exudate of each of the five sorghum cultivars.

Table 2.2. Model terms for the minimum adequate models of percentage germination for each of the three West African populations of *S. hermonthica* in response to five different sorghum cultivars. The three fitted models were quasi-binomial generalised linear models (section 2.2.5.1). D.f. = degrees of freedom; $p < 0.05$ = *; $p < 0.01$ = **; $p < 0.001$ = ***.

***Striga* population and model terms**	**D.f.**	**Deviance**	**Residual d.f.**	**Residual deviance**	***F***
Kouare, Burkina Faso					
Host cultivar	4	904.61	110	2724.7	38.63***
Dilution	1	1551.26	109	1173.5	264.94***
Root dry weight	1	52.21	108	1121.2	8.92**
Host cultivar × Dilution	4	333.70	104	787.5	14.25***
Host cultivar × Dilution × Root dry weight	5	214.37	99	573.2	7.32***
Samanko, Mali					
Host cultivar	4	2323.47	120	4442.2	63.27***
Dilution	1	1616.13	119	2826.1	176.06***
Root dry weight	1	86.65	118	2739.5	9.44**
Host cultivar × Dilution	4	1204.36	114	1535.1	32.80***
Host cultivar × Dilution × Root dry weight	5	436.36	109	1098.7	9.51***
Sotuba, Mali					
Host cultivar	4	2075.56	120	2860.6	89.82***
Dilution	1	1435.94	119	1424.6	248.55***
Root dry weight	1	36.19	118	1388.4	6.26*
Host cultivar × Dilution	4	538.44	114	850.0	23.30***
Dilution × Root dry weight	4	186.21	110	663.8	8.06***

2.3.1.3 Germination responses to neat root exudate: *Striga* population × host interactions

The preceding analysis indicates that the germination sensitivities of the three *Striga* populations to the five sorghum hosts were similar: apart from the dilution factor, the identity of the sorghum cultivar appeared to be the main determinant of how the *Striga* populations reacted to host root exudate (Fig. 2.7 a–c; Table 2.2). In order to further investigate the effect of *Striga* population on germination in response to the different sorghum hosts, and to enable a comparison of the strength of any pre-attachment *Striga*

population × sorghum host interaction with any such interactions occurring at the post-attachment stage (section 2.3.2), the germination data from only one of the dilution points (1×; i.e. neat host root exudate) were examined across all three *Striga* populations (Table 2.3; Fig. 2.8a–c). These data were modelled using a quasi-binomial generalised linear model, with a standard logit link function (Crawley 2007). Tukey's HSD *post hoc* contrasts were used within each *Striga* population to compare the germination responses to neat root exudate between the five host cultivars; the *post hoc* contrasts were carried out on percentage germination once these had been adjusted for the effect of root dry weight (Fig. 2.8a–c). Although there was a statistically detectable interaction of host cultivar × *Striga* population for the neat root exudate data (Table 2.3), the graphs of these data, and the *post hoc* tests (Fig. 2.8a–c), suggested that this interaction effect was not of a size sufficient to change the order in which the different *Striga* populations responded to the five sorghum cultivar root exudates: the two categories of high (CSM; Malisor; MG) and low (Brhan; SRN) germination stimulant-producing sorghum cultivars remained broadly the same between the three *Striga* populations (Fig. 2.8a–c).

Table 2.3. Model terms for the minimum adequate model of percentage germination of three *S. hermonthica* populations in response to the neat root exudates of five different sorghum cultivars. D.f. = degrees of freedom; $p < 0.05$ = *; $p < 0.01$ = **; $p < 0.001$ = ***.

Model terms	**D.f.**	**Deviance**	**Residual d.f.**	**Residual deviance**	***F***
Host cultivar	4	1112.45	68	1068.90	47.85***
Root dry weight	1	107.22	67	961.68	18.45***
Striga population	2	238.94	65	722.74	20.56***
Host cultivar × Root dry weight	4	95.51	61	624.23	4.24**
Host cultivar × *Striga* population	8	261.01	53	363.22	5.61***
Root dry weight × *Striga* population	2	46.92	51	316.30	4.04*

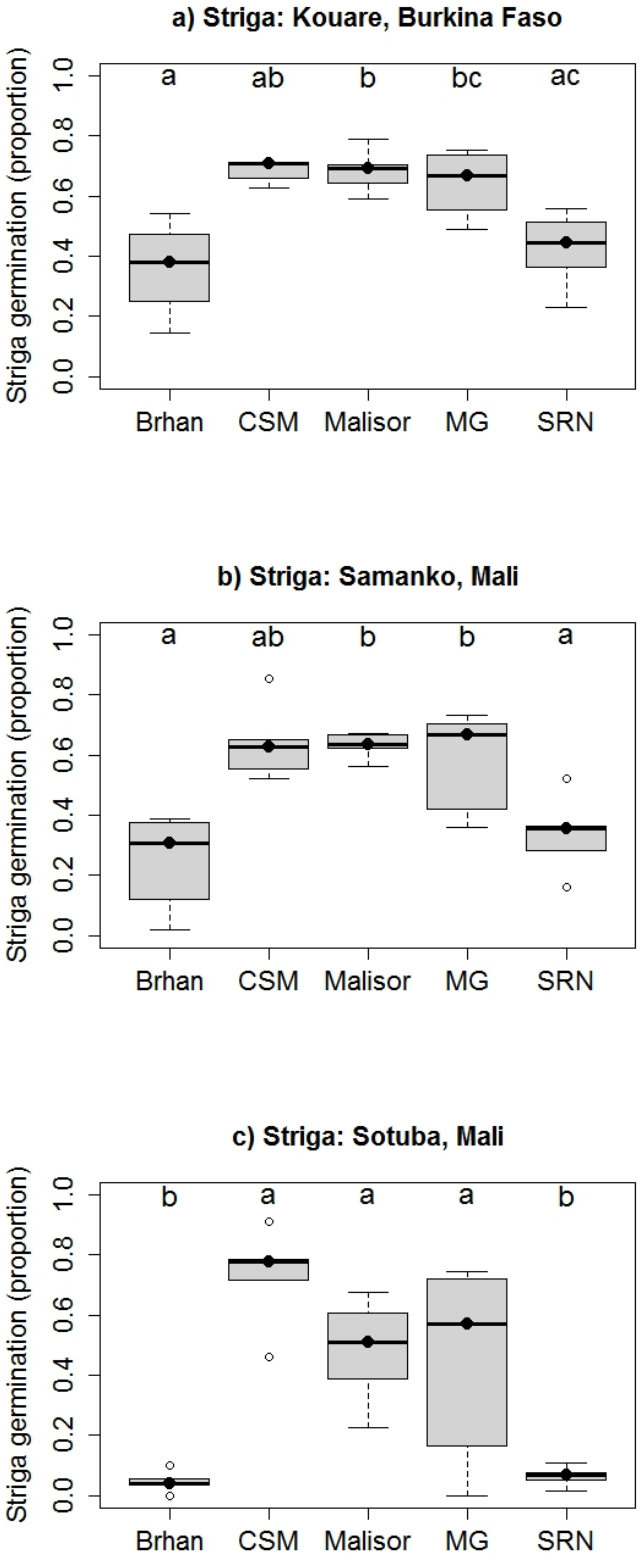

Figure 2.8a–c. *Striga* population percentage germination after exposure to neat root exudate of each of five sorghum cultivars. Tukey's HSDs ($p < 0.05$, d.f. = 1) are shown on each graph.

2.3.2 Post-attachment resistance assays

Post-attachment resistance was quantified for all *Striga* population/sorghum cultivar combinations in three different ways: as a count of the number of attached *Striga* (Fig. 2.9a–c), as the total dry weight of attached *Striga* (Fig. 2.10a–c), and as the median length of attached *Striga* (Fig. 2.11a–c). The number of attached *Striga* parasites was significantly affected by sorghum cultivar ($F = 40.30$, $p < 0.0001$, d.f. = 4), and by the *Striga* population ($F = 51.41$, $p < 0.0001$, d.f. = 2; Fig. 2.9a–c). There was also evidence for an interaction between host cultivar and the *Striga* population ($F = 16.15$, $p < 0.0001$, d.f. = 8; Fig. 2.9a–c), as for pre-attachment resistance (section 2.3.1), indicating the importance of G × G interactions in post-attachment *Striga* virulence/host resistance outcomes. Inspection of Figure 2.9a–c indicated that the differential resistance of sorghum cultivars Brhan and MG between the three different *Striga* populations was likely to be driving the significant interaction: Brhan and MG were largely resistant to *S. hermonthica* from Samanko and Sotuba, but susceptible to the *Striga* from Kouare. There was also evidence for Malisor having lower resistance to the Kouare *Striga* (Fig. 2.9a).

Total dry weight of attached *Striga* was also significantly affected by sorghum cultivar ($F = 59.94$, $p < 0.0001$, d.f. = 4; Fig. 2.10a–c), *Striga* population ($F = 53.13$, $p < 0.0001$, d.f. = 2; Fig. 2.10a–c), and again by an interaction between them ($F = 11.94$, $p < 0.0001$, d.f. = 8; Fig. 2.10a–c). Figure 2.10a–c suggests a similar pattern driving the G × G interaction to that for the number of attached *Striga*: Brhan and MG most strongly displayed differential resistance between Kaoure and the two *Striga* populations from Mali (Samanko and Sotuba); and, again, Malisor appeared to have lower resistance to Kouare *Striga* (Fig. 2.10a).

Finally, median length of *Striga* attachments was significantly affected by host (F = 13.86, $p < 0.0001$, d.f. = 4; Fig. 2.11a–c), but not, in this case, by *Striga* population (F = 1.88, p = 0.159, d.f. = 2; Fig. 2.11a–c). Correspondingly, Figure 2.11a–c reveals little difference in the way in which the five sorghum cultivars interact with the three *Striga* populations, whilst the differences between cultivars appear to be consistent between populations. The interaction between host and *Striga* population was again significant (F = 2.33, p = 0.027, d.f. = 8; Fig. 2.11a–c), although this interaction was weaker than for the *Striga* count and dry weight data. This, and the absence of a main effect of *Striga* population, suggests that the median length of attached *Striga* is a less sensitive metric for revealing differential interactions between host cultivars and *Striga* populations.

Post hoc Tukey's HSD contrasts (1 degree of freedom) between host cultivars were performed within *Striga* populations for each post-attachment metric of parasite virulence, and are given in Figures 2.9a–c, 2.10a–c and 2.11a–c, where a difference in letter indicates a significant difference at the 5% level. For one virulence metric/*Striga* population combination (median length of *Striga* for the Samanko population; Fig. 2.11b), heteroscedasticity was detected (Breusch-Pagan test, $p < 0.01$); heteroscedasticity can affect the standard errors of estimates, and so the conclusions of *post hoc* tests (Westfall *et al.* 2011). Therefore for the median length of *Striga* for the Samanko population, the *post hoc* tests were adjusted for the observed heteroscedasticity using a covariance matrix estimated from the data using functions in the R package 'sandwich' v. 2.2-9 (Zeileis 2004). The *post hoc* Tukey's HSD tests were implemented in R using the package 'multcomp' v. 1.2-14 (Hothorn *et al.* 2008).

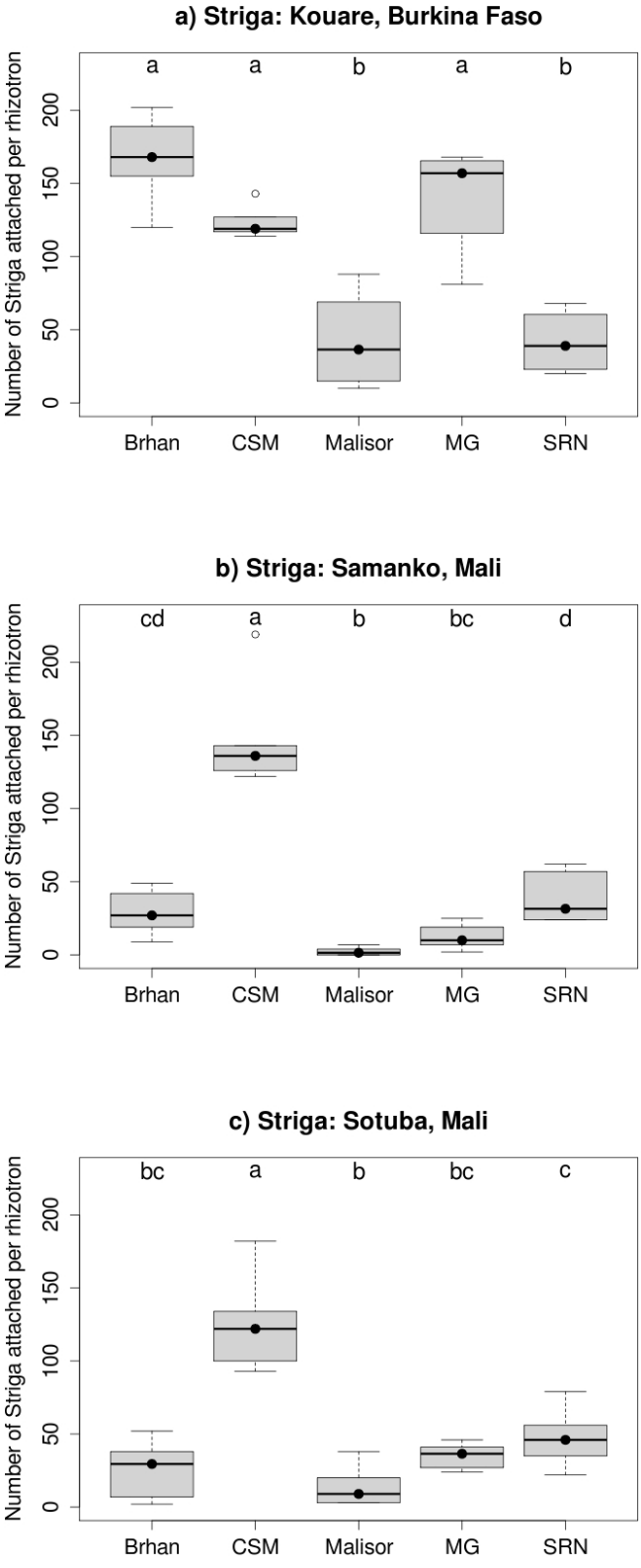

Figure 2.9a–c. Evaluation of the resistance of the five sorghum cultivars against the three *Striga* populations, as measured by the number of attached *Striga*. Tukey's HSDs ($p < 0.05$, d.f. = 1) are shown on each graph.

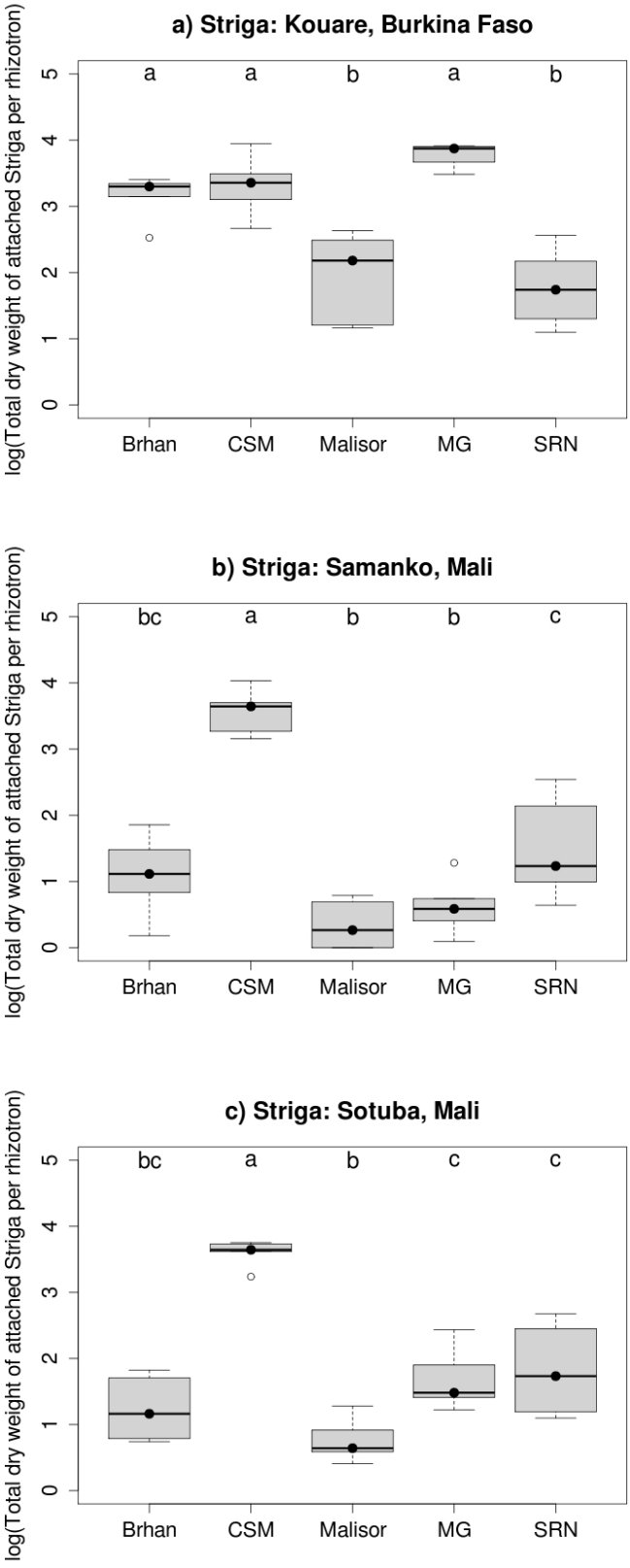

Figure 2.10a–c. Evaluation of the resistance of the five sorghum cultivars against the three *Striga* populations, as measured by the total dry weight of *Striga* harvested from a rhizotron. Tukey's HSDs ($p < 0.05$, d.f. = 1) are shown on each graph.

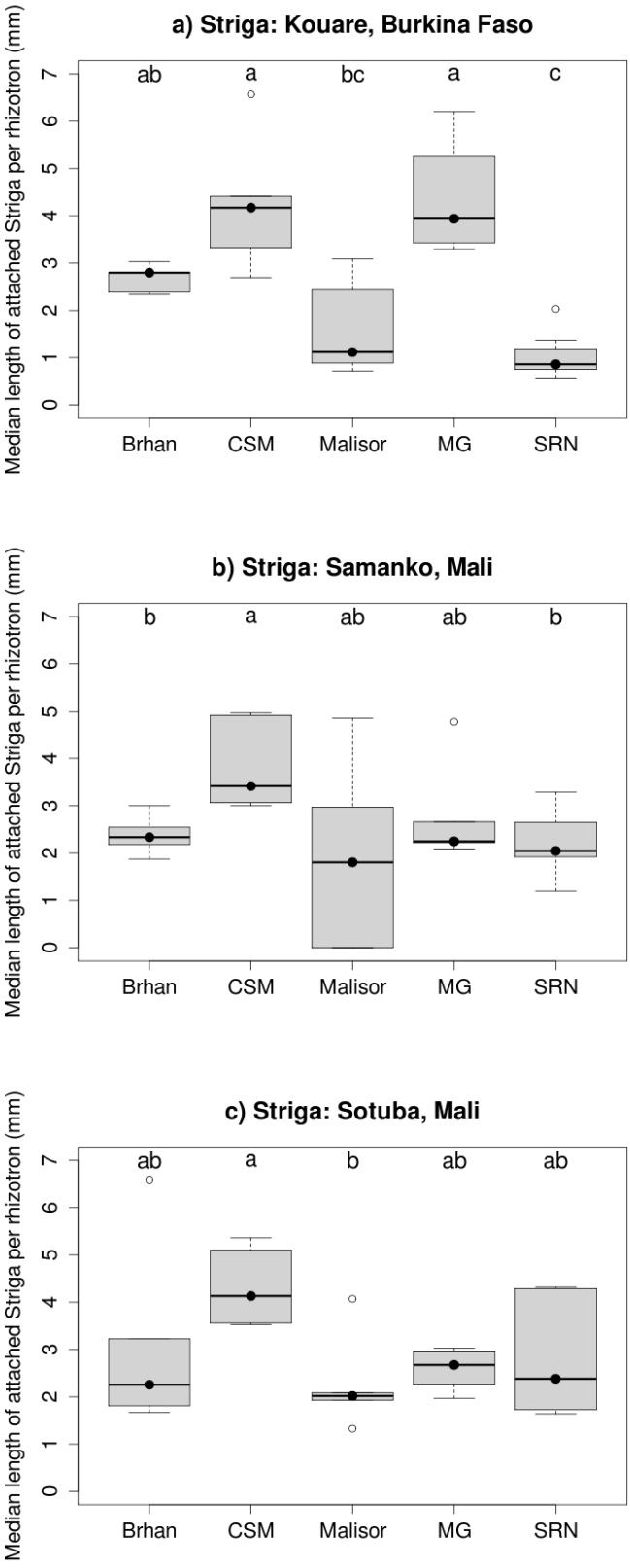

Figure 2.11a–c. Evaluation of the resistance of the five sorghum cultivars against the three *Striga* populations, as measured by the median length of the *Striga* harvested from a rhizotron. Tukey's HSDs ($p < 0.05$, d.f. = 1) are shown on each graph.

2.3.3 Field trials of sorghum cultivar resistance

2.3.3.1 Sorghum cultivar resistance over multiple sites and years

The amount of *S. hermonthica* parasitising the five sorghum cultivars in the field trials varied significantly between years, sites and sorghum host cultivars (Fig. 2.12; Table 2.4). Surprisingly, in contrast to the G × G interaction revealed by the lab assay data (sections 2.3.1 and 2.3.2), the field emergence data provided no evidence for a host cultivar × site interaction (Table 2.4). However, the overall emergence at a site varied strongly between years (Fig. 2.12), and this interaction effect was the strongest factor in the minimum adequate model (Table 2.4). At Samanko in 2008 and 2010, and Sotuba in 2008 and 2011, *Striga* emergence was generally low, and subsequently there was less information on the relative resistances of the sorghum cultivars evaluated (Fig. 2.12). The low emergence was presumed to be due to unmeasured environmental variation that also affected sorghum grain production (Dr T. van Mourik, pers. comm.) Even though *Striga* emergence was low at Samanko in 2008 and 2010 (Fig. 2.12), the grain harvest per sorghum plant per sub-plot from these field trials was significantly lower for these years than in 2009 (Kruskal-Wallis $\chi^2 = 58.83$, $p < 0.0001$, d.f. = 2, Fig. 2.13; data provided by Dr T. van Mourik, ICRISAT, Mali) when there was higher *Striga* emergence (Fig. 2.12), indicating that environmental factors, such as rainfall, affected both the crop and the parasite. In some sites and years (Kouare in 2008 and Samanko in 2009) there was greater *Striga* emergence, and distinctions in parasite resistance between cultivars then became clearer (Fig. 2.12). However, Tukey's HSD *post hoc* contrasts within sites and years revealed few significant differences in *Striga* emergence between host cultivars: the exception was for Samanko in 2009 (Fig. 2.12), where the resistance of Malisor was significantly different from all of the other cultivars except SRN (all differences $p < 0.05$), and, additionally, SRN was different from CSM ($p =$

0.01).

Table 2.4. Statistics for the model terms included in the minimum adequate linear model of emerged *Striga* at 90 d host plant^{-1} sub-plot^{-1}. D.f. = degrees of freedom; $p < 0.001$ = ***.

Model term	D.f.	Sum of squares	Mean square	*F*
Host cultivar	4	14.87	3.72	6.26***
Site	2	21.31	10.65	17.94***
Year	3	20.12	6.71	11.29***
Site × year	2	52.17	26.08	43.92***

Figure 2.12. Emerged *Striga* at 90 d host plant^{-1} sub-plot^{-1} for the three West African sorghum resistance field trials conducted between 2008 and 2011. Within field sites and years, the lines connect the means of the sub-plot replicates for particular sorghum cultivars.

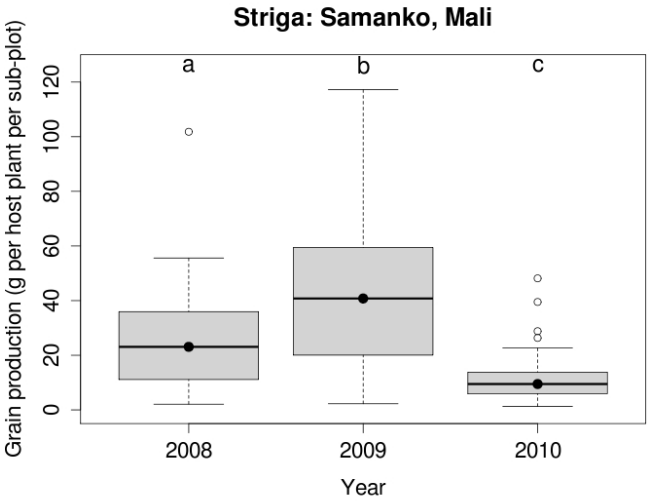

Figure 2.13. Grain production (g host plant^{-1} sub-plot^{-1}) across all sorghum cultivars in the three years of the sorghum *S. hermonthica* resistance trials at Samanko, Mali. Different letters indicate significant *post hoc* differences at the $p < 0.05$ (d.f. = 1) level following a significant global Kruskal-Wallis test ($p < 0.0001$).

2.3.3.2 Testing the explanatory power of pre- and post-attachment resistance estimates from lab-assays

The different estimates of pre- and post-attachment resistance (as defined in section 2.2.5.4) of the five sorghum cultivars to the three *Striga* populations varied in their ability to explain the observed variation in sorghum cultivar resistance to *S. hermonthica* across the field trials. If either of pre- or post-attachment resistance (or their combination), as observed in the lab, were strong influences on sorghum resistance as observed in the field, then estimates of these metrics would be expected to display strong explanatory power as covariates in statistical models of *Striga* field emergence, i.e. cultivars with high post-attachment resistance in rhizotron assays would show low field emergence, and *vice versa*.

An assessment of the relative explanatory power of the lab-assay metrics for the *Striga* field emergence data for each site is given in Table 2.5; each model contained the resistance metric specified (table rows) and 'year' as covariates. One commonly used

rule-of-thumb suggests that models within 2 AIC units of the best model can be thought of as not significantly different from the best model (Gelman & Hill 2006; Bolker 2008); these models are given in bold in Table 2.5. Even if all models are poor, AIC will still provide a relative ranking of the various models which have been hypothesised to have some explanatory power (Anderson 2008).

Table 2.5. The relative ability (as measured by AIC) of pre- and post-attachment lab-derived metrics of resistance to explain the observed variation in the emerged *Striga* at 90 d host plant^{-1} sub-plot^{-1} at three different West African field sites across multiple years. AIC results in bold within a field site are within 2 AIC units of the best model. All models contained 'year' as an explanatory factor, as well as one of the covariates given in the first column.

	AIC estimates (and rank)		
Covariate	**Burkina Faso**	**Samanko**	**Sotuba**
Post-attachment estimates			
Striga number	89.90 (7)	132.10 (2)	102.40 (6)
Striga total dry weight	**88.21 (5)**	132.40 (=4)	102.30 (5)
Striga median length	**87.58 (3)**	**123.50 (1)**	103.50 (7)
Germination estimates (pre-attachment)	**89.10 (6)**	139.90 (7)	**98.19 (1)**
Weighted post-attachment estimates (post-attachment × germination)			
Striga number	**87.67 (4)**	132.20 (3)	100.80 (4)
Striga total dry weight	**87.10 (1)**	132.40 (=4)	**100.00 (3)**
Striga median length	**87.18 (2)**	133.10 (6)	**99.29 (2)**

The best explanatory metric depended on the *Striga* population (i.e. field site). At Burkina Faso no single resistance metric stood out as the best, whereas at Samanko the unweighted *Striga* length metric captured the patterns in the field data well. The measure of seed germination sensitivity (pre-attachment resistance) was the top predictor of emerged *Striga* at Sotuba, with the germination-weighted post-attachment metrics close behind. Inspection of the within-year model R^2 values for each site indicated considerable inter-annual variation in the variance explained by any one resistance metric, suggesting a considerable environmental component to the *Striga*

field emergence data (Table 2.6). In some sites and years (Samanko in 2008 and 2009), the post-attachment resistance metrics showed reasonably good correspondence to the field data, but the germination pre-attachment metrics were poor. In other places and years (e.g. Sotuba in 2010 and 2011) germination sensitivity was a good predictor, and correspondingly, the weighted metrics also outperformed the unweighted ones. This appears to have been driven by the low sensitivity of the Sotuba *Striga* to the Brhan and SRN root exudates matching the low *Striga* field emergence for these cultivars.

It was also notable that even in years with low *Striga* emergence, such as Samanko in 2008 and Sotuba in 2011 (Fig. 2.12), the resistance metrics could still explain a considerable amount of the variance in *Striga* field emergence (Table 2.6), suggesting that some of the patterns of sorghum cultivar resistance observed in the lab experiments were in evidence in the field, even when parasite emergence was apparently limited by environmental conditions. Conversely, the strong G × G interactions suggested by the post-attachment rhizotron metrics, such as the differences in resistance for Brhan and MG between the Burkina Faso and Mali *Striga* populations (Figs 2.9a–c and 2.10a–c), were not clearly shown in the field emergence data (Fig. 2.12).

Table 2.6. The relative ability (as measured by R^2) of pre- and post-attachment lab-derived metrics of resistance to explain the observed variation in the emerged *Striga* at 90 d host plant^{-1} sub-plot^{-1} at three different West African field sites in different years.

	Burkina Faso		**Samanko**			**Sotuba**		
Model term	**2008**	**2010**	**2008**	**2009**	**2010**	**2009**	**2010**	**2011**
Post-attachment estimates								
Striga number	0.104	0.001	0.258	0.239	0.057	0.169	0.000	0.250
Striga total dry weight	0.158	0.020	0.304	0.229	0.048	0.184	0.001	0.231
Striga median length	0.176	0.035	0.244	0.526	0.102	0.115	0.002	0.220
Germination estimates (pre-attachment)	0.026	0.109	0.000	0.015	0.010	0.076	0.116	0.405
Weighted post-attachment estimates (post-attachment × germination)								
Striga number	0.180	0.028	0.204	0.237	0.061	0.133	0.018	0.361
Striga total dry weight	0.144	0.085	0.212	0.234	0.057	0.136	0.031	0.382
Striga median length	0.159	0.065	0.076	0.243	0.059	0.114	0.054	0.410

2.4 Discussion

Interactions between species may often be dependent on the genetic composition of the populations or individuals involved (Whitham *et al.* 2006). Understanding when this is important can provide insights into the dynamics of particular communities, which may be of fundamental interest for ecology and its applications (Hersch-Green *et al.* 2011). Sorghum and *S. hermonthica* have a long coevolutionary history (Welsh & Mohamed 2011), and some research has already been performed on the variable interactions between these species (Omanya *et al.* 2000). Given the numerous varieties of sorghum, and the fact that *S. hermonthica* is geographically widespread and genetically diverse (Mohamed *et al.* 2007), these two species provide many opportunities to investigate the importance of G × G interactions for plant-plant parasitism. Their agricultural importance increases the likelihood that large, multi-site experiments can be run,

providing valuable insights for crop breeding, as well as more general insights into host-parasite ecology.

This chapter has shown that G × G interactions can be highly significant in controlled host-parasite experiments, and may depend both on which stage of the interaction is examined, and on the virulence/resistance metric used, indicating the complexities of these species' interactions. The high inter-annual and inter-site variability of *Striga* emergence in field experiments suggests the presence of G × G × E interactions, and indicates that direct extrapolations from the lab to the field may not always be merited in this system.

For pre-attachment sorghum resistance, or, conversely, *Striga* sensitivity to germination stimulants, G × G interactions were detected, but *post hoc* tests suggested that the three populations of *S. hermonthica* exhibited the same rank order of germination sensitivity to the five sorghum cultivars tested. If biological variation is likely to influence the concentration of root exudates in the soil in the field (e.g. soil nutrient concentrations; Yoneyama *et al.* 2007), or other influences on germination (e.g. soil microorganisms or decaying plant matter; Haussmann, Hess, Welz, *et al.* 2000; Haussmann *et al.* 2001) are taken into account, it is likely that the G × G interaction detected here at the pre-attachment stage has little biological significance in the field. Two broad categories of *Striga* germination stimulant activity (CSM, Malisor and MG as high stimulant cultivars and SRN and Brhan as low stimulant cultivars) seem all that the evidence from the current study of three *Striga* populations would support. However, some inter-population variability was observed: compared to the Kouare and Samanko *Striga* populations, the Sotuba population appeared to have lower sensitivity to all of the cultivars except CSM. A similar overall pattern was found in a lab study by Haussmann

et al. (2001): these authors also tested the germination of three populations of *S. hermonthica* exposed to root exudate from five sorghum cultivars. They found that host identity was the most important factor influencing *Striga* germination, followed by *Striga* population, with only a small interaction effect. In my experiments, the effect of root dry weight was also important (the variation in dry weight was around ten-fold: range = 0.02-0.18 mg), suggesting that workers should make a greater effort to quantify the effects of root biomass variation in assays for germination stimulation and germination stimulant production (cf. Hess *et al.* 1992; Vasey *et al.* 2005; Jamil *et al.* 2011).

The rhizotron studies of sorghum post-attachment resistance indicated the presence of significant G × G interactions for all three metrics; this was supported by different resistance rankings and *post hoc* results between the three *Striga* populations, especially for the cultivars Brhan and MG. For the number of *Striga* attachments and the total dry weight, both Brhan and MG appeared less resistant to the *Striga* from Burkina Faso than they did to the two neighbouring populations of *S. hermonthica* from Mali, suggesting a geographic component to the genetics underlying *Striga* virulence.

The three different post-attachment metrics used in this study may reflect different underlying aspects of the host-parasite interaction. Rhizotron studies of *Striga*-host interactions have used a variety of scoring methods that have either categorised attachments into different parasite developmental stages, and have interpreted these directly and/or analysed derivatives of these (Gurney *et al.* 2003, 2006; Huang *et al.* 2012), or have used parasite length and/or dry weight (Gurney *et al.* 2003; Cissoko *et al.* 2011). Different cellular and molecular interactions may lie behind different phenotypic measures of a parasite's success on its host. For example, in rhizotrons, the

number of attachments on a host may be high, but individual parasites may not develop to any great size (Cissoko *et al.* 2011); indeed, Drs M. Cissoko & A. Boisnard (pers. comm.) found that using *Striga* total dry weight as the response metric in a quantitative genetic study of rice resistance led to clearer identification of QTL than when using the number of *Striga* attachments. The results reported here indicate that parasite number and total dry weight provide a greater separation in parasite resistance between the different hosts than the median length of a parasite, although all three responses differentiated hosts to some extent; this suggests the choice of a response metric is important when reaching conclusions about host-parasite interactions. Some metrics may not reflect important aspects of parasite fitness in the field, just as some measures of parasite field virulence correlate better with measures of host fitness than others (Omanya *et al.* 2000). Whilst rhizotron-derived metrics have been used to identify resistant host germplasm (Gurney *et al.* 2003) and QTL for *Striga* resistance (Gurney *et al.* 2006; Swarbrick *et al.* 2009), correlations with field responses have not been investigated, as has been the case for germination assays or pot studies of *Striga* virulence (Omanya *et al.* 2000). Of course, if different rhizotron metrics represent different underlying aspects of the host-parasite interaction, it is possible that they may correlate better with field metrics other than *Striga* emergence.

In the field, a wide range of measures have been used to quantify the *Striga* virulence/host resistance relationship (Omanya *et al.* 2004; Rodenburg *et al.* 2005), and studies have arrived at broadly similar conclusions as to which measures are the most stable for evaluating the heritable component of a host response. In the current study, *Striga* field emergence at 90 days, found to be a reasonably stable measure by Omanya *et al.* (2004), was used. However, considerable variability in response was still

observed, mainly caused by site, year and their interaction, but, perhaps surprisingly given the rhizotron results for post-attachment resistance, no host × site interaction was detected, suggesting that environmental conditions overrode the G × G interactions detected in the lab. The variable R^2 values for the models testing the explanatory power of the lab-assay measures between years, but within sites, also suggested that unmeasured environmental variables were the main determinants of *Striga* emergence.

Although the most important focus for field studies of host resistance to *Striga* will often be the stability of host resistance (because stable genetic variation is required for breeding programs; Rodenburg *et al.* 2005), in other areas, such as for more general research into host-parasite coevolution (Wolinska & King 2009) or community genetics (Rowntree *et al.* 2011), the effects of enviromental variation on host-parasite interactions may be at least as important as G × G interactions and host/parasite genotype main effects (Tack *et al.* 2012). Indirectly assessing the predictive value of lab-derived G × G interactions is one way to investigate the importance of the environment (Omanya *et al.* 2000), especially where reciprocal common garden experiments are challenging to perform. Here, the use of AIC values to compare models of *Striga* field emergence indicated that the best lab predictor of the field response varied between sites and years. At Burkina Faso there was no clear predictive advantage to any of the metrics tested, whilst at Sotuba germination sensitivity was the best predictor, no doubt because of the correspondence of the low *Striga* emergence on Brhan and SRN with the low sensitivity of the Sotuba *Striga* to the root exudates of these cultivars. Whilst length was selected as the best predictor for Samanko by AIC across all years, inspection of the yearly R^2 values at this site showed that this measure was often high for several of the different metrics, and that the selection of length as the

best explanatory variable seemed to be driven by its particularly close correspondence to the *Striga* field emergence in 2009. Despite reasonable R^2 values (e.g. for Burkina Faso in 2008, Samanko in 2008 and 2009, and Sotuba in 2011) for virulence metrics that resulted in significant differences between cultivars in the lab, only for Samanko in 2009 did *post hoc* tests indicate significant differences between host cultivars for *Striga* field emergence. This demonstrates the importance of multi-year testing when evaluating the significance of lab-derived information on G × G interactions for real field outcomes. Realism in experimentation is clearly of great importance when attempting to extrapolate inferences to field studies of species' interactions (Fenton *et al.* 2012).

Omanya *et al.* (2004) reported R^2 values of 0.000-0.281 (mean = 0.10, their Table 7) for the relationship between pot studies and the field metric 'area under the *Striga* number progress curve' (ASNPC; obtained by integrating the curve of *Striga* emergence over time); whilst they reported R^2 values of 0.002-0.397 (mean = 0.091, their Table 8) for the relationship between their germination assay (maximum germination distance; Hess et al., 1992) and ASNPC. Omanya *et al.* (2004) also report larger correlations between ASNPC and germination distance (R^2 = 0.608 and 0.818, their Table 9), although these higher values were derived from correlating ASNPC averaged across two Mali sites with the germination data averaged between a site from Mali (Samanko) and one from Niger (Bengou), so these R^2 values do not truly represent the relationship between a *Striga* population that has been simultaneously tested in the field and the lab; that is, the change in metric was partially confounded with a change in *Striga* population. The mean R^2 value of the rhizotron and germination metrics presented in the current study is 0.147, very close to the mean values from Omanya *et al.* (2004). I found no differences

between the average R^2 values of the different metrics: all of them varied in their predictive ability, and no single metric consistently outperformed the others. In contrast, Omanya *et al.* (2000, 2004) contended that their germination assay results were superior to their pot study in explaining *Striga* field emergence, although their confidence in this measure was due to good results at one site, and even these results were not very different from some of the better results that they obtained with pot studies.

Recent research on a selection of new upland 'NERICA' rice cultivars has suggested reasonable correspondence between rhizotron-based measures of post-attachment resistance and field studies (Cissoko *et al.* 2011; Atera *et al.* 2012); however, the field study (Atera *et al.* 2012) also showed significant year-to-year variation, and indicated that one rice cultivar, NERICA 4, that has been found to be resistant in rhizotron and germination stimulant assays (Cissoko *et al.* 2011; Jamil *et al.* 2011), was very susceptible in the field site in Western Kenya where the study was conducted; it is not known whether is this due to differences in *S. hermonthica* genotypes or to environmental factors. Omanya *et al.* (2004) outlined some reasons why pot experiments may sometimes be an unrealistic representation of the field environment for the *Striga*-host interaction, and it is clear that further knowledge of the environmental influences on host-parasite interactions will be required if predictions of G × G interactions in the field are to be consistently realistic (Wolinska & King 2009).

Lab assays may be useful for assessing G × G interactions in some instances: they often represent a cost-effective way of screening large amounts of germplasm, or of minimising the variability in environmental influences on species' interactions. However, the data presented here, and the results of other workers on parasitic plants (Omanya *et al.* 2000; Omanya *et al.* 2004), suggest that interactions observed under

assay conditions may deviate considerably from the field situation, or vary so much with environmental influences across field sites that any G × G effect found under controlled conditions may be biologically insignificant. Mitchell *et al.* (2005) came to a similar conclusion after finding strong effects of temperature on G × G interactions between the crustacean *Daphnia magna* and its bacterial parasite *Pasteuria ramosa*. Alternatively, the fact that G × G interactions can be demonstrated between *S. hermonthica* and sorghum, and that rice and sorghum hosts have been shown to select for certain genotypes of *Striga* in the lab (Koyama 2000a; Huang *et al.* 2012), also suggests that, even though environmental influences on species' interactions can be large, over time the high fecundity of *Striga*, coupled with repeated exposure to a particular host cultivar, may still lead to strong G × G interactions of the type that are often thought of in terms of host specificity (Vasudeva Rao & Musselman 1987; Mohamed *et al.* 2007). Indeed, the presence of host specialisation among species in many parasitic plant genera (Parker & Riches 1993; Stace 2010) suggests that this must be so: variation among coevolution in demes across space and time is likely to lead to differing degrees of host adaptation (Thompson 2005), and understanding the variables that influence this at different temporal and spatial scales is a major challenge (Burdon & Thrall 2009).

The absence of a strong G × G interaction at the pre-attachment stage of the parasite life-cycle suggests that populations of *Striga* may be less likely to evolve higher sensitivity against low-germination stimulant hosts (i.e. populations seem less likely to become locally adapated at this life-stage), although germination itself was not universally better at predicting field responses than the post-attachment metrics in the current study. However, it seems likely that even the low germination stimulant

cultivars in this study, Brhan and SRN, still produce enough stimulant, under some circumstances, to lead to levels of parasite infection that could easily maintain levels of soil infestation (Van Delft *et al.* 1997; Van Mourik *et al.* 2008), reduce crop yields (Gurney *et al.* 1999), and lead to adaptation at the post-attachment life-stage.

2.4.1 Conclusions

Here I have shown the existence of strong G × G interactions between *S. hermonthica* populations and sorghum cultivars at the post-attachment life-stage. Evidence for interactions at the pre-attachment life-stage was less convincing, suggesting that the different steps necessary for successful parasitism (or resistance) may evolve at different rates. Field trials of sorghum cultivar resistance showed that environmental influences can mask G × G interactions found in lab experiments; further to this, there was little evidence for any particular lab-derived resistance metric being a universally good predictor of sorghum resistance in the field. Temporal environmental variation in selection pressure for host resistance or parasite virulence may constrain the ability of pathogens to adapt to their hosts (Kisdi 2002; Kawecki & Ebert 2004); in the current agricultural system, periodic changes of crop may already place constraints on the adaptation of *S. hermonthica* to particular sorghum genotypes. The discovery of G × G interactions in the current work does not demonstrate the adaptation of *Striga* to specific crop genotypes because historical information on previous host-parasite exposure is not available. However, the presence of G × G interactions based on pre-existing genetic diversity suggests that host genotype-specific parasite adaptation is very likely in this system. Further work in this area should seek to correlate historical knowledge of crop hosts planted in a location with the existence of host-parasite G × G interactions, and seek to quantify the effects of environmental variation on the development of *Striga*

adaptation to crop host genotypes.

Chapter 3. Detecting locus-specific host selection in the field

3.1 Introduction

Striga hermonthica has a wide distribution in Africa (Mohamed *et al.* 2001) and is an obligate outbreeder (Safa *et al.* 1984). Variation in corolla shape and colour, and in plant architecture, suggests that the species is genetically variable (Parker & Riches 1993; O. Pescott, personal observations). Indeed, it is now well-established that *S. hermonthica* is a species containing a high level of genetic variation (Table 3.1), as would be expected for an outbreeding and weedy annual plant (Loveless & Hamrick 1984; Hamrick & Godt 1996). The other main motivating factor for population genetic work on *S. hermonthica* has been the question of whether genetic variation correlating with host specificity can be observed (Bharathalakshmi *et al.* 1990; Olivier *et al.* 1998; Kuiper *et al.* 1998; Koyama 2000a; Ali *et al.* 2009; Yoshida *et al.* 2010; Estep *et al.* 2011; Welsh & Mohamed 2011; Huang *et al.* 2012; Table 3.1).

Due to its wide distribution, and its ability to parasitise the commonest cereal crops in Africa, researchers have long been interested in the relationship between *S. hermonthica* genetic diversity and host range (Vasudeva Rao & Musselman 1987; Mohamed *et al.* 2007). Reports of *S. hermonthica* populations specific to sorghum and pearl millet in the Sudan (Musselman & Hepper 1986), and experimental work on the physiological basis of these interactions (Parker & Reid 1979), have led to a now considerable body of genetic research on variation in the host specificity of *S. hermonthica* across Africa (Table 3.1). One often stated aim of this research is to identify geographic regions containing *S. hermonthica* populations with particular host specificities in order to inform the testing and deployment of resistant host germplasm,

and the provision of advice to farmers (e.g. Estep *et al.* 2011). However, very little evidence for host adaptation at the genetic level has been found to date (Table 3.1). This is in contrast to physiological work which has found host specificity (which is more clearly thought of as an interaction between host and parasite genotypes i.e. a G × G interaction) at both the pre-attachment (germination stimulation) (Parker & Reid 1979; Bebawi *et al.* 1986; Vasudeva Rao & Musselman 1987; Jamil *et al.* 2011; Chapter 2), and parasite attachment levels (Kim *et al.* 1994; Freitag *et al.* 1996; Cissoko *et al.* 2011; Huang *et al.* 2012; Chapter 2).

.

Table 3.1. A summary of studies that have investigated genetic variation and population differentiation in *S. hermonthica*. H_{exp} is the average expected heterozygosity.

Study	Location & host(s) of *S. hermonthica* sample(s)	Sampling method	Molecular markers	Mean H_{exp} or H_{obs}	F_{ST} (or analogue) range	Conclusions	Comments
Bharathalaksmi *et al.* (1990)	Burkina Faso: Sorghum; millet. Sudan: Sorghum. (Cultivars not specified).	Field sampling.	9 allozyme loci	H_{exp} = 0.317	0.011-0.214	Some genetic differentiation; geographic factors appeared more important than host adaptation.	-
Kuiper *et al.* (1996)	Benin: Maize; wild grasses. Burkina Faso: Pearl millet. Mali: Sorghum. Kenya: Maize; sorghum; wild sorghum; wild grasses (Cultivars not specified).	Field sampling of seed, followed by axenic culture.	14 allozyme loci	H_{obs} = 0.180	0.116-0.552	East and West African *S. hermonthica* clustered together; no indication of host specificity.	UPGMA cluster analysis not bootstrapped.
Olivier *et al.* (1996)	-	-	-	-	-	-	Identical to Olivier *et al.* (1998).
Olivier *et al.* (1998)	Burkina Faso: Sorghum; maize; millet; wild grasses. Mali: Sorghum; millet. Niger: Millet. (Cultivars not specified).	Field sampling.	2 allozyme loci	H_{exp} = 0.400 (calculated from data presented)	-0.038-0.065 (calculated from data presented)	No signal of either geography or host adaptation detected.	-
Koyama (2000b)	Kenya: Sorghum. Mali: Sorghum. Nigeria: Sorghum. (Cultivars not specified).	Lab-based sampling on one sorghum cultivar (ICSV 111).	10 allozyme loci 33 RAPD loci	-	-	Geographic genetic variation found, both within and between populations.	1. Selection within field samples by host in the lab means underlying population variation in the field not exactly known. 2. Host adaptation not addressed. 3. UPGMA cluster analysis not bootstrapped. 4. Discriminant analysis technique seeks to separate populations by maximising

							the variance between pre-defined groups, and is therefore, arguably, inappropriate for testing a null hypothesis that groups are not distinct.
Koyama (2000a)	Kenya: Sorghum. (Cultivar not specified).	Lab-based sampling on sorghum cultivars (cvs ICSV 111; ICSV 400; Serena; ICSV 1007; SRN39).	10 allozyme loci 33 RAPD loci	-	-	Strong clustering of *S.hermonthica* sub-samples selected by sorghum cultivars in pot experiments.	1. UPGMA cluster analysis not bootstrapped. 2. Discriminant analysis technique seeks to separate populations by maximising the variance between pre-defined groups, and is therefore, arguably, inappropriate for testing a null hypothesis that groups are not distinct.
Gethi *et al.* (2005)	Kenya: Sorghum; maize. (Cultivars not specified).	Field sampling.	349 AFLP loci	-	D = 0.007-0.025	Generally low differentiation; no relationship between pairwise genetic distances and geographic distance.	Host adaptation not addressed.
Ali *et al.* (2009)	Sudan: Sorghum (cvs Wad Ahmed and Abu-70); millet (cvs Ashana and Sudanil); maize (cvs Hudalba and Banar).	Lab-based sampling using the same hosts that the *Striga* seed was sampled from in the field.	23 AFLP loci	-	-	Some evidence that *Striga* from maize and sorghum were more closely related to each other than to a population collected on millet.	1. UPGMA cluster analysis not bootstrapped. 2. Possible partial confounding of host-identity with geographic distance.
Yoshida *et al.* (2010)	Kenya: Maize. Sudan: Maize; sorghum; pearl millet. (Cultivars not specified).	Field sampling.	10 microsatellite loci	H_{exp} = 0.549	-	No correlation between genetic distance and host specificity found.	-

Estep *et al.* (2011)	Mali: Sorghum; pearl millet. (Cultivars not specified).	Field sampling.	12 microsatellite loci	H_{exp} = 0.715	R_{ST} = 0.048	Little between-population genetic differentiation, but two population clusters identified. These clusters did not correlate with host-identity or with obvious environmental factors or geographic distance.	-
Welsh & Mohamed (2011)	Ethiopia: Maize (farmer's variety); millet (land race); sorghum (land races, and the improved varieties Gubiye and Meko); tef (an improved variety and a land race).	Field sampling.	385 AFLP loci	H_{exp} = 0.204	0.032–0.293	High levels of differentiation between some populations; geographic factors, not host adaptation, appeared to be responsible.	-
Huang *et al.* (2012)	Kenya: Maize (cv. H511).	Lab-based sampling on rice cultivars (IAC165; Kasalath; Nipponbare).	191 AFLP loci	H_{exp} = 0.234	0.013 (only one seed population sampled)	Different rice hosts shown to select for different genotypes within one *Striga* population.	-

In general, for genetic work, investigating the host adaptation of *Striga* to species rather than to genotypes within species (e.g. cultivars) has been the preferred approach (Table 3.1). This is because crop hosts can be instantly grouped by species, and associations between groupings of related *Striga* populations and the species that these most successfully parasitise (or have been found to be parasitising) can be quickly assessed (e.g. Bharathalaksmi *et al.* 1990; Welsh & Mohamed 2011). However, investigating associations between *Striga* population relatedness and host cultivar genotype relatedness would also require the genotyping of host cultivars. These approaches essentially rely on establishing the degree of congruence between parasite and host population-level phylogenies in order to demonstrate host adaptation (cf. Nieberding & Olivieri 2007). This approach can also be seen from a G × G interaction perspective: parasite populations with similar reactions against related sets of host genotypes would be expected to be genetically similar. However, at the intra-specific host genotype level, G × G interactions may have diverse foundations, and the hypothesis of congruence between host and parasite phylogenies may deserve less weight *a priori* (cf. Huyse *et al.* 2005). In other words, there may be multiple independent genetic mechanisms underlying the parasitism of a specific host genotype (species or cultivar) by *S. hermonthica*, and therefore *Striga* populations that have similar success in parasitising a particular host genotype(s) need not be closely related. Indeed, circumstantial evidence for this being the case is provided by the fact that African cereal crops can be successfully parasitised by both *S. hermonthica* and *S. asiatica* (Parker & Riches 1993).

The lack of a strong genetic signal for host-parasite G × G interactions to date (i.e. strong congruence between *Striga* relatedness and host relatedness), but abundant physiological evidence for G × G interactions across host species and genotypes (Parker

& Reid 1979; Bebawi *et al.* 1986; Vasudeva Rao & Musselman 1987; Kim *et al.* 1994; Freitag *et al.* 1996; Cissoko *et al.* 2011; Jamil *et al.*, 2011; Huang *et al.* 2012; Chapter 2), strongly suggests that *S. hermonthica* can rapidly adapt to new hosts, and this conclusion has been reached by several authors (Welsh & Mohamed 2011; Huang *et al.* 2012).

Another potential reason for this discrepancy lies in the different experimental designs that physiological and genetic studies have used. Physiological approaches generally expose different host genotypes to populations of *S. hermonthica*, either in the lab or in a common field environment, using an orthogonal design to ensure all combinations of host and parasite are investigated (Haussmann *et al.* 2004; Dr T. van Mourik, pers. comm.) Genetic approaches have, for the most part, used opportunistic sampling regimes. This has meant that different host species or genotypes have not been represented at all *Striga* sampling sites (Table 3.1); this results in some degree of confounding between geographic variation and variation due to host selection. Ideally, the same set of hosts would be planted at all sites: that is, a replicated, randomised field trial across multiple sites would provide the greatest ability to partition *Striga* genetic variation into host-specific and geographic components (see Manel *et al.* 2009 for an example of this type of approach). The opportunistic sampling that has dominated research into the genetic basis of host adaptation in *S. hermonthica* is only likely to uncover long established, broad patterns of population differentiation, whereas, in fact, rapid, subtle micro-evolutionary responses of *Striga* to recently encountered hosts may underlie host adaptation (cf. Hendry *et al.* 2007). The uncovering of patterns of relatively large-scale population differentiation may fulfil the aim of characterising broad genetic structure across *Striga* populations (e.g. Estep *et al.* 2011), but it is not at

all certain that *S. hermonthica* will have uniform responses to different host genotypes within the populations so delimited; especially when, locally, populations are likely to contain the raw genetic material for adaptation to novel hosts (Koyama 2000a; Huang *et al.* 2012), and physiological G × G interactions are common (see above). Looking at within-population genetic diversity (Koyama 2000a; Huang *et al.* 2012) is therefore likely to be the most effective way of uncovering the genetic basis of host adaptation. I emphasise that this is a shift from the prevailing approach to host adaptation in *Striga*, in which it has been assumed that patterns of specificity in the landscape are well enough established (or well enough differentiated in the seed bank) to be uncovered by simply assessing the association of host identity with some measure of genetic distance (Barathalaksmi *et al.* 1990; Kuiper *et al.* 1996; Olivier *et al.* 1998; Ali *et al.* 2009; Yoshida *et al.* 2010; Estep *et al.* 2011; Welsh & Mohamed 2011; Table 3.1). The prevailing approach essentially assumes that genome-wide patterns of diversity are sufficient to reveal host adaptation, whereas a focus on locating specific loci underlying specificity is likely to be more successful in the early stages of host adaptation (Strasburg *et al.* 2012).

Koyama (2000a) and Huang *et al.* (2012) have both shown that, in the lab, different hosts can select for different genotypes from a single population of *S. hermonthica*. The work of Koyama (2000a) shows this in a particularly striking manner: using 33 random amplified polymorphic DNA (RAPD) loci (checked for Mendelian segregation), unweighted pair group-mean averaging (UPGMA) clustering (based on a Euclidean distance matrix calculated from RAPD fragment presences and absences) clustered sorghum cultivar-selected *Striga* individuals into groups that showed no overlap. No formal permutation test or bootstrapping was applied to this analysis, but the results

appear, at least, to be clear. Huang *et al.* (2012) used a similar approach, and found a low level of differentiation between three rice cultivar-selected sub-populations of *S. hermonthica* (F_{ST} = 0.013); they also used outlier tests to identify 24 AFLP loci that were significantly differentiated between the host-selected *Striga* sub-populations; these loci were postulated to represent, or be linked to, specific areas of the *S. hermonthica* genome involved in host adaptation.

The aim of this chapter is to extend this within-population approach to the genetics of host specificity to the field, and to provide further insight into the frequency of occurrence of within-population selection of *Striga* parasites by their hosts. Originally, the intention was to sample from three field sites, and to investigate the occurrence of outliers within *Striga* populations, to compare these between populations. However in the event, genomic DNA usable for AFLP was only available from one field population; this removed the possibility of obtaining insights into variation in within-population host specificity in the field, but has still allowed for the investigation of several points of interest. Specifically, the following questions have been addressed: (1) In the field, are host-selected *S. hermonthica* sub-populations genetically differentiated, and, if so, at what level? (2) Can individual AFLP loci (outliers) that are driving this differentiation be identified? And, (3), are outlier loci associated with parasite virulence, as measured by the emergence of *S. hermonthica* in the field?

3.2 Materials and methods

3.2.1 Plant materials and field experiments

I utilised an African field experiment which was primarily intended to to characterise *Striga*-resistant and *Striga*-tolerant varieties of sorghum (Dr T. van Mourik, pers.

comm.; see Chapter 2 for further information). This assessment took place across three sites in West Africa; these sites are part of the International Sorghum and Millet Collaborative Research Support Program (INTSORMIL CRSP; http://intsormil.org). One site was at Kouare, Burkina Faso (11°95'N:00°30'E), and two sites were in Mali, at Sotuba (12°66'N:07°91'W) and Samanko (12°52'N:08°07'W). The field trials evaluated 15 sorghum genotypes. The intention was to sample *S. hermonthica* growing on 9 of these 15 genotypes; the 9 sorghum hosts were initially chosen for their differential reactions to *Striga* across the field sites (Table 3.2). The trials used a randomized complete block design with either three or four replicates. Each replicate consisted of a row of sub-plots, one sub-plot for each of the 15 sorghum cultivars (see Chapter 2 for further details). Ten *Striga* plants were harvested from each replicate sub-plot; all sampling took place in November 2009. To promote uniform parasite pressure, either the planting holes (Sotuba and Kouare) or the planting ridges (Samanko) were artificially infested with *S. hermonthica* seed previously harvested at the test site.

Table 3.2. Average reactions of 9 sorghum genotypes to *S. hermonthica* in 2009 at the INTSORMIL trial sites, as qualitatively assessed by Dr T. van Mourik (pers. comm.) R = Resistant; S = Susceptible; N = Neutral.

Sorghum genotype	Kouare, Burkina Faso	Samanko, Mali	Sotuba, Mali	Overall assessment
Brhan	R	N	R	Resistant
CEF 322/35-1-2	S	N	R	Contrasting
CSM388	S	S	S	Susceptible
Malisor 92-1	R	R	R	Resistant
Mota Galmi	R	S	N	Contrasting
Lina3	N	S	S	Susceptible
Sariaso9	S	S	N	Susceptible
SRN39	R	R	N	Resistant
Wassa	N	N	R	Intermediate

The harvesting of *Striga* was coordinated by Drs Tom van Mourik (ICRISAT, Mali) and

Hamidou Traore (INERA, Burkina Faso). Low *Striga* emergence at Sotuba led to a decision being taken by the field coordinators not to sample at this site. Leaves from individual *S. hermonthica* plants were sampled from the Kouare and Samanko sites and were oven-dried at 40°C (Drs T. van Mourik and H. Traore, pers. comm.) In total, dried leaf tissue samples of 10 *S. hermonthica* plants × 4 sub-plot replicates × 9 sorghum genotypes × 2 sites = 720 were received at the University of Sheffield for DNA extraction.

3.2.2 High-throughput DNA extraction

Between 4–6 mg of dried *Striga* tissue was placed in 2 mL grinding tubes with 5 mm steel ball bearings (Qiagen) and ground in a tissue homogeniser (Qiagen, 5 min at 25 Hz). Two-hundred and fifty µL of pre-heated extraction buffer (100 mM Tris-HCl (pH 7.4); 500 mM NaCl; 50 mM EDTA; 0.7% sodium dodecyl sulphate; 52 mM sodium sulphite; 1.6 µg RNAse A; 16 µg Proteinase K) was added; tubes were vortexed thoroughly and incubated on a rotating rack at 55°C for 30 min. Two-hundred and fifty µL of precipitation buffer (3.6 M potassium; 6 M acetate) was added to each sample tube, whereupon samples were placed at -20°C for 15 min, with occasional inversions. The samples (with the tubes still containing the grinding bearings) were then spun at 13,000 rpm for 5 min in a desktop centrifuge (Hawk 15/05, Sanyo Ltd, Japan). The supernatants were transferred to the wells of Millipore 'DV' filter plates. Samples had already been assigned a number, and these numbers were randomly assigned to wells across eight 96-well plate templates; the transfer of supernatants to filter plates matched these templates, and from this point forward all steps were performed in 96-well plates. Millipore filter plates were placed atop new 1.2 mL storage plates (ABgene, UK). A plate centrifuge (Hermle Z 300K, Hermle Labortechnik Ltd, Germany) was used at

3500 rpm for 5 min to filter the samples into the storage plates. Six-hundred and fifty μL of binding buffer (NaI 6 M; pH 7.0) was added to the filtered samples and mixed in the tips of an automatic, 1000 μL multichannel pipette (StarLab, UK). The samples were then transferred in 225 μL aliquots to Millipore silica 'FB' filter plates, atop old 1.2 mL storage plates for waste collection. These plates were spun at 500 rpm for 1 min. This was repeated until all of the sample mixture had been passed through the silica filter plate. The DNA samples, now bound to the silica filters, were then washed twice (1000 rpm for 1 min) using 225 μL of freshly prepared, cold, washing buffer (10 mM Tris; 0.5 mM EDTA; 50 mM NaCl; 50% ethanol). An extra spin was performed at 3500 rpm for 2 min to remove residual wash buffer; the plates were then dried at 37°C to ensure any remaining ethanol had evaporated. Two lots of 40 μL of elution buffer (10 mM Tris-HCl pH 8.0; 0.1 mM EDTA), preheated to 80°C, were sequentially applied to the plates to elute the DNA into new, 96-well, v-bottomed tissue culture plates (VWR International Ltd, UK). Samples (5 μL) were subsequently loaded onto 1% (w/v) TAE-agarose gels (4 mm thick; 40 mM Tris-acetate; 1 mM EDTA; 0.1 μg mL^{-1} ethidium bromide) for DNA quality checking via gel electrophoresis (gels ran at 70 V for 40 min). DNA size-marker ladders (Bioline, UK) were used for comparison. A representative gel is shown in Figure 3.1.

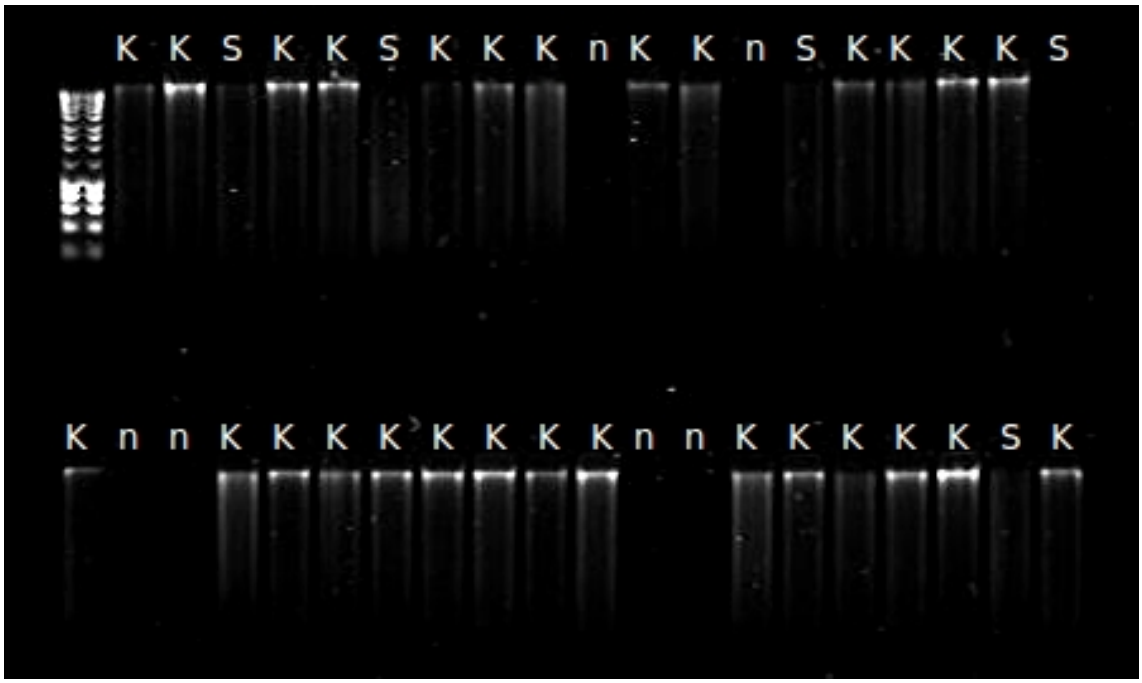

Figure 3.1. A representative gel from the DNA extraction procedure. Generally, samples from Kouare were intact, with an obvious band of intact genomic DNA, while samples from Samanko show a smear without evidence of a high molecular weight band. K = Kaoure, Burkina Faso; S = Samanko, Mali; n = negative controls. The top left lane contains Hyperladder I, with nucleotide fragments running from 10,000 base pairs to 200 base pairs.

From Figure 3.1 it can be seen that several samples were degraded and failed to show any evidence of high molecular weight DNA. These degraded samples originated overwhelmingly from Samanko, Mali. Over all the samples, in excess of 75% of the Samanko samples were degraded; re-extractions were performed on subsets of the degraded Samanko samples, but no improvement in DNA quality was found. Because intact, high molecular weight DNA is a prerequisite for reliable AFLP (Meudt & Clarke 2007), the Samanko samples were not taken forward for molecular marker generation. Therefore, from this point onwards, all analyses were restricted to *Striga* from Kouare, Burkina Faso.

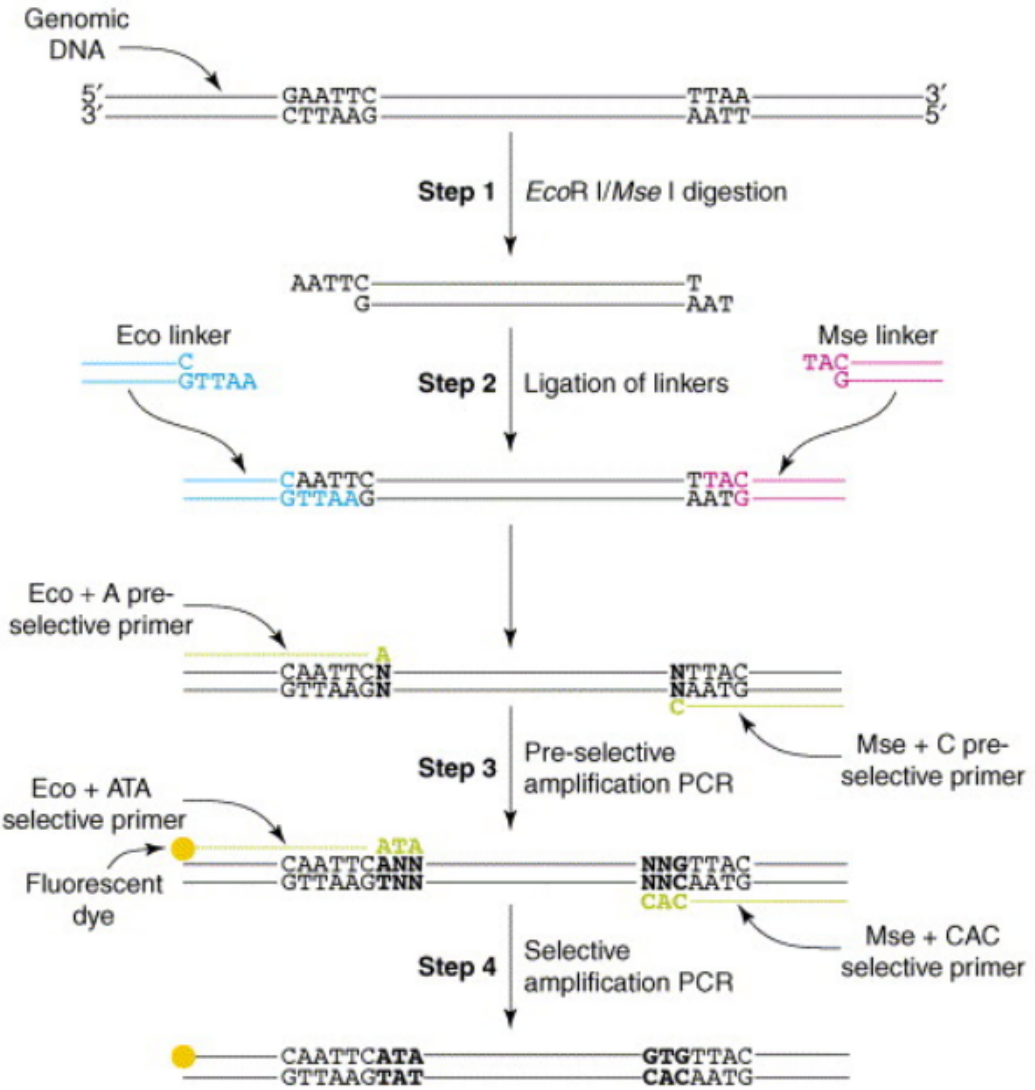

Figure 3.2 Overview of the fluorescent AFLP procedure. From Meudt & Clake (2007).

3.2.3 AFLP analysis

AFLP analysis involves the digestion of genomic DNA using two restriction enzymes followed by polymerase chain reaction (PCR)-based amplication of a subset of the resulting fragments, thus producing a reduced, 'fingerprint'-like representation of an individual's genome (Vos *et al.* 1995; Figure 3.2). Polymorphisms between individuals are created either by genetic changes creating or destroying restriction sites in the genome, or from nucleotide insertions or deletions between restriction sites. Normally, the two restriction enzymes comprise a 'rare cutter' which has a larger cutting motif,

such as *EcoR*I, and a frequent cutter with a smaller cutting motif, such as *Mse*I (Figure 3.2 Step 1). This typically creates a fragment size distribution with lots of smaller (< 600 bp) fragments that are subsequently amenable to PCR. Double-stranded DNA adapters are then ligated to the ends of the DNA fragments (Figure 3.2 Step 2). These adapters contain a sequence which serves as a binding site for PCR primers which are then used to amplify selected fragment subsets. The fragment subsets are selected by adding extra base pairs to the primers, which then only amplify those fragments having the complementary base next in their sequence; for example, only 1 in 16 fragments (assuming all bases are equally likely) will be amplified if one extra base is added to each of the two primers. This subset selection is usually done in two stages, first a 'pre-amplification' stage selects a subset using just one extra overhanging base on each primer (Figure 3.2 Step 3), then a second 'selective amplification' selects for the final subset of fragments by the addition of an extra two selective bases to each primer (Figure 3.2 Step 4). At the selective amplification stage, one of the selective primers is labelled with a 5' fluorescent dye; this ensures that any fragments amplified with this primer will be detectable using capillary electrophoresis. In theory, different numbers of selective bases can be used at each of the two amplification stages, and research has been carried out on the number of selective bases that are likely to be optimal for different genome sizes (Fay *et al.* 2005). *S. hermonthica* is diploid, with n = 19 (Aigbokhan *et al.* 1998); the genome has been estimated at 1,801 Mbp ± 321 (2C value; Yoshida *et al.*, 2010). This is 1.84 pg ± 0.33, which is within the genome size range suggested by Fay *et al.* (2005) as being appropriate for an AFLP approach using 3 selective bases for both selective amplification primers. Indeed, this '+3/+3' approach has been successfully used for work with *S. hermonthica* (Gethi *et al.* 2005; Welsh &

Mohamed 2011; Huang *et al.* 2012). Table 3.3 reports the nucleotide sequences used for AFLP marker generation.

Table 3.3. PCR primers used in the AFLP procedure.

Primer name	Primer sequence 5'-3'			5' fluorescent dye
	Core	Restriction site	Selective bases	
Pre-selective				
Eco+T	GACTGCGTACC	AATTC	T	-
Eco+A	GACTGCGTACC	AATTC	A	-
Mse+C	GATGAGTCCTGAG	TAA	C	-
Mse+G	GATGAGTCCTGAG	TAA	G	-
Selective				
Eco+AGA	GACTGCGTACC	AATTC	AGA	6-FAM
Eco+ATC	GACTGCGTACC	AATTC	ATC	NED
Eco+TGA	GACTGCGTACC	AATTC	TGA	HEX
Mse+CGT	GATGAGTCCTGAG	TAA	CGT	-
Mse+CGC	GATGAGTCCTGAG	TAA	CGC	-
Mse+CAG	GATGAGTCCTGAG	TAA	CAG	-
Mse+CAC	GATGAGTCCTGAG	TAA	CAC	-
Mse+GAC	GATGAGTCCTGAG	TAA	GAC	-
Mse+GAG	GATGAGTCCTGAG	TAA	GAG	-
Mse+GCT	GATGAGTCCTGAG	TAA	GCT	-

3.2.4 DNA restriction and AFLP adaptor ligation

For each sample, genomic DNA (approximately 300 ng) was digested at 37°C for 3 h using 1 U *Eco*RI (New England Biolabs) and 1 U *Mse*I (Roche) in 100 mM Tris-acetate pH 7.9, 100 mM magnesium acetate, 500 mM potassium acetate and 10 mM

dithiothreitol, with 0.26 μg μL^{-1} BSA. Then, 5.5 μL of an AFLP adaptor ligation mastermix (0.9× T4 DNA ligase buffer; 0.5 U T4 DNA ligase; 25 pmol *EcoR*I adaptor; 25 pmol *Mse*I adaptor) was added to each sample, and plates were incubated overnight in a water-bath at 16°C. The ligated samples were subsequently diluted by the addition of 50 μL double-distilled (dd) H_2O.

3.2.5 AFLP fragment PCR amplifications

Three sets of pre-selective amplifications were carried out: *Eco*+A/*Mse*+C; *Eco*+A/*Mse*+G; and, *Eco*+T/*Mse*+C. The PCR reactions were carried out in 10 μL reactions consisting of: 2 μL of diluted ligation mixture; 2% deionised formamide; 1× Thermoprime PCR buffer; 2 mM $MgCl_2$; 0.2 mM dNTPs (Bioline); 3 μg μL^{-1} BSA; 0.175 U Thermoprime Taq DNA polymerase; and 2.5 pmol of each pre-selective primer. Reactions were overlaid with mineral oil (Sigma) to stop evaporation. The PCR program was: 2 min at 72°C (extension); 20 cycles of: 20 s at 94°C (denaturing), 30 s at 56°C (annealing), 2 min at 72°C (extension); and a final extension of 10 min at 72°C. A subset of pre-amplified samples was run on 1% agarose gels in order to check that the previous steps had been successful; a faint smear should be observed (Figure 3.3). Although this is not a necessary part of the AFLP procedure, it confirms that the prior steps were successful before embarking on selective amplification and capillary electrophoresis of the resulting DNA fragments.

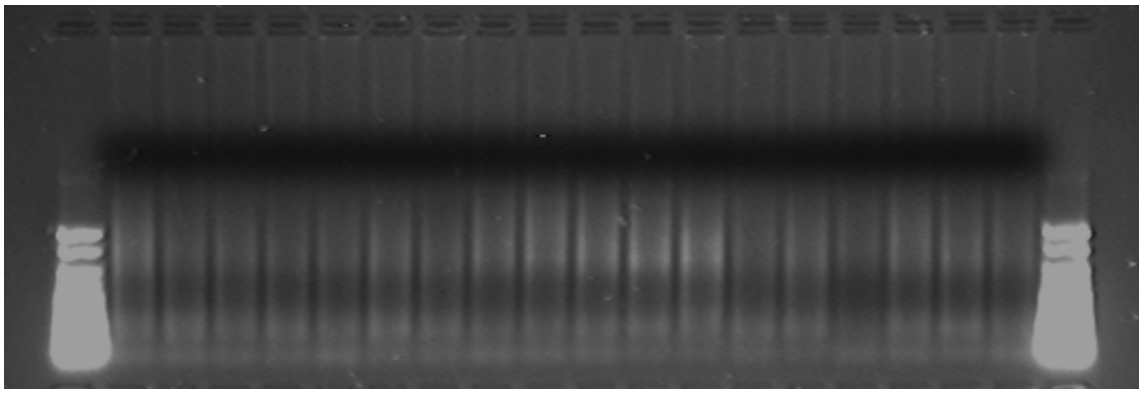

Figure 3.3 An agarose gel showing AFLP pre-amplification success. The faintness of the pre-amplified samples in this case is indicated by the over-exposure of the ladders (Hyperladder V, 500–25 base pairs).

The pre-selective amplifications were diluted by a factor of 1:4 by the addition of ddH_2O. The final selective amplifications were carried out in 10 μL reactions consisting of: 2 μL diluted pre-amplification product; 2% deionised formamide; 1× Thermoprime PCR buffer; 2 mM $MgCl_2$; 0.2 mM dNTPs (Bioline); 0.175 U Thermoprime Taq DNA polymerase; and 2.5 pmol of each selective primer. Reactions were again overlaid with mineral oil. A touch-down PCR was used to ensure high stringency of fragment amplification (Table 3.4; Bonin *et al.* 2005). Selective amplifications were finally diluted by a factor of 1:30 by the addition of ddH_2O.

Table 3.4. Touch-down PCR for AFLP selective amplification.

	Step	**Temp (ºC)**	**Time (min:sec)**	**Cycles**	**Touch-down**
1	Denature	94	2:00	1	-
2	Denature	94	0:20		
	Anneal	66	0:30	10	-1.0 ºC/cycle
	Extension	72	2:00		
3	Denature	94	0:30		-
	Anneal	56	0:30	25	
	Extension	72	3:00		
4	Adenylation	72	10:00	1	-

The selective amplification primer pairs used were: *Eco*+AGA & *Mse*+CGT; *Eco*+ATC

& *Mse*+CGC; *Eco*+TGA & *Mse*+CAG; *Eco*+TGA & *Mse*+CAC; *Eco*+TGA & *Mse*+CGT; *Eco*+AGA & *Mse*+GAG; *Eco*+ATC & *Mse*+GCT; *Eco*+AGA & *Mse*+GAC. The first five of these pairings were used by Huang *et al.* (2012) in their study of genetic variability within an East African *S. hermonthica* population. The last three were chosen as additional pairings after testing 19 *Eco*/*Mse* primer combinations on a subset of 8 samples. The new selective amplification pairings were chosen based on an assessment of the base pair range over which peaks appeared, the number of peaks, the signal-to-noise ratio and the evenness of the peaks' heights.

3.2.6 DNA fragment electrophoresis

Capillary electrophoresis of fluorescent AFLP fragments was carried out on an ABI3730 DNA sequencer. One μL of a diluted selective amplification was added to 9 μL of HiDi deionised formamide (ABI) containing 1 μL of the ROX 500 internal sizing standard. Plates were then heated at 95°C for 3 min to denature the DNA fragments (necessary for the entry of fragments into the capillaries), and were then quenched on ice for 1 min, helping to stop fragments re-annealing. The intensity of the detected fluorescence of the fragments is a function of how much sample is injected into the capillaries at the commencement of electrophoresis, if too much sample is injected background noise can interfere with the detection of the DNA fragments (Trybush *et al.* 2006). Tests of different injection periods led to a 10 s injection being selected.

3.2.7 Locus selection

The data files produced by the ABI3730 were viewed in GeneMapper v. 3.7 (Applied Biosystems, CA, USA). Samples for which the fluorescent traces were judged to have been poorly amplified (using the same criteria as detailed above for primer pair selection) were separated again by capillary electrophoresis. For samples where this did

not improve traces, the selective amplification stage PCR was performed again. AFLP loci for each primer pair were chosen manually by using GeneMapper to inspect the presence or absence of bands at specific fragment sizes (loci) across all *Striga* individuals.

3.2.8 AFLP error rate

Once the loci have been selected in GeneMapper, a *Striga* sample-by-locus matrix of peak-height intensities can be exported. However, attributing fragment presence/absence scores at loci across samples is not necessarily straightforward: some loci may show fluorescence peak-height variation which does not obviously show a clear cut-off point distinguishing fragment absence homozygotes from fragment presence heterozygotes and homozygotes. For this reason, and to guard against other sources of error such as labelling mistakes or sample cross-contamination, it is best practice to include duplicate samples in order to check the reproducibility of amplification at the selected loci (Bonin *et al.* 2004). These duplicates can be used to assess the error rate associated with any given fluorescence cut-off point that may be used to distinguish AFLP fragment presence from absence at a locus (the 'allele-calling' threshold). Here, 20 duplicated samples were used to assess the genotyping error rate. The error rate is defined as the number of mismatching genotype calls between replicates divided by the total number of loci for the 20 duplicated samples; that is, the percentage of replicated loci that do not have matching genotype calls (Bonin et al., 2004). The R script AFLPScore (Whitlock, Hipperson, Mannarelli, Butlin, *et al.* 2008) was used to assess the error rate, and to produce the resulting fragment presence/absence matrix. The published version of this program, v. 1.4 (Whitlock, Hipperson, Mannarelli, Butlin, *et al.* 2008), allowed the user to specify a range of

fluorescence thresholds, either as an absolute fluorescence value or as a percentage of the maximum peak-height at a locus, and calculated the corresponding error rates. Because of potential variation in PCR amplification between primer pairs, this meant that it was preferable to assess primer pairs separately, to allow for the fact that different primer pairs may have different scoring optima, thus producing the most accurate AFLP genotype matrix. However, this method does not allow for the fact that individual loci within primer pairs may have different scoring optima due to differential PCR amplification success. A developmental version of AFLPScore (v. 2.0) seeks to rectify this, and was used here with the permission of the author, Dr Raj Whitlock (University of Liverpool).

AFLPScore v. 2.0 operates on a similar principle to its predecessor, but, instead of just using the duplicated samples to assess the relationship between the allele-calling threshold and the error rate, an optimisation algorithm is run across all samples at a locus. The AFLPScore v. 2.0 algorithm still considers the error rate across the duplicated samples, but also considers the distribution of all peak-heights at a locus and the proportion of peaks included as the fragment present genotype. These three variables are combined into a cost function, which, when minimised, has been found to provide sensible results for test datasets (Dr R. Whitlock, pers. comm.) Graphs of the value of the cost function compared to the error rate are assessed by eye to choose a cost function value that trades-off the error rate across all loci with the number of loci included in the final dataset. AFLPScore also performs a peak-height normalisation (to the median height of a sample's fingerprint) to ensure that variation between samples in total AFLP fingerprint intensity does not bias genotype calls. This is an important part of AFLP scoring regardless of the scoring method used, and other scoring programs

normally incorporate a normalisation step, or require normalised data. The final genotyping error rate was 4.65%, which is around the level typically reported for AFLP studies (Bonin *et al.* 2004). The final numbers of AFLP loci analysed for each primer pair are reported in Table 3.5.

Table 3.5. Primer pairs used for AFLP analysis and the final number of loci used in genetic analyses.

Selective primer pair	Number of loci
Eco+AGA and *Mse*+CGT	217
Eco+AGA and *Mse*+GAC	210
Eco+AGA and *Mse*+GAG	166
Eco+ATC and *Mse*+CGC	146
Eco+ATC and *Mse*+GCT	70
Eco+TGA and *Mse*+CAC	94
Eco+TGA and *Mse*+CGT	203
Eco+TGA and *Mse*+CAG	169
	Total: 1275

3.2.9 Analyses of genetic variation and differentiation

AFLP-surv v. 1.0 (Vekemans *et al.* 2002) was used to calculate the percentage of loci that were polymorphic (PLP) at the 5% level; this is an estimate of polymorphism that only counts a locus as polymorphic if the least frequent AFLP genotype exceeds the 5% level. This is designed to screen out polymorphic loci caused by genotyping error; for example, low-frequency fragments may appear occasionally due to PCR mis-amplification, or, conversely, high-frequency fragments may drop-out due to PCR failure. Assessing PLP at the 5% level should also reduce any correlation between PLP and sample size (rare alleles are more likely to appear in larger samples; Avise 2004).

AFLP-surv was also used to calculate the average expected heterozygosity (H_{exp}) of the total *Striga* population sample, as well as for each of the 9 host-associated sub-

populations. H_{exp} is a measure that requires the estimation of allele frequencies from the AFLP fragment presence/absence data. Recalling that the AFLP fragment presence band is assumed to represent both fragment present homozygotes and heterozygotes, this means that the frequency of the fragment absence allele (q) can be calculated by taking the square root of the number of fragment absence homozgotes. In practice this has been found to result in biased estimates of *q (Lynch & Milligan 1994)*, and appropriate adjustments have been formulated (Lynch & Milligan 1994; Zhivotovsky 1999). The AFLP-surv implementation of Zhivotovsky's (1999) Bayesian method was used to estimate the population and sub-population allele frequencies, and the method of Lynch & Milligan (1994) was used to calculate H_{exp}.

Wright's (1951) measure of genetic differentiation, F_{ST}, is a standard measure of population structure. F_{ST} can be interpreted as 'the proportion of genetic variation distributed among (as opposed to within) [demes of a] subdivided population' (Avise 2004). One often used formulation (due to Nei 1973) is: $F_{ST} = (H_T–H_S)/H_T$, where H_S is the mean expected heterozygosity within sub-populations under Hardy-Weinberg equilibrium, and H_T is the overall expected heterozygosity across the total population, also assuming Hardy-Weinberg equilibrium (Avise 2004). An F_{ST} of 1 would indicate that sub-populations were fixed for different alleles, whilst 0 would mean that sub-populations were identical (Avise 2004). AFLP-surv v. 1.0 (Vekemans *et al.* 2002) implements a permutation routine, where individuals are randomly permuted between sub-populations, to test whether the observed value of F_{ST} is significantly different from zero. After simulating a given number of datasets (here 1000), the observed F_{ST} value is compared to the 95th and 99th percentiles of the distribution of the simulated F_{ST} values to assess its significance. F_{ST} among sub-population pairs ('pairwise' F_{ST}) was also

calculated in AFLP-surv.

A non-metric multidimensional-scaling (NMDS) ordination was also used to visualise the relationships between *Striga* individuals and sub-populations. NMDS is used to ordinate objects for which (dis)similarity indices have been calculated based on the presence or absence of descriptors, e.g., ordinating sites using dissimilarities calculated between sites based on the presence or absence of species. NMDS uses a ranked distance method that seeks to represent the distances between objects in the dissimilarity matrix in (usually) 2 dimensions. Here, a Jaccard dissimilarity matrix was calculated from AFLP fragment presence/absence (Bonin *et al.* 2007) in order to ordinate *Striga* individuals in genetic space. This approach is a useful complement to measures, such as F_{ST}, that require the extra step of allele-frequency estimation. A permutation-based multivariate ANOVA (PerMANOVA; 999 permutations) was used to assess the amount of genetic variation between *Striga* sub-populations explained by sorghum host identity. These multivariate analyses were implemented in R using the 'vegan' package v. 2.0-5 (Oksanen *et al.* 2011).

3.2.10 Outlier tests for differentiated loci

Outlier tests, or genome scans, have rapidly increased in popularity over the past few years (Bierne *et al.* 2011). They aim to identify individual loci that are significantly differentiated between populations. One of the first proposals that loci subject to selection could be differentiated from 'neutral' loci was that of Lewontin & Krakauer (1973). They postulated that whilst demographic processes (e.g. drift, gene flow etc.) should have random effects on neutral loci across the genome, selected loci should display a recognisable pattern differentiating them from neutral loci. Specifically, those loci that have been recently selected for, or are under current selection, should have an

F_{ST} value that significantly exceeds the genome-wide average for neutral polymorphisms (Storz 2005), although the decision on what constitutes a 'significant' departure from neutrality is not uncomplicated (Butlin 2010; Strasburg *et al.* 2012). As neutral genetic variation can 'hitch-hike' with loci that are directly under selection, such closely linked neutral markers can show differentiation (i.e. be identified as 'outlier' loci) that suggests linkage to genomic sites under selection. At the time, the Lewontin & Krakauer (1973) approach was quickly abandoned after it was shown by several critics that the assumptions which it made were likely to be violated in natural populations (Beaumont 2005). However, their ideas have recently been resurrected in the form of 'population genomics' (Luikart *et al.* 2003) and are proving of considerable value in the study of adaptation and speciation (Storz 2005; Butlin 2010; Strasburg *et al.* 2012). A variety of simulation approaches now suggest that the ideas of Lewontin & Krakauer (1973), and more modern methods based upon similar reasoning, are relatively robust to violations of the assumptions underlying them (Beaumont 2005; Butlin 2010; but see Hermisson 2009). A potential strength of the population genomic approach is that it is not necessary for a researcher to have any *a priori* knowledge of what an adaptive phenotype might be within a target species (Storz 2005); however, the assumption that an outlier locus is the result of extrinsic selective forces should be carefully evaluated in each situation (Bierne *et al.* 2011).

In this study the program BayeScan v. 1.0 (Foll & Gaggiotti 2008) was used to assess AFLP loci for signatures of selection. BayeScan estimates the likelihood of selection at any given locus by using a logistic regression to decompose overall F_{ST} coefficients into a population-specific component (beta) shared by all loci and a locus-specific component (alpha) shared by all populations (Foll 2012). The presence of a locus-

specific component suggests either that there is diversifying selection (alpha is greater than zero) or balancing or purifying selection (alpha is less than zero). The posterior probabilities of the model with the locus-specific component (that is, alpha significantly different from zero) and the model without this component are compared in order to produce a Bayesian measure of the strength of evidence for accepting that alpha is not zero for any given locus (Foll 2012).

3.2.11 Tests of correlation between field virulence and sub-population allele frequencies

BayeScan can also be requested to output sub-population allele frequencies (Foll 2012) estimated using the method of Zhivotovsky (1999). Given the hypothesis that host selection is the underlying cause of any locus-specific differentiation between host-selected sub-populations, a complementary approach to the identification of differentiated loci by BayeScan is to seek to correlate sub-population allele frequencies with some indicator of host selection pressure (Bonin *et al.* 2007). Field data on average *Striga* emergence were available for Burkina Faso in 2009, the year in which the samples were harvested (provided by Dr T. van Mourik), therefore I investigated whether there were any significant correlations between *Striga* field emergence and allele frequency across the 9 *Striga* sub-populations. This was performed by writing an R program that, for every AFLP locus, permuted the estimated allele frequencies between sub-populations 1000 times, and, for each permutation, calculated Pearson's *r* between the allele frequencies and the average number of emerged *Striga* across sub-populations. The observed correlations were compared to the 2.5^{th} and 97.5^{th} percentiles of this null distribution to assess significance (the quantiles were calculated using the R function 'quantile', which uses interpolation to estimate the continuous distribution of the sample statistic, and is approximately unbiased regardless of the underlying

distribution of the statistic) (Frohne & Hyndman 2012). Because of the number of loci being tested (1275), the Benjamini-Hochberg false-discovery rate (FDR) correction (Benjamini *et al.* 2001) was implemented at the 5% level in a spreadsheet program (LibreOffice Calc v. 3.5) to reduce the number of false positives, whilst also controlling the false negative rate.

3.2.12 Population structure simulations for a differentiated virulence locus

BayeScan produces estimates of locus-specific F_{ST}. For loci which are determined to be under selection by BayeScan, we can inspect the magnitude of F_{ST} at these loci compared to those loci which are estimated to be neutral. If hosts are driving the selection of outlier loci, simulations may be useful for investigating if the observed F_{ST} at outlier loci is what would be expected if a simple hypothesis of differential host selection were true. For this purpose I used the statistical software R (R Core Team 2012) to simulate selection at a single AFLP locus involved in differential *Striga* virulence on hosts of differing susceptibilities. The sorghum hosts were divided into two groups (Table 3.2), resistant (Malisor 92-1; SRN39; Brhan; and Mota Galmi) and susceptible (CSM388; Sariaso9; Lina3; CEF 322/35-1-2; and Wassa). The mean (± s.e.) *Striga* field emergence for the resistant group was 23.3 ± 3.4, for the susceptible group it was 110.0 ± 8.7.

The first step in the simulation used a Bernoulli distribution, with the probability determined by the underlying fragment frequency, to produce the distribution for an AFLP fragment across 360 individuals in the *Striga* population (40 *Striga* × 9 sorghum hosts). A second round of selection was then implemented within the resistant sub-populations. In these sub-populations the fragment was assumed to be associated with successful parasitism, therefore plants with the fragment survived. For those plants that

did not receive the fragment, the probability of survival was 0.5. If the plant was lost, it was replaced by a random draw from the global pool of individuals where fragment presence was at the global underlying frequency for that simulation. Neither the fragment presence or absence states were assumed to have any effect on plants parasitising susceptible sorghum hosts (i.e. there was no cost to having the virulence fragment on the susceptible host). For these sub-populations the distribution of the fragment was simply determined by random picks from the underlying global fragment frequency distribution. The simulation was performed 5 times for each initial frequency of the AFLP fragment ranging from 0.1–1.0 in increments of 0.1. The resulting data were formatted in R for input to AFLP-surv using code from AFLPDAT (Ehrich 2006). AFLP-surv v. 1.0 (Vekemans *et al.* 2002) was used to estimate F_{ST} at the locus using the Bayesian method of Zhivotovsky (1999); 1000 permutations were used to assess the significance of these estimates. For each fragment frequency the mean F_{ST} of the 5 simulation runs is presented. The resulting F_{ST} values provide a point of comparison for the estimates produced by BayeScan for outlier loci.

3.3 Results

3.3.1 Genetic diversity within the field-collected Burkina Faso S. hermonthica

Ten AFLP primer pairs produced 1275 scored loci, with fragment sizes between 50 and 500 base pairs, across 356 *S. hermonthica* plants harvested from 9 sorghum cultivars. The AFLP analysis was restricted to plants harvested in Burkina Faso, due to problems with sample harvesting and DNA quality at the two Mali sites (see above). Data on population genetic variation, analysed in terms of the percentage of polymorphic loci (PLP) and average expected heterozygosities (H_{exp}), are presented in Table 3.6 for host-

associated *Striga* sub-populations and the overall *Striga* population sample.

Table 3.6. Measures of genetic variation for sorghum host-associated *S. hermonthica* from Burkina Faso, estimated from AFLP data. PLP = percentage of polymorphic loci; H_{exp} = average expected heterozygosity.

***S. hermonthica* host**	**Number of samples**	**PLP at 5%**	**H_{exp} ± s.e.**
Malisor 92-1	38	29.7	0.097 ± 0.004
SRN39	40	36.3	0.106 ± 0.004
Brhan	40	37.5	0.114 ± 0.004
CSM388	40	37.3	0.116 ± 0.004
Sariaso9	40	38.3	0.114 ± 0.004
Lina3	39	35.5	0.116 ± 0.004
CEF 322/35-1-2	39	42.5	0.126 ± 0.004
Mota Galmi	40	40.0	0.118 ± 0.004
Wassa	40	36.7	0.106 ± 0.004
Total	356	37.1	0.113 ± 0.003

3.3.2 Genetic differentiation between host-associated sub-populations of S. hermonthica

Genetic differentiation (F_{ST}) between host-associated sub-populations was estimated at 0.004. The 0.5th and 99.5th percentiles of the null distribution of F_{ST}, generated in AFLP-surv by permuting individuals between sub-populations 1000 times, were -0.0027 and 0.0020. This indicates that the observed value of F_{ST} was significantly different from zero at the 1% confidence level. The pairwise F_{ST} estimates between sub-populations were variable, with 9 of the 36 estimates at 0.000 (Table 3.7). The highest estimate was between the CSM388- and Malisor 92-1-associated sub-populations (pairwise F_{ST} = 0.011). There was no clear indication that sub-populations were more closely related to each other within the categories of field resistance estimated at Kouare (Table 3.2).

Table 3.7. Pairwise F_{ST} estimates between all sorghum cultivar-associated sub-populations of *S. hermonthica.* (Sorghum cultivar names are given here in abbreviated form.)

Cultivar	Malisor	SRN	Brhan	CSM	Sariaso	Lina	CEF	M.G.	Wassa
Malisor	-	-	-	-	-	-	-	-	-
SRN	0.000	-	-	-	-	-	-	-	-
Brhan	0.007	0.006	-	-	-	-	-	-	-
CSM	0.011	0.009	0.000	-	-	-	-	-	-
Sariaso	0.002	0.001	0.000	0.002	-	-	-	-	-
Lina	0.010	0.008	0.000	0.000	0.002	-	-	-	-
CEF	0.010	0.007	0.000	0.000	0.000	0.003	-	-	-
M.G.	0.009	0.008	0.002	0.002	0.001	0.007	0.000	-	-
Wassa	0.007	0.005	0.008	0.009	0.003	0.016	0.005	0.003	-

A NMDS ordination was also used to visualise the relationships between *Striga* individuals collected from the different sorghum hosts (Fig. 3.5). The stress measure of this ordination was 0.20; values of less than 0.30 are considered acceptable (Kent 2012). Stress is here a measure of the difference between the actual multivariate distances between objects and the distances as represented in 2 dimensions (Kent 2012). The NMDS ordination also suggested very little genetic differentiation between the host-associated *Striga* sub-populations; indeed, the PerMANOVA found that host identity, although statistically significant, explained less than 1% of the genetic variation in the population ($r^2 = 0.007$, $p = 0.001$).

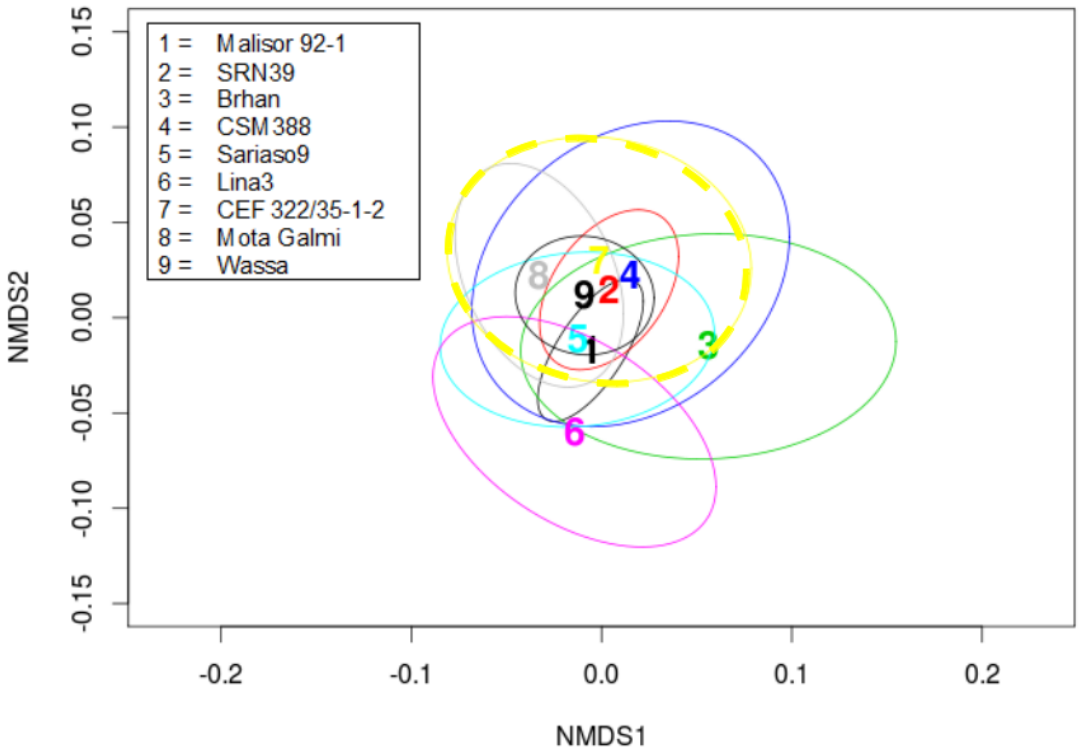

Figure 3.5. Non-metric multidimensional scaling (NMDS) ordination showing the effects of the 9 sorghum hosts on *S. hermonthica* sub-populations. For clarity, the 95% confidence ellipses of each host-associated sub-population, but not the individual *Striga* plants, have been shown. The number of each sorghum host is at the centre of its 95% confidence ellipse, and numbers and ellipses are shown in the same colour. (NMDS stress = 0.20.)

3.3.3 Putative outlier loci between host-associated sub-populations of S. hermonthica

The BayeScan outlier analysis identified 14 loci for which the evidence for the inclusion of locus-specific selection was deemed at least 'strong' on the Jeffreys scale of evidence for Bayes factors (Foll, 2012; Table 3.8). The Bayes factor provides an estimate of the evidence for the choice of one model over another; here, the choice was between a model including a locus-specific selection effect and one without it. Jeffreys scale of evidence provides a simple interpretation of Bayes factors: 'strong' evidence, in this case, indicates that P(alpha $\neq$ 0) is at least 0.91 (where alpha is the locus-specific selection effect; Foll 2012).

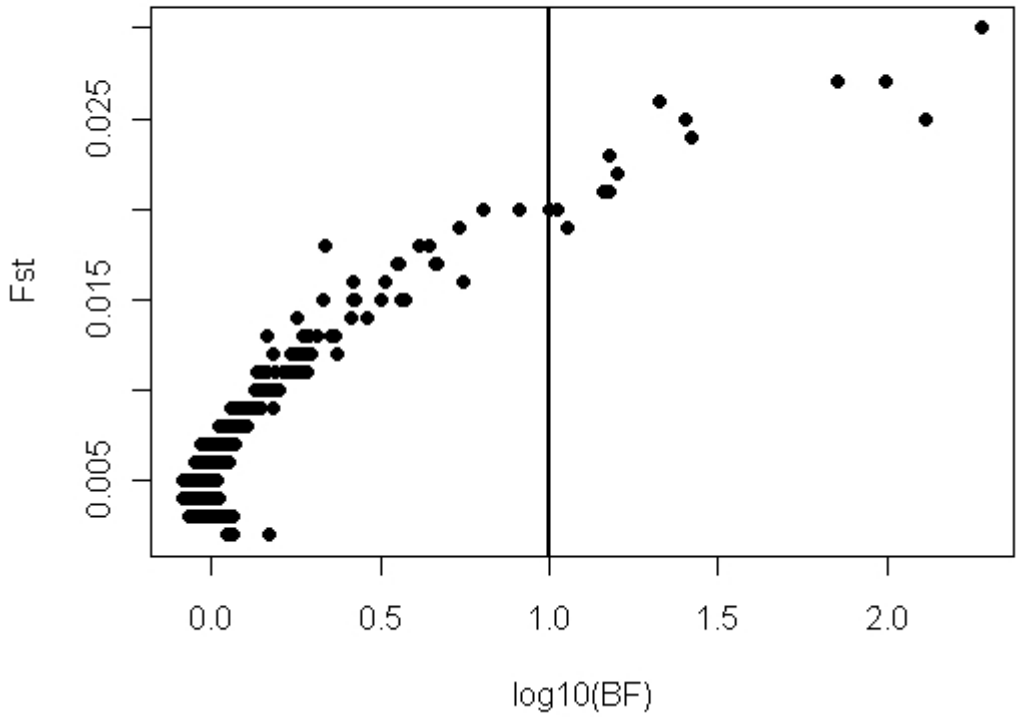

Figure 3.6. Locus-specific F_{ST} plotted against the logarithm of the Bayes factor for each locus. The bold vertical line at log(BF) = 1 designates the cut-off point for which all loci to the right-hand side of the line have P(alpha ≠ 0) > 0.91, where alpha is the locus-specific selection effect.

Table 3.8. Numbers of AFLP loci and mean locus-specific F_{ST} for the different strength of evidence categories used by BayeScan to estimate the probability that a locus-specific selection effect does not equal zero.

P(alpha ≠ 0)	**Jeffreys scale of evidence**	**Number of AFLP loci**	**Mean locus-specific F_{ST}**
0.91–0.97	Strong	10	0.022
0.97–0.99	Very strong	2	0.027
0.99–1.00	Decisive	2	0.029

3.3.4 Test of association between sub-population allele frequencies and Striga *emergence*

No loci were found to have significant correlations with average *Striga* field emergence at the 5% level after using the Benjamini-Hochberg FDR adjustment. Without the multiple test adjustment, 53 loci were significant at the 5% level, and 15 at the 1% level. However, none of these loci was among those selected by BayeScan as outliers.

3.3.5 Simulation of locus-specific genetic differentiation under a model of host selection

Simulations of host selection at a virulence locus with different intial AFLP fragment frequencies indicated that population differentiation (as measured by F_{ST}) became significant when the virulence fragment was at an underlying frequency of 0.2 in the total population (Table 3.9). Initial AFLP fragment frequencies of 0.4–0.5 resulted in F_{ST} values around the size estimated by BayeScan for outlying loci (~ 0.03; Table 3.8).

Table 3.9. Mean F_{ST} values for 5 simulated experiments with varied initial global frequencies of a virulence fragment of differential importance between resistant (fragment favoured for parasite growth) and susceptible (fragment has no effect on parasite growth) host plants. Significance is given as the proportion of the individual simulations that were significant at the given level.

Initial fragment frequency	Mean F_{ST}	Significant at 5%	Significant at 1%
0.1	0.003	0/5	0/5
0.2	0.008	2/5	2/5
0.3	0.024	3/5	2/5
0.4	0.024	3/5	3/5
0.5	0.048	4/5	3/5
0.6	0.037	3/5	2/5
0.7	0.054	4/5	3/5
0.8	0.071	4/5	4/5
0.9	0.039	4/5	1/5
1.0	-0.005	0/5	0/5

3.4 Discussion

Although a number of studies have now investigated the population genetics of *S. hermonthica* (Table 3.1), this is the first detailed study of within-population genetic variation as it relates to differential host selection under field conditions. I have investigated the genetic component of host range within a population of *S. hermonthica*, and assessed whether loci with outlying values of F_{ST} correlate with simple measures of host resistance in the field, and whether simulations based on simple models of pathogen virulence produce locus-specific values of F_{ST} of a similar magnitude to observed outliers.

3.4.1 Genetic diversity within a S. hermonthica *population from Burkina Faso*

My results provide more evidence of within-population variability for *S. hermonthica.* The Burkina Faso population sampled had an average expected heterozygosity (H_{exp}) of 0.113 and a percentage of polymorphic loci (PLP) of 37.1% at the 5% level. These figures are lower than the findings of other comparable investigations using AFLP (Welsh & Mohamed 2011: mean (among-population) H_{exp} = 0.204; mean (among-population) absolute PLP = 60.9%; Huang *et al.*, 2012: (within-population) H_{exp} = 0.234; (within-population) PLP at 5% = 72.8%). Comparisons are only made with other studies utilising AFLP, because the molecular marker system used can be a strong determinant of results (Nybom 2004). Hamrick & Godt (1996), in a review of 98 population genetic studies using allozymes, reported similar values for annual, outcrossing species (mean among-study H_{exp} = 0.186; mean among-study absolute PLP = 59.1%). Nybom (2004), investigating the effect of molecular marker system on measures of genetic variation, found that estimates of H_{exp} from dominant RAPD data were lower than for microsatellites, and gave mean RAPD-derived H_{exp} values of 0.13

for annual species and 0.27 for outcrossers.

Compared to these informal meta-analytical studies (Hamrick & Godt 1996; Nybom 2004), the slightly lower value of H_{exp} in the current study may simply be due to variation between study species, the number of populations of a species investigated in any one primary study, and the loss of information on among-population variance in H_{exp} within any particular species. Specifically focusing on *S. hermonthica*, the lower variation found in this chapter could be an artefact of the artificial planting-hole infestation used in the field-trials to help promote even parasite pressure across the field experiment: if the harvested seed came from a small field, inter-breeding between close relatives might have depressed the variation in the *S. hermonthica* sample used for infestation. Alternatively, the discrepancy may arise from technical issues associated with the AFLP methodology. The number of loci per primer pair scored in the current study is at the high end reported for AFLP studies (Caballero *et al.* 2008); this is likely to mean that more markers with low polymorphism have been scored. In addition, larger samples of individuals (356 here, compared to 99 in Huang *et al.* 2012 and 107 in Welsh & Mohamed 2011) will detect more low-frequency loci in a population. Given the error rate reported here (4.65%), it is assumed that the scored loci are reliable; it should be noted however that low-frequency loci are by definition less likely to appear in samples replicated for error rate analyses, therefore, it is possible that some loci with a very low occurrence of the presence or absence fragment may be erroneous without inflating the reported error rate. If the resources are available, it is recommended that AFLP scoring should be repeated independently by another researcher (Bonin *et al.* 2004), or software that seeks to make the locus selection step, as well as the allele-calling step, more repeatable could be used (Arrigo *et al.* 2009). However, whilst these strategies may

mean that a smaller number of markers are scored, which is likely to reduce the variation resulting from different scoring decisions between observers (i.e. increasing the precision of estimates that could be made from a single data-set), it does not guarantee increased accuracy: the 'true' underlying state of the system remains unknown. Indeed, different methods of AFLP data analysis have been shown to affect population genetic parameters and the error rate (Herrmann *et al.* 2010), and this should always be borne in mind when assessing statistics derived from AFLP data, especially where studies make inferences based on the distribution of alleles at individual loci (Herrmann *et al.* 2013).

3.4.2 Genetic differentiation between host-selected sub-populations

Overall, F_{ST} across the sorghum host-selected sub-populations was lower (0.004), than that reported by Huang *et al.* (2012) for rice host-selected sub-populations (0.013). These low values may correspond to a situation in which only a low percentage of loci are determining host adaptation; F_{ST} can be interpreted as the percentage of variation accounted for by a between-population component, and in that sense these figures agree with the PerMANOVA in attributing less than 1% of genetic variation to a between-population source (host adaptation in this case). However, it also possible that host-adaptation is dependent on many loci, each with a small effect. In this situation the changes in allele frequencies between host-adapted sub-populations would also be low, again producing low values of F_{ST}.

Compared to Huang *et al.* (2012), the lower value of F_{ST} found here may also be due to some of the technical issues described above that may have lowered H_{exp} and PLP (section 3.4.1). Pairwise F_{ST} values did not reveal any evidence for the clustering of resistant hosts, contrary to the situation found for the three rice cultivars of Huang *et al.*

(2012). *Striga* from Malisor and SRN39, two of the most resistant cultivars, showed no differentiation (pairwise F_{ST} = 0.000), although this was also true for the resistant Brhan when compared to several of the susceptible cultivars. The lack of clear differentiation between the *Striga* harvested from resistant and susceptible hosts may be due to variation in the underlying resistance mechanisms between resistant cultivars (e.g. pre-attachment versus post-attachment; Chapter 2), or because the *Striga* plants are subject to other sources of selection, e.g. unmeasured environmental variation was found to affect both sorghum hosts and *Striga* emergence in the field trials analysed in Chapter 2, or strong stochastic effects. These situations would essentially mean that interactions between species' genomes would not be the only factors determining the growth of individual *S. hermonthica* plants on resistant hosts, meaning that the detection of loci with small effects is likely to be less effective than in the controlled environment lab study of Huang *et al.* (2012).

3.4.3 Outlier analyses and putative host selection on loci

Even with low differentiation between host-selected sub-populations, it may still be the case that individual loci are responding to host identity. Outlier analyses were conducted in order to identify loci displaying a signature of selection. The BayeScan outlier analysis is relatively conservative in that it allows for a background level of differentiation within the population under investigation (Foll & Gaggiotti 2008; Huang *et al.* 2012). Fourteen loci were identified by BayeScan v. 1.0 as potentially responding to host selection. The two loci with the strongest level of support for an effect of selection had an average locus-specific F_{ST} of 0.029. Simulation models of a single locus at which host adaptation was occurring in a common response to resistant hosts, but not susceptible ones, indicated that a locus-specific F_{ST} value around this level could

be produced by an AFLP fragment with a global frequency of 0.4 (indicating an underlying allele frequency of $p = 0.33$). Whilst a single gene was found to underly the virulence of a population of the autogamous *S. gesnerioides* on cowpea (Li & Timko 2009), and point mutations can often have large effects in plant-pathogen interactions (Sacristán & García-Arenal 2008), it is far from certain that a locus in *S. hermonthica* would respond to selection in this simple fashion between resistant and susceptible hosts. The simulation, however, does indicate that alleles potentially underlying differential host responses, such as reported here and in Huang *et al.* (2012), may exist at intermediate frequencies in populations of *S. hermonthica*.

The attempt to correlate underlying population allele frequencies with average *Striga* field emergence did not yield any clear result. This may again be due to environmental variation in field measures of virulence (Chapter 2), or to the fact that the global allele frequencies used in this analysis were estimated, and are likely to be subject to some degree of uncertainty. Alternatively, virulence may be the result of interactions between many loci of small effect, reducing correlations between any one locus and virulence; another possibility is that different loci may be responsible for virulence on different hosts, especially if resistance mechanisms vary between cultivars (Yoder & Scholes 2010).

Perhaps even more than for the measures of genetic variation and differentiation discussed above, errors in AFLP bin selection and allele calling are likely to affect the outcomes of outlier analyses, because accurate conclusions depend entirely on inferences based on the distribution of alleles at single loci. Again, ideally, analyses should be subject to checking by re-scoring raw data, especially if an investment of effort in the sequencing of outlier AFLP loci is intended. A similar study of the

population genetics of host adaptation in *S. hermonthica* starting today would almost certainly utilise sequenced loci, such as single-nucleotide polymorphisms located from restriction enzyme-associated next-generation sequencing (Baird *et al.* 2008). Whilst such systems are not error free (Davey *et al.* 2011), they do allow the genotyping of thousands of loci, which can be checked against increasing amounts of genomic data from related species, allowing a stronger probability of finding links between signatures of selection and predictions of molecular function. However, such approaches still have to contend with the problem of distinguishing intrinsic from extrinsic selection if the aim of a study is to ultimately understand the actual role of selected loci in adaptation (Bierne *et al.* 2011).

3.4.4 Conclusions

This chapter has demonstrated that variation within *S. hermonthica* populations in the field can lead to the host selection of loci, as has been demonstrated for lab experiments (Koyama 2000a; Huang *et al.* 2012). This supports the idea that virulence alleles may accumulate in populations in response to host selection, resulting, locally, in the adaptation of *Striga* to host genotypes, and the gradual erosion of host resistance. Further, well-replicated studies would help to shed light on the question of whether outlier loci are the same over time or across the landscape. The use of sequenced loci should make these questions more amenable to accurate genetic analyses in the future.

Chapter 4. Epistasis, segregation distortion, and maternal identity are all linked to virulence in the parasitic plant *Striga hermonthica*

4.1 Introduction

Ecotypic differentiation is the phenomenon whereby environmental heterogeneity results in genetically differentiated populations within a species, the hypothesis usually being that this has been caused by the adaptation of populations to their local environment (Linhart & Grant 1996; Hufford & Mazer 2003). Additionally, if environmental variability is high at the scale of the population, we may also predict the presence of (pre-)adaptive genetic variation (Byers 2005). The evolution of phenotypic plasticity is also a possibility, and whether evolution leads to genetic polymorphism or plasticity may depend on the 'grain' of environmental variation relative to generation time (Meyers & Bull 2002; Wennersten & Forsman 2012). Between-population variation in parasite adaptation to different hosts can be thought of in terms of a population-level genotype-by-genotype (G × G) interaction, just as ecotypic differentiation could be considered the result of a population-level genotype-by-environment (G × E) interaction, with the mean fitness of populations varying between locations. G × G host-parasite interactions may be due to local (mal)adaptation (Koskela *et al.* 2000; Gandon 2002), but this may not always be the most appropriate conceptual approach. For example, obligate parasitic weeds of agroecosystems, such as *Striga hermonthica*, may regularly encounter different host genotypes, with the rotation of crop-hosts by the farmer potentially working against the evolution of local adaptation of the parasite to a single host population. This may be especially true if a parasite species is an obligate outcrosser (Hufford & Mazer 2003; Gandon & Nuismer 2009), like *S.*

hermonthica (Safa *et al.* 1984). Annual changes of crop-host could be considered an example of a fine-grained environment (i.e. an enviroment where the scale of heterogeneity is similar to the generation time; Rodríguez 2012), whereas if a farmer grew the same crop for many seasons, this could be considered a coarse-grained environment. At the individual genotype level, Rodríguez (2012) found some evidence for G × E interactions being stronger in heterogeneous environments with a coarse-grain, and predicted that this was due to the increased time available for adaptation to a particular environmental state. However, even if environmental heterogeneity is fine-grained, populations of allogamous parasites may still be genetically differentiated due to isolation-by-distance, genetic drift, population history, differential selection by different farming regimes (including particular sequences of crop-hosts), or other biotic or physical habitat differences. This can lead to genetic variation within and between populations, potentially resulting in pre-adaptation to unencountered host genotypes (Wennersten & Forsman 2012). Population-level host-parasite G × G interactions may not be unusual in such a situation (Lambrechts *et al.* 2006; Wolinska & King 2009) because genetically diverse populations are more likely to contain pre-adapted genotypes that can take advantage of changes in the environment to differing degrees (Hughes *et al.* 2008; Wennersten & Forsman 2012).

For *S. hermonthica*, successful parasitism of its host has been shown to depend, at least partially (Chapter 2), on a population-level G × G interaction: different *Striga* populations may have different responses to a set of host genotypes, which could be different crop species, or cultivars of that species (Kim *et al.* 1994; Freitag *et al.* 1996; Cissoko *et al.* 2011; Huang *et al.* 2012). This indicates the presence of genetic variation for virulence between *Striga* populations, and has important consequences for the

resilience of resistant hosts. The call for a better understanding of the relationship between variability in *Striga* virulence and the resistance of different host genotypes has been repeated regularly in recent times (Vasudeva Rao & Musselman 1987; Parker & Riches 1993; Haussmann, Hess, Welz, *et al.* 2000; Mohamed *et al.* 2007; Scholes & Press 2008). However, none of these authors have drawn upon the extensive literature on plant or parasite population differentiation and adaptation to speculate on the possible genetic bases of population-level G × G interactions between *S. hermonthica* and its hosts, or to suggest that investigations of this topic are broadened to incorporate insights or techniques used in these areas of research.

If assessed in the field, perceived population-level G × G interactions between *Striga* populations and their hosts may be partially due to environmental factors, such as infestation differences (seed bank density and distribution), or soil and other environmental factors which affect the resistance of the host (Vasudeva Rao & Musselman 1987). Host resistance may also depend on phenotypes which work by avoiding parasite seed, for example, by root architecture variation, or by having low levels of root-exuded germination stimulants (Haussmann *et al.* 2004), meaning that there may be several gene-based resistance mechanisms at work. Understanding these various factors is likely to aid the development and deployment of resistant crop-hosts. However, here I focus on examining post-attachment resistance to *Striga* under controlled conditions. This has been found to be a reliable way of evaluating what part of the population-level host specificity of *S. hermonthica* has a genetic basis (Lane *et al.* 1993; Huang *et al.* 2012), and allows the attribution of this specificity to a particular, well-studied stage (Yoder & Scholes 2010) of the host-parasite interaction (cf. Chapter 2).

Studies of genetic variation in *S. hermonthica* across Africa (Chapter 3; Bharathalakshmi *et al.* 1990; Kuiper *et al.* 1996; Olivier *et al.* 1998; Koyama 2000b; Gethi *et al.* 2005; Ali *et al.* 2009; Yoshida *et al.* 2010; Welsh & Mohamed 2011) together suggest that geographic distance, not host genotype, is the main factor structuring genetic variation; however, as expected for an outcrossing plant (McDonald & Linde 2002), variation within populations can be considerable (Welsh & Mohamed 2011), and a host selection pressure for particular parasite genotypes has been demonstrated (Koyama 2000a; Huang *et al.* 2012). This suggests that crop-host resistance is likely to be highly vulnerable to parasite virulence evolution, and indeed, this has apparently been the case (Yoder *et al.* 2007; Parker 2009). The potential for many different population-level interactions between parasite genepools and host genotypes clearly exists (Huang *et al.* 2012), irrespective of the failure of the above studies to find strong evidence for host-determined genetic structure among populations.

Population-level virulence can be thought of as the chance that a randomly selected individual from a particular population has of parasitising a particular host. The pool of hosts that a *S. hermonthica* population can parasitise may be very broad: for example, the following studies all used the same *S. hermonthica* seed collection: Gurney *et al.* (2003): infecting maize and the maize wild relative *Tripsacum dactyloides*; Gurney *et al.* (2006): infecting rice cultivars; Huang *et al.* (2012): infecting rice cultivars; this study: infecting sorghum cultivars. Even if a host is relatively resistant to a population of *S. hermonthica*, there are always likely to be individual parasites capable of overcoming host resistance: even the most resistant sorghum genotypes in this study will support some parasites. Indeed, a host genotype that is completely immune to *S. hermonthica* has not yet been observed. All these observations support the contention

that genetically diverse populations of *S. hermonthica* are likely to be pre-adapted to a range of different host genotypes, with the 'host range potential' of these adaptations potentially differing between genepools. The fact that different populations of *S. hermonthica* can exhibit G × G interactions against sets of host genotypes offers an opportunity to investigate the genetic architecture of virulence through experimental crossing and linkage mapping. Experimental crossing between differently adapted (or pre-adapted) populations can be used to reveal the genetic architecture of adaptive traits (Lynch 1991; Hufford & Mazer 2003; Dybdahl *et al.* 2008; Crémieux *et al.* 2010).

Trait genetic architectures are important for any discussion of differentiation between populations, and are a part of the raw material of evolution (Wade & Goodnight 1998; Thompson 1999). Heterosis (hybrid vigour) and epistasis (interaction between loci) may be important characteristics of the underlying genetic architecture of any particular trait (Falconer & Mackay 1996); epistasis may be particularly important for the evolution of incompatibilities between populations, potentially leading to speciation (Whitlock *et al.* 1995; Wade & Goodnight 1998; Demuth & Wade 2005). A number of studies have investigated the presence of heterotic and epistatic components for a variety of traits in wild animals (e.g. Bradshaw & Holzapfel 2000; Meffert 2000) and plants (e.g. Fenster & Galloway 2000; Rhode & Cruzan 2005), and, in the study of pathogenic interactions, it is increasingly realised that epistasis can contribute to host resistance (Wilfert & Schmid-Hempel 2008; Wegner *et al.* 2009; Lambrechts 2010). The importance of individual-level G × G interactions in pathogenic interactions has also received attention (Lambrechts *et al.* 2006; Lambrechts 2011). If individual-level G × G interactions are common, and the genetic architecture of host resistance can depend on parasite genotype (Little *et al.* 2006; Wegner *et al.* 2009), then it is reasonable to suppose that

the converse may also sometimes be true: that is, in some pathogenic interactions, the genetic architecture of a pathogen's virulence traits may depend on host genotype. As recently highlighted by Detwiler & Criscione (2010), there has been surprisingly little work on the genetic architecture of virulence traits in parasites, despite a burgeoning literature on the modelling of host-parasite systems, much of which has predicted an important role for parasite epistasis (Peters & Lively 1999; Fenton & Brockhurst 2007; Dybdahl *et al.* 2008). In the field of animal parasitology, Detwiler & Criscione (2010) list several exceptions to this general trend, including the work of Wright & Ross (1980), Trouvé *et al.* (1998), Pagès *et al.* (2002), Volf *et al.* (2007) and Dybdahl *et al.* (2008). These studies have investigated the virulence of F_1 or later generation parasite hybrids in different host species or genotypes (Pagès *et al.* 2002; Volf *et al.* 2007; Dybdahl *et al.* 2008) or in a single host background (Wright & Ross 1980; Trouvé *et al.* 1998). Wright & Ross (1980), Pagès *et al.* (2002) and Volf *et al.* (2007) all found evidence for F_1 heterosis in traits related to fecundity and virulence in different species of trematode flatworms (Wright & Ross 1980; Pagès *et al.* 2002) and a protozoan (Volf *et al.* 2007). Unfortunately, the studies of Wright & Ross (1980) and Volf *et al.* (2007) did not investigate later hybrid generations, which is often necessary for detecting epistasis (Lynch 1991). Trouvé *et al.* (1998) and Pagès *et al.* (2002) both found epistasis (hybrid breakdown) for fecundity in late-generation hybrids between two different trematode flatworms; Trouvé *et al.* (1998), however, did not find evidence for heterosis in the F_1. In a particularly notable study, Dybdahl *et al.* (2008) were able to find evidence for hybrid breakdown in F_1 populations of a trematode flatworm (*Microphallus* sp.); these authors were able to relate the observed outbreeding depression to local adaptation using sympatric and allopatric snail hosts, demonstrating

that non-additive gene effects underlaid host adaptation.

The presence of individual- or population-level G × G interactions between hosts and parasites, and of epistasis for host resistance or parasite virulence, has implications for host-parasite coevolution, as it can create the necessary conditions for a 'geographic mosaic' of coevolution, where different populations are undergoing different reciprocal evolutionary interactions, with host and parasite traits varying across the species' distributions (Thompson 1999). However, it should be recalled that the *Striga*-host system differs from systems involving two wild species due to the absence of host evolution in the usual sense (Huang *et al.* 2012): the action of *Striga* upon its human-managed crop-host is unlikely to cause micro-evolutionary change, although farmers may save the seed of open-pollinated landraces, possibly allowing some selection by the parasite on the host, and reciprocal selection with wild grass hosts may also occur in fallow years, or at the edges of fields. However, there may still be a type of selection of *Striga* on its crop-host, in that if the intensity of parasitism on a host genotype becomes so intense that the farmer changes crop cultivar or species, variation in host resistance through time will be created, albeit in a less gradual way than for a wild host-parasite interaction. It seems likely therefore that many of the conditions that Thompson (1999) sets out as necessary for a geographic mosaic of coevolution may occur in the *Striga*-host system: given G × G interactions between *S. hermonthica* and its hosts, reciprocal selection is likely to differ geographically, and virulence traits are also likely to differ between populations, at the cellular level for example (Yoshida & Shirasu 2009; Ali *et al.* 2009). As noted above, the genetic architectures of phenotypes that might underlay these variable population-level interactions are an under-explored subject area, particularly for parasite virulence (Detwiler & Criscione 2010). Understanding the

genetic architecture of the *Striga*-sorghum host-parasite interaction may then help to shed light on the molecular basis of plant-plant parasitism, and to better understand its evolution.

As well as the results of crossing experiments providing insights into the genetic bases of traits (e.g. heterosis and epistasis), experimental crosses also allow insights into segregation distortion and cyto-nuclear interactions, and into marker arrangement through the creation of linkage maps. Here, I investigate the basis of host-parasite population-level G × G interactions, via the first examination of the genetic architecture of *S. hermonthica* virulence, using experimental crosses and a linkage map created with codominant fluorescent AFLP markers. A pseudo-backcross F_1 (BCF_1) population of recombinant parasitic plants was created by crossing individuals from two populations of *S. hermonthica* divergent in their virulence reactions against three sorghum cultivars. The BCF_1 plants were assessed under these host conditions, as well as in the absence of a host, with the aim of investigating the following questions: (1) Does *S. hermonthica* virulence exhibit heterosis or epistasis, and does this depend on host? (2) Do cyto-nuclear interactions affect virulence? (3) How much segregation distortion is there, and does this depend on host? (4) Is there any evidence for segregation distortion at different levels (allelic, zygotic, or genotypic) in response to different hosts? (5) Are markers which are differentially selected between host types clustered or separated on the linkage map?

It is also intended that the markers identified in this study will be useful in future experimental genome scans (e.g. Huang *et al.* 2012), either using AFLP, or using the same restriction enzymes and selective nucleotide bases to generate co-localised sequenced single nucleotide polymorphism-based markers, such as sequenced

restriction enzyme associated DNA (sRAD) markers (Baird *et al.* 2008).

4.2 Materials and methods

4.2.1 Plant materials

The individuals crossed to create the pedigree used in this study were from two *S. hermonthica* seed populations collected from different host species and from different regions of Africa, increasing the potential for genetic variation between the populations (Koyama 2000b). The seed populations were collected from *S. hermonthica* plants parasitising either maize (H5 hybrid) in Kibos, Kisumu, Western Province, Kenya (collected 1997, called '*Sh*K97' hereafter), or rice in northern Ivory Coast (collected 1997, called '*Sh*IC97' hereafter). The approximate distance between these sites is 4,600 km. These two *Striga* seed populations have been previously noted to exhibit differing virulence reactions against the sorghum hosts E36 and N13 (A. Boisnard, pers. comm.)

4.2.2 Experimental crosses and growing conditions

A three generation, pseudo-backcross pedigree was originally chosen to maximise the number of markers segregating in the 'testcross' configuration (Grattapaglia & Sederoff 1994), thereby increasing linkage disequilibrium between alleles of the donor parent genome and the number of recombination events that can be detected with dominant markers (Kirst *et al.* 2004). Three generations of crossing can improve the mapping of dominant markers by increasing the proportion of phase-known marker pairs (Williams 1998).

The parental populations *Sh*K97 and *Sh*IC97 were grown in pots (1:1 mixture of sand:M3 John Innes potting compost) in a controlled environment growth room (irradiance: 500 μmol m^{-2} s^{-1} at plant height; 12 h photoperiod; relative humidity 60%).

Thirty mg of *S. hermonthica* seed was mixed with sand, and distributed in a layer 60 mm below the final soil surface; the hosts were planted two weeks after the soil was inoculated to allow for seed conditioning. Pots were hand-watered as required with 40% Long Ashton solution containing 2 mM ammonium nitrate (Hewitt 1966). The hosts in the parental generation were either the highly susceptible maize hybrid WH502 (for *Sh*K97), or the highly susceptible rice cultivar IAC165 (for *Sh*IC97). Different plant hosts were used to ensure that the *S. hermonthica* plants selected for crossing were divergent for any genomic regions involved in differential virulence; using susceptible hosts at this stage was also important to ensure parasite growth. Two parental individuals from *Sh*IC97 were crossed with each other to create a *Sh*IC97 F_1 (*Sh*IC97/*Sh*IC97; all crosses given in Purdy notation); the cross in the other half of the pedigree was the hybridisation of individuals from *Sh*K97 and *Sh*IC97, with *Sh*K97 providing the mother plant (*Sh*K97/*Sh*IC97) (Fig. 4.1). The *Sh*IC97 F_1 plant was also grown on the rice cultivar IAC165, whilst the hybrid *Sh*K97/*Sh*IC97 F_1 was grown on the highly susceptible maize hybrid cultivar H511, again, this was intended to promote the retention of *Sh*K97 alleles for host-specific virulence. The two F_1 plants were crossed reciprocally to create two BCF_1 seed populations: *Sh*K97/*Sh*IC97//*Sh*IC97/*Sh*IC97 (hybrid F_1 as the mother, hereafter called BCF_1HyMo) and *Sh*IC97/*Sh*IC97//*Sh*K97/*Sh*IC97 (hybrid F_1 as the father, hereafter called BCF_1HyFa) (Fig. 4.1). (The *Sh*IC97/*Sh*IC97 F_1 cross was originally included with the intention of increasing the homozygosity of recurrent parent alleles, and so the number of testcross loci informative for the donor parent, before the usage of codominant AFLP [see below, section 4.2.4] was arrived at.) All crosses were performed by emasculating the pollen donor by removing the corolla tube and transferring the pollen using a fine

paint brush to the dorsal (upper) surface of the flattened stigma of the pollen recipient (Safa *et al.* 1984). Separate brushes were used for different cross-pollinations and kept in individual sealed bags between crossing events. Capsules were harvested when they were observed to be dry on the mother plant and were subsequently stored at 27˚C for 8 weeks of ripening. In agreement with research demonstrating *S. hermonthica* to be obligately outcrossing (Safa *et al.* 1984), no mature capsules were observed to form at flowers that had not been manually outcrossed. Tissue samples of all parental and F_1 plants were dried on silica gel in air-tight tubes and stored in the dark until required for DNA extraction. All germination tests were conducted after 12 days conditioning at 27˚C by the addition of 2 mL of the germination stimulant GR-24 at 0.1 mg L^{-1} (Gurney *et al.* 2006).

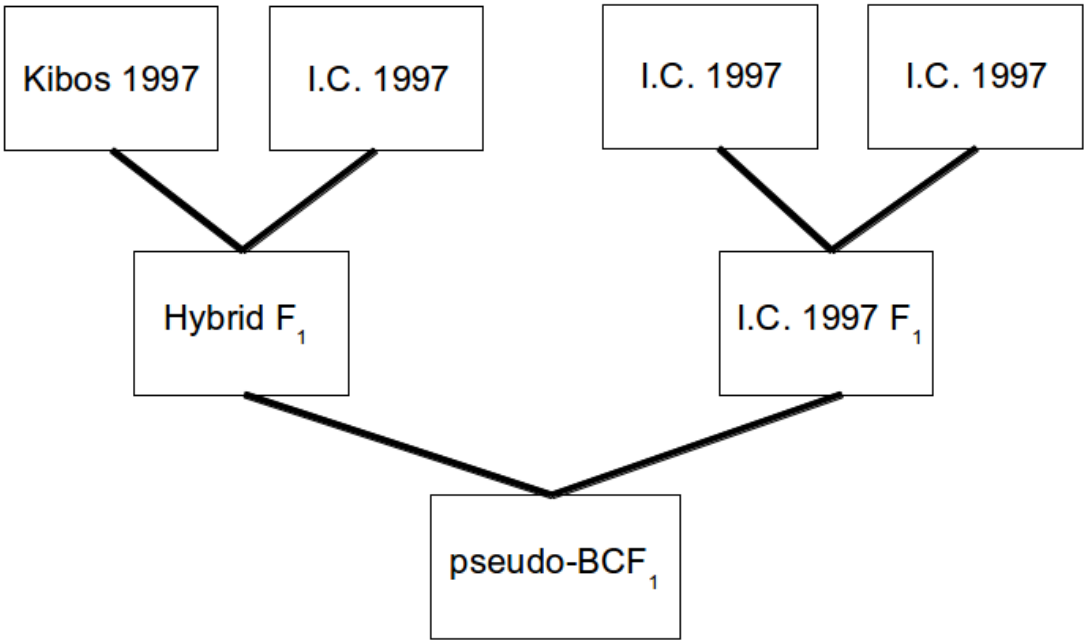

Figure 4.1. Crossing design used to produce the F_1 and BCF_1 populations. One-way crosses were performed for the first (parental) generation; for the *Sh*K97/*Sh*IC97 cross the Kibos plant was the mother. The second round of crossing was reciprocal, producing both *Sh*K97/*Sh*IC97//*Sh*IC97/*Sh*IC97 and *Sh*IC97/*Sh*IC97//*Sh*K97/*Sh*IC97 pseudo-backcross populations.

Each of the two reciprocal BCF_1 seed populations was grown in three different ways: on agar without a host; on the relatively susceptible sorghum cultivar E36; and on the relatively resistant sorghum cultivar Malisor. *S. hermonthica* BCF_1 seeds for axenic culture were surface sterilised with 70% ethanol for 2 min, then with fresh 20% bleach solution (with 2 drops Tween-20/500 mL water) for 10 min, followed by washing at least 6 times with sterile distilled water. Seeds were transferred to 0.9% water agar for conditioning. Conditioning on agar, as suggested by Yoshida *et al.* (2010), was found to be the best way of avoiding contamination by microorganisms originating in the *Striga* testa. Yoshida *et al.* (2010) added sucrose (1%) to their conditioning medium, however, here it was found that pure agar was best for discouraging microbial growth. Germination was triggered after 12 days at 27˚C by the addition of approximately 2 mL filter-sterilised GR-24 (0.1 mg L^{-1}) to the agar plates. After 3−5 days germinated seeds were transferred, using sterile technique, to a 0.9% agar medium (pH 5.7) containing 1% sucrose and 4.3 g L^{-1} MS salts (Berner *et al.* 1997). As the seedlings outgrew the Petri dishes they were transferred to Phytatray containers (Sigma) with the same medium (Fig. 4.2). After two weeks in the dark at 27˚C Phytatrays were transferred to a growth room under the environmental conditions described above at the beginning of this section (4.2.2). No fungal or bacterial contamination was observed in any of the Phytatrays from which *Striga* plants were harvested for genotyping.

Figure 4.2. *Striga hermonthica* plants growing on water agar medium in a Phytatray (top), and, the hostless *Striga* plants immediately after harvesting for DNA extraction (bottom).

Sorghum cultivars E36 and Malisor were established in root observation chambers (rhizotrons) and maintained in a growth room; again, under the conditions described above (section 4.2.2). Sorghum host roots were inoculated 12 days after sowing with pre-germinated seed of one of the two reciprocal *S. hermonthica* BCF_1 seed populations (BCF_1HyMo or BCF_1HyFa); at least 10 rhizotrons were used for each reciprocal/host combination. *S. hermonthica* BCF_1 plants were harvested from the sorghum hosts 34−40

days post-infection, and from agar 8−10 weeks post-germination. All harvested *Striga* plants were taken forward directly for DNA extraction, described in the section immediately below (4.2.3).

4.2.3 DNA extraction and AFLP

For DNA extraction and AFLP genotyping, the total numbers of plants in each growing condition were: agar: n = 163 (BCF_1HyMo = 108; BCF_1HyFa = 55); E36: n = 179 (BCF_1HyMo = 128; BCF_1HyFa = 51); Malisor: n = 189 (BCF_1HyMo = 134; BCF_1HyFa = 55). The reciprocal population sizes across growing conditions were therefore: BCF_1HyMo: n = 370; BCF_1HyFa: n = 161. The lower sample size of the BCF_1HyFa population was due to less seed being produced from that direction of the reciprocal cross.

For all three growing conditions, 10−60 mg of *S. hermonthica* tissue was harvested from individual plants directly into grinding tubes and kept on ice until drying. Tissue samples were dried for 5 h at 45˚C, whereupon they were ground in a tissue homogeniser (Qiagen, 5 min at 25 Hz). Extraction buffer containing: 100 mM Tris-HCl (pH 7.4); 500 mM NaCl; 50 mM EDTA; 0.7% sodium dodecyl sulphate; 52 mM sodium sulphite; 1.6 µg RNAse A; and 16 µg Proteinase K (Mogg & Bond 2003; Whitlock, Hipperson, Mannarelli, & Burke 2008) was added to each sample. Sample tubes were vortexed thoroughly and incubated overnight at 55˚C on a rotating rack. The rest of the extraction followed Whitlock, Hipperson, Mannarelli, & Burke (2008) except for the following pH adjustments: the 6 M NaI chaotropic DNA binding buffer was at pH 7.0; the 10 mM Tris-HCl elution buffer was at pH 8.0. All steps were performed manually (cf. Whitlock, Hipperson, Mannarelli, & Burke 2008).

DNA solutions were quality and quantity checked using a Nanodrop spectrophotometer

(ND-1000, Thermo Scientific) and on 1% agarose gels with uncut Lambda DNA standards (New England BioLabs). DNA samples were normalised to approximately 1 ng μL^{-1}. Dilution factors for normalisation were calculated from the agarose gel estimated concentrations and subsets of the normalised samples were re-checked for approximate equality. Ten ng of DNA were then digested at 37˚C for 3 h with either 1 U *Eco*RI/1U *Mse*I in 1× Tris-Acetate buffer with 3 μg BSA (Whitlock, Hipperson, Mannarelli, & Burke 2008), or 1 U *Pst*I-HF/1 U *Mse*I in 1× NEBuffer 4 with 3 μg BSA (all restriction enzymes from New England BioLabs). The decision to use an additional restriction enzyme pair (*Pst*I/*Mse*I), in addition to the *Eco*RI/*Mse*I pairing used in Chapter 3, was due to evidence that *Pst*I may target gene-rich genomic regions due to its sensitivity to cytosine-methylated DNA (Young *et al.* 1999).

Ligation, preselective amplification and selective amplification followed Whitlock, Hipperson, Mannarelli, & Burke (2008) except for the following modifications: preselective amplifications were performed in 10 μL volumes using 2 μL of diluted, ligated DNA with both primers at 0.5 μM, these were diluted 1:4 with sterile, distilled water; selective amplifications were performed likewise. Optimal selective amplification dilutions and capillary electrophoresis injection times were selected by testing a range of both variables (Trybush *et al.* 2006); a 1:10 dilution with a 5 s injection time was determined to give good signal to noise ratio, peak shape, and amplification range for peak scoring. A fuller description of the DNA extraction and AFLP procedures used in this chapter is given in Chapter 3, section 3.2.

4.2.4 Genotyping and segregation

I used 10 primer pair combinations to genotype the mapping population pedigree (Table 4.1). The four original parental and two F_1 plants (Fig. 4.1) were genotyped in triplicate

to check the reproducibility of fragments. Initial bin sets were generated automatically using the following settings in GeneMapper v. 3.7 (GM; Applied Biosystems): peak-height threshold: 100 relative fluorescence units; max bin-width: 1.0 base pair (bp); smoothing = 'no smoothing' (see Holland *et al.* 2008 for an example of how these settings can affect inferences made from fluorescent AFLP data, indicating that they should be explicitly stated). Bin sets for each primer combination were screened for polymorphic fragments (considering both presence/absence and peak amplification intensity differences) in the BCF_1 agar population in the range 65−600 bp ('Primary GM bin set'; Table 4.1, section 4.3.1). These polymorphic fragments were then analysed using the pedigree, and were kept in the marker set if they could be assigned to a Mendelian segregation type, whilst also allowing for the possibility of segregation distortion in the BCF_1 ('Manual pedigree edit'; Table 4.1, section 4.3.1). The GM 'sum of signal' normalisation option was applied across markers within the edited bin set to create the matrix of normalised peak heights to be exported for further analysis.

Scoring AFLP codominantly means that there are six different ways in which markers can segregate in the BCF_1, where *a* is the fragment presence allele and *o* the absence allele: (♀)*ao* × *ao*(♂) (1:2:1; B3.7); *ao* × *ao* (3:1; C.8); *ao* × *aa* (1:1; D1.10); *ao* × *oo* (1:1; D1.13); *aa* × *ao* (1:1; D2.15); *oo* × *ao* (1:1; D2.18). The codes following the segregation ratios are those used in Wu *et al.* (2002); types B3.7 and C.8 are in 'intercross' configuration, whilst all D-types are in 'testcross' configuration (Kirst *et al.* 2004). The R function CodomAFLP (Gort & van Eeuwijk 2010) was used at individual loci to classify bands into genotypic categories within segregation types. As well as allocating segregating marker bands across individuals into the genotypic categories listed above, CodomAFLP also allows for bands to be classified as *a*- or *o*-, that is, not

homozygous absent (not *oo*) or not homozygous present (not *aa*). Classification was informed by knowledge of the segregation type inferred from inspecting the three generations of the pedigree. A conservative posterior probability threshold of 0.98 was used to semi-manually classify bands into genotypic categories (Gort & van Eeuwijk 2010). Examples of two loci classified in this manner are given in Figure 4.3. At individual marker loci, allelic intensity distributions that did not clearly fit the segregation type indicated by the pedigree were discarded ('Manual CdAFLP edit'; Table 4.1, section 4.3.1). Allowing genotypes with missing information to be classified partially may bias our ability to accurately detect segregation distortion. However, where the amplification intensity distributions of different genotypes overlap, and the partial *a*- or *o*- genotype calls are most likely to be made, it is not possible to say *a priori* whether these missing-information genotypes are most likely to be outliers from the lower-intensity or higher-intensity fully-assigned genotypes. Given that my primary purpose was to assess comparative segregation distortion between the growing conditions of the BCF_1, I assume that the partially classified *a*- or *o*- calls were randomly drawn from the true lower- and higher-intensity genotype distributions that they deviate from, and so removed them from the segregation distortion estimates.

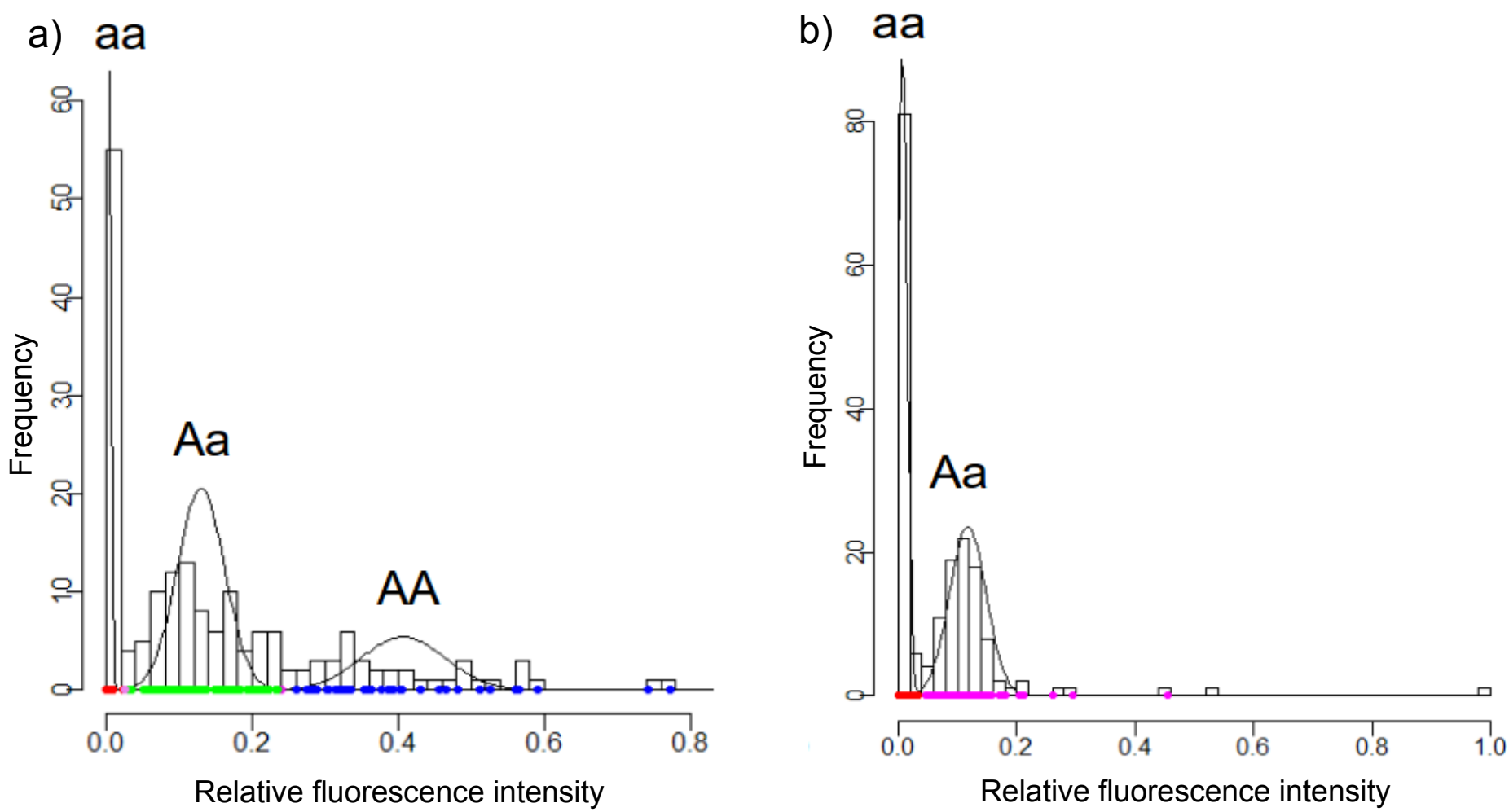

Figure 4.3. Screenshots of two loci scored in CodomAFLP. a) shows a B3.7-type loci (*ao* × *ao*) that has been scored using mixed normal distributions; individuals at the junctions of aa/Aa and Aa/AA may receive partial genotype scores. b) shows a D.13-type (*ao* × *oo*) locus that has been scored using mixed quasi-Poisson distributions. In b), two intensely amplified samples have been manually excluded from the automatic classification in order to achieve a better fit.

Deviations from Mendelian segregation ratios at individual loci were assessed by chi-squared goodness-of-fit tests with 1 or 2 degrees of freedom (d.f.) depending on whether the loci were in testcross (1:1), intercross (3:1), or intercross (1:2:1) configuration. I used a 'locus-by-locus' significance level of $\alpha = 0.05$ (essentially a Bonferroni correction), as is the current standard for analyses of segregation distortion (e.g. Schwarz-Sommer *et al.* 2003; Kuittinen *et al.* 2004; Koevoets *et al.* 2012). Comparisons of the proportion of distorted loci between growing conditions and reciprocal familes were conducted using binomial GLMs; *a priori*, orthogonal GLM contrasts compared distortion between the two plant hosts {E36 vs Malisor}, and between the two plant hosts and agar {0.5 E36 + 0.5 Malisor vs agar}. Spearman's *rho* was also used to compare trends in segregation distortion between growing conditions and reciprocal families; a non-parametric measure of correlation was used because the distribution of *p* values across loci was not always normally distributed (as assessed by Kolmogorov-Smirnov tests), and Pearson's *r* is not robust to departures from normality (Dytham 2003). All other statistical tests were *post hoc* investigations and are detailed in the results as appropriate.

4.2.5 Linkage mapping

To perform the linkage analyses within reciprocal families I used CRI-MAP version 2.4 (Green *et al.* 1990), as well as the 'improved' CRI-MAP v. 2.503 (Evans & Maddox 2009) and the CRIGEN utilities package (Liu & Grosz 2006). CRI-MAP v. 2.4 (TWOPOINT option) and the AUTOGROUP option of the CRIGEN package were used to calculate two-point recombination fractions between markers and to create linkage groups, whilst CRI-MAP v. 2.503 was used to test marker positions. Linkage groups were determined with a logarithmic odds (LOD) score threshold of 3. LOD scores

assess the probability that two loci are linked using a likelihood ratio test; this compares the estimated recombination fraction (θ) between two loci to the null hypothesis that $\theta = 0.5$, which would be the case if there was no linkage (Saccone 2011). A LOD threshold of 3 is equivalent to a chi-square of 13.8 with 1 d.f., this equals a *p* value of 0.0001 for a one-sided test of the null hypothesis that $\theta = 0.5$ (Saccone 2011). Linked markers within groups were ordered using the BUILD, FLIPS*n* and ALL options of CRI-MAP (Featherston 1995). For the larger linkage groups (> 6 markers) I tested marker orders by running BUILD several times, starting each time with different subsets of markers (Featherston 1995); the function CHROMPIC was also used to check for unlikely genotypes, such as triple recombinants (Featherston 1995). CRI-MAP accepts markers with partial information (that is, *a*- and *o*-) therefore all genotype information generated by CodomAFLP was used directly in linkage mapping. CRI-MAP can also accept information from three generations of a pedigree, enabling linkage between testcross markers in repulsion to be more accurately assessed (Williams 1998; Sham & McGuffin 2002).

4.2.6 Screening for virulence differences across the pedigree

S. hermonthica seed populations (*Sh*K97; *Sh*IC97; *Sh*IC97 F_1; Hybrid F_1; BCF_1HyMo; and BCF_1HyFa) were conditioned, germinated, and used to inoculate rhizotrons as described above for the plants grown for DNA extraction. The resistant sorghum cultivar N13 was also used for virulence screening, in addition to E36 and Malisor. At least 5 rhizotrons were used for each population/host combination. Two different measures of virulence were used: the number of *Striga* plants attached to the host roots (Cissoko *et al.* 2011), and the height of the host plant. Host height, relative to uninfected control plants, is used here as an easily measured proxy for parasite fitness impacts on

the host (Kim *et al.* 1994); Graves *et al.* (1989) also found height, as well as grain yield and other biomass components of *Striga*-infected sorghum plants, to be reduced relative to uninfected controls. The attached *Striga* plants were harvested from their hosts at 30 days after infection using a stereomicroscope; counts were performed subsequently from digital photographs of Petri dishes containing the harvested *Striga* plants using ImageJ v. 1.45 (http://rsb.info.nih.gov/ij/). Sorghum heights were measured from the base of the main stem to the uppermost exposed ligule.

Striga attachment counts on the different hosts were analysed in a negative binomial generalised linear model (GLM); the negative binomial distribution was chosen over the Poisson because of aggregation and over-dispersion in the dataset (Crawley 2007). Likelihood ratio deletion tests were used to calculate the minimum adequate model in all cases (Crawley 2007). The negative binomial GLMs used the function 'glm.nb' from the R package 'MASS' (Venables & Ripley 2002). *A priori* contrasts, specified in the GLM (Crawley 2007), were used to test for heterotic or epistatic effects on virulence for each host, and to compare the parental *Striga* populations. The three contrasts used were: {*Sh*K97 vs *Sh*IC97}; {0.5 *Sh*K97 + 0.5 *Sh*IC97 vs hybrid F_1}; and, {0.5 *Sh*IC97 + 0.5 hybrid F_1 vs BCF_1} (Falconer & Mackay 1996; Galloway & Fenster 2001). The first contrast is the parental comparison. The second contrast compares the F_1 to the mid-parent value as a test for heterosis (sometimes known as mid-parent heterosis; Lynch 1991; Falconer & Mackay 1996 p. 255; Lamkey & Edwards 1999); whilst the third contrast compares the BCF_1 to the average of the F_1 and the line used as the recurrent parent in the pseudo-backcross, this is a test for epistasis (Lynch 1991; Falconer & Mackay 1996 p. 285; Kelly 2005). These two tests are part of the set of 'scaling tests' originally developed by Mather (1949). It is worth noting that some authors restrict the

use of the terms 'heterosis' and 'hybrid vigour' to those situations where F_1 trait values exceed those of the parents; however, the use of these words to describe non-additive departures in the F_1 from a mid-parent average is well established in the quantitative genetics literature (Lynch 1991; Falconer & Mackay 1996; Lamkey & Edwards 1999; Kelly 2005). Due to the fact that only certain crosses were performed in my pedigree (e.g. the backcross was only performed to one of the original parents, and no F_2 or later generation plants were produced), the additional scaling tests required to decompose deviations from the non-epistatic model into particular components, such as additive × additive or additive × dominant interactions, could not be carried out (cf. Kelly 2005). Additionally, it should be noted that the test for heterosis used here differs from the normal scenario in that we are not comparing individual-level genotype traits, but are comparing the parental population genepools to an F_1 population formed from two individual representatives of these populations. The heterosis tests are therefore only accurate if the population-level fitness is equal to the trait value (here, the probability of successful attachment, or pathogenicity) expectation for any randomly chosen genotype from the population; this is unlikely for the highly heterozygous *S. hermonthica*, and, ideally, multiple replicates of the pedigree examined here would be produced. This would reveal whether the virulence of the F_1 populations averaged around the mid-parent value, with highly virulent F_1 populatons appearing merely as the result of crosses between fit genotypes, or whether the mean of the multiple F_1 populations was significantly above the mid-parent value, suggesting that heterosis was a general feature of crosses between the two parental populations.

Host height data, normalised for each host to the average height of the uninfected control plants, were analysed in a general linear model. Model checking was carried out

by inspecting plots of residuals for heteroscedasticity and by using the R package 'gvlma', which performs a suite of tests of linear model assumptions (Peña & Slate 2006).

4.3 Results

4.3.1 DNA extraction and AFLP genotyping

Estimated DNA concentrations ranged from 0.5−20 ng μL^{-1}. The correlation between tissue fresh weight and final DNA concentration was low (Pearson's $r = 0.33$, $p < 0.01$). It is probable that this reflects different drying and grinding efficiencies between tissue samples. *S. hermonthica* tissue bruises and blackens very quickly on handling (personal observations); because blackening is often attributed to nucleolytic degradation via the production of polyphenols (e.g. Angeles *et al.* 2005), differences in handling and drying time are likely to lead to different amounts of genomic DNA degradation between samples of a similar fresh weight. The relatively low amount of genomic DNA used in this study (10 ng per sample) did not prohibit the production of strong, reproducible fluorescent AFLP traces, as has been found elsewhere (Vos *et al.* 1995; Trybush *et al.* 2006). After the three rounds of checking and selection of loci described above (section 4.2.4), 173 AFLP loci remained (Table 4.1). The segregation classes of these loci are given in Table 4.2.

Table 4.1. Numbers of polymorphic loci at each stage of analysis for each primer pair in the AFLP data set. (GM = GeneMapper; CdAFLP = CodomAFLP; *Ec* = *Eco*RI; *Ms* = *Mse*I; *Ps* = *Pst*I). The stages of analysis used here are described in section 4.2.4.

Primer pair	Primary GM bin set	Manual pedigree edit	Manual CdAFLP edit
*Ec*AGA.*Ms*GAG	68	32	30
*Ec*AGA.*Ms*GAC	72	42	30
*Ec*AGA.*Ms*CGT	62	38	32
*Ec*ATC.*Ms*GCT	33	18	17
*Ec*ATC.*Ms*CGT	43	13	11
*Ec*ATC.*Ms*CGC	28	9	8
*Ps*TCC.*Ms*GCG	34	9	9
*Ps*TCC.*Ms*GCT	45	10	9
*Ps*TCC.*Ms*GGC	51	13	11
*Ps*TCC.*Ms*GGG	41	19	16
Total	477	203	173

Table 4.2. Numbers of AFLP loci in each segregation class.

B3.7 (1:2:1) (*ao* × *ao*)	C.8 (3:1) (*ao* × *ao*)	D1.10 (1:1) (*ao* × *aa*)	D1.13 (1:1) (*ao* × *oo*)	D2.15 (1:1) (*aa* × *ao*)	D2.18 (1:1) (*oo* × *ao*)	Total
28	4	14	69	8	50	173

4.3.2 Segregation patterns in the BCF_1 generation

4.3.2.1 Segregation distortion

Segregation distortion at the locus-by-locus significance level $\alpha = 0.05$ was high under all three growing conditions (Table 4.3). There was a significant effect of host (deviance = 30.28, $p < 0.0001$, d.f. = 2) and of reciprocal population (deviance = 16.33, $p < 0.0001$, d.f. = 1) on the proportion of distorted loci, as assessed by analyses of deviance of binomial GLMs, but no interaction between them (deviance = 0.328, $p = 0.849$, d.f. = 2). The binomial GLM containing reciprocal population and host as covariates revealed significant differences in segregation distortion, both between the two plant hosts ($z = 2.58$, $p = 0.01$, d.f. = 1) and between the combined plant hosts and agar ($z = -4.71$, $p < 0.0001$, d.f. = 1). The contrast between the two reciprocal populations was also

significant ($z = 4.02$, $p < 0.0001$, d.f. = 1).

Table 4.3. Numbers (and percentages) of distorted loci for each growing condition and reciprocal population.

Growing condition	**BCF_1HyMo**	**BCF_1HyFa**
Agar	51 (29%)	35 (20%)
E36	73 (42%)	48 (28%)
Malisor	87 (50%)	67 (39%)
Plant (E36 + Malisor)	97 (56%)	71 (41%)

Comparing distorted loci across all three growing conditions, 32 loci (19%) were distorted in common; this changed to 48 (28%) and 13 (8%) within BCF_1HyMo and BCF_1HyFa respectively. Although there was more absolute distortion at the locus-by-locus 0.05 level in BCF_1HyMo than in BCF_1HyFa (Table 4.3), the sizes of the distortions at individual loci, as estimated by the *p* values, were correlated between reciprocals within growing conditions (agar: Spearman's $rho = 0.62$, $p < 0.001$; E36: $rho = 0.68$, $p < 0.001$; Malisor: $rho = 0.76$, $p < 0.001$). A strong correlation of the deviations from Mendelian expectation was also found between reciprocals in the plant-grown subset (Spearman's $rho = 0.77$, $p < 0.001$). However, the segregation distortion patterns between growing conditions were not strongly correlated: between the agar and plant data sets (reciprocal populations combined) Spearman's $rho = 0.38$ ($p < 0.001$); this was the same for correlations performed within reciprocal populations (BCF_1HyMo: Spearman's $rho = 0.31$, $p < 0.001$; BCF_1HyFa: $rho = 0.39$, $p < 0.001$). Between the two plant hosts these correlations were of a similar size, again, both for the combined data and within reciprocals (combined: Spearman's $rho = 0.37$, $p < 0.001$; BCF_1HyMo: $rho = 0.36$, $p < 0.001$; BCF_1HyFa: $rho = 0.28$, $p < 0.001$).

4.3.2.2 Relative differences in segregation between growing conditions

The relative difference in segregation ratio between growing conditions was also considered; this is simply the comparison of observed segregation ratios (at any given locus) between populations, without reference to Mendelian expectations. For the agar/plant comparison 59 loci (34%) were segregating differently between the growing conditions (chi-squared tests with 1 or 2 d.f. depending on locus type, using a locus-by-locus significance level of $\alpha = 0.05$); for the plant/plant (E36 vs Malisor) comparison this was 63 (36%). Thirty-one loci (18%) were found to be differentiated in common between these two comparisons, indicating that around half of the loci behaving differently between agar and plant-grown *Striga* also did so within the plant-grown subset, that is, between E36 and Malisor. This suggests that, in these cases, only one of the two plant hosts was driving the difference between the agar and plant-grown populations, rather than a general plant/agar difference being responsible. However, overall, the correlation between the sizes of the differences in segregation ratio was low between the agar/plant and plant/plant comparisons (Spearman's $rho = 0.32$, $p < 0.001$).

4.3.2.3 Segregation distortion at intercross loci

For intercross loci (B3.7-type; Table 4.2) there are three ways in which distortion can occur: allelic, zygotic and genotypic distortion (Leppälä *et al.* 2008). Allelic distortion is the distortion of observed allele frequencies away from the expected frequencies; zygotic distortion is the distortion of observed genotypes away from the expected genotype frequencies given the observed allele frequencies; and genotypic (or 'overall') distortion is the distortion of observed genotype frequencies away from those expected under a Mendelian model. Genotypic distortion can be caused by either allelic distortion, zygotic distortion, or both. Testcross loci (D-types; Table 4.2) can only be investigated for overall distortion, as it is not possible to separate allelic and zygotic

effects (Bechsgaard *et al.* 2004). The numbers of distorted intercross loci for each growing condition (combined across reciprocals) are given in Table 4.4. The sample of intercross loci in Table 4.4 suggests that allelic distortion may have contributed to differences between growing conditions more than zygotic distortions. This conclusion was also supported if the different types of distortion were examined between reciprocals within growing conditions (Table 4.5).

Table 4.4. Numbers (and percentages) of intercross loci distorted at the allelic, zygotic and genotype levels across growing conditions. Distortions were judged to be significant if they exceeded the experiment-wide locus-by-locus $\alpha = 0.05$ level ($p = 0.05/173$).

	Number (and percentage) distorted		
Growing condition	**Allelic**	**Zygotic**	**Genotypic**
Agar	5 (18%)	5 (18%)	9 (32%)
E36	11 (39%)	4 (14%)	14 (50%)
Malisor	13 (46%)	4 (14%)	14 (50%)
Plant (E36 + Malisor)	15 (54%)	5 (18%)	15 (54%)

Table 4.5. Numbers (and percentages) of intercross loci distorted at the allelic, zygotic and genotypic levels across within reciprocal populations across growing conditions. Distortions were judged to be significant if they exceeded the experiment-wide locus-by-locus $\alpha = 0.05$ level ($p = 0.05/173$).

Growing condition	**Allelic**		**Zygotic**		**Genotypic**	
	BCF_1HyMo	**BCF_1HyFa**	**BCF_1HyMo**	**BCF_1HyFa**	**BCF_1HyMo**	**BCF_1HyFa**
Agar	4 (14%)	2 (7%)	3 (11%)	2 (7%)	7 (25%)	4 (14%)
E36	10 (36%)	7 (25%)	3 (11%)	3 (11%)	11 (39%)	8 (29%)
Malisor	13 (46%)	9 (32%)	3 (11%)	5 (18%)	13 (46%)	9 (32%)
Plant (E36 + Malisor)	14 (50%)	11 (39%)	2 (7%)	1 (4%)	14 (50%)	10 (36%)

4.3.2.4 Segregation distortion and parental alleles at testcross loci

Where distorted testcross loci (D-types; Table 4.2) had the hybrid plant *Sh*K97/*Sh*IC97 as the heterozygous parent in the F_1, it was possible in some cases to trace back the heterozygous allele to the parental generation, and determine if the allelle originated in

the *Sh*K97 or *Sh*IC97 plant (Table 4.6). There was no evidence for the distortion of loci with *Sh*K97 alleles being associated with growing condition, either for loci at which the allele was in excess (Fisher's exact test: $p = 0.999$, d.f. = 2), or deficient (Fisher's exact test: $p = 0.655$, d.f. = 2) (Table 4.6).

Table 4.6. Proportions of distorted testcross loci in BCF_1 populations for which the heterozygous allele could be traced back to one of the two *Striga* parental populations.

	Proportion of loci with the hybrid as the heterozygous F_1 plant		**Proportion of loci traced to the *Sh*K97 or *Sh*IC97 parent**		**Proportion of loci with *Sh*K97 allele**	
Growing condition	**Heterozygote in excess**	**Heterozygote deficient**	**Heterozygote in excess**	**Heterozygote deficient**	**Heterozygote in excess**	**Heterozygote deficient**
Agar	14/22 (63%)	8/22 (37%)	12/14 (86%)	7/8 (88%)	11/12 (92%)	4/7 (57%)
E36	20/33 (61%)	13/33 (39%)	18/20 (90%)	11/13 (85%)	16/18 (89%)	6/11 (55%)
Malisor	22/42 (52%)	20/42 (48%)	19/22 (86%)	17/20 (85%)	16/19 (84%)	12/17 (71%)
Overall	-	-	-	-	43/49 (88%)	22/35 (63%)

However, the *Sh*K97 parent was found to be the source of the heterozygous allele at loci with an over-representation of heterozygotes more frequently than at loci with an under-representation; across all growing conditions this trend was significant (heterozygotes in excess: 43/49 (88%); heterozygotes deficient: 22/35 (63%); chi-squared = 5.88, $p = 0.015$, d.f. = 1) (Table 4.6).

Where distorted loci with an excess of a *Sh*K97 heterozygous allele could be assigned to a linkage group (agar: 8 loci; E36: 10 loci; Malisor: 12 loci), I tested for epistasis between loci (e.g. McDaniel *et al.* 2007; Li *et al.* 2011). Here, epistasis indicates the presence of evidence for digenic interactions between particular loci; that is, loci appearing together in the same individual more often than would be expected by independent assortment alone. Pair-wise comparisons for digenic epistatic interactions were only performed for markers on separate linkage groups (markers on the same group are likely to show non-independence due to linkage, irrespective of epistasis).

Seventeen, 40 and 56 chi-squared tests with one d.f. were performed for agar-, E36-, and Malisor-grown *Striga* plants respectively. Using an uncorrected one per cent significant level, one significant interaction was found on agar, 9 for Malisor, but none on the E36-grown plants. The significant interactions between distorted loci with an excess of the *Sh*K97 allele on Malisor involved linkage groups 2, 3, 4, 5 and 7 (Figs 4.4 & 4.5), whilst the interaction between the distorted loci for the agar-grown plants was between linkage groups 4 and 7 (Figs 4.4 & 4.5).

4.3.3 Linkage mapping

Linkage groups were created using all testcross and intercross markers in one data set for each reciprocal BCF_1 population. The reciprocal maps were constructed separately because of differences in recombination frequency: the mean pairwise marker recombination frequencies for BCF_1Mo were 0.177, 0.171, and 0.169 for the agar, E36 and Malisor growing conditions respecitively; and for BCF_1Fa these figures were 0.097, 0.102 and 0.132. Due to the non-independence of genetic material between reciprocal populations, it is difficult to directly assess the significance of the apparently lower recombination in the BCF_1Fa population from mean values, but pairwise comparisons using shared marker pairs can be used to make comparisons (Beukeboom *et al.* 2010). The Wilcoxon signed rank test was used to compare the subset of pairwise recombination frequencies shared between reciprocal families for a particular growing conditions, although for Malisor the number of shared linked marker pairs was low, meaning that the negative result may be due to low power (agar: Wilcoxon's V = 7735, $p < 0.0001$, $n = 46$; E36: Wilcoxon's V = 87, $p = 0.015$, $n = 31$; Malisor: Wilcoxon's V = 37, $p = 0.89$, $n = 13$).

The inclusion of distorted loci in these datasets appeared to seriously affect map

construction: their inclusion resulted in the linkage groupings being dominated by one large linkage group (90/173 loci, 52%), probably resulting from pseudo-linkage caused by segregation distortion in the dataset (Table 4.3). This remained the case even when maps were constructed for each growing condition separately. Therefore, for each reciprocal population, I removed those subsets of genotype data that were distorted within particular growing conditions; for example, if a locus was distorted within the agar-grown population, but not within the E36- or Malisor-grown populations, then the data from the agar population were recoded as missing data. In this way I hoped to retain as many loci as possible in the final map, whilst minimising the effects of segregation distortion; a comparable approach was recently used to construct a consensus map from six mapping populations containing distorted loci in the hybrid cereal triticale (Alheit *et al.* 2011).

After removing the distorted subsets, the BCF_1HyMo reciprocal population retained 125 markers with at least some genotypic information, whilst the BCF_1HyFa population retained 160. However, there was little difference between reciprocals in the number of markers finally mapped: the BCF_1HyMo reciprocal map had 92 markers grouped in 17 linkage groups, whilst BCF_1HyFa had 94 markers in 21 linkage groups (Figs 4.4 & 4.5). In total the aligned maps have 116 unique markers (67% of all markers scored), and the linkage groups align across reciprocals into 17 groups, slightly less than the estimated haploid number of 19 chromosomes for *S. hermonthica* (Aigbokhan *et al.* 1998), although linkage groups with gaps > 50 cM may represent artefactual joins due to remaining segregation distortion (Schwarz-Sommer *et al.* 2003). There were no conflicts in the linkage group assignments between the populations, and the groups

could be aligned; however, the groups were fragmented between the reciprocal maps, and some minor changes in marker order were also observed (Figs 4.4 & 4.5). Markers that segregated differently between the agar and plant growing conditions, and/or between E36 and Malisor (see section 4.3.2.2), are displayed on the maps where those markers were assigned to a linkage group (Figs 4.4 & 4.5). For markers that only segregated differently between the plant and agar growing conditions, 6 of 28 markers were placed on the map; for E36 and Malisor, this was 20 of 32; whilst for those markers that segregated significantly differently in both of these comparisons, 16 out of 31 makers were mapped (Figs 4.4 & 4.5).

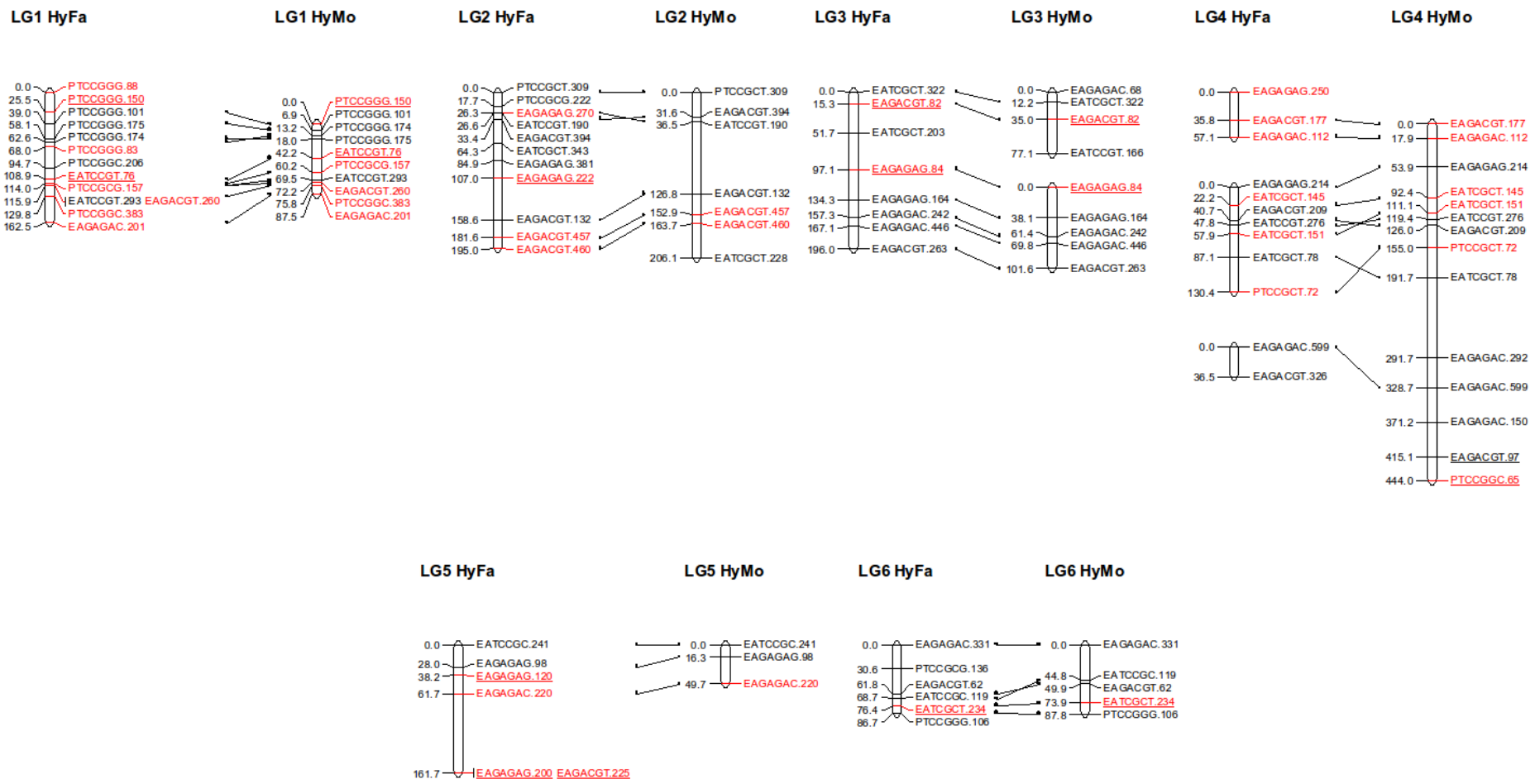

Figure 4.4. AFLP linkage map for *S. hermonthica* derived from an intraspecific cross between a West African (*Sh*IC97) and an East Africa population (*Sh*K97). Linkage groups from the two directions of the reciprocal cross are aligned, and the positions of common markers between reciprocal linkage groups are joined with a line. Distances are given in centiMorgans (Kosambi). Loci that segregated significantly differently between the agar and plant growing conditions are underlined; the colour red indicates loci that segregated differently between E36 and Malisor; loci that segregated differently in both situations are both red and underlined (see section 4.3.2.2 for details of the analysis of differences in segregation). HyFa = BCF_1HyFa; HyMo = BCF_1HyMo; LG = linkage group.

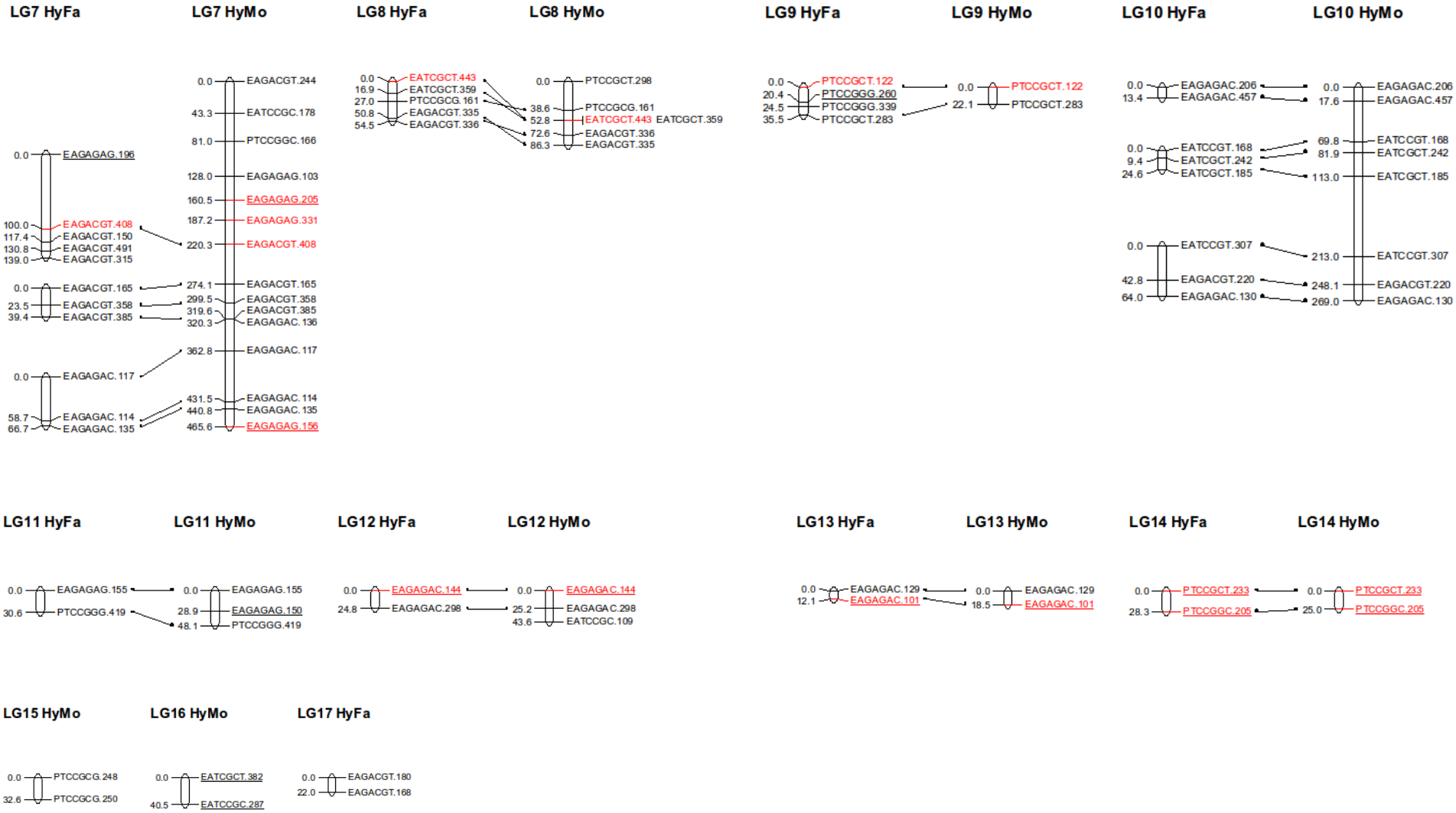

Figure 4.5. See Figure 4.4 for details.

4.3.4 Heterosis and epistasis of S. hermonthica *traits in the pedigree*

4.3.4.1 *S. hermonthica* germination across the pedigree

Using the artificial germination stimulant GR-24, there were no significant differences in germination between any of the seed populations in the pedigree: a null model was preferred to one including seed population as a covariate (quasi-binomial GLM: $F = 0.240$, $p = 0.913$, d.f. = 4; Table 4.7).

Table 4.7. Mean (± binomial standard error) germination across the generations in the *S. hermonthica* pedigree, as assessed by stimulation with the artificial germination stimulant GR-24.

Pedigree generation	**Mean ± s.e.**
*Sh*K97	53.2% ± 4.7
*Sh*IC97	54.9% ± 4.1
*Sh*IC97 F_1	51.3% ± 5.0
Hybrid F_1	51.5% ± 5.2
BCF_1HyMo	52.7% ± 4.5
BCF_1HyFa	48.9% ± 4.7

4.3.4.2 *S. hermonthica* virulence across the pedigree

The number of attachments on the three host genotypes varied strongly depending on the *Striga* generation assessed (Fig. 4.6). Across all *Striga* generations, the susceptible host E36 usually had more attachments, but the hybrid F_1 generation was also very virulent on the normally resistant hosts N13 and Malisor (Fig. 4.6). The likelihood ratio deletion tests revealed significant effects of both host genotype (deviance = 100.65, $p < 0.001$, d.f. = 2) and *Striga* population (deviance = 115.91, $p < 0.001$, d.f. = 4) on parasite virulence, as estimated by the number of *Striga* attached (Fig. 4.6). The interaction between host genotype and *Striga* population was also significant (deviance = 52.93, $p < 0.001$, d.f. = 8; Fig. 4.6). (For this analysis, the reciprocal populations of the BCF_1 were not differentiated, this was due to there being either no difference, or

only very small differences between reciprocals; see section 4.3.5.) The GLM contrasts showed that the two parental seed populations (*Sh*K97 and *Sh*IC97) were significantly different in virulence on N13 ($z = 10.11$, $p < 0.001$, d.f. = 1) and Malisor ($z = 2.92$, $p = 0.004$, d.f. = 1), but not on E36 ($z = -0.69$, $p = 0.491$, d.f. = 1; Fig. 4.6); this result highlights the strong population-level G × G interaction for which this combination of parasite and host populations was chosen. The scaling test contrasts revealed that the hybrid F_1 displayed (putative) heterosis on N13 ($z = -7.68$, $p < 0.001$, d.f. = 1) and Malisor ($z = -2.31$, $p = 0.021$, d.f. = 1), but not on the susceptible host E36 ($z = -1.51$, $p = 0.130$, d.f. = 1; Fig. 4.6). The final scaling test contrast showed that significant epistasis in the BCF_1 was observed on N13 ($z = 7.05$, $p < 0.001$, d.f. = 1) and E36 ($z = 4.15$, $p < 0.001$, d.f. = 1), but was only marginal for Malisor ($z = 1.80$, $p = 0.072$, d.f. = 1; Fig. 4.6).

The effect of *Striga* on the height of the host plant (relative to controls) depended strongly on the host genotype (Fig. 4.7). The varying levels of *Striga* attachment on E36, shown in Figure 4.6, caused a constant reduction in height of around 50% (Fig. 4.7). The effect was more variable for N13 and Malisor (Fig. 4.7), apparently due to the low virulence displayed by the *Sh*IC97, *Sh*IC97 F_1 and BCF_1 generations (Fig. 4.6). Again, the hybrid F_1 generation appeared to have an unsually large effect on host height when parasitising the resistant cultivars N13 and Malisor (Fig. 4.7), commensurate with the increased *Striga* attachment observed (Fig. 4.6).

Models of the host height data failed to meet several of the assumptions of a linear model (heteroscedasticity and skewness were detected). Inspecting Cook's distances (Crawley 2005) for the model showed that there was one very influential outlier. After re-running the model with this datapoint removed, the model met all of the necessary

assumptions. The removed datapoint was from population *Sh*IC97 growing on Malisor: the host plant was unusually tall. Compared to other replicates of this host-parasite combination the number of *Striga* growing on this host plant was not atypical, however, the roots of the host sorghum did display a high degree of colouring, noted during the experiment, possibly suggesting an atypical host response.

The final model for host height found host (F = 38.92, p < 0.001, d.f. = 2), *Striga* population (F = 30.02, p < 0.001, d.f. = 4), and the interaction between these (F = 16.54. p < 0.001, d.f. = 8) to be significant (Fig. 4.7). The contrasts were implemented for each host as before. The two parental populations (*Sh*K97 and *Sh*IC97) had significantly different effects on N13 (t = -13.33, p < 0.001, d.f. = 1) and Malisor (t = -5.83, p < 0.001, d.f. = 1), but not on E36 (t = -1.07, p = 0.295, d.f. = 1; Fig. 4.7). Heterosis, realised here as a reduction in host height compared to the mid-parent value (that is, increased parasite virulence) was significant on the resistant host N13 (t = 6.45, p < 0.001, d.f. = 1), marginal on the resistant host Malisor (t = 1.77, p = 0.096, d.f. = 1), but not significant on the susceptible host E36 (t = -0.39, p = 0.697, d.f. = 1; Fig. 4.7). Epistasis, realised as a reduced impact of *Striga* on host height compared to that expected from the parental values, was significant on the resistant hosts N13 (t = -11.66, p < 0.001, d.f. = 1) and Malisor (t = -3.41, p = 0.003, d.f. = 1), but not on E36 (t = 0.304, p = 0.764, d.f. = 1; Fig. 4.7).

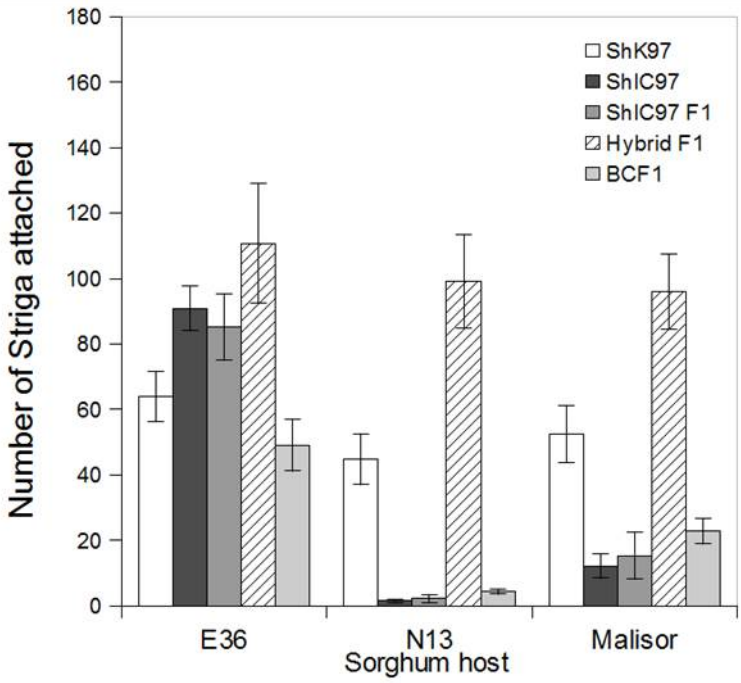

Figure 4.6. Mean (± s.e.) *S. hermonthica* attachment in rhizotrons on three sorghum hosts across all generations of the *Striga* pedigree. The BCF_1 is BCF_1HyMo and BCF_1HyFa combined. Heterosis can be seen by the degree to which the hybrid F_1 exceeds the average of the parental populations {0.5 *Sh*K97 + 0.5 *Sh*IC97 vs hybrid F_1}; epistasis can be seen by the degree to which the BCF_1 exceeds the average of the hybrid F_1 and *Sh*IC97 {0.5 hybrid F_1 + 0.5 ShIC97 vs BCF_1}.

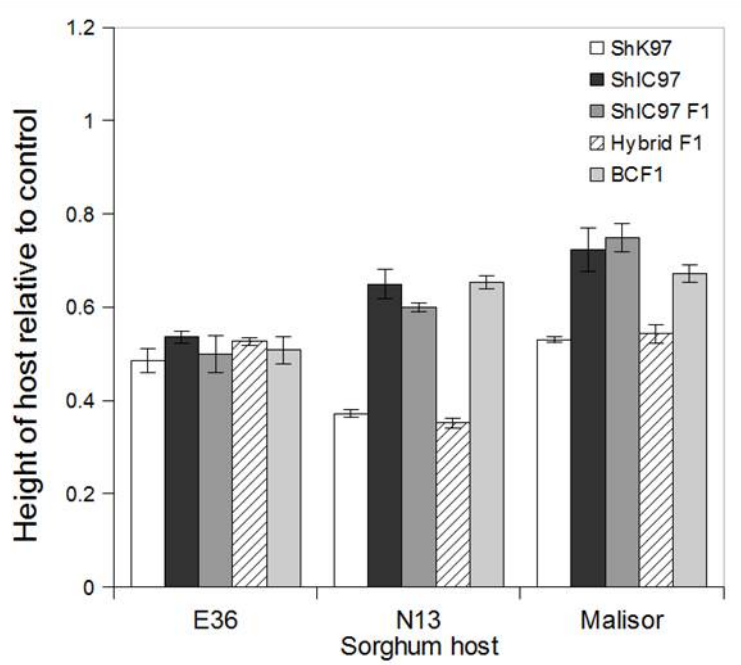

Figure 4.7. Mean (± s.e.) relative heights of three sorghum hosts parasitised by different generations of the *Striga* pedigree. The BCF_1 is BCF_1HyMo and BCF_1HyFa combined. Heterosis is displayed if the height of a host parasitised by the hybrid F_1 deceeds the average of the parental populations {0.5 *Sh*K97 + 0.5 *Sh*IC97 vs hybrid F_1}; epistasis can be seen by the degree to which the BCF_1 deceeds the average of the hybrid F_1 and *Sh*IC97 {0.5 hybrid F_1 + 0.5 ShIC97 vs BCF_1}.

4.3.5 F_1 maternal identity and virulence

Reciprocal populations of the BCF_1 (BCF_1HyMo and BCF_1HyFa) were tested separately in the post-attachment virulence experiments, in order to test for an effect of maternal identity on virulence. On the resistant host N13 there was a significant effect of maternal identity: BCF_1HyMo (mean *Striga* attachment ± s.e. = 6.11 ± 1.97) was more virulent than BCF_1HyFa (mean ± s.e. = 2.60 ± 0.64) (negative binomial GLM: z = 3.06, p = 0.002, d.f. = 1). However, this difference was not found when growing the reciprocal populations on the resistant Malisor (BCF_1HyMo mean ± s.e. = 27.40 ± 14.57; BCF_1HyFa mean ± s.e. = 16.64 ± 6.39; z = 1.17, p = 0.242, d.f. = 1) or the susceptible E36 (BCF_1HyMo mean ± s.e. = 52.00 ± 17.60; BCF_1HyFa mean ± s.e. = 44.75 ± 11.39; z = 0.52, p = 0.606, d.f. = 1).

4.4 Discussion

4.4.1 Heterosis and epistasis for S. hermonthica *virulence*

I have exploited a population-level G × G interaction between two *S. hermonthica* populations and three sorghum cultivars to investigate the genetic architecture of two *Striga* virulence traits. The use of an outbred line cross analysis and simple scaling tests, which provide insights into the genetic basis of differences in traits between populations and/or individuals (Falconer & Mackay 1996; Fenster *et al.* 1997; Kelly 2005; Demuth & Wade 2006), revealed the existence of epistasis for *Striga* host attachment and for parasite effects on host height, and indicated the possible existence of heterosis. The use of three different sorghum genotypes to evaluate the different generations in the *Striga* pedigree also indicated that the inferred architecture of these traits can depend on host genotype, suggesting that the presence of a population-level G × G interaction between host and parasite can also affect the genetic architecture of virulence within a parasite

population, and so its coevolution (Detwiler & Criscione 2010).

The F_1 hybrid (*Sh*K97/*Sh*IC97) did not show any difference in germination from the other generations in the pedigree, but on the two resistant hosts, N13 and Malisor, there was possible evidence for a strong heterotic effect on parasite attachment, although this depends on the individual-level distribution of pathogenicity in the base populations (see section 4.2.6). The effect of the F_1 hybrid on host height displayed a similar pattern: the impact of *Striga* on the susceptible host E36 did not show a significant heterotic effect, in contrast to N13 and Malisor. The lack of the potential heterotic effect of *Striga* on E36 height may be due to the non-linear relationship between *Striga*-attachment and host physiology, where small increases in parasite attachment rapidly lead to the maximum reduction in host biomass components (Gurney *et al.* 1999). Allelic dominance and overdominance are often given as the two main potential causes of heterosis (Charlesworth & Willis 2009). Empirical evidence has generally favoured dominance as the main explanation of heterosis (Fenster & Galloway 2000; Waser & Williams 2001), suggesting that the masking of deleterious recessive alleles in the F_1 often underlies hybrid vigour. Heterosis has been observed in a number of studies of plant inter-population crosses, for example, in *Chamaecrista* (Fenster & Galloway 2000), *Phlox* (Levin 1984), *Piriqueta* (Rhode & Cruzan 2005) and *Scabiosa* (Treuren *et al.* 1993), leading some authors to suggest that the fixation of deleterious alleles within populations due to drift may be a general phenomenon (Fenster & Galloway 2000). However, plant populations are often small and partially inbred (Charlesworth & Charlesworth 1987), and so high levels of inbreeding may be the reason for the frequency with which dominance has been found to be the main cause of heterosis in plants.

Heterosis has, to my knowledge, not been demonstrated for a parasitic plant, although both heterosis and epistasis have been shown to occur in animal parasites (see section 4.1); therefore it would be of interest to replicate the pedigree generated in this study to see if the potential heterosis found here can be supported. If the heterosis is a real effect of interaction between genomes, rather than the result of crossing particularly fit individuals from the parental populations, then it is particularly strong on the resistant hosts N13 and Malisor, where the amount of parasite attachment outperfomed not only the mid-parent value but also both of the parents. This would suggest that within the parental *Striga* populations, *Sh*K97 and *Sh*IC97, the loss of virulence due to deleterious recessives might be considerable. However, both dominance and overdominance as explanations of heterosis assume that the F_1 has higher heterozygosity than the parents (i.e. the parents must be homozygous for some deleterious recessives in order for hybridization to cause heterosis). In the current case, with *S. hermonthica* being self-incompatible and within-population genetic variability being high (e.g. Welsh & Mohamed 2011), and with it being a highly fecund annual plant with seeds that may survive in the seed-bank for over 10 years (Yoder & Musselman 2006), it seems unlikely that the *Striga* parents used in this cross were inbred to any great extent. It is, therefore, not clear that there will have been an increase in heterozygosity in the F_1 sufficient to cause strong heterosis on the resistant sorghum hosts.

An alternative explanation is that epistasis plays an important part in heterosis (Lynch 1991; Lamkey & Edwards 1999). Lynch (1991) states that, in the absence of inbreeding, and when only considering epistasis between two loci, "outcrossing enhancement in the F_1 can be caused by dominance of favorable genes isolated in the two parental populations …, or by the existence of favorable additive × additive epistatic interactions

between genes from different sources …, or both". Lynch (1991) reanalysed grain yield data from line crosses between maize genotypes of differing levels of genetic divergence, demonstrating the existence of favourable additive × additive epistatic effects underlying heterosis. This is a particularly interesting result, because it is normally assumed that epistasis is due the existence of within, and not among, population coadapted gene complexes (Lynch 1991). In the current situation, where the parental populations were harvested from different host crops (*Sh*IC97 from rice and *Sh*K97 from maize), it is possible that the parental populations contain host-adapted alleles, that, in combination, are beneficial on particular sorghum cultivars. Given that *S. hermonthica* is thought to have coevolved with sorghum and wild grasses, and later adapted to rice and maize (Weber *et al.* 1995), the chance of favourable interactions between population genepools may be increased by the retention of ancestral mechanisms of parasitism. Complementary molecular pathways for parasitism, or the concealment of avirulence genes from the host, could underlay the between-population component of heterosis on the resistant hosts N13 and Malisor. It should also be noted that the interpretation of the genetic architectures uncovered in this *Striga* pedigree is complicated by the host-genotype dependence of the results. Typically, theories of heterosis have been based on unconditional increases in fitness, although genotype × environment interactions for heterosis have been observed, and techniques for partitioning the contribution of 'intrinsic' heterosis have been developed for certain line cross scenarios (Xu & Zhu 1999).

The BCF_1 generation displayed a striking breakdown of both *Striga* attachment and parasite effects on host-height for the majority of the host-parasite combinations, although, again, there was no effect on germination. The effect of the BCF_1 on E36

height did not display hybrid breakdown, presumably due to there still being enough attached *Striga* to have the maximum phenotypic effect on this highly susceptible sorghum host (Gurney *et al.* 1999). BCF_1 *Striga* attachment on Malisor did not show a significant effect of hybrid breakdown, but there was still a departure from the average value of the recurrent parent (*Sh*IC97) and the F_1 in the expected direction. The occurrence of hybrid breakdown is normally taken as evidence for the existence of genetic interactions (coadapted gene complexes) that affect the trait for which the breakdown is observed, i.e. there is epistasis for the trait (Falconer & Mackay 1996). Due to the backcross design, any additive × additive between-population coadaptation that contributed to F_1 heterosis would be likely to be disrupted, especially if these interactions involved several *Sh*K97 loci. Alternatively, the hybrid breakdown of *Striga* attachment observed on N13 may not be due to the disruption of coadaptation, but could be due to the reconstituting of deleterious recessive homozygotes (e.g. host-detected avirulence factors) in the *Sh*IC97 genome. Potentially, both disruption of coadaptation and the reconstitution of deleterious recessives could be occurring together; this has a parallel in the composite explanations of Haldane's rule (hybrid breakdown in the heterogametic sex) that are now often favoured (Koevoets *et al.* 2012).

Where hybrid breakdown was found, it was found in both of the BCF_1 populations derived from the reciprocal cross, suggesting that the hybrid breakdown itself was not due to a cytoplasmic effect. However, the size of this breakdown did differ significantly for reciprocal populations of the BCF_1 *Striga* when grown on the resistant host N13. The BCF_1 with the hybrid mother plant (having *Sh*K97 cytoplasm) was found to be more virulent. Other authors have found a cytoplasmic component to the genetic architecture of hybrid breakdown in inter-population crosses in plants (Campbell &

Waser 2001; Galloway & Fenster 2001; Rhode & Cruzan 2005; Etterson *et al.* 2007) and other organisms (e.g. Pritchard *et al.* 2011), suggesting that this phenomenon may not be uncommon. However, even though a significant difference was found for host N13, in absolute terms the difference in average *Striga* attachment between the reciprocals on N13 was small (although the effect size would be classed as large: Cohen's d = 1.08). Replicate crosses of the type made here, as well as reciprocal crossing at the F_1 stage (Rhode & Cruzan 2005), would provide a more thorough understanding of the subtle, and apparently host-genotype specific, effects of cytoplasmic background on *S. hermonthica* virulence. Pritchard *et al.* (2011) give examples of where cyto-nuclear incompatibilities between copepod populations have been traced to particular combinations of genes involved in ATP production; this suggests the possibility that cyto-nuclear effects on virulence in *Striga* may not be directly tied to cellular machinery involved in parasitism, but may simply be a small, though significant, indirect effect of generally lower cellular efficiency.

4.4.2 Segregation distortion

Segregation, or transmission ratio, distortion is an increasing focus for genomic studies (Lexer & Widmer 2008). It is of particular importance to studies of interspecific hybridisation, where it is often linked to the existence of Bateson-Dobzhansky-Muller interactions (e.g. Koevoets *et al.* 2012). Likewise, it has also been identified in intraspecific crosses (Jenczewski *et al.* 1997; Hall & Willis 2005; McDaniel *et al.* 2007; Leppälä *et al.* 2008; Bikard *et al.* 2009; Li *et al.* 2011; Pritchard *et al.* 2011). Segregation distortion has generally been found to be lower for intraspecific than for interspecific crosses (Jenczewski *et al.* 1997), but intraspecific levels can still be high. For example, Hall & Willis (2005) found 48% of molecular markers to be distorted in

an inter-population cross of the North American herb *Mimulus guttatus*; this was later found to be largely due to a chromosomal inversion containing genes involved in local adaptation to soils (Lowry & Willis 2010). The amount, and patterns, of segregation distortion can be a useful guide to the architecture of intrinsic incompatibilities that may exist between populations (McDaniel *et al.* 2007), although proving whether the observed genetic incompatibilities arose in response to, or as a by-product of, extrinsic (e.g. local adaptation) or intrinsic (i.e. intra-genomic) factors may be harder to achieve (Bierne *et al.* 2011). Distorted loci have also been implicated in digenic epistasis (Schwarz-Sommer *et al.* 2003; McDaniel *et al.* 2007; Li *et al.* 2011; Alheit *et al.* 2011), lending support to the Wrightian view of populations as coadapted complexes of genes with significant non-additive interactions leading to inter-population hybrid breakdown (Wade & Goodnight 1998). Knowledge of whether distortion acts primarily at the allelic or zygotic level may also provide information about the causative mechanism (Fishman *et al.* 2001; Bechsgaard *et al.* 2004).

Here, for the first time, I have revealed a relationship between segregation distortion in a parasite and host resistance. Segregation distortion levels for *S. hermonthica* on agar were not dissimilar to other intraspecific crosses (Jenczewski *et al.* 1997), despite the considerable geographic distance between the populations crossed (4,600 km); however, distortion was significantly higher when assessed in *Striga* grown on plant hosts. Growth on a host exerted a selection pressure that resulted in alleles at certain loci being over- or under-represented in those *Striga* plants that could successfully parasitise a host compared to growth on agar. Once *Striga* seedlings were in the sterile agar environment all germinated seed transplanted survived, unless there was contamination of the growth medium, indicating minimal selective pressure for growth from the agar environment

itself. This suggests that distortion observed in the agar-grown population is a reasonable proxy for intrinsic distortion, operating pre-attachment, between the two *Striga* populations, and that the difference in distortion between the agar and the plant-grown populations (8–21%, depending on reciprocal population and plant host; Table 4.3) represents extrinsic host selection pressure. Non-parametric, rank-based correlations were used to assess the similarity of distortion between growing conditions and reciprocals because assessing distortion at a locus as a binary variable using the $\alpha = 0.05$ locus-by-locus cut-off could miss similarities due to the different sample sizes, and so power, between the reciprocal populations. The correlations also provide different information on the patterns in the magnitude of segregation distortion across loci between growing conditions and reciprocals. The correlations showed that, despite a significant difference between reciprocals using the $\alpha = 0.05$ level, the patterns of distortion across loci between reciprocals were similar within growing conditions. The correlations also showed a lack of similarity between growing conditions, both between agar and the two plants combined, and between the two different plant hosts. Relative differences in segregation ratios between growing conditions again suggested a similar pattern, with clear indications that a number of loci were responding differently to the different host environments.

The results from the 28 loci in intercross configuration (B3.7-type) that could be analysed for distortion at the allelic and zygotic levels indicated that allelic distortion may have contributed more to the differences between growing conditions than zygotic distortion. Allelic distortion does not necessarily mean that the distorting mechanism occurs pre-zygotically, selection against particular alleles could occur in the sporophyte independently of their heterospecific allelic partner (i.e. an additive effect). Overall, the

results from the novel approach used here to distinguish segregation distortion with an apparently intrinsic basis from that with a host-related, extrinsic cause, supports the hypothesis of the host-specific nature of differential G × G interactions in *S. hermonthica*–sorghum parasitism. This provides direct genetic evidence to support the indirect evidence from phenotypic means used in the line cross analysis (section 4.4.1).

In the small number of cases where the parental origin of alleles could be determined, there appeared to be a general bias towards the over-representation of *Sh*K97 alleles in all growing conditions. It is possible that the asymmetric introgression of *Sh*K97 alleles is a result of inbreeding depression caused by the backcross of the F_1 hybrid to a *Sh*IC97 F_1 plant that was itself the result of crossing two individuals from the *Sh*IC97 parental population. However, the *Sh*IC97 parents were both randomly chosen from a very large batch of field-collected seed, and *S. hermonthica* individuals have been found to be highly heterozygous (Bharathalakshmi *et al.* 1990; Huang 2007; Welsh & Mohamed 2011), as expected for a self-incompatible, annual plant (Hamrick & Godt 1996). The *Sh*IC97/*Sh*IC97 F_1 seed population also displayed identical reactions to the parental population in rhizotron assays, although it is possible that fitness reductions in this generation may not have been manifest in the two bio-assays performed. Furthermore, although there were slightly fewer testcross loci informative for the *Sh*IC97 F_1 parent (D2-types; Table 4.2) in the BCF_1 compared to the F_1 hybrid (D1-types; Table 4.2), the number was of a similar magnitude, suggesting that the *Sh*IC97/*Sh*IC97 F_1 plant was not significantly less heterozygous than the *Sh*K97/*Sh*IC97 F_1 hybrid. However, the AFLP loci assayed may only represent a small portion of the *S. hermonthica* genome, and it remains possible that deleterious recessives from the *Sh*IC97 population may be responsible for some part of the observed distortion towards the *Sh*K97 parent. An

alternative explanation, not requiring the invocation of *Sh*IC97 deleterious recessives, implicates interactions between *Sh*K97 alleles: epistasis for general fitness, operating pre-attachment, could have meant that BCF_1 individuals were more likely to survive if they retained a larger fraction of the *Sh*K97 genome, regardless of the host environment. However, mitigating against this hypothesis, tests of digenic interactions between a small number of loci with excesses of *Sh*K97 alleles found most evidence for significant interactions restricted to Malisor, although higher-order interactions were not evaluated.

4.4.3 The linkage map

The comparison of the 17 linkage groups constructed here with the estimated 19 chromosomes of *S. hermonthica* is suggestive, but, given that several of the linkage groups formed only consisted of small numbers of markers, it seems likely that these may, in fact, be due to spurious linkage, possibly because of remaining segregation distortion in reponse to the different host environments. As such, any inferences drawn from the linkage map should probably be tempered with considerable caution.

Due to apparent large-scale pseudo-linkage, possibly created by segregation distortion (Schwarz-Sommer *et al.* 2003), a linkage map including all markers could not be resolved, with the end result that I cannot comment on the distribution of absolute segregation distortion across linkage groups. Plotting loci that showed relative differences in segregation onto the linkage map was a compromise between plotting no information relating to host environments, and plotting absolute segregation distortion for each growing condition, for which many of the most distorted markers would be absent due to their not having been mapped. It was considered that relative segregation differences would provide a less biased picture of the distribution of host-specific differences in segregation effects, because it is not necessarily associated with strong

absolute segregation distortion in any one host environment. The final picture portrayed by this information was of an absence of strong clustering between the mapped markers that segregated differently for the agar/plant comparison or between the plant/plant host comparison, suggesting again that the genetic architecture of virulence is likely to be complex, and that loci involved generally in parasitism are not necessarily the same as those that respond differently to different hosts.

4.4.4 Conclusions

Segregation distortion is not an unusual phenomenon in inter-population crosses, but very few studies have examined whether, for a particular cross, distortion varies with the environment (but see Ellstrand & Devlin 1989), and, to my knowledge, none have looked at host genotype effects. Launey & Hedgecock (2001), Rogers & Bernatchez (2006), Niehuis *et al.* (2008) and Pritchard *et al.* (2011) have all identified life-stage specific distortion by comparing genetic maps generated from different life-stages of the oyster *Crassostrea gigas*, the lake whitefish *Coregonus clupeaformis,* the wasp *Nasonia* and the copepod *Tigriopus californicus* respectively. The closest type of investigation to the segregation distortion approach has involved replicated analyses of genetic variation for hybridisation compatibility within species: low intraspecific variation for compatibility was found in an analysis of two *Helianthus* species (Buerkle & Rieseberg 2001), whereas some loci involved in hybridisation success for two *Mimulus* species were found to be polymorphic across populations (Sweigart *et al.* 2007). If the genetic architecture of a trait, such as virulence, can be shown to depend on host genotype or the environment, and distortion is linked to epistasis and Bateson-Dobzhansky-Muller incompatibilities between hybridizing gene pools (whether intraspecific or interspecific), then examining the effects of host or the environment on segregation

distortion would appear to be an important part of understanding the variability that exists in inter-population compatibility within species (Lexer & Widmer 2008; Cutter 2012), potentially also providing insights into the early stages of speciation (Pritchard *et al.* 2011). The importance of this variability is of clear importance for parasites, where variation in host adaptation can determine the outcome of a parasite challenge, and potentially drive speciation (Huyse *et al.* 2005; Thorogood *et al.* 2008, 2009).

A final point should be made regarding the replicability of the results presented in this chapter: the putative heterosis, epistasis and levels of segregation distortion observed here are all the result of one set of crosses between particular individuals. High levels of variation between individuals within populations has been found for *S. hermonthica* (Welsh & Mohamed 2011; Huang *et al.* 2012), and therefore it should be emphasised that a different set of crosses from the populations used here may have yielded different results. Johansen-Morris & Latta (2006) found considerable variation in hybrid breakdown between F_6 recombinant inbred lines of the wild oat *Avena barbata*, and studies of this type would be useful in putting the results found here for *S. hermonthica* into a broader context. Moreover, studies of crosses between different populations separated by different spatial, temporal or genetic distances would also assist in uncovering how pervasive heterosis, epistasis and segregation distortion are in *S. hermonthica*, and so provide further insight into the genetic basis of G × G interactions between *Striga* and its hosts.

Chapter 5: Investigating the links between segregation distortion and locus-specific selection in *S. hermonthica* using an intra-population outlier analysis

5.1 Introduction

In Chapter 4 I posited that epistasis may partly explain the heterosis and hybrid breakdown of *Striga hermonthica* virulence observed in a backcross F_1 (BCF_1) population on three different sorghum host genotypes. Segregation distortion (SD) at individual loci was also shown to be an important feature of these interactions: BCF_1 *Striga* growing on the sorghum cultivar Malisor exhibited significantly more SD than those growing on the sorghum cultivar E36. Additionally, non-recurrent Kibos 1997 *Striga* alleles appeared to constitute the majority of the over-represented alleles at distorted loci in the BCF_1 population for all hosts, potentially indicating a general favouring of Kibos 1997 alleles (although the disfavouring of I.C. 1997 alleles remains a possibility). Despite these indicators of different parasite genetic responses to different host genotypes (*viz.* high SD with an over-representation of Kibos 1997 alleles on Malisor; lower SD, but also with an over-representation of Kibos 1997 alleles on E36), no difference in virulence was observed at the parasite attachment stage between the parental *S. hermonthica* Kibos 1997 on the hosts E36 and Malisor (Chapter 4, Fig. 4.6). In contrast, in Chapter 2, I demonstrated in lab-based rhizotron experiments that the sorghum cultivar Malisor had consistently strong post-attachment resistance to three separate populations of *S. hermonthica* collected in West Africa. This evidence for the existence of a *Striga*-genotype-dependent resistance mechanism in Malisor, coupled with the genetic evidence that Kibos 1997 alleles appeared to be over-represented more often in the BCF_1 population parasitising Malisor (due to the higher SD), suggests the

hypothesis that, at some loci, certain Kibos 1997 alleles may be favoured on the Malisor host.

The *S. hermonthica* population Kibos 1997 has been shown to contain within-population genetic variation that can respond to hosts of widely differing resistance levels (Huang *et al.* 2012). Therefore, it is possible that the similar attachment phenotypes shown by this *Striga* population on the hosts Malisor and E36 in Chapter 4 are actually masking genetic variation for virulence within the Kibos 1997 population. That is, similar population-level virulence phenotypes may be underlain by different individual-level genetic mechanisms. The strong post-attachment resistance shown by Malisor to some *Striga* populations (Chapter 2), in contrast to the highly susceptible host E36 (Grenier *et al.* 2007), again suggests the hypothesis that subsets of *S. hermonthica* Kibos 1997 parasitising Malisor and E36 may differ genetically at virulence loci, overcoming different types of resistance mechanisms, despite similar levels of overall population-level attachment success (cf. Huang *et al.* 2012).

The methodology of the outlier analysis, or genome scan, has generally been used to investigate genetic differentiation in the context of speciation or adaptation, where these processes can create islands of differentiation between hybridising genomes in their early stages (Butlin 2010; Strasburg *et al.* 2012). However, as shown in Chapter 3, and in Huang *et al.* (2012), the outlier analysis approach can also be applied to experimental situations. If there is a large amount of genetic variation within a population, and a strong selective pressure is applied, then an outlier analysis methodology can be used to identify differentially selected loci between sub-populations (Huang *et al.* 2012). Under certain conditions, large populations, large samples, and low F_{ST}, for example, the ability to detect outliers should be enhanced (Pérez-Figueroa *et al.* 2010; Strasburg *et*

al. 2012). The experimental approach to detecting host-selected loci within populations of *S. hermonthica* meets, or can be made to meet, all of these requirements.

In this chapter I use a rhizotron-based host selection experiment to investigate whether the two sorghum hosts Malisor and E36, both appearing alike in resistance to *S. hermonthica* Kibos 1997, are actually selecting for different subsets of parasites. This question is addressed using AFLP markers and an outlier analysis approach to identify loci which are differentially selected between two host-selected *Striga* populations. The set of AFLP loci scored codominantly in Chapter 4 form the basis of the loci investigated in this study. However, the loci scored in Chapter 4 segregated in a backcross population with an individual from the Kibos 1997 *Striga* population as the donor parent; this means that some of the loci scored in Chapter 4 will have originated in the recurrent (I.C. 1997) parent. In addition, Kibos 1997 loci that did not segregate in the mapping cross are also likely to occur in the present analysis of a much larger population sample. In light of these two facts, whilst the set of loci used in Chapter 4 forms the basis of the loci scored in this chapter, a number of extra loci were also selected and scored.

If the current experiment identifies outlier loci, it may therefore be possible in some cases to also assess the following points: whether these loci were distorted in the BCF_1, and under which growing conditions; if loci segregated differently between growing conditions; whether the loci were placed on the genetic map; and, if more than one locus is found, whether these loci cluster together on the map. The mapping of outliers should increase their value by allowing researchers to see whether multiple outliers mark multiple selected genomic regions, or, in fact, cluster in one area of the genome (Strasburg *et al.* 2012). Information about the SD, or relative segregation differences, at

outlier loci may also provide more confidence that a locus is involved in host adaptation (Lexer & Widmer 2008), especially if differences have been noted between host growing conditions.

5.2 Materials and methods

5.2.1 Plant materials

The *Sorghum bicolor* cultivar Malisor was obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Mali, courtesy of Drs T. van Mourik and H. Traore; *S. bicolor* cultivar E36 was sourced from stock held by Prof. J.D. Scholes in the Department of Animal and Plant Sciences, University of Sheffield. The *S. hermonthica* seed population ('Kibos 1997') was collected from plants parasitising maize (H5 hybrid) in Kibos, Kisumu, Western Province, Kenya in 1997.

5.2.2 Growth and harvesting of S. hermonthica

The rhizotron experiments conducted here follow the experimental set-up described in Chapter 2. Ten replicate rhizotrons for each sorghum cultivar were infected with 12.5 mg of preconditioned *Striga* seed at 12 days after sowing. Rhizotrons were placed in growth rooms under the following conditions: irradiance: 500 μmol m^{-2} s^{-1} at plant height; 12 h photoperiod; relative humidity 60%. The sorghum plants were watered automatically 4 times a day with approximately 100 mL of 40% Long Ashton solution, containing 2 mM ammonium nitrate (Hewitt 1966). Eighty and 93 parasites were harvested for DNA extraction from E36 and Malisor respectively, from across the ten replicate rhizotrons, at 30 days post-inoculation. Twenty samples (11%) were replicated at this stage in order to assess the genotyping error rate inherent in the generation of AFLPs (Bonin *et al.* 2004).

5.2.3 DNA extractions

The DNA extraction method followed the method described in Chapter 4. The extracted DNA was visualised on agarose gels (1%) to check the quality and to estimate the quantity by comparison with Lamda DNA standards (New England Biolabs; Fig. 5.1). A representative gel is shown in Figure 5.1. DNA samples were subsequently normalised to 1 ng μL^{-1} by the addition of sterile ddH_2O.

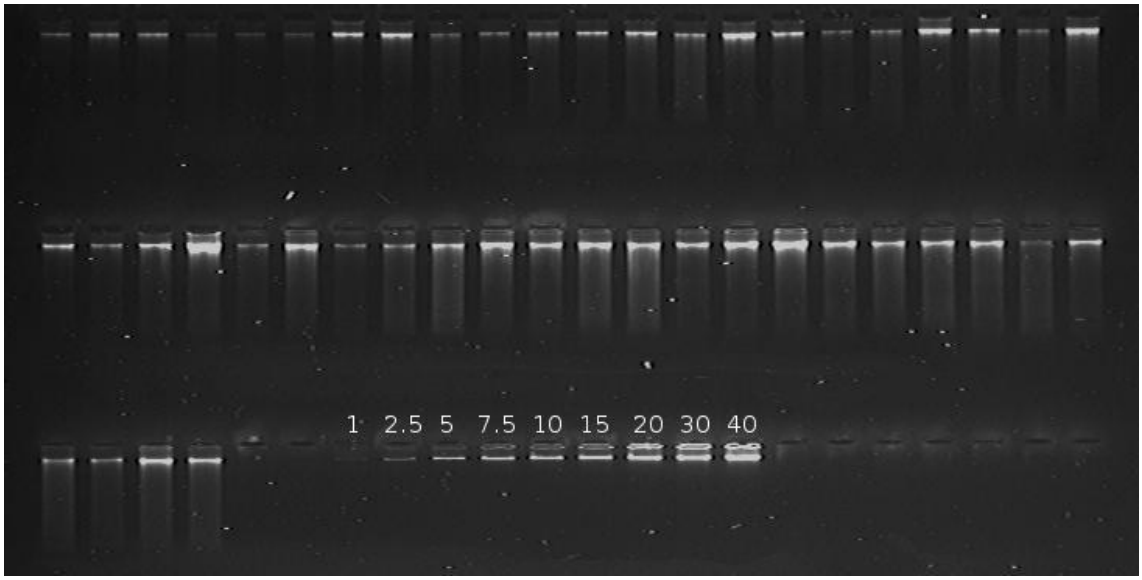

Figure 5.1. A representative agarose gel (1%) image showing intact genomic DNA. The Lamda DNA standards for quantification are labelled with their concentrations in ng μL^{-1}.

5.2.4 AFLP generation and analysis

The generation of AFLPs also followed the method described in Chapter 4; the same ten primer pairs were used. The final locus sets used in Chapter 4 were used here as the starting point for choosing AFLP loci to score. In the final instance, 101 of 173 (58%) of the loci used in Chapter 4 were scored for this analysis; an extra 213 loci were also scored, bringing the final number of loci analysed to 314 (Table 5.1). AFLPScore v. 2.0 (Chapter 3, section 3.2.8) was used to check the error rate associated with scoring these loci dominantly: the final error rate was 2.1%.

Table 5.1. Numbers of mapped loci and additional loci used for each primer pair in the AFLP data set. (*Ec* = *Eco*RI; *Ms* = *Mse*I; *Ps* = *Pst*I.)

Primer pair	Final mapping population loci set	Final mapping population loci scored in outlier analysis	Additional loci scored in outlier analysis	Total
*Ec*AGA.*Ms*GAG	30	18	17	35
*Ec*AGA.*Ms*GAC	30	19	47	66
*Ec*AGA.*Ms*CGT	32	23	28	51
*Ec*ATC.*Ms*GCT	17	15	15	30
*Ec*ATC.*Ms*CGT	11	5	18	23
*Ec*ATC.*Ms*CGC	8	4	14	18
*Ps*TCC.*Ms*GCG	9	2	22	24
*Ps*TCC.*Ms*GCT	9	0	19	19
*Ps*TCC.*Ms*GGC	11	7	12	19
*Ps*TCC.*Ms*GGG	16	8	21	29
Total	173	101	213	314

5.2.5 Genetic variation, differentiation, and outlier analyses of dominant AFLP data

AFLP-surv v. 1.0 (Vekemans *et al.* 2002) was used to calculate the percentage of polymorphic loci at the 5% level (PLP) and the average expected heterozygosity (H_{exp}), both for the two host-selected sub-populations and for the total population (Table 5.2). AFLP-surv was also used to calculate F_{ST}; the significance of the estimate was assessed using a null distribution of F_{ST} values calculated from 999 permutations of individuals between the two sub-populations (Vekemans *et al.* 2002).

The outlier analysis for the dominant AFLP data set was conducted using BayeScan v. 2.1 (Fischer et al. 2011; Foll 2012) (see Chapter 3, section 3.2.10 for an overview of the workings of this method). BayeScan v. 2.1 allows the user to set the prior odds of the model with a selection effect at a locus compared to the model without selection, where previously in BayeScan v. 1.0 (Foll & Gaggiotti 2008) they were set at 1:1. Fischer *et*

al. (2011) chose a prior probability of 1/10 for the model with selection, thereby reducing the chance that BayeScan will predict an effect of selection at any given loci. Fischer *et al.* (2011) and Foll (2012) state that this choice, set as the default in BayeScan v. 2.1, is intended to adjust for multiple testing, and that it reflects a reasonable prior belief that only around 10% of loci will be under selection. Following other recent genome scan analyses using BayeScan v. 2.1 (Girard & Angers 2011; Kautt *et al.* 2012; Deagle *et al.* 2012), I present the results from analyses using both a prior odds of selection of 10:1 and 1:1.

An additional feature of BayeScan v. 2.1 is that the program calculates *q* values for each locus. A *q* value is the False Discovery Rate (FDR) analogue of the *p* value; that is, for any individual locus it is the "minimum FDR that can be attained when calling that feature significant" (Storey & Tibshirani 2003). The FDR used will depend on the 'cost' of making false discoveries (Stephens & Balding 2009): here, the main aim is to establish the presence or absence of outliers, and the presence of some false positives is tolerable: therefore FDRs of 10 and 25% were used.

5.2.6 Outlier analysis of codominant AFLP data

As of version 2.0, BayeScan can also use AFLP fluorescence data (Fischer *et al.* 2011) to assess information on AFLP codominance represented by the amount of amplification observed at a locus. Fifty-eight percent of the loci scored in the current study were chosen due to their codominant information content identified in Chapter 4 using the method of Gort & Van Eeuwijk (2010) and knowledge of segregation class gained from the experimental pedigree. Here, the use of a larger population sample of *S. hermonthica* Kibos 1997 means that homoplasy of AFLP loci may be higher than in the pedigree analysed in Chapter 4 (the two generations of crossing in the pedigree should

have resulted in a reduction in homoplasy at loci due to the fixation of some alleles; and, the larger population samples used here will increase the probability of homoplasy occurring). However, Foll *et al.* (2010) and Fischer *et al.* (2011) demonstrated that AFLP codominance could be inferred and exploited even when precise knowledge of individual genotypes are lacking. Foll *et al.* (2010) used a Bayesian approach to estimate population-specific F_{IS} and F_{ST} coefficients using the distribution of amplification intensities at a locus to estimate the frequencies of heterozygous and homozygous dominant individuals. Foll *et al.* (2010) showed, by investigating simulated data, that both high quality and low quality fluorescence information on amplification intensity at loci considerably improves the estimation of *F*-statistics, compared to treating AFLP loci dominantly. 'Low quality' information indicates the situation where there is poor separation of heterozygous and dominant homozygous fluorescence intensities at a locus (Foll *et al.* 2010). Indeed, the method of Foll *et al.* (2010) was shown to be almost as accurate, and to have as low variance, as the use of single-nucleotide polymorphisms (SNPs) in estimating population-specific F_{IS} and F_{ST} coefficients. This improved inference of *F*-statistics from AFLP markers can also improve the identification of outlier loci (Fischer *et al.* 2011). The codominant scoring method of Foll *et al.* (2010) is different to that of Gort & Van Eeuwijk (2010) (utilised in Chapter 4 to produce individual genotypes for genetic mapping) because genotype predictions for individuals are not made, instead, an assessment of the relative frequencies of genotypes are made at each locus. This information is then used in a hierarchical Bayesian framework for the estimation of *F*-statistics (Foll *et al.* 2010). Fischer *et al.* (2011) emphasize the importance of overcoming AFLP technical issues for ensuring that systematic biases in AFLP amplification intensity do not occur: these

include the need to equalise DNA starting concentrations, and the need to ensure that all samples for a particular primer pair are processed together (and randomised with respect to DNA extraction and PCR plates), thus removing variation associated with instrumentation or laboratory reagents as far as possible. These precautions have been standard procedure for AFLP generation throughout the work presented in this thesis.

AFLP amplification intensities were exported from GeneMapper v. 3.7 (Applied Biosystems) for direct use in BayeScan v. 2.1 (Foll 2012). Fischer *et al.* (2011) recommend the removal of extreme outlier values from the amplification intensity matrix, defining extreme values as those more than three times the 95% quantile of the amplification intensity distribution at a locus: this operation was performed using an R script provided with BayeScan v. 2.1 for this purpose (Foll 2012). For codominant data, BayeScan v. 2.1 designates homozygous recessive genotypes as those intensities that fall below 10% of the maximum amplification intensity (post-extreme value processing) at a locus; however, this parameter can be altered, and, after inspecting histograms of the distribution of intensities at a selection of loci, 5% of the maximum amplification intensity was determined to be a better cut-off point for assigning homozygous-recessive genotypes in this dataset. For loci that appear to be monomorphic in the sense of all individuals having band-presence genotypes, a 'DIP' test for bimodality (Hartigan & Hartigan 1985) was performed to assess whether the distribution of amplification intensities at a locus is bimodal. This is intended to test whether such a locus might actually be polymorphic in the codominant sense, i.e. having a mix of heterozygotes and homozygous-dominant genotypes. This test for bimodality was performed using an R script provided with BayeScan v. 2.1 (Foll 2012). Fischer *et al.* (2011) suggest that even a small number of AFLP markers with a band-presence component that can be

interpreted codominantly is sufficient to considerably improve the estimation of *F*-statistics and population allele frequencies. As with the dominant BayeScan analyses, I present the results from codominant analyses using both a prior odds of selection of 10:1 and 1:1.

5.3 Results

5.3.1 Genetic diversity and differentiation estimated from dominant AFLP data

Ten AFLP primer pairs produced 314 scored loci, with fragment sizes between 65 and 550 base pairs, across 173 *S. hermonthica* plants harvested from 2 sorghum cultivars. Data on population genetic variation analysed in terms of the percentage of polymorphic loci (PLP) and average expected heterozygosities (H_{exp}), estimated from dominant scoring of AFLPs, are presented in Table 5.2 for the two host-associated *Striga* sub-populations and the overall *Striga* population sample.

Table 5.2. Measures of genetic variation for sorghum host-associated *S. hermonthica* Kibos 1997, estimated from dominant AFLP data. PLP = percentage of polymorphic loci; H_{exp} = average expected heterozygosity.

***S. hermonthica* host**	**Number of samples**	**PLP at 5%**	**H_{exp} ± s.e.**
E36	93	81.0	0.258 ± 0.009
Malisor	80	82.4	0.258 ± 0.009
Total population	173	81.7	0.259 ± 0.000

Genetic differentiation (F_{ST}) between the two host-associated sub-populations was estimated at 0.003 ($p = 0.01$).

5.3.2 Outlier analyses using dominant AFLPs

The BayeScan v. 2.1 outlier analysis using dominantly scored AFLP loci found no outlier loci when the ratio of prior probabilities of non-selection to selection was set to 10:1 or 1:1 (Figures 5.2 and 5.3). However, one locus under the 1:1 priors scenario had

a *q* value of 0.255, suggesting that this locus has a 1/4 chance of being a false positive result.

5.3.3 Outlier analyses using codominant AFLPs

The BayeScan v. 2.1 outlier analysis using codominantly scored AFLP loci found no outlier loci when the ratio of prior probabilities of non-selection to selection was set to 10:1 (Figure 5.4); however, when the ratio was set to 1:1, 6 outlier loci were detected (Figure 5.5) with *q* values below 0.25. This suggests that 1–2 of these detected loci are likely to be false positives.

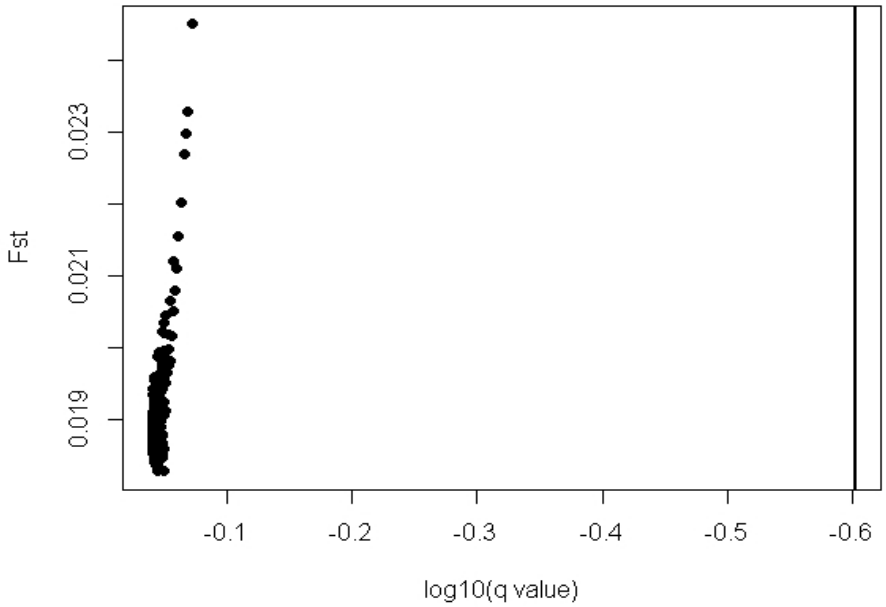

Figure 5.2. Locus-specific F_{ST} plotted against the logarithm of the q value for each dominantly scored AFLP locus; data estimated using a 10:1 odds ratio of priors. No outlier loci were detected using q values of either 0.10 or 0.25; the q = 0.25 cut-off point is shown on the right-hand side of the graph.

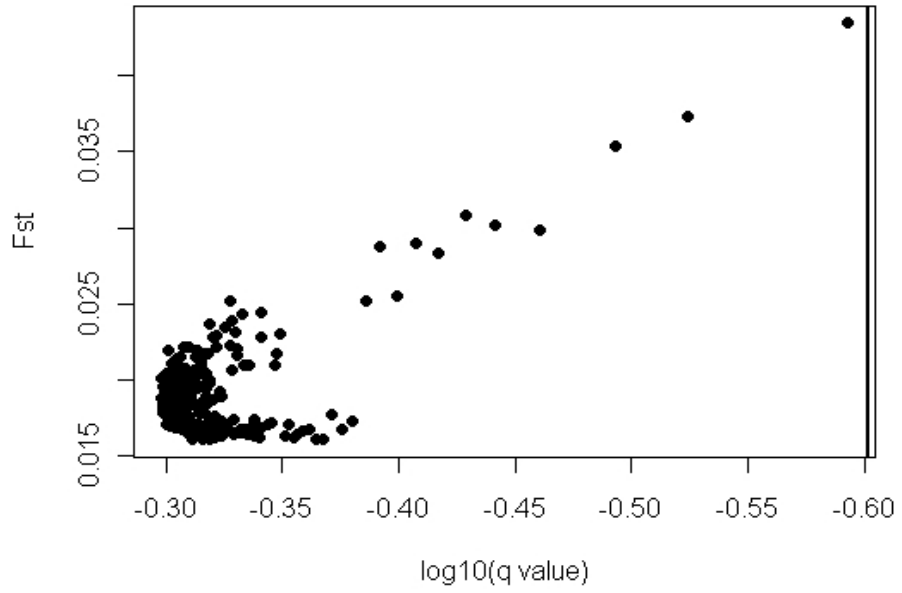

Figure 5.3. Locus-specific F_{ST} plotted against the logarithm of the q value for each dominantly scored AFLP locus; data estimated using a 1:1 odds ratio of priors. No outlier loci were detected using either a q value of 0.10 or 0.25; however, one outlier (q = 0.255) was very close to q = 0.25; this cut-off point is shown on the right-hand side of the graph.

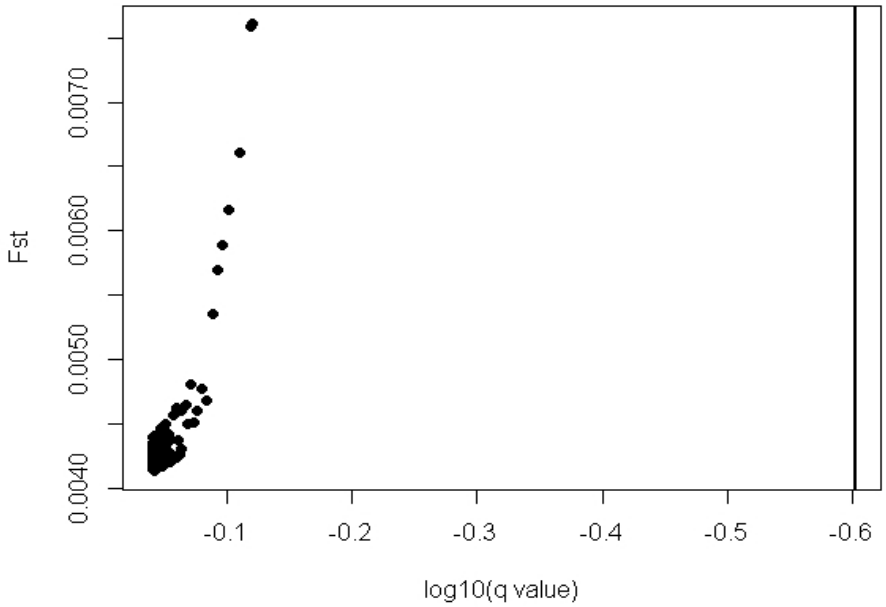

Figure 5.4. Locus-specific F_{ST} plotted against the logarithm of the q value for each codominantly scored AFLP locus; data estimated using a 10:1 odds ratio of priors. No outlier loci were detected using a q value of either 0.10 and 0.25; the q = 0.25 cut-off point is shown on the right-hand side of the graph.

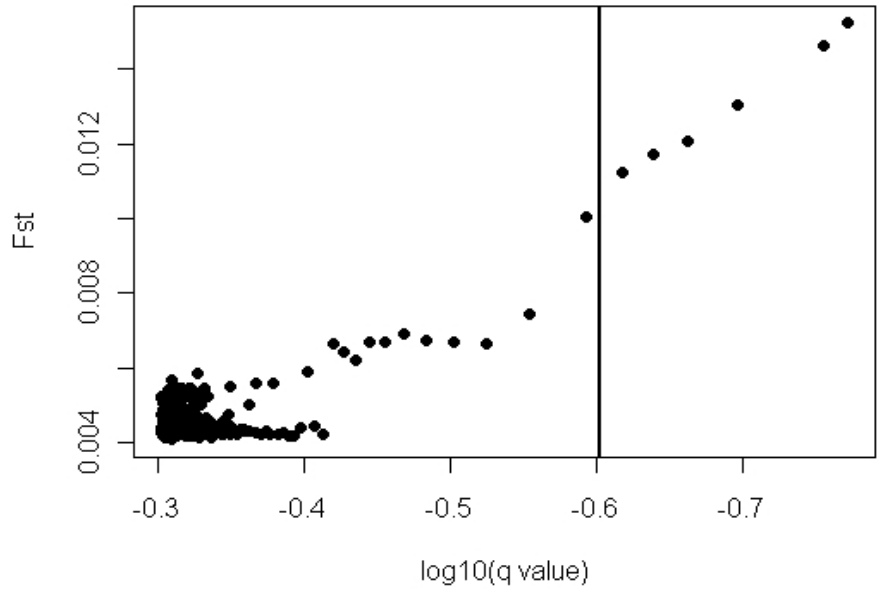

Figure 5.5. Locus-specific F_{ST} plotted against the logarithm of the q value for each codominantly scored AFLP locus; data estimated using a 1:1 odds ratio of priors. No outlier loci were detected using a q value of 0.10; 6 outliers were detected with a q value < 0.25; this cut-off point is shown on the graph.

5.3.4 Characteristics of detected outlier loci

Table 5.3 lists the 7 outlier loci detected, reporting their *q* values, the posterior probabilities of selection, and the alpha selection coefficient as estimated by BayeScan v. 2.1 (Foll, 2012). Three of these loci were found to have also been scored in the experimental pedigree reported in Chapter 4. These three loci were investigated for their segregation behaviour under the three experimental growth conditions used in Chapter 4; that is, when the BCF_1 *S. hermonthica* were grown on agar without a host, on the sorghum cultivar E36 and on the sorghum cultivar Malisor (Table 5.4). All three outlier loci were found to have been distorted in the BCF_1 under at least one growth condition: locus *Ec*ATC.*Ms*CGC.178 was distorted within the populations growing on agar, E36 and Malisor; locus *Ec*ATC.*Ms*GCT.203 was distorted only within the BCF_1 population growing on E36; and locus *Ps*TCC.*Ms*GGC.288 was distorted for all conditions (Table 5.4). Generally, few differences were found for these three loci between reciprocal populations for a given growing condition, although locus *Ec*ATC.*Ms*CGC.178 was found to be distorted within both reciprocal populations but not overall. This was due to the distortion being in different directions in the separate reciprocal populations. Two outlier loci (*Ec*ATC.*Ms*CGC.178 and *Ec*ATC.*Ms*GCT.203) were found to have significantly different segregation patterns between the E36 and Malisor growing conditions, whilst no locus was found to segregate differently between the agar and plant conditions (Table 5.4).

Table 5.3. Outlier AFLP loci detected in genome scans using BayeScan v. 2.1 with 1:1 prior odds ratio. (*Ec* = *Eco*RI; *Ms* = *Mse*I; *Ps* = *Pst*I.)

AFLP outlier loci	*P*(selection)	*q* value	α	Scored in pedigree?	Mapped? (linkage group)
Dominant scan					
*Ec*ATC.*Ms*GCT.109	0.745	0.255	0.793	N	-
Codominant scan					
*Ec*ATC.*Ms*CGC.178	0.831	0.169	1.079	Y	Y (LG7)
*Ec*ATC.*Ms*GCT.203	0.818	0.175	1.038	Y	N
*Ec*ATC.*Ms*CGT.141	0.748	0.201	0.890	N	-
*Ec*ATC.*Ms*CGT.226	0.734	0.217	0.829	N	-
*Ps*TCC.*Ms*GGC.288	0.723	0.229	0.795	Y	N
*Ec*AGA.*Ms*CGT.347	0.700	0.241	0.727	N	-

Table 5.4. Outlier AFLP loci also scored in the BCF_1 population, as assessed for segregation distortion (SD) and relative differences in segregation pattern using a Bonferroni-corrected cut-off of $p = 0.00028$ (locus-by-locus $\alpha = 0.05$). The data are presented for the overall BCF_1 population, with the situation for SD in the reciprocal populations given in parentheses as (BCF_1HyMo/ BCF_1HyFa). (*Ec* = *Eco*RI; *Ms* = *Mse*I; *Ps* = *Pst*I.)

AFLP outlier loci	Locus showed segregation distortion in BCF_1?					Relative difference in BCF_1 segregation pattern at locus?	
	Across all conditions	Agar	Plant (E36 + Malisor)	E36	Malisor	Agar v. Plant	E36 v. Malisor
Codominant scan							
*Ec*ATC.*Ms*CGC.178	N (Y/Y)	Y (Y/Y)	N (Y/Y)	Y (N/Y)	Y (Y/Y)	N	Y
*Ec*ATC.*Ms*GCT.203	N (N/N)	N (N/N)	N (N/N)	Y (Y/N)	N (N/N)	N	Y
*Ps*TCC.*Ms*GGC.288	Y (Y/Y)	Y (Y/Y)	Y (Y/Y)	Y (Y/Y)	Y (Y/Y)	N	N

5.4 Discussion

5.4.1 Genetic variation and differentiation in the S. hermonthica *Kibos 1997 population*

In this chapter I have used an experimental outlier analysis to investigate the potential host selection of AFLP markers within one population of *S. hermonthica*. This population was previously investigated with respect to both segregation and virulence in relation to the two sorghum hosts used in the current experiment (Chapter 4). The genetic diversity (H_{exp} = 0.259; PLP at 5% = 81.7%) of the *S. hermonthica* Kibos 1997 population sample genotyped here was very similar to the results obtained by Huang *et al.* (2012) for samples of the same population parasitising three rice cultivars (H_{exp} = 0.234; PLP at 5% = 72.8%). F_{ST} was found to be lower in the current study (0.003) than for the experiment reported by Huang *et al.* (2012) (F_{ST} = 0.013), but was very close to that reported in Chapter 3 of this thesis for a West African *S. hermonthica* field population sampled across 9 sorghum cultivars (F_{ST} = 0.004).

Restricting the comparison to the two rhizotron-based lab experiments using the same *Striga* population, the difference in F_{ST} may be due to the higher selective pressure exerted on the *Striga* population by the rice cultivars in Huang *et al.* (2012) compared to the sorghum cultivars used here. This suggests that the resistance mechanisms which act in Malisor against the populations of West African *Striga* (Chapter 2) still allow for the parasitism of a larger variety of individual genotypes from the East African Kibos 1997 population compared to the resistant rice cultivar Nipponbare; that is, the higher population F_{ST} in Huang *et al.* (2012) may indicate that host selection is affecting a larger proportion of the genome compared to the sorghum hosts in the current chapter.

The similarity in F_{ST} between the West African field population sampled in Chapter 3 and the Kibos 1997 population sampled in the current chapter is surprising given that

some of the cultivars in the West African experiments, including Malisor, showed strong resistance to the *S. hermonthica* populations used, both in the lab and the field (Chapter 2). The low F_{ST} in the field experiment may be due to environmental variation swamping any host effect, or due to the technical issues associated with scoring a larger number of AFLP loci discussed in Chapter 3.

5.4.2 The host selection of outlier loci

Seven loci with reasonable evidence of selection were detected from outlier analyses of AFLPs between *S. hermonthica* Kibos 1997 samples growing on the sorghum cultivars E36 and Malisor. This indicates that, despite no overall difference in *Striga* attachment between the cultivars (Chapter 4), some differential genetic response was still occurring. The differential SD results obtained in Chapter 4 for the BCF_1 population are therefore likely to have been at least partially driven by differential interactions between the sorghum cultivars and Kibos 1997 alleles, even if there was also active selection against I.C. 1997 alleles by Malisor. Further evidence for the differential selection of the identified Kibos 1997 outlier loci is provided by the fact that, of the three outlier loci that were also scored in the BCF_1 population, two were found to have significantly different segregation patterns between the E36- and Malisor-grown BCF_1 populations.

Despite the generally high levels of SD on the plant hosts (28–50%, depending on the reciprocal family), and the significant difference in SD observed between E36 and Malisor (Chapter 4), the identification here of a maximum of 7 positively selected outliers (2.2% of loci scored) appears to indicate that strong directional selection is rare, but that weaker selective forces may be common across the genome (although, again, the caveat that some of the SD difference between the hosts is likely to be due to host/I.C. 1997 allelic interactions should be stated). Genome scans in plants have, on

average, identified 8.9% of loci (range = 0.4–35.5%) as outliers in any given set of markers (Strasburg *et al.* 2012), whereas investigations of intraspecific SD have often found much higher proportions of affected loci, similar to the levels reported in this thesis (e.g. Jenczewski *et al.* 1997; Bratteler *et al.* 2006; McDaniel *et al.* 2007). Strasburg *et al.* (2012) suggest that "loci identified in genome scans for selection probably reflect an unrepresentative subset of traits whose genetic architecture lends itself more easily to such detection". These authors point out that differences between hard sweeps (selection on newly arisen mutations) and soft sweeps (selection on pre-existing genetic variation) may lead to different probabilities of detection in outlier analyses (also see Prezeworski *et al.* 2005; Barrett & Schluter 2008). Pritchard *et al.* (2010) suggest that traits based on polygenic standing variation could also allow for rapid adaptation whilst not producing the classical signatures of selection characteristic of hard sweeps. The epistasis found in the BCF_1 population (Chapter 4) suggests that this may be the case in the current *Striga*-sorghum system. Recent work on plant host-associations in the apple maggot fly *Rhagoletis pomonella* has also indicated that selection may be widespread across the genome, even when few F_{ST} outliers are detected by standard methods (Michel *et al.* 2010). Selection acting via soft sweeps seems likely to underlie the host selection found in the current chapter and in Chapter 4; these experiments, by their nature as experimental tests of host-selection conducted on an undifferentiated population over only one generation, will clearly favour adaptive responses based on pre-existing genetic variation (Barrett & Schluter 2008). This conclusion allows for the integration of the considerable host-driven SD, and differential segregation, observed in Chapter 4 with the small number of outlier loci observed in the current chapter; it also supports the contention that selection may be

much more widespread across genomes than the 8.9% average found in plant outlier analyses suggests (Strasburg *et al.* 2012). The idea of outlier versus non-oulier loci may, then, actually misrepresent the genomic reality of selection (Butlin 2010).

5.4.3 The prior probability of selection and the detection of outlier loci

An important feature of the analysis reported here is that the detection of outliers was dependent both on the way in which AFLP markers were scored and on the prior probability assigned to the model including an effect of selection at a locus: the most outliers were detected for codominantly-scored AFLP using a 1:1 odds ratio for the prior probabilities of selection to non-selection at a locus. Codominant AFLP scoring has previously been found to provide more information for the estimation of allele frequencies, and so to improve the detection of outliers (Fischer *et al.* 2011), but there has been little clear discussion on the choice of a prior probability for an effect of selection within the outlier analysis literature.

Beaumont & Balding (2004) were the first authors to put forward a Bayesian method for outlier analyses, with the probability of selection at a locus being estimated by the posterior distribution of the locus-specific effect alpha. These authors also noted that an extension of their method, using Bayesian model selection, would allow for the use of different prior probabilities for the opposing models (selection v. non-selection), thus allowing researchers to address the issue of multiple testing, to include prior information about loci, and to directly make probability statements about selection at a locus (Beaumont & Balding 2004). However, the outlier detection method described in Beaumont & Balding (2004), and incorporated into the popular Bayesian outlier detection software 'BayesFst', did not include this extension. BayesFst attempts to address the issue of multiple testing by the setting of a Gaussian prior with a mean of

zero and a standard deviation of 1 for the locus-specific effect alpha. For an effect of selection to be accepted for a locus, the posterior density of alpha has to exclude zero at some specified two-tailed level *P*; this can be thought of as a Bayesian *p* value, *P*(data | null model), rather than a Bayesian posterior probability, *P*(selection | data) (Beaumont & Balding 2004; Riebler *et al.* 2008). Although BayesFst does not allow for a formal statement of the posterior probability of selection at a locus, it appears to yield results that are marginally better (in terms of the type 1 and 2 error rates) than the earlier frequentist method 'Fdist' (Beaumont & Balding 2004).

The alternative Bayesian model selection method briefly put forward in the discussion of Beaumont & Balding (2004) was first implemented by Riebler *et al.* (2008). Riebler *et al.* (2008) used a Bernoulli-distributed auxiliary variable to estimate the posterior probability of selection at a locus; a beta prior was used for this variable, which allowed the authors to formally address the problem of multiple testing, setting the parameters of the beta distribution such that only a small fraction of loci (10%) were expected to be under selection *a priori*. Using simulated data, the model selection method of Riebler *et al.* (2008) was found to correctly identify all loci under directional selection, with a lower false positive rate (higher specificity) for any given power (true positive rate), compared to the Bayesian method without model selection. However, when analysing a *Drosophila melanogaster* allozyme dataset previously investigated by Beaumont & Balding (2004), the new method appeared to be much more conservative, identifying only 1 of the 10 loci selected by the method of Beaumont & Balding (2004) as being under selection (Riebler *et al.* 2008); of course, for these empirical data, the true situation is not known.

Independently, a similar Bayesian model selection method was developed by Foll &

Gaggiotti (2008), who incorporated their method into the now popular software 'BayeScan'. Under the Foll & Gaggiotti (2008) method, the posterior probabilities of both models, with and without selection, are estimated for a locus. The final estimate of selection at a locus, $P(\text{alpha} \neq 0)$, is estimated from the number of times that the selective effect is incorporated into the model at a locus. This approach also allows for the setting of an odds ratio for the prior probability of selection to non-selection, which was set at 1:1 for BayeScan v. 1.0 (Foll & Gaggiotti 2008); in common with Beaumont & Balding (2004), a normal prior with a mean of zero and a standard deviation of 1 was used by Foll & Gaggiotti (2008) for the locus-specific effect alpha.

The recent simulation-based study of Pérez-Figueroa *et al.* (2010) compared BayeScan v. 1.0 with two other outlier analysis methods (Dfdist and DetselD). Dfdist is the dominant marker version of Fdist (Beaumont & Nichols 1996); DetselD is an unpublished version of Detsel (Vitalis *et al.* 2003), also for dominant data (Pérez-Figueroa *et al.* 2010). Pérez-Figueroa *et al.* (2010) found that BayeScan v. 1.0 performed in a similar way to the FDR-corrected Dfdist; for example, for a neutral F_{ST} of 0.025, and mean selection coefficients of 0.5 and 0.05, BayeScan and Dfdist had similar false positive and false negative rates, whilst DetselD had a higher rate of false negatives. The results of Pérez-Figueroa *et al.* (2010) suggest that, in common with the BayesFst/Fdist comparison performed by Beaumont & Balding (2004), the prior distribution for alpha used in BayeScan corrected for multiple testing at a similar level to the FDR-corrected frequentist method Dfdist.

Fischer *et al.* (2011) introduced a modified version of BayeScan, version 2.0, which included the ability for the user to change the prior odds ratio of selection versus non-selection. These authors set the default ratio to 10:1, comparable to the strategy used by

Riebler *et al.* (2008); this approach should minimise the number of false positives, for a given power (Riebler *et al.* 2008). The decision of Fischer *et al.* (2011) to set their default odds ratio to 10:1, with little discussion of the choice, is perhaps unfortunate, as recent studies now appear to be in two minds as to which prior odds ratio is the most appropriate. For example, Girard & Anger (2011), in their analysis of the mechanisms driving patterns of diversity of functional genes in the longnose dace *Rhinichthys cataractae*, presented separate analyses of selection using BayeScan v. 2.01 with prior odds of both 1 and 10 for the model with selection, stating that, "[i]ncreasing the prior odds value will tend to eliminate [false positives] but does so at the cost of reducing the power of the method", even though Riebler *et al.* (2008) showed in their simulations that reducing the prior odds of selection actually increased the specificity (i.e. reduced the proportion of false positives) without reducing the power to detect directional selection. Deagle *et al.* (2012), in their analysis of parallel evolution in the three-spined stickleback *Gasterosteus aculeatus*, also used separate BayeScan v. 2.01 tests with prior odds of both 1 and 10, using the more stringent prior odds for the global analysis, whilst using the 1:1 prior odds test for the multiple independent local comparisons. In an investigation into an adaptive radiation of the Midas cichlid fish (*Amphilophus citrinellus*) species complex in Nicaragua, Kautt *et al.* (2012) followed the strategy of Deagle *et al.* (2012), although adding the qualification that the outliers detected in the local analyses would not be followed up unless they were also detected in the more conservative global analysis. This brief survey of recent Bayesian outlier analysis strategies confirms the conclusion of Butlin (2010) and Strasburg *et al.* (2012) that different approaches make generalisations about the extent of genomic selection across taxa hazardous.

In their overview of Bayesian statisical methods in genetic association studies, Stephens & Balding (2009) recommend the use of the prior distribution to correct for multiple testing, pointing out that if the prior probability of an effect at a locus is assumed to be constant across loci, then this probability can also be interpreted as an estimate of the overall proportion of loci that are associated with a phenotype. This is the approach taken by Riebler *et al.* (2008) and Fischer *et al.* (2011) in setting the prior probability of selection at 10:1; that is, around 10% of loci are expected, *a priori*, to be under selection. However, given that the simulation studies discussed above (Beaumont & Balding 2004; Pérez-Figueroa *et al.* 2010) have found that Bayesian methods without a prior odds adjustment seemingly already correct for multiple testing, due to the prior distribution used for the selection effect alpha (at least in comparison with FDR-corrected frequentist methods) it appears that further investigation of the necessity for the adjustment of the prior odds ratio in Bayesian outlier analysis methods is required.

Bayesian statisticians often incorporate 'loss functions' into their analyses (Scott & Berger 2006), where the cost of following up a false positive can be explicitly included in decision making. Given the declining cost of population genomic analyses, the relative absence of ethical issues in ecological and evolutionary studies, the relative or absolute lack of prior information concerning loci in non-model organisms, and the ability to choose outliers based on an FDR cut-off (Foll 2012), the use of a 1:1 prior odds ratio currently still seems to be the best approach for locating candidate outlier loci in non-model species. False positives seem unlikely to have very serious consequences in genic tests for selection, especially if, in general, there is more likely to be a spectrum of selective effects rather than a cut-off, due to the action of soft sweeps (Pritchard *et al.* 2010), and where the existence of multiple algorithmic approaches make between-

species comparisons of the proportion of sites affected by selection potentially highly misleading and of dubious utility.

5.4.4 Conclusions

This chapter has provided further evidence for the presence of genetic variation for host adaptation within populations of *S. hermonthica*. Surprisingly, even when phenotypic evidence does not indicate the presence of differential interactions at the population level, subtle host adaptation may still be occuring. The use of a *Striga* population previously used to produce an inter-population pedigree (Chapter 4) has also provided a link between evidence for widespread effects of host identity on segregation patterns and the strong directional selection detected by an intra-population outlier analysis. In particular, the contrast between the apparent extent of these effects was revealed. The effects of the importance of the prior odds ratio for detecting selection was also demonstrated.

6. General Discussion

6.1 Conclusions

Striga hermonthica has been considered the largest biological constraint to food production in sub-Saharan Africa (Parker & Riches 1993), and its impact as a weed of subsistence agriculture in that region is thought to be increasing (Ejeta 2007; Parker 2009). Currently, relatively little resistant germplasm is available in host cereal crops, therefore it is important that what resistance is available is deployed with an awareness of how genetic variation within and between populations of *S. hermonthica* may affect its durability (Scholes & Press 2008). The main research paradigm in this area has been that of host specificity, where populations of *S. hermonthica* have been investigated for particular virulence reactions against sets of host species or cultivars. I have earlier emphasised (Chapter 1, section 1.4) the conceptual problems associated with using this framework for an obligate outbreeding pathogen in an agroecosystem: there is doubtless value in characterising geographic zones containing races of host-specific *Striga* for the strongly inbreeding species *S. asiatica* (Botanga *et al.* 2002) and *S. gesnerioides* (Botanga & Timko 2006), but for the rapidly adaptable and outcrossing *S. hermonthica*, we may doubt the utility of this approach. This doubt is amplified by the fact that, in order to test for host adaptation, many studies have simply assumed that observed field hosts of sampled *Striga* represent those hosts to which the parasites are uniquely adapted, and because *ad hoc* sampling regimes can easily confound geographic variability with host identity (e.g. Olivier *et al.* 1998; Yoshida *et al.* 2010; Welsh & Mohamed 2011; Estep *et al.* 2011). Additionally, neutral genetic variation is, by definition, unable to provide an insight into adaptation at functional genetic loci in

situations where these are restricted in genomic extent, and regular exchange of non-adaptive genetic material occurs between incipient host-adapted *Striga* sub-populations (i.e. host-related lineage sorting is only present at adaptive loci; Bierne *et al.* 2013).

In this thesis, I have expanded the toolbox of approaches used to investigate the patterns and process of host adaptation in *S. hermonthica*. I achieved this via several routes:

(1) Via lab-based investigations of differential virulence between populations of *S. hermonthica* and genotypes of its sorghum host; I developed this approach to consider the relative extent of adaptation at different life history stages, and the extent to which such population-level G × G interactions examined in the lab can be used to predict actual outcomes in the field over several years.

(2) Via the investigation of the within-population component of host adaptation; showing, for the first time, that locus-specific signatures of selection associated with host genotype can be identified in the field.

(3) Via the creation of an experimental pedigree, allowing a preliminary investigation of heterosis, epistasis, segregation distortion and maternal identity as genetic mechanisms associated with host adaptation in *S. hermonthica*.

And, (4), via the use of an outlier analysis to investigate the correspondence of the results from this methodology with data on locus-specific segregation distortion obtained from the experimental pedigree. This last experiment also indicated that locus-specific selection can occur at the individual-level, even when studies of differential virulence do not indicate population-level G × G interactions.

Overall, these experiments suggest a complex basis to virulence in *S. hermonthica*, that seems unlikely to be resolved into a simple, qualitative, genetic mechanism, such as has been found to be the case between certain races of *S. gesnerioides* and cultivars of its

cowpea host (Timko *et al.* 2012). However, the results presented in this thesis do not prove the absence of resistance gene-based, qualitative interactions between *S. hermonthica* and its hosts. Indeed, the distinction between qualitative and quantitative virulence may be blurred in some pathosystems (Burdon *et al.* 1996). Host populations can contain multiple resistance genes: Burdon and colleagues (1996, 2006) report that there is evidence for up to 10 resistance genes, and possibly more, in some pathosystems; at the corresponding, outbred, parasite population level this may appear as a form of quantitative virulence, as with the parasite attachment measure often used for *S. hermonthica*. This seems especially likely to be the case for genetically diverse pathogen populations, and could lead to pathosystems of 'considerable genotypic and phenotypic complexity' (Burdon *et al.* 2006). Ecological predictions concerning the occurrence of qualitative resistance have been put forward suggesting that endemic pathogens (that is, continuously present ones, like *S. hermonthica*) may be less likely to exhibit gene-for-gene interactions with their hosts than pathogens causing disease in a boom-and-bust, epidemic fashion (Burdon *et al.* 1996). Burdon *et al.* (1996) suggested that, in wild pathosystems, qualitative resistance is most likely to occur where the host only encounters particular pathogen genotypes periodically, with qualitative resistance acting to reduce the chance of a pathogen establishing in a particular host population. The contention of Burdon *et al.* (1996) is that, where pathogens undergo boom-and-bust dynamics in local host populations, different pathogen genotypes may be present in the landscape, and, therefore, a population with qualitative resistance to particular pathogen genotypes may escape a proportion of infection events, with a corresponding increase in the fitness of the host population proportional to the frequency of the relevant resistance genes in the population, and to the frequency of particular pathogen genotypes in the

landscape. Notably, some of the conditions put forward by Burdon *et al.* (1996) as promoting qualitative resistance overlap with those predicted to decrease pathogen effective population size and increase genetic structure (Chapter 1, section 1.6; Huyse *et al.* 2005; Barrett *et al.* 2008); for example, the presence of epidemics, and regular pathogen local extinction events. Whilst *S. hermonthica* exhibits many life history features thought to be associated with low genetic structure and large effective population size (e.g. multiple hosts, sexual reproduction, human-aided long-distance dispersal, stable populations due to the ability to survive as seeds until hosts are available; Huyse *et al.* 2005; Barrett *et al.* 2008), the impacts of the environment on *Striga* and its hosts may cause occasional reductions in effective population size; one study has shown that adverse environmental conditions can considerably reduce the *S. hermonthica* seed bank (Gbèhounou *et al.* 2003); therefore, in some areas, conditions may make the occurrence of qualitative resistance interactions more likely. Furthermore, qualitative, gene-for-gene systems have been discovered in other plant pathogens with obligately outbreeding life histories (e.g. the biotrophic wheat smut fungus *Ustilago hordei*; Martínez-Espinoza *et al.* 2002).

Overall, it seems that considerably more work on obligately outbreeding parasites will be required before strong generalisations can be made about the nature of their expected interactions with hosts in different pathosystems. The theory concerning the evolution of qualitative resistance put forward by Burdon *et al.* (1996), and the observations of McDonald & Linde (2002), Huyse *et al.* (2005), and Barrett *et al.* (2008) on the relationships between parasite life history and population genetics, together suggest that gene-for-gene systems may be relatively uncommon in outbred parasites, or, of a level of complexity such that the appearance of a quantitative interaction is generated at the

population-level through the presence of multiple genotypes. In theory, if the within-population diversity of *S. hermonthica* virulence were founded upon different constituent sub-populations exhibiting different qualitative, gene-for-gene interactions with different hosts, then we might have expected to find very strong locus-specific outliers in our analyses. If an outlier were linked to a virulence gene, targeted by a resistance gene in one host-selected sub-population but not another, then we should expect that parasite locus to be completely differentiated between sub-populations. This is the standard, strong, trade-off implied by qualitative, gene-for-gene resistance, where the presence of a virulence gene in a particular pathogen genotype is detected by a corresponding host resistance gene (thereby becoming an avirulence gene instead; Agrios 2005). In a monocultural agroecosystem, host resistance genes may be rapidly overcome by parasites, rendering them useless (Thompson & Burdon 1992). In this case, we might expect that alternative, unnecessary, virulence mechanisms would be lost from the population if they imposed some cost on the parasite. For *S. hermonthica*, the presence of a seed bank, and regular changes of host crop, especially for populations from experimental stations, might mean that multiple virulence mechanisms could be retained in the longer-term, especially if they were not detected by the resistance mechanisms of the hosts encountered. However, evidence for a qualitative basis to differential virulence in *S. hermonthica* was not found in the current work. Very highly differentiated loci were not identified (Chapters 3 and 5), and neither were they reported by Huang *et al.* (2012); it should however be recalled that the detection of outliers is dependent on how thoroughly the molecular markers used cover the genome: qualitative virulence loci could easily have been missed if they were not in linkage disequilibrium with the molecular markers scored.

The presence of population-level differential virulence between *S. hermonthica* and its hosts at a particular life-stage, which can be influenced by the environment (Chapter 2), the discovery of host-selected loci in the field that do not map straightforwardly to commonly used metrics of *Striga* population-level virulence (Chapter 3), and the strong evidence for epistasis and host-specific segregation distortion (Chapter 4), which can partly be linked to locus-specific selection (Chapter 5), all suggest that the genetic basis of the *S. hermonthica*-sorghum pathosystem involves interactions among many loci in the parasite and host genomes. As previously observed (Chapter 5, section 5.4), approaches based on F_{ST} outlier analyses may only detect a subset of loci that have historically undergone, or are currently undergoing selection (Michel *et al.* 2010; Strasburg *et al.* 2012), and additional approaches, such as those presented in this thesis, that investigate the potential for phenomena such as standing genetic variation (Pritchard *et al.* 2010) and epistasis (Wolf *et al.* 2000) to contribute to adaptation, may broaden our understanding of this most complex process.

6.2 Future Directions for Research

It is unfortunate that true genotype-by-genotype interaction studies cannot be conducted easily for *S. hermonthica*; even though genotypes could be asexually propagated using similar techniques to those used in Chapter 4 of this thesis, initiating functional parasitism for clones cultivated *in vitro* would be extremely challenging, and the impacts of single parasites on host fitness may be very low (Gurney *et al.* 1999), and of little relevance for agroecosystems. Population genetic methods are therefore likely to remain the best way forward for the characterisation and discovery of genetic variation for virulence in *S. hermonthica*. The use of modern spatial genetic methods to characterise significant barriers to gene flow between demes (e.g. Welsh & Mohamed

2011), and the use of multiple, hierarchical levels of *F* statistics to apportion genetic variation within and between populations at different spatial scales (Goudet 2005), will increase our understanding of the population structures exhibited by *S. hermonthica.* Human movement of genotypes may confuse this picture, but it is also possible that the movement of migrants between populations could be revealed using genetic 'assignment' methods (Manel *et al.* 2005). However, for the identification of functional variation, and an understanding of the genetic basis of differential virulence, sampling regimes using replicated field trials and sequenced markers, rather than PCR-based, anonymous ones, are likely to be the most effective approach for identifying virulence loci, particularly if detailed information on the molecular basis of resistance in the host cultivars used is available. Despite calls for a better understanding of the genetic structure of *S. hermonthica* (Mohamed *et al.* 2007), practically, for the deployment of resistant cultivars, it seems that the most important requirement is for the widespread appreciation by plant breeders of the fact that different *S. hermonthica* populations may differ in virulence, and that this requires a precautionary approach to cultivar deployment, as well as comprehensive monitoring for host resistance breakdown. As previously stated, knowledge of broad population genetic structure of *S. hermonthica* may not be a guarantee of uniform virulence characteristics within populations (e.g., see Burdon & Thrall 2000 for an exemplary review of how spatial scale influences host-pathogen interactions in the *Linum marginale-Melampsora lini* pathosystem). Functional sites for virulence, and the interaction of virulence traits with the environment, may vary across the landscape where neutral markers do not (Barrett *et al.* 2008; Bierne *et al.* 2013).

In the absence of detailed historical information about different populations of *S.*

hermonthica and the hosts that have been grown in different places at different times, and of knowledge of the impacts of the environment upon these interactions, studies that seek to unpick the effects of different hosts and environments on virulence via well designed, replicated field trials and supporting lab analyses, and that attempt to associate particular genetic loci with these extrinsic phenomena, seem the best way forward. The rapid development of sophisticated next-generation sequencing techniques that generate thousands of single-nucleotide polymorphism markers (Baird *et al.* 2008; Seeb *et al.* 2011), and the development of multivariate ordination techniques that can partition genetic variation between spatial and environmental causes (Manel *et al.* 2010, 2012), suggest that, with the correct infrastructure, support and collaborations, rapid progress could be made on the genetic basis of *S. hermonthica* virulence in the coming years.

7. References

Agrawal AF, Lively CM (2003) Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 323–334.

Agrios GN (2005) *Plant Pathology*. Elsevier Academic Press, USA.

Aigbokhan EI, Berner DK, Musselman LJ (1998) Reproductive ability of hybrids of *Striga aspera* and *Striga hermonthica*. *Phytopathology*, **88**, 563–567.

Alheit KV, Reif JC, Maurer HP *et al.* (2011) Detection of segregation distortion loci in triticale (× *Triticosecale* Wittmack) based on a high-density DArT marker consensus genetic linkage map. *BMC Genomics*, **12**, 380.

Ali RAMA, El-Hussein AA, Mohamed KI, Babiker AGT (2009) Specificity and genetic relatedness among *Striga hermonthica* strains in Sudan. *Life Sciences International Journal*, **3**, 1159–1166.

Anderson DR (2008) *Model Based Inference in the Life Sciences: A Primer on Evidence*. Springer.

Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, **22**, 699–708.

Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics*, **27**, 258–266.

Angeles J, Laurena A, Tecson-Mendoza E (2005) Extraction of genomic DNA from the lipid-, polysaccharide-, and polyphenol-rich coconut (*Cocos nucifera* L.). *Plant Molecular Biology Reporter*, **23**, 297–298.

Arrigo N, Tuszynski JW, Ehrich D, Gerdes T, Alvarez N (2009) Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics*, **10**, 33.

Atera E, Itoh K, Asuma T, Ishii T (2012) Response of NERICA rice to *Striga hermonthica* infections in western Kenya. *International Journal of Agriculture & Biology*, **14**, 271–275.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*. Sinauer Associates, Sunderland, MA.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barkman TJ, McNeal JR, Lim S-H *et al.* (2007) Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology*, **7**, 248.

Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J (2009) Continua of specificity and virulence in plant host–pathogen interactions: causes and consequences. *New Phytologist*, **183**, 513–529.

Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008) Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends in Ecology & Evolution*, **23**, 678–685.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Beaumont MA (2005) Adaptation and speciation: What can F_{ST} tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **263**, 1619–1626.

Bebawi FF, Awad AE, Khalid SA (1986) Germination, host preference, and phenolic content of witchweed (*Striga hermonthica*) seed populations. *Weed Science*, **34**, 529–532.

Bechsgaard J, Bataillon T, Schierup MH (2004) Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*. *Journal of Evolutionary Biology*, **17**, 554–561.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, **125**, 279–284.

Berner DK, Cardwell KF, Faturoti BO, Ikie FO, Williams OA (1994) Relative roles of wind, crop seeds, and cattle in dispersal of *Striga* spp. *Plant Disease*, **78**, 402.

Berner DK, Kling JG, Singh BB (1995) *Striga* research and control - A perspective from Africa. *Plant Disease*, **79**, 652–660.

Berner DK, Winslow MD, Awad AE *et al.* (Eds.) (1997) *Striga Research Methods - A Manual*. International Institute of Tropical Agriculture, Ibadan, Nigeria.

Beukeboom LW, Niehuis O, Pannebakker BA *et al.* (2010) A comparison of recombination frequencies in intraspecific versus interspecific mapping populations of *Nasonia*. *Heredity*, **104**, 302–309.

Bharathalakshmi, Werth CR, Musselman LJ (1990) A study of genetic diversity among host-specific populations of the witchweed *Striga hermonthica* (Scrophulariaceae) in Africa. *Plant Systematics and Evolution*, **172**, 1–12.

Bierne N, Gagnaire PA, David P (2013) The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, **59**, 72–86.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Bikard D, Patel D, Metté CL *et al.* (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science*, **323**, 623–626.

Bolker BM (2008) *Ecological Models and Data in R*. Princeton University Press, USA.

Bonin A, Bellemain E, Bronken Eidesen P *et al.* (2004) How to track and assess

genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: A toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Bonin A, Pompanon F, Taberlet P (2005) Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. *Methods in Enzymology*, **395**, 145–161.

Botanga CJ, Kling JG, Berner DK, Timko MP (2002) Genetic variability of *Striga asiatica* (L.) Kuntz based on AFLP analysis and host-parasite interaction. *Euphytica*, **128**, 375–388.

Botanga CJ, Timko MP (2005) Genetic structure and analysis of host and nonhost interactions of *Striga gesnerioides* (witchweed) from central Florida. *Phytopathology*, **95**, 1166–1173.

Botanga CJ, Timko MP (2006) Phenetic relationships among different races of *Striga gesnerioides* (Willd.) Vatke from West Africa. *Genome*, **49**, 1351–1365.

Bradshaw WE, Holzapfel CM (2000) The evolution of genetic architectures and the divergence of natural populations. In: *Epistasis and the Evolutionary Process* (eds Wolf JB, Brodie III ED, Wade MJ), pp. 245–263. Oxford University Press, New York, NY.

Bratteler M, Lexer C, Widmer A (2006) A genetic linkage map of *Silene vulgaris* based on AFLP markers. *Genome*, **49**, 320–327.

Buerkle CA, Rieseberg LH (2001) Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution*, **55**, 684–691.

Burdon JJ, Thrall PH (2000) Coevolution at multiple spatial scales: *Linum marginale–Melampsora lini* – from the individual to the species. *Evolutionary Ecology*, **14**, 261–281.

Burdon JJ, Thrall PH (2009) Coevolution of plants and their pathogens in natural

habitats. *Science*, **324**, 755–756.

Burdon JJ, Thrall PH, Ericson L (2006) The current and future dynamics of disease in plant communities. *Annual Review of Phytopathology*, **44**, 19–39.

Burdon JJ, Wennström A, Elmqvist T, Kirby GC (1996) The role of race specific resistance in natural plant populations. *Oikos*, **76**, 411.

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, **65**, 23–35.

Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409–418.

Byers D (2005) Evolution in heterogeneous environments and the potential of maintenance of genetic variation in traits of adaptive significance. *Genetica*, **123**, 107–124.

Caballero A, Quesada H, Rolán-Alvarez E (2008) Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, **179**, 539–554.

Campbell DR, Waser NM (2001) Genotype-by-environment interaction and the fitness of plant hybrids in the wild. *Evolution*, **55**, 669–676.

Chang M, Lynn DG (1986) The haustorium and the chemistry of host recognition in parasitic angiosperms. *Journal of Chemical Ecology*, **12**, 561–579.

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, **18**, 237–268.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783–796.

Cissoko M, Boisnard A, Rodenburg J, Press MC, Scholes JD (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**, 952–963.

Clay K, Kover PX (1996) The Red Queen hypothesis and plant/pathogen interactions.

Annual Review of Phytopathology, **34**, 29–50.

Crawley MJ (2005) *Statistics: An Introduction using R*. John Wiley & Sons, Chichester, England.

Crawley MJ (2007) *The R Book*. John Wiley & Sons, Chichester, England.

Crémieux L, Bischoff A, Müller-Schärer H, Steinger T (2010) Gene flow from foreign provenances into local plant populations: Fitness consequences and implications for biodiversity restoration. *American Journal of Botany*, **97**, 94–100.

Cutter AD (2012) The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in Ecology & Evolution*, **27**, 209–218.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Deagle BE, Jones FC, Chan YF *et al.* (2012) Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. *Proceedings of the Royal Society. Series B: Biological Sciences*, **279**, 1277–1286.

Van Delft G-J, Graves JD, Fitter AH, Pruiksma MA (1997) Spatial distribution and population dynamics of *Striga hermonthica* seeds in naturally infested farm soils. *Plant and Soil*, **195**, 1–15.

Demuth JP, Wade MJ (2005) On the theoretical and empirical framework for studying genetic interactions within and among species. *The American Naturalist*, **165**, 524–536.

Demuth JP, Wade MJ (2006) Experimental methods for measuring gene interactions. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 289–316.

Detwiler JT, Criscione CD (2010) An infectious topic in reticulate evolution: Introgression and hybridization in animal parasites. *Genes*, **1**, 102–123.

Druce GC (1930) *Hayward's Botanist's Pocket Book*. 19th Ed. G. Bell and Sons, Ltd., London, UK.

Dybdahl MF, Jokela J, Delph LF, Koskella B, Lively CM (2008) Hybrid fitness in a locally adapted parasite. *The American Naturalist*, **172**, 772–782.

Dytham C (2003) *Choosing and Using Statistics: A Biologist's Guide*. Wiley-Blackwell, Chichester, England.

Ehrich D (2006) AFLPDAT: A collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes*, **6**, 603–604.

Ejeta G (2007) The *Striga* scourge in Africa: A growing pandemic. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 3–16. World Scientific.

Ejeta G, Rich PJ, Mohamed A (2007) Dissecting a complex trait to simpler components for effective breeding of sorghum with a high level of *Striga* resistance. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 87–98. World Scientific.

El Hiweris SO (1987) Nature of resistance to *Striga hermonthica* (Del.) Benth. parasitism in some *Sorghum vulgare* (Pers.) cultivars. *Weed Research*, **27**, 305–311.

Ellstrand N, Devlin B (1989) Transmission genetics of isozyme loci in *Raphanus sativus* (Brassicaceae): Stress-dependent non-Mendelian segregation. *American Journal of Botany*, **76**, 40–46.

Emechebe AM, Ellis-Jones J, Schulz S *et al.* (2004) Farmers' perception of the *Striga* problem and its control in nothern Nigeria. *Experimental Agriculture*, **40**, 215–232.

Estep MC, Mourik TA, Muth P *et al.* (2011) Genetic diversity of a parasitic weed, *Striga hermonthica*, on sorghum and pearl millet in Mali. *Tropical Plant Biology*, **4**, 91–98.

Etterson JR, Keller SR, Galloway LF (2007) Epistatic and cytonuclear interactions govern outbreeding depression in the autotetraploid *Campanulastrum americanum*. *Evolution*, **61**, 2671–2683.

Evans I, Maddox J (2009) CRI-MAP, version 2.503. http://www.animalgenome.org/tools/share/crimap/.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman, Malaysia.

Fay MF, Cowan RS, Leitch IJ (2005) The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Annals of Botany*, **95**, 237–246.

Featherston DW (1995) CRI-MAP Tutorial. http://www.ch.embnet.org/Embnetut/Crimap/.

Fenster CB, Galloway LF (2000) Population differentiation in an annual legume: Genetic architecture. *Evolution*, **54**, 1157–1172.

Fenster CB, Galloway LF, Chao L (1997) Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology & Evolution*, **12**, 282–286.

Fenton A, Antonovics J, Brockhurst MA (2012) Two-step infection processes can lead to coevolution between functionally independent infection and resistance pathways. *Evolution*, **66**, 2030–2041.

Fenton A, Brockhurst MA (2007) Epistatic interactions alter dynamics of multilocus gene-for-gene coevolution. *PLoS ONE*, **2**, e1156.

Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, **20**, 1450–1462.

Fishman L, Kelly AJ, Morgan E, Willis JH (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics*, **159**, 1701–1716.

Flor HH (1956) The complementary genic systems in flax and flax rust. In: *Advances in Genetics* (ed M. Demerec), pp. 29–54. Academic Press.

Foll M (2012) BayeScan v2.1 User Manual. http://cmpg.unibe.ch/software/bayescan/.

Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. *Molecular Ecology*, **19**, 4638–4647.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.

Frankel OH, Soulé ME (1981) *Conservation and Evolution*. CUP Archive, Cambridge, UK.

Freitag J, Hess DE, Welz HG, Geiger HH (1996) Pearl millet and sorghum specific races of *Striga hermonthica* in Niger. In: *Advances in Parasitic Plant Research* (eds Moreno MT, Cubero JI, Berner D, *et al.*), pp. 471–478. Junta de Andalucia, Cordoba, Spain.

Frohne I, Hyndman RJ (2012) R: Sample Quantiles. *http://stat.ethz.ch/R-manual/R-patched/library/stats/html/quantile.html*.

Galloway LF, Fenster CB (2001) Nuclear and cytoplasmic contributions to intraspecific divergence in an annual legume. *Evolution*, **55**, 488–497.

Galvani AP (2003) Epidemiology meets evolutionary ecology. *Trends in Ecology & Evolution*, **18**, 132–139.

Gandon S (2002) Local adaptation and the geometry of host–parasite coevolution. *Ecology Letters*, **5**, 246–256.

Gandon S, Michalakis Y (2000) Evolution of parasite virulence against qualitative or quantitative host resistance. *Proceedings of the Royal Society. Series B: Biological Sciences*, **267**, 985–990.

Gandon S, Nuismer SL (2009) Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. *The American Naturalist*, **173**, 212–224.

Gbèhounou G, Pieterse AH, Verkleij JAC (2003) Longevity of *Striga* seeds reconsidered: Results of a field study on purple witchweed (*Striga hermonthica*) in Bénin. *Weed Science*, **51**, 940–946.

Gelman A, Hill J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY.

Gethi JG, Smith ME, Mitchell SE, Kresovich S (2005) Genetic diversity of *Striga hermonthica* and *Striga asiatica* populations in Kenya. *Weed Research*, **45**, 64–73.

Gilman RT, Nuismer SL, Jhwueng D-C (2012) Coevolution in multidimensional trait space favours escape from parasites and pathogens. *Nature*, **483**, 328–330.

Girard P, Angers B (2011) The functional gene diversity in natural populations over postglacial areas: The shaping mechanisms behind genetic composition of longnose dace (*Rhinichthys cataractas*) in northeastern North America. *Journal of Molecular Evolution*, **73**, 45–57.

Gort G, van Eeuwijk F (2010) Codominant scoring of AFLP in association panels. *TAG Theoretical and Applied Genetics*, **121**, 337–351.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics*, **137**, 1121–1137.

Graves JD, Press MC, Stewart GR (1989) A carbon balance model of the sorghum-*Striga hermonthica* host-parasite association. *Plant, Cell & Environment*, **12**, 101–107.

Grech K, Watt K, Read AF (2006) Host–parasite interactions for virulence and resistance in a malaria model system. *Journal of Evolutionary Biology*, **19**, 1620–1630.

Green P, Falls K, Crooks S (1990) CRI-MAP, version 2.4. http://compgen.rutgers.edu/old/multimap/crimap/.

Grenier C, Ibrahim Y, Haussmann BIG, Kiambi D, Ejeta G (2007) Marker assisted selection for *Striga* resistance in sorghum. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 3–16. World Scientific.

Gressel J, Hanafi A, Head G *et al.* (2004) Major heretofore intractable biotic constraints to African food security that may be amenable to novel biotechnological solutions. *Crop Protection*, **23**, 661–689.

Gurney AL, Grimanelli D, Kanampiu F *et al.* (2003) Novel sources of resistance to *Striga hermonthica* in *Tripsacum dactyloides*, a wild relative of maize. *New Phytologist*, **160**, 557–568.

Gurney AL, Press MC, Scholes JD (1999) Infection time and density influence the response of sorghum to the parasitic angiosperm *Striga hermonthica*. *New Phytologist*, **143**, 573–580.

Gurney AL, Slate J, Press MC, Scholes JD (2006) A novel form of resistance in rice to the angiosperm parasite *Striga hermonthica*. *New Phytologist*, **169**, 199–208.

Hall MC, Willis JH (2005) Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*. *Genetics*, **170**, 375 –386.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **351**, 1291–1298.

Hartigan JA, Hartigan PM (1985) The DIP test of unimodality. *The Annals of Statistics*, **13**, 70–84.

Haussmann BIG, Hess DE, Koyama ML *et al.* (Eds.) (2000) *Breeding for striga resistance in cereals. Proceedings of a workshop held at IITA, Ibadan, Nigeria, from 18-20 August 1999.* Margraf Verlag, Weikersheim, Germany.

Haussmann BIG, Hess DE, Omanya GO *et al.* (2001) Major and minor genes for stimulation of seed germination in sorghum, and interaction with different *Striga* populations. *Crop Science*, **41**, 1507–1512.

Haussmann BIG, Hess DE, Omanya GO *et al.* (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *TAG Theoretical and Applied Genetics*, **109**, 1005–1016.

Haussmann BIG, Hess DE, Welz H-G, Geiger HH (2000) Improved methodologies for breeding striga-resistant sorghums. *Field Crops Research*, **66**, 195–211.

Hearne SJ (2009) Control—the *Striga* conundrum. *Pest Management Science*, **65**, 603–614.

Hendry AP, Nosil P, Rieseberg LH (2007) The speed of ecological speciation. *Functional Ecology*, **21**, 455–464.

Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity*, **103**, 283–284.

Herrmann M, Holderegger R, Van Strien MJ (2013) Influence of parameter settings in automated scoring of AFLPs on population genetic analysis. *Molecular Ecology Resources*, **13**, 128–134.

Herrmann D, Poncet BN, Manel S *et al.* (2010) Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates. *Genome*, **53**, 302–310.

Hersch-Green E., Turley NE, Johnson MTJ (2011) Community genetics: What have we accomplished and where should we be going? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 1453–1460.

Hess DE, Ejeta G, Butler LG (1992) Selecting sorghum genotypes expressing a quantitative biosynthetic trait that confers resistance to *Striga*. *Phytochemistry*, **31**, 493–497.

Hewitt EJ (1966) *Sand and Water Methods Used in the Study of Plant Nutrition*. Commonwealth Agricultural Bureaux, Wallingford, UK.

Holland BR, Clarke AC, Meudt HM (2008) Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology*, **57**, 347–366.

Hosmani MM, Parker C (1980) Host specificity of *Striga hermonthica* and *Striga asiatica* to some maize, sorghum and pearl millet varieties. In: *Annual Meeting of the Indian Weed Science Society* . Tocklai Experimental Station, Jorhat,

Assam.

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biometrical Journal*, **50**, 346–363.

Huang K (2007) The Molecular Basis of Resistance and Susceptibility to Striga Species. PhD Thesis. University of Sheffield, Sheffield, UK.

Huang K, Whitlock R, Press MC, Scholes JD (2012) Variation for host range within and among populations of the parasitic plant *Striga hermonthica*. *Heredity*, **108**, 96–104.

Hufford KM, Mazer SJ (2003) Plant ecotypes: Genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution*, **18**, 147–155.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. *Ecology Letters*, **11**, 609–623.

Huyse T, Poulin R, Théron A (2005) Speciation in parasites: A population genetics approach. *Trends in Parasitology*, **21**, 469–475.

Jamil M, Rodenburg J, Charnikhova T, Bouwmeester HJ (2011) Pre-attachment *Striga hermonthica* resistance of New Rice for Africa (NERICA) cultivars based on low strigolactone production. *New Phytologist*, **192**, 964–975.

Jarosz AM, Davelos AL (1995) Effects of disease in wild plant populations and the evolution of pathogen aggressiveness. *New Phytologist*, **129**, 371–387.

Jenczewski E, Gherardi M, Bonnin I *et al.* (1997) Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). *TAG Theoretical and Applied Genetics*, **94**, 682–691.

Jerome CA, Ford BA (2002) The discovery of three genetic races of the dwarf mistletoe *Arceuthobium americanum* (Viscaceae) provides insight into the evolution of parasitic angiosperms. *Molecular Ecology*, **11**, 387–405.

Johansen-Morris AD, Latta RG (2006) Fitness consequences of hybridization between ecotypes of *Avena barbata*: Hybrid breakdown, hybrid vigor, and transgressive segregation. *Evolution*, **60**, 1585–1595.

Jones JDG, Dangl JL (2006) The plant immune system. *Nature*, **444**, 323–329.

Kaewchumnong K, Price AH (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**, 206–216.

Kaltz O, Shykoff JA (1998) Local adaptation in host–parasite systems. *Heredity*, **81**, 361–370.

Kapran I, Grenier C, Ejeta G (2007) Introgression of genes for *Striga* resistance into African landraces of sorghum. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 129–142. World Scientific.

Kautt AF, Elmer KR, Meyer A (2012) Genomic signatures of divergent selection and speciation patterns in a "natural experiment", the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Molecular Ecology*, **21**, 4770–4786.

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

Kelly JK (2005) Epistasis in monkeyflowers. *Genetics*, **171**, 1917–1931.

Kent M (2012) *Vegetation Description and Data Analysis: A Practical Approach*. Wiley-Blackwell.

Keyes WJ, Palmer AG, Erbil WK *et al.* (2007) Semagenesis and the parasitic angiosperm *Striga asiatica*. *The Plant Journal*, **51**, 707–716.

Keyes WJ, Taylor JV, Apkarian RP, Lynn DG (2001) Dancing together. Social controls in parasitic plant development. *Plant Physiology*, **127**, 1508–1512.

Kim SK, Akintunde AY, Walker P (1994) Responses of maize, sorghum and millet host plants to infestation by *Striga hermonthica*. *Crop Protection*, **13**, 582–590.

Kim D, Kocz R, Boone L, Keyes WJ, Lynn DG (1998) On becoming a parasite: Evaluating the role of wall oxidases in parasitic plant development. *Chemistry & Biology*, **5**, 103–117.

Kirst M, Myburg A, Sederoff R (2004) Genetic mapping in forest trees: Markers,

linkage analysis and genomics. In: *Genetic Engineering: Principles and Methods* (ed Setlow JK), pp. 105–41. Springer.

Kisdi É (2002) Dispersal: Risk spreading versus local adaptation. *The American Naturalist*, **159**, 579–596.

Koevoets T, Niehuis O, van de Zande L, Beukeboom LW (2012) Hybrid incompatibilities in the parasitic wasp genus *Nasonia*: Negative effects of hemizygosity and the identification of transmission ratio distortion loci. *Heredity*, **108**, 302–311.

Koskela T, Salonen V, Mutikainen P (2000) Local adaptation of a holoparasitic plant, *Cuscuta europaea*: Variation among populations. *Journal of Evolutionary Biology*, **13**, 749–755.

Koyama ML (2000a) Genetic variability of *Striga hermonthica* and effect of resistant sorghum cultivars on population dynamics. In: *Breeding for striga resistance in cereals. Proceedings of a workshop held at IITA, Ibadan, Nigeria, from 18-20 August 1999.* (eds Haussmann BIG, Hess DE, Koyama ML, *et al.*), pp. 247–260. Margraf Verlag, Weikersheim, Germany.

Koyama ML (2000b) Molecular markers for the study of pathogen variability: Implications for breeding resistance to *Striga hermonthica*. In: *Breeding for striga resistance in cereals. Proceedings of a workshop held at IITA, Ibadan, Nigeria, from 18-20 August 1999.* (eds Haussmann BIG, Hess DE, Koyama ML, *et al.*), pp. 227–246. Margraf Verlag, Weikersheim, Germany.

Kuijt J (1977) Haustoria of phanerogamic parasites. *Annual Review of Phytopathology*, **15**, 91–118.

Kuiper E, Groot A, Noordover EC, Pieterse AH, Verkleij JA (1998) Tropical grasses vary in their resistance to *Striga aspera*, *Striga hermonthica*, and their hybrids. *Canadian Journal of Botany*, **76**, 2131–2144.

Kuiper E, Koevoets PLM, Verkleij JAC, Pieterse AH (1996) Genetic variability of *Striga aspera* and *Striga hermonthica*. Are they really two distinct species? In:

Advances in Parasitic Plant Research (eds Moreno MT, Cubero JI, Berner D, *et al.*), pp. 123–134. Junta de Andalucia, Cordoba, Spain.

Kuittinen H, Haan AA de, Vogl C *et al.* (2004) Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics*, **168**, 1575–1584.

Laine A-L, Burdon JJ, Dodds PN, Thrall PH (2011) Spatial variation in disease resistance: From molecules to metapopulations. *Journal of Ecology*, **99**, 96–112.

Lambrechts L (2010) Dissecting the genetic architecture of host–pathogen specificity. *PLoS Pathogens*, **6**, e1001019.

Lambrechts L (2011) Quantitative genetics of *Aedes aegypti* vector competence for dengue viruses: Towards a new paradigm? *Trends in Parasitology*, **27**, 111–114.

Lambrechts L, Fellous S, Koella JC (2006) Coevolutionary interactions between host and parasite genotypes. *Trends in Parasitology*, **22**, 12–16.

Lamkey KR, Edwards JW (1999) The quantitative genetics of heterosis. In: *The Genetics and Exploitation of Heterosis in Crops* (eds Coors JG, Pandey S), pp. 31–48. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI.

Lane JA, Bailey JA, Butler RC, Terry PJ (1993) Resistance of Cowpea [*Vigna unguiculata* (L.) Walp.] to *Striga gesnerioides* (Willd.) Vatke, a parasitic angiosperm. *New Phytologist*, **125**, 405–412.

Lane JA, Moore THM, Child DV *et al.* (1994) Virulence characteristics of a new race of the parasitic angiosperm, *Striga gesnerioides*, from southern Benin on cowpea (*Vigna unguiculata*). *Euphytica*, **72**, 183–188.

Lane JA, Moore THM, Child DV, Cardwell KF (1996) Characterization of virulence and geographic distribution of *Striga gesnerioides* on cowpea in West Africa. *Plant Disease*, **80**, 299.

Lannou C (2012) Variation and selection of quantitative traits in plant pathogens. *Annual Review of Phytopathology*, **50**, 319–338.

Launey S, Hedgecock D (2001) High genetic load in the Pacific oyster *Crassostrea*

gigas. *Genetics*, **159**, 255–265.

Leppälä J, Bechsgaard JS, Schierup MH, Savolainen O (2008) Transmission ratio distortion in *Arabidopsis lyrata*: Effects of population divergence and the S-locus. *Heredity*, **100**, 71–78.

Levin DA (1984) Inbreeding depression and proximity-dependent crossing success in *Phlox drummondii*. *Evolution*, **38**, 116–127.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Lexer C, Widmer A (2008) The genic view of plant speciation: Recent progress and emerging questions. *Philosophical Transactions of the Royal Society. Series B: Biological Sciences*, **363**, 3023–3036.

Li J, Timko MP (2009) Gene-for-gene resistance in *Striga*-cowpea associations. *Science*, **325**, 1094–1094.

Li X, Wang X, Wei Y, Brummer E (2011) Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. *TAG Theoretical and Applied Genetics*, **123**, 667–679.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **27**, 237–277.

Little TJ, Watt K, Ebert D (2006) Parasite-host specificity: Experimental studies on the basis of parasite adaptation. *Evolution*, **60**, 31–38.

Liu X, Grosz M (2006) *CRIGEN*. Animal Genomics and Breed Group, Monsanto, St Louis, MO, USA.

Lively CM (1999) Migration, virulence, and the geographic mosaic of adaptation by parasites. *The American Naturalist*, **153**, S34–S47.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65–95.

Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive

isolation. *PLoS Biology*, **8**, e1000500.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lynch M (1991) The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution*, **45**, 622–629.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Mach J (2010) A shot in the dark: How parasitic plants find host roots. *The Plant Cell*, **22**, 995.

Maiti RK, Ramaiah KV, Bisen SS, Chidley VL (1984) A comparative study of the haustorial development of *Striga asiatica* (L.) Kuntze on sorghum cultivars. *Annals of Botany*, **54**, 447–457.

Manel S, Conord C, Després L (2009) Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology*, **9**, 288.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136–142.

Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, **21**, 3729–3738.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, **19**, 3824–3835.

Martínez-Espinoza AD, García-Pedrajas MD, Gold SE (2002) The Ustilaginales as plant pests and model systems. *Fungal Genetics and Biology*, **35**, 1–20.

Mather K (1949) *Biometrical Genetics*. Methuen, London.

McDaniel SF, Willis JH, Shaw AJ (2007) A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. *Genetics*, **176**, 2489–2500.

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, **40**, 349–379.

De Meaux J, Mitchell-Olds T (2003) Evolution of plant resistance at the molecular level: Ecological context of species interactions. *Heredity*, **91**, 345–352.

Meffert LM (2000) The evolutionary potential of morphology and mating behaviour: The role of epistasis in bottlenecked populations. In: *Epistasis and the Evolutionary Process* (eds Wolf JB, Brodie III ED, Wade MJ), pp. 177–196. Oxford University Press, New York, NY.

Meudt HM, Clarke AC (2007) Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends in Plant Science*, **12**, 106–117.

Meyers LA, Bull JJ (2002) Fighting change with change: Adaptive variation in an uncertain world. *Trends in Ecology & Evolution*, **17**, 551–557.

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. *Current Opinion in Plant Biology*, **8**, 129–134.

Mhedbi-Hajri N, Darrasse A, Pigné S *et al.* (2011) Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evolutionary Biology*, **11**, 67.

Michel AP, Sim S, Powell THQ *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, **107**, 9724–9729.

Mitchell SE, Rogers ES, Little TJ, Read AF (2005) Host-parasite and genotype-by-environment interactions: Temperature modifies potential for selection by a sterilizing pathogen. *Evolution*, **59**, 70–80.

Mogg RJ, Bond JM (2003) A cheap, reliable and rapid method of extracting high-quality DNA from plants. *Molecular Ecology Notes*, **3**, 666–668.

Mohamed KI, Bolin JF, Musselman LJ, Peterson AT (2007) Genetic diversity of *Striga* and implications for control and modeling future distributions. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 71–84. World Scientific.

Mohamed A, Ellicott A, Housley TL, Ejeta G (2003) Hypersensitive response to *Striga* infection in *Sorghum*. *Crop Science*, **43**, 1320.

Mohamed KI, Musselman LJ, Riches CR (2001) The genus *Striga* (Scrophulariaceae) in Africa. *Annals of the Missouri Botanical Garden*, **88**, 60–103.

Van Mourik TA, Bianchi F, Van der Werf W, Stomph TJ (2008) Long-term management of *Striga hermonthica*: Strategy evaluation with a spatio-temporal population model. *Weed Research*, **48**, 329–339.

Musselman LJ, Bahrathalakshmi, Safa SB *et al.* (1991) Recent research on the biology of *Striga asiatica*, *S. gesnerioides* and *S. hermonthica*. In: *Proceedings, International Workshops organized by IITA, ICRISAT and IDRC.* (ed Kim SK), pp. 31–41.

Musselman LJ, Hepper FN (1986) The Witchweeds (*Striga*, Scrophulariaceae) of the Sudan Republic. *Kew Bulletin*, **41**, 205–221.

Musselman LJ, Parker C, Dixon N (1982) Notes on autogamy and flower structure in agronomically important species of *Striga* (Scrophulariaceae) and *Orobanche* (Orobanchaceae). *Beitrage zur Biologie der Pflanzen*, **56**, 329–43.

Musselman LJ, Yoder JI, Westwood JH (2001) Parasitic plants major problem to food crops. *Science*, **293**, 1434.

Mutikainen P, Salonen V, Puustinen S, Koskela T (2000) Local adaptation, resistance, and virulence in a hemiparasitic plant-host plant interaction. *Evolution*, **54**, 433–440.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321–3323.

Neuhauser C, Andow DA, Heimpel GE *et al.* (2003) Community genetics: Expanding

the synthesis of ecology and genetics. *Ecology*, **84**, 545–558.

Nickrent DL, Musselman LJ (1979) Autogamy in the American strain of witchweed, *Striga asiatica* (Scrophulariaceae). *Brittonia*, **31**, 253–256.

Nieberding CM, Olivieri I (2007) Parasites: Proxies for host genealogy and ecology? *Trends in Ecology & Evolution*, **22**, 156–165.

Niehuis O, Judson AK, Gadau J (2008) Cytonuclear genic incompatibilities cause increased mortality in male F_2 hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics*, **178**, 413–426.

Nuismer SL, Gandon S (2008) Moving beyond common-garden and transplant designs: Insight into the causes of local adaptation in species interactions. *The American Naturalist*, **171**, 658–668.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.

Oksanen J, Blanchet F, Kindt R *et al.* (2011) vegan: Community Ecology Package. http://CRAN.R-project.org/package=vegan.

Olivier A, Glaszmann J-C, Lanaud C, Leroux GD (1998) Population structure, genetic diversity and host specificity of the parasitic weed *Striga hermonthica* (Scrophulariaceae) in Sahel. *Plant Systematics and Evolution*, **209**, 33–45.

Olivier A, Glaszmann J-C, Lanaud C, Salle G, Leroux GD (1996) An insight into the population structure and genetic diversity of *Striga hermonthica* in West Africa. In: *Advances in Parasitic Plant Research* (eds Moreno MT, Cubero JI, Berner D, *et al.*), pp. 113–122. Junta de Andalucia, Cordoba, Spain.

Omanya GO, Haussmann BIG, Hess DE *et al.* (2000) Evaluation of laboratory, pot, and field measures of striga resistance in sorghum. In: *Breeding for striga resistance in cereals. Proceedings of a workshop held at IITA, Ibadan, Nigeria, from 18-20 August 1999.* (eds Haussmann BIG, Hess DE, Koyama ML, *et al.*), pp. 59–72. Margraf Verlag, Weikersheim, Germany.

Omanya GO, Haussmann BIG, Hess DE *et al.* (2004) Utility of indirect and direct

selection traits for improving *Striga* resistance in two sorghum recombinant inbred populations. *Field Crops Research*, **89**, 237–252.

Pagès JR, Southgate VR, Tchuem Tchuenté LA, Jourdane J (2002) Experimental evidence of hybrid breakdown between the two geographical strains of *Schistosoma intercalatum*. *Parasitology*, **124**, 169–175.

Park J-M, Manen J-F, Colwell AE, Schneeweiss GM (2008) A plastid gene phylogeny of the non-photosynthetic parasitic *Orobanche* (Orobanchaceae) and related genera. *Journal of Plant Research*, **121**, 365–376.

Parker C (2009) Observations on the current status of *Orobanche* and *Striga* problems worldwide. *Pest Management Science*, **65**, 453–459.

Parker C, Reid DC (1979) Host specificity in *Striga* species: Some preliminary observations. In: *Proceedings of the 2nd International Symposium on Parasitic Weeds* (eds Musselman LJ, Worsham AD, Eplee RE), pp. 79–90. Raleigh, N.C., USA.

Parker C, Riches CR (1993) *Parasitic Weeds of the World: Biology and Control*. CAB International, Wallingford, UK.

Peña EA, Slate EH (2006) Global validation of linear model assumptions. *Journal of the American Statistical Association*, **101**, 341.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Peters AD, Lively CM (1999) The Red Queen and fluctuating epistasis: A population genetic analysis of antagonistic coevolution. *The American Naturalist*, **154**, 393–405.

Phoenix GK, Press MC (2005) Linking physiological traits to impacts on community structure and function: The role of root hemiparasitic Orobanchaceae (ex-Scrophulariaceae). *Journal of Ecology*, **93**, 67–78.

Van der Plank JE (1968) *Disease Resistance in Plants*. Academic Press, New York.

Poulin R (2008) *Evolutionary Ecology of Parasites*. Princeton University Press, Princeton, NJ.

Press MC, Graves J (Eds.) (1995) *Parasitic Plants*. Chapman & Hall, London.

Press MC, Gurney AL, Scholes JD (1999) The control of *Striga* on maize and sorghum: nitrogen × crop genotype interactions. *Haustorium*, **36**.

Prezeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. *Evolution*, **59**, 2312–2323.

Pritchard VL, Dimond L, Harrison JS *et al.* (2011) Interpopulation hybridization results in widespread viability selection across the genome in *Tigriopus californicus*. *BMC Genetics*, **12**, 54.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–215.

R Core Team (2012) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramaiah KV (1984) Physiological specialisation of *Striga hermonthica* and crop specificity. In: *Proceedings of the Third International Symposium on Parasitic Weeds* (eds Parker C, Musselman LJ, Polhill RM, Wilson AK), pp. 58–65. International Center for Agricultural Reseach in the Dry Areas, Aleppo, Syria.

Ransom JK, Babiker AG, Odhiambo GD (2007) Integrating crop management practices for *Striga* control. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 213–228. World Scientific.

Reiss GC, Bailey JA (1998) *Striga gesnerioides* parasitising cowpea: Development of infection structures and mechanisms of penetration. *Annals of Botany*, **81**, 431–440.

Rhode JM, Cruzan MB (2005) Contributions of heterosis and epistasis to hybrid fitness. *The American Naturalist*, **166**, E124–E139.

Rich PJ, Ejeta G (2007) Biology of host-parasite interactions in *Striga* species. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-*

Hunt (eds Ejeta G, Gressel J), pp. 3–16. World Scientific.

Riebler A, Held L, Stephan W (2008) Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics*, **178**, 1817–1829.

Rodenburg J, Bastiaans L, Weltzien E, Hess DE (2005) How can field selection for *Striga* resistance and tolerance in sorghum be improved? *Field Crops Research*, **93**, 34–50.

Rodríguez RL (2012) Grain of environment explains variation in the strength of genotype × environment interaction. *Journal of Evolutionary Biology*, **25**, 1897–1901.

Rogers SM, Bernatchez L (2006) The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology*, **19**, 1979–1994.

Rowntree JK, Cameron DD, Preziosi RF (2011) Genetic variation changes the interactions between the parasitic plant-ecosystem engineer *Rhinanthus* and its hosts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **366**, 1380–1388.

Saccone NL (2011) Genetic mapping of complex traits. In: *Genomics: Essential Methods* (eds Starkey M, Elaswarapu R), pp. 67–90. John Wiley & Sons, Chichester, England.

Sacristán S, García-Arenal F (2008) The evolution of virulence and pathogenicity in plant pathogen populations. *Molecular Plant Pathology*, **9**, 369–384.

Safa SB, Jones BMG, Musselman LJ (1984) Mechanisms favouring outbreeding in *Striga hermonthica* [Scrophulariaceae]. *New Phytologist*, **96**, 299–305.

Schmid-Hempel P (2011) *Evolutionary Parasitology: The Integrated Study of Infections, Immunology, Ecology, and Genetics*. OUP Oxford.

Scholes JD, Press MC (2008) *Striga* infestation of cereal crops – an unsolved problem in resource limited agriculture. *Current Opinion in Plant Biology*, **11**, 180–186.

Scholes JD, Swarbrick PJ, Slate J, Press MC (2007) The molecular genetic basis of host resistance to *Striga* species: A way forward. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 173–184. World Scientific.

Schulz S, Hussaini MA, Kling JG, Berner DK, Ikie FO (2003) Evaluation of integrated *Striga hermonthica* control technologies under farmer management. *Experimental Agriculture*, **39**, 99–108.

Schwarz-Sommer Z, Silva E de A, Berndtgen R *et al.* (2003) A linkage map of an F_2 hybrid population of *Antirrhinum majus* and *A. molle*. *Genetics*, **163**, 699–710.

Scott JG, Berger JO (2006) An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, **136**, 2144–2162.

Seeb JE, Carvalho G, Hauser L *et al.* (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, **11**, 1–8.

Sham P, McGuffin P (2002) Linkage and association. In: *Psychiatric Genetics And Genomics* (eds McGuffin P, Owen MJ, Gottesman II), pp. 55–73. Oxford University Press, New York, NY.

Singleton V, Orthofer R, Lamuela-Raventos R (1999) Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent. *Methods in Enzymology*, **299**, 152–179.

Soetaert K, Herman PMJ (2009) *A Practical Guide to Ecological Modelling: Using R as a Simulation Platform*. Springer.

Stace C (2010) *New Flora of the British Isles*. Cambridge University Press, New York, USA.

Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, **10**, 681–690.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, **100**, 9440–9445.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Strasburg JL, Sherman NA, Wright KM *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society. Series B: Biological Sciences*, **367**, 364–373.

Swarbrick PJ, Scholes JD, Press MC, Slate J (2009) A major QTL for resistance of rice to the parasitic plant *Striga hermonthica* is not dependent on genetic background. *Pest Management Science*, **65**, 528–532.

Sweigart AL, Mason AR, Willis JH (2007) Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution*, **61**, 141–151.

Tack AJM, Thrall PH, Barrett LG, Burdon JJ, Laine A-L (2012) Variation in infectivity and aggressiveness in space and time in wild host–pathogen systems: Causes and consequences. *Journal of Evolutionary Biology*, **25**, 1918–1936.

Thompson JN (1999) Specific hypotheses on the geographic mosaic of coevolution. *The American Naturalist*, **153**, S1–S14.

Thompson JN (2005) *The Geographic Mosaic of Coevolution*. University of Chicago Press, USA.

Thompson JN, Burdon JJ (1992) Gene-for-gene coevolution between plants and parasites. *Nature*, **360**, 121–125.

Thorogood CJ, Rumsey FJ, Harris SA, Hiscock SJ (2008) Host-driven divergence in the parasitic plant *Orobanche minor* Sm. (Orobanchaceae). *Molecular Ecology*, **17**, 4289–4303.

Thorogood CJ, Rumsey FJ, Hiscock SJ (2009) Host-specific races in the holoparasitic angiosperm *Orobanche minor*: Implications for speciation in parasitic plants. *Annals of Botany*, **103**, 1005–1014.

Timko MP, Huang K, Lis KE (2012) Host resistance and parasite virulence in *Striga*–host plant interactions: A shifting balance of power. *Weed Science*, **60**, 307–315.

Treuren R van, Bijlsma R, Ouborg NJ, Delden W van (1993) The significance of genetic erosion in the process of extinction. IV. Inbreeding depression and heterosis effects caused by selfing and outcrossing in *Scabiosa columbaria*. *Evolution*, **47**, 1669–1680.

Trouvé S, Renaud F, Durand P, Jourdane J (1998) Experimental evidence of hybrid breakdown between genetically distinct populations of *Echinostoma caproni*. *Parasitology*, **117**, 133–135.

Trybush S, Hanley S, Cho K-H *et al.* (2006) Getting the most out of fluorescent amplified fragment length polymorphism. *Canadian Journal of Botany*, **84**, 1347–1354.

Vasey RA, Scholes JD, Press MC (2005) Wheat (*Triticum aestivum*) is susceptible to the parasitic angiosperm *Striga hermonthica*, a major cereal pathogen in Africa. *Phytopathology*, **95**, 1294–1300.

Vasudeva Rao MJ, Musselman LJ (1987) Host specificity in *Striga* spp. and physiological strains. In: *Parasitic Weeds in Agriculture: Striga* (ed Musselman LJ), pp. 13–25. CRC Press, Boca Raton, FL.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, USA.

Vitalis R, Dawson K, Boursot P, Belkhir K (2003) DetSel 1.0: A computer program to detect markers responding to selection. *Journal of Heredity*, **94**, 429–431.

Volf P, Benkova I, Myskova J *et al.* (2007) Increased transmission potential of *Leishmania major*/*Leishmania infantum* hybrids. *International Journal for Parasitology*, **37**, 589–593.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wade MJ, Goodnight CJ (1998) The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. *Evolution*, **52**, 1537–1553.

Warton DI, Hui FKC (2011) The arcsine is asinine: The analysis of proportions in ecology. *Ecology*, **92**, 3–10.

Waser NM, Williams CF (2001) Inbreeding and outbreeding. In: *Evolutionary Ecology: Concepts and Case Studies* (eds Fox CW, Roff DA, Fairburn DJ). Oxford University Press.

Weber G, Elemo K, Lagoke STO, Awad A, Oikeh S (1995) Population dynamics and determinants of *Striga hermonthica* on maize and sorghum in savanna farming systems. *Crop Protection*, **14**, 283–290.

Wegner KM, Berenos C, Schmid-Hempel P (2009) Host genetic architecture in single and multiple infections. *Journal of Evolutionary Biology*, **22**, 396–404.

Welsh AB, Mohamed KI (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**, 773–782.

Wennersten L, Forsman A (2012) Population-level consequences of polymorphism, plasticity and randomized phenotype switching: A review of predictions. *Biological Reviews of the Cambridge Philosophical Society*, **87**, 756–767.

Westfall P, Tobias R, Wolfinger R (2011) *Multiple Comparisons and Multiple Tests Using SAS*. SAS Institute Inc., Cary, NC.

Whitham TG, Bailey JK, Schweitzer JA *et al.* (2006) A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics*, **7**, 510–523.

Whitham TG, Young WP, Martinsen GD *et al.* (2003) Community and ecosystem genetics: A consequence of the extended phenotype. *Ecology*, **84**, 559–573.

Whitlock R, Hipperson H, Mannarelli M-E, Burke T (2008) A high-throughput protocol for extracting high-purity genomic DNA from plants and animals. *Molecular*

Ecology Resources, **8**, 736–741.

Whitlock R, Hipperson H, Mannarelli M-E, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, **8**, 725–735.

Whitlock MC, Phillips PC, Moore FB-G, Tonsor SJ (1995) Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics*, **26**, 601–629.

Wilfert L, Schmid-Hempel P (2008) The genetic architecture of susceptibility to parasites. *BMC Evolutionary Biology*, **8**, 187.

Williams CG (1998) QTL mapping in outbred pedigrees. In: *Molecular Dissection of Complex Traits* (ed Paterson A.) CRC Press.

Wilson-Jones K (1955) Further experiments on witchweed control. II. The existence of physiological strains of *Striga hermonthica*. *Empire Journal of Experimental Agriculture*, **23**, 206–213.

Wolf JB, Brodie III ED, Wade MJ (Eds.) (2000) *Epistasis and the Evolutionary Process*. Oxford University Press, New York, NY.

Wolinska J, King KC (2009) Environment can alter selection in host–parasite interactions. *Trends in Parasitology*, **25**, 236–244.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wright CA, Ross GC (1980) Hybrids between *Schistosoma haematobium* and *S. mattheei* and their identification by isoelectric focusing of enzymes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **74**, 326–332.

Wu R, Ma C-X, Painter I, Zeng Z-B (2002) Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theoretical Population Biology*, **61**, 349–363.

Xu ZC, Zhu J (1999) An approach for predicting heterosis based on an additive, dominance and additive × additive model with environment interaction.

Heredity, **82**, 510–517.

Yee TW (2010) The VGAM package for categorical data analysis. *Journal of Statistical Software*, **32**, 1–34.

Yoder JI, Musselman LJ (2006) *Striga*: A subterranean parasitic angiosperm (witchweed). In: *Encyclopedia of Plant and Crop Science* (ed Goodman R). Taylor & Francis Group, New York.

Yoder JI, Reagan R, Tomilov A, Tomilova N, Torres M (2007) Host detection by root parasites: Insights from transcriptome profiles. In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* (eds Ejeta G, Gressel J), pp. 3–16. World Scientific.

Yoder JI, Scholes JD (2010) Host plant resistance to parasitic weeds; recent progress and bottlenecks. *Current Opinion in Plant Biology*, **13**, 478–484.

Yoneyama K, Xie X, Kusumoto D *et al.* (2007) Nitrogen deficiency as well as phosphorus deficiency in sorghum promotes the production and exudation of 5-deoxystrigol, the host recognition signal for arbuscular mycorrhizal fungi and root parasites. *Planta*, **227**, 125–132.

Yoshida S, Ishida JK, Kamal NM *et al.* (2010) A full-length enriched cDNA library and expressed sequence tag analysis of the parasitic weed, *Striga hermonthica*. *BMC Plant Biology*, **10**, 55.

Yoshida S, Shirasu K (2009) Multiple layers of incompatibility to the parasitic witchweed, *Striga hermonthica*. *New Phytologist*, **183**, 180–189.

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. *TAG Theoretical and Applied Genetics*, **99**, 785–792.

Zeileis A (2004) Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, **11**, 1–17.

Zhan J, Mundt CC, Hoffer ME, McDonald BA (2002) Local adaptation and effect of host genotype on the rate of pathogen evolution: An experimental test in a plant pathosystem. *Journal of Evolutionary Biology*, **15**, 634–647.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, **8**, 907–913.

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, **1**, 3–14.