



**Genetic approaches to identify resistance in rice to
the parasitic weed, *Striga***

Nat Phon-or

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department of Animal and Plant Sciences

March 2021

Abstract

Striga asiatica (L.) Kuntze and *S. hermonthica* (Delile) Benth. are obligate hemiparasitic weeds that parasitise the roots of cereal crops, such as maize, sorghum and upland rice in sub-Saharan Africa. *Striga* infestation causes 20–100 % yield loss in low-input farming in Africa, affecting the livelihood of millions of subsistent farmers. Upland rice has gained popularity as a staple crop in the region, but production is increasingly constrained by *Striga* infestation. Control of *Striga* is difficult because the parasites inflict irreversible damage to the host before they emerge above ground and the parasite seed bank is widespread and genetically diverse. The use of crop varieties with genetic resistance to *Striga* is affordable and effective but knowledge of the mechanisms and the identity of resistance genes/loci is sparse, impeding the development of resistant rice cultivars with durable resistance. Thus, the aims of this thesis are (i) to identify and characterise resistance phenotypes in *Oryza sativa* L. to *S. asiatica* and to identify genetic loci and candidate genes underlying the resistance phenotype and (ii) to perform a genome-wide association study (GWAS) using an existing dataset of resistance/susceptibility levels of an *O. glaberrima* Steud. diversity panel to *S. hermonthica*, to identify loci underlying resistance.

Parental genotypes of three *O. sativa* recombinant inbred line (RIL) populations were initially phenotyped to determine their susceptibility to an accession of *S. asiatica* from Ethiopia. Parental genotypes of one RIL population exhibited differential susceptibility; rice genotype IR64 was susceptible, whereas CT8556-37-2-3-1-M had few attachments and parasites that attached grew more slowly and were smaller than those on IR64. A detailed microscopic analysis of the early stages of infection of the roots of the two rice genotypes (2–8 d after inoculation of the roots) revealed that the early infection process and the formation of host–parasite xylem connections were not significantly different on the two hosts. However, subsequent growth (length and biomass accumulation) of attached parasites was significantly slower on CT8556-37-2-3-1-M than on IR64. Also, a pot-based experiment revealed that very few parasites emerged above ground on CT8556-37-2-3-1-M compared to IR64 and the emergence was delayed.

The RIL population derived from the cross between the two rice genotypes were phenotyped for levels of *S. asiatica*-resistance, by measuring the biomass of the parasites per host plant in

the rhizotron system. Six significant quantitative trait loci (QTL) were identified on chromosomes 1, 5 and 11, using different QTL mapping methods: simple interval mapping; inclusive composite interval mapping; multiple interval mapping; and multiple QTL mapping. All six QTL had small effect sizes (each explaining 7.5–17.5 % of the phenotypic variance) and additively contributed to host resistance. Three QTL on chromosomes 1, 5 and 11 each, with the largest effect sizes, were consistently mapped by multiple mapping methods.

To identify candidate genes and process underlying the resistance mechanism, genomic sequences within these three QTL regions (of both rice genotypes) were analysed by gene prediction software and annotated. All three QTL regions contained genes involved in the process such as nutrient acquisition (e.g. amino acid and ammonium transporter genes), cell wall biogenesis and modification (e.g. expansin gene and genes involved in the biosynthesis of cell wall components), hormones involved in growth and development (e.g. abscisic acid receptor, auxin transporter and auxin-amido synthetase genes) and host biotic stress responses (e.g. genes in the salicylic acid and jasmonic acid signalling pathways).

Finally, the existing resistance/susceptibility scores of an *O. glaberrima* diversity panel that had been screened against an accession of *S. hermonthica* were used in a GWAS. No significant association was identified between the single nucleotide polymorphisms and the level of resistance/susceptibility as, unexpectedly, most *O. glaberrima* genotypes were susceptible to the accession of *S. hermonthica* from Kibos, Kenya used in this study.

Table of Contents

Abstract.....	i
Table of Contents	iii
List of Figures.....	vii
List of Tables	x
Acknowledgement.....	xii
Declaration.....	xiii
Abbreviations.....	xiv
Chapter 1 General introduction	1
1.1 Parasitic plants.....	3
1.2 The genus <i>Striga</i>	5
1.3 Life cycle of <i>Striga</i>	8
1.4 Current controls for the management of <i>Striga</i> weeds	12
1.5 Different phenotypes of <i>Striga</i> resistance	15
1.6 Dissection of the genetic basis of <i>Striga</i> resistance	18
1.7 Expanding rice production in Africa is threatened by the infestation of <i>Striga</i> weeds.....	23
1.8 Aims of Thesis.....	24
Chapter 2 Selection of a rice recombinant inbred line (RIL) population for mapping resistance to <i>Striga asiatica</i>: Characterisation of the resistance/susceptibility phenotype of the parents of three RIL populations	27
2.1 Introduction	29
2.1.1 Aims and objectives	31
2.2 Methodology	32
2.2.1 Plant materials.....	32
2.2.2 Growth and infection of rice genotypes with <i>Striga asiatica</i>	32
2.2.3 Quantification of resistance/susceptibility of the four parental genotypes of three RIL populations.....	33

2.2.4	Characterisation of the development of <i>Striga asiatica</i> on IR64 and CT8556-37-2-3-1-M during the early phase of the life cycle.....	34
2.2.5	Characterisation of changes in biomass, length and number of <i>Striga asiatica</i> plants from 12 to 42 d after inoculation	38
2.2.6	Characterisation of the emergence of <i>Striga asiatica</i> plants growing on IR64 and CT8556-37-2-3-1-M in pots	39
2.3	Results.....	40
2.3.1	The four parental genotypes of three rice RIL populations exhibited different levels of susceptibility to <i>Striga asiatica</i>	40
2.3.2	Why are <i>Striga asiatica</i> plants smaller on CT8556-37-2-3-1-M compared to IR64?	42
2.3.3	<i>Striga asiatica</i> plants grew more slowly on CT8556-37-2-3-1-M than on IR64 from 8 d after inoculation.	44
2.3.4	Fewer <i>Striga asiatica</i> plants emerged more slowly when growing on CT8556-37-2-3-1-M compared to IR64.....	49
2.4	Discussion	52
2.4.1	The number of parasites was lower on CT8556-37-2-3-1-M than on IR64....	52
2.4.2	Parasites grew more slowly on CT8556-37-2-3-1-M compared to IR64.....	53
2.4.3	How useful is this slower growth phenotype of <i>Striga asiatica</i> for <i>Striga</i> control?	56
Chapter 3 Identification of quantitative trait loci underlying resistance to <i>Striga asiatica</i> in rice		59
3.1	Introduction	61
3.1.1	Aim and objectives	67
3.2	Methodology.....	68
3.2.1	Plant materials	68
3.2.2	Phenotypic screen, data preparation and statistical analyses	69
3.2.2.1	Phenotypic scoring of the recombinant inbred line population using a two-step phenotypic scoring strategy	69
3.2.2.2	Elimination of escapes.....	71
3.2.2.3	Selection of the trait to represent the phenotype	73
3.2.2.4	Assessment of the phenotypic consistency of the genotypes between batches.....	74
3.2.2.5	Did the biomass of parasites meet the assumptions of LMM?.....	75
3.2.2.6	Phenotypic inputs for QTL analysis from the LMM.....	78
3.2.3	QTL analyses	78
3.2.3.1	Linkage map construction	78
3.2.3.2	Simple interval mapping (SIM)	79

3.2.3.3 Inclusive composite interval mapping (ICIM).....	81
3.2.3.4 Multiple interval mapping (MIM).....	81
3.2.3.5 Multiple QTL mapping (MQM).....	81
3.3 Results	85
3.3.1 Phenotypes of RILs from Screen 1	85
3.3.2 Simple interval mapping (SIM) of phenotypic data from Screen 1 revealed one significant QTL on chromosome 1.....	86
3.3.3 Phenotypes of RILs in the combined dataset.....	86
3.3.4 QTL analyses of combined dataset detected many QTL underlying resistance to <i>Striga asiatica</i>	89
3.4 Discussion.....	101
3.4.1 How do these QTL for <i>Striga asiatica</i> -resistance compare to QTL for <i>Striga</i> -resistance from other studies?	101
3.4.2 Differences and similarities between QTL mapping methods.....	103
3.4.3 What might have affected the QTL analyses?.....	105
3.4.4 Identification of candidate genes underlying the resistance to <i>Striga asiatica</i>	107
Chapter 4 Identification of candidate genes underlying resistance in rice to <i>Striga asiatica</i>	111
4.1 Introduction	113
4.1.1 Aim and objectives.....	116
4.2 Methodology	117
4.2.1 Genome sequences.....	117
4.2.2 Conversion of positions of markers within the QTL regions from Nipponbare coordinates to IR64 coordinates.....	117
4.2.3 Prediction and annotation of IR64 genes.....	120
4.2.4 Assembling CT8556-37-2-3-1-M reads using reference-guided <i>de novo</i> assembly.....	122
4.2.5 Prediction and annotation of CT8556-37-2-3-1-M genes	125
4.3 Results	126
4.3.1 Locating the physical positions of three major QTL on the IR64 genome sequence	126
4.3.2 Gene models and transcriptome of IR64.....	128
4.3.3 Mapping of the reads and assembled contigs of CT8556-37-2-3-1-M onto the IR64 genome sequence.....	134
4.3.4 Predicted genes within 2 cM distance from QTL peaks	135
4.4 Discussion.....	151

4.4.1	Could limitations in the acquisition of nutrients contribute to the slow growth of parasites on CT8556-37-2-3-1-M?	152
4.4.2	Could differences in host hormone status/regulation indirectly affect the growth of the parasite?	154
4.4.3	Is there any evidence for the production of toxic compounds in rice genotype CT8556-37-2-3-1-M?.....	157
4.4.4	Host defence responses against <i>Striga</i> infection	157
4.4.5	Lack of a complete genome sequence for CT8556-37-2-3-1-M hindered the comparison of the two rice genomes.	159
4.4.6	Prospects	160
Chapter 5 A genome-wide association study to identify resistance to <i>Striga hermonthica</i> in the African rice species, <i>Oryza glaberrima</i>		163
5.1	Introduction	165
5.1.1	Aim and objectives	167
5.2	Methodology.....	167
5.2.1	Plant materials	167
5.2.2	Preparation of phenotypic data	168
5.2.2.1	Selection of the traits to represent the resistance phenotype.....	170
5.2.2.2	Removal of <i>Oryza sativa</i> checks from analyses.....	170
5.2.2.3	Assessment of the phenotypic consistency of the <i>Oryza glaberrima</i> genotypes between batches.....	172
5.2.3	Preparation of genotypic data.....	177
5.2.4	Association analyses	180
5.3	Results.....	181
5.3.1	<i>Oryza glaberrima</i> genotypes were mostly susceptible to an accession of <i>Striga hermonthica</i> and had low repeatabilities.	181
5.3.2	There was no significant association between the molecular markers and the biomass or the number of parasites.	183
5.4	Discussion	183
Chapter 6 General discussion		191
6.1	Introduction	193
6.2	Slow growth resistance to <i>Striga asiatica</i>	194
6.3	Genetic basis of the slow growth resistance to <i>Striga asiatica</i>	198
6.4	Genome-wide association study for the detection of loci for resistance to <i>Striga hermonthica</i> in rice	204
References.....		206
Appendix.....		227

List of Figures

Figure 1.1	Parasitic plants can be categorised by the degree of parasitism and the site of attachment.	4
Figure 1.2	Distributions of the three <i>Striga</i> species that cause the most damage to agriculture in Africa: <i>Striga hermonthica</i> , <i>S. asiatica</i> and <i>S. gesnerioides</i>	7
Figure 1.3	Life cycle of <i>Striga</i>	9
Figure 1.4	Distribution of the collection sites of herbarium samples of <i>Striga</i> species and rain-fed rice-producing areas in Africa.	24
Figure 2.1	Developmental stages of <i>Striga asiatica</i> individuals between 2 and 8 d after inoculation.....	36
Figure 2.2	Susceptibility/resistance of the common parental line (IR64) and three selected diversity donors (CT8556-37-2-3-1-M, CT10037-56-6-M-M-1 and ITA164) of a nested association mapping population.	41
Figure 2.3	Distributions of <i>Striga asiatica</i> individuals in different developmental stages between 2 and 8 d after inoculation (DAI).....	43
Figure 2.4	Growth of <i>Striga asiatica</i> individuals between 12 and 42 d after inoculation (DAI).	47
Figure 2.5	Images of <i>Striga asiatica</i> plants on the root systems of CT8556-37-2-3-1-M and IR64 from 12 to 42 d after inoculation (DAI).	48
Figure 2.6	Above ground growth of <i>Striga asiatica</i> plants on IAC165, IR64 and CT8556-37-2-3-1-M rice genotypes in pots.	49
Figure 2.7	<i>Striga asiatica</i> plants parasitising on IAC165, IR64 and CT8556-37-2-3-1-M rice genotypes in pots.....	51
Figure 3.1	Flowchart of experimental design, data preparation and statistical analyses prior to quantitative trait locus (QTL) analyses.	70
Figure 3.2	Scans of rhizotrons of the rice recombinant inbred line (RIL) number 176 before the <i>Striga asiatica</i> plants were harvested at 28 d after inoculation.	72
Figure 3.3	Relationships between the measurements of traits related to <i>Striga asiatica</i> -resistance in Screen 1.	73
Figure 3.4	Diagnostic plots of the untransformed and log-transformed biomass of <i>Striga asiatica</i> per host from Screen 1 after the data was fitted to a linear mixed-effect model.	76

Figure 3.5	Diagnostic plots of the untransformed and log-transformed biomass of <i>Striga asiatica</i> per host from Screen 1 and Screen 2 after the data was fitted to a linear mixed-effect model.	77
Figure 3.6	Genotypic data of the recombinant inbred line (RIL) population derived from CT8556-37-2-3-1-M × IR64 for interval mapping.	80
Figure 3.7	Genotypic data of the recombinant inbred line (RIL) population derived from CT8556-37-2-3-1-M × IR64 for multiple QTL mapping (MQM).	84
Figure 3.8	Biomass of <i>Striga asiatica</i> per rice plant on 48 randomly selected recombinant inbred lines (RILs), derived from CT8556-37-2-3-1-M × IR64, in Screen 1.	85
Figure 3.9	Simple interval mapping (SIM) of quantitative trait loci (QTL) for <i>Striga asiatica</i> -resistance in Screen 1.....	87
Figure 3.10	The biomass of <i>Striga asiatica</i> per host plant of a population of recombinant inbred lines (RILs) derived from CT8556-37-2-3-1-M × IR64.	88
Figure 3.11	Simple interval mapping (SIM) of quantitative trait loci (QTL) for <i>Striga asiatica</i> -resistance in the combined the dataset.	90
Figure 3.12	LOD scores of model comparisons when two quantitative trait loci were fitted on chromosomes 1, 5 and 11.....	93
Figure 3.13	Epistatic interaction between two loci on chromosome 11 at 19 cM (Chr11_3608272 maker) and chromosome 12 at 31 cM (Chr12_19525818 marker).....	94
Figure 3.14	Multiple QTL mapping (MQM) of quantitative trait loci (QTL) for <i>Striga asiatica</i> -resistance in the combined dataset.	96
Figure 3.15	Inclusive composite interval mapping (ICIM) of quantitative trait loci (QTL) for <i>Striga asiatica</i> -resistance in the combined dataset.	98
Figure 3.16	Multiple interval mapping (MIM) of quantitative trait loci (QTL) for <i>Striga asiatica</i> -resistance in the combined dataset.	99
Figure 4.1	Alignments of chromosomes 1, 5 and 11 of the rice genotypes, Nipponbare and IR64.....	119
Figure 4.2	Physical locations of the logarithm of the odds ratio (LOD) support intervals of <i>Striga asiatica</i> -resistance QTL on the IR64 genome sequence...	127
Figure 4.3	Genes predicted by different gene prediction programs within the <i>Striga asiatica</i> -resistance quantitative trait locus (QTL), <i>qSaB1.1</i> , on the IR64 genome.	129
Figure 4.4	Genes predicted by different gene prediction programs within the <i>Striga asiatica</i> -resistance quantitative trait locus (QTL), <i>qSaB5.1</i> , on the IR64 genome.	131

Figure 4.5	Genes predicted by different gene prediction programs within the <i>Striga asiatica</i> -resistance quantitative trait locus (QTL), <i>qSaB11.1</i> , on the IR64 genome.....	132
Figure 4.6	Comparison of predicted genes on the IR64 genome sequence and CT8556-37-2-3-1-M contigs within 2 cM from the peak of the <i>Striga asiatica</i> -resistance quantitative trait locus (QTL), <i>qSaB1.1</i>	136
Figure 4.7	Comparison of predicted genes on the IR64 genome sequence and CT8556-37-2-3-1-M contigs within 2 cM from the peak of the <i>Striga asiatica</i> -resistance quantitative trait locus (QTL), <i>qSaB5.1</i>	137
Figure 4.8	Comparison of predicted genes on the IR64 genome sequence and CT8556-37-2-3-1-M contigs within 2 cM from the peak of the <i>Striga asiatica</i> -resistance quantitative trait locus (QTL), <i>qSaB11.1</i>	139
Figure 4.9	Functional categorisation of proteins encoded by genes predicted within the 2 cM distance from peak markers of three major <i>Striga asiatica</i> -resistance quantitative trait loci (QTL), <i>qSaB1.1</i> , <i>qSaB5.1</i> and <i>qSaB11.1</i>	143
Figure 5.1	Flowchart of experimental design, data preparation, statistical and association analyses.	169
Figure 5.2	Relationships between the biomass, the cumulative length and the number of <i>Striga hermonthica</i> individuals per host.	171
Figure 5.3	Comparison of the susceptibility of the <i>Oryza glaberrima</i> genotypes and the two <i>O. sativa</i> checks (IAC165 and Koshihikari) to <i>Striga hermonthica</i> . .	173
Figure 5.4	Diagnostic plots of the GLMMs for the biomass of <i>Striga hermonthica</i> plants per host.	175
Figure 5.5	Diagnostic plots of the GLMMs for the number of <i>Striga hermonthica</i> individuals per host.	176
Figure 5.6	Assessment of the number of ancestral populations of the 149 <i>Oryza glaberrima</i> genotypes.	178
Figure 5.7	The population stratification of the single nucleotide polymorphism data....	179
Figure 5.8	Resistance to <i>Striga hermonthica</i> of a panel of 156 <i>Oryza glaberrima</i> genotypes.	182
Figure 5.9	Genome-wide association studies of resistance to <i>Striga hermonthica</i> in <i>Oryza glaberrima</i>	184
Figure 6.1	Overview of work carried out in this PhD programme, findings and future directions from the work in this thesis.	195

List of Tables

Table 2.1 Results of the Mann–Witney’s <i>U</i> tests on the differences between the proportions of parasites at different developmental stages on the two rice genotypes, CT8556-37-2-3-1-M and IR64, at 2, 3, 5 and 8 d after inoculation (DAI).....	44
Table 3.1 Quantitative trait loci (QTL) for resistance to <i>Striga hermonthica</i> identified in rice.	63
Table 3.2 Summary of repeatability of the phenotypic scores from Screen 1.	86
Table 3.3 Summary of repeatability of the phenotypic scores from the combined dataset	89
Table 3.4 Summary of all significant results from the two-dimensional scan, whether epistatic interactions between loci were significant or not, using MQM-1.....	94
Table 3.5 Locations of putative quantitative trait loci (QTL) identified from the two-dimensional scan in MQM-1, and the results of QTL model selection procedure using a stepwise search algorithm.	95
Table 3.6 Summary of results from different methods of mapping quantitative trait loci (QTL) for the slow growth resistance to <i>Striga asiatica</i>	100
Table 4.1 Positions of the three major quantitative trait loci (QTL) for the slow growth resistance phenotype of <i>Striga asiatica</i> on the genome sequences of the rice genotypes, Nipponbare and IR64.....	120
Table 4.2 k-mer lengths and their corresponding minimum k-mer count thresholds used in <i>de novo</i> assemblies of CT8556-37-2-3-1-M reads with a Bloom filter and a de Bruijn graph.....	124
Table 4.3 Summary statistics of the IR64-guided assembly of CT8556-37-2-3-1-M contigs (only contigs longer than 200 bp are considered).	124
Table 4.4 Locations, peak positions, the 2 cM distance from the peaks and lengths of three major quantitative trait loci (QTL) for the slow growth resistance phenotype of <i>Striga asiatica</i> on the IR64 genome sequence.....	126
Table 4.5 Number of genes within the whole quantitative trait locus (QTL) regions and the 2 cM distance from the QTL peaks of three major <i>Striga asiatica</i> -resistance QTL on chromosomes 1, 5 and 11 of rice, predicted by four gene prediction programs, Eugene, FGenesh, Augustus and GeneMark. ...	128

Table 4.6	Number of genes predicted by FGenesH on the sequences of the IR64 genome and CT8556-37-2-3-1-M contigs (longer than 2442 bp) within 2 cM distance from the peaks of the three major <i>Striga asiatica</i> -resistance QTL.....	135
Table 4.7	Annotations of genes within the 2 cM distance from peak markers of three major <i>Striga asiatica</i> -resistance QTL, <i>qSaB1.1</i> , <i>qSaB5.1</i> and <i>qSaB11.1</i> , categorised into functional groups.	144
Table 5.1	Intraclass (intra-batch) correlation coefficient for resistance to <i>Striga hermonthica</i> of <i>Oryza sativa</i> check varieties, Koshihikari and IAC165...	171
Table 5.2	Summary of repeatabilities of the <i>Striga hermonthica</i> -resistance.....	181

Acknowledgement

First of all, I would like to thank my supervisors, Professor Julie Scholes and Dr Mathias Lorieux, for agreeing to supervise me and accepting me onto the project. I am grateful for their patience, guidance, encouragement and support during this PhD programme. I would like to also thank them for reading my work and their suggestions for improvements.

I would like to also thank Professor Jon Slate for his advice on the handling of data variability and the calculation of repeatability.

I would like to acknowledge the Agricultural Research Development Agency (ARDA), Thailand for the financial support which enables my studentship in the UK, not only for this PhD programme but also from the very beginning back in 2010. I would like to also thank the teams at ARDA, the Office of Educational Affairs at the Royal Thai Embassy, London and the Office of the Civil Service Commission, Thailand for their support.

I would like to thank everyone who has helped me during my time as a PhD student, especially the members of C45. I am very grateful to Dr Mamadou Cissoko, Peijun Zhang, Dr Suo Qiu, Dr Emily Beardon, Dr James Bradley, David Rapley and Rachel Greenhill for their friendships, knowledge, technical assistance on microscopy work and bioinformatics, and tremendous help on *Striga* infection. I would also like to thank Nichola White, Maggi Killion and technicians for their help and organisation of the laboratory.

I am grateful to many friends who have supported and encouraged me throughout the years and made my whole journey far away from home, not just the 1471 days of PhD journey, memorable and enjoyable. Also, I would like to thank Sheffield and its people for providing a welcoming environment that will truly miss.

Last but not least, I would like to thank my family who always asks about my wellbeing and provides me with their support, for which I am grateful to have.

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Abbreviations

ABA	Abscisic acid
ABC	ATP-binding cassette
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
ASNPC	Area under <i>Striga asiatica</i> number progress curve
BC _N (generation)	N th backcross generation
bHLH	Basic helix-loop-helix
BIL	Backcross inbred line
BLUP	Best linear unbiased predictor
CC	Coiled-coil
CIAT	Centro Internacional de Agricultura Tropical (International Center for Tropical Agriculture)
CIM	Composite interval mapping
CPM	Read <u>counts</u> per one <u>million</u> reads
DAI	Days after inoculation
DAP	Days after planting
DArT	Diversity arrays technology
<i>df</i>	Degree of freedom
DH	Double haploid
DMBQ	2,6- <u>dimethoxy</u> -1,4- <u>benzoquinone</u>
dpi	Dots per inch (scan resolution)
DUF	Domains of unknown function
F _N (generation)	N th filial generation
FDR	False discovery rate
GBS	Genotyped by sequencing
GLMM	<u>Generalised</u> <u>linear</u> <u>mixed-effects</u> <u>model</u>
GO	Gene ontology

GWAS	Genome-wide association study
HIF	Haustorium-inducing factor
HR	Hypersensitive response
IAA	Indole-3-acetic acid
ICIM	Inclusive composite interval mapping
IRD	Institut de Recherche pour le Développement (Research Institute for Development)
IRGSP-1.0	International Rice Genome Sequencing Project Nipponbare reference genome
JA	Jasmonic acid
<i>K</i>	Number of ancestral populations
K-matrix	Kinship (relatedness) matrix
<i>LGS</i>	<i>LOW GERMINATION STIMULANT</i>
<i>Lhf</i>	Low haustorial factor
LMM	General <u>l</u> inear <u>m</u> ixed-effects <u>m</u> odel
LOD	<u>L</u> ogarithm of the <u>o</u> dds ratio
LRR	Leucine-rich repeat
MeJA	<u>M</u> ethyl <u>j</u> asmonate
MIM	Multiple interval mapping
MLM	Mixed linear model
MQM	Multiple QTL mapping
MR	Marker regression
MSU	Michigan State University
NAM (population)	Nested-association mapping
NAM (protein)	No apical meristem
NB	Nuclear-binding
NCBI	National Center for Biotechnology Information
NERICA	<u>N</u> ew <u>R</u> ice for <u>A</u> frica
NIAS	National Institute of Agrobiological Sciences
nr	Non-redundant (protein sequence)
PCA	Principal component analysis
PVE	Phenotypic variance explained
Q-matrix	Population stratification structure matrix
Q-Q	Quantile-quantile

ABBREVIATIONS

<i>qSaB</i>	QTL for <i>Striga asiatica</i> <u>b</u> iomass
<i>qShB</i>	QTL for <i>Striga hermonthica</i> <u>b</u> iomass
<i>qShN</i>	QTL for <i>Striga hermonthica</i> <u>n</u> umber
<i>qShR</i>	QTL for <i>Striga hermonthica</i> survival <u>r</u> ate (<u>r</u> esistance)
<i>qShTFE</i>	QTL for <i>Striga hermonthica</i> <u>T</u> FE
QTL	Quantitative trait locus (<i>singular</i>) / loci (<i>plural</i>)
R/R	Resistance gene/protein
RAP-DB	Rice Annotation Project Database
RFLP	Restriction fragment length polymorphism
RGR	Relative growth rate
RIL	Recombinant inbred line
RLK	Receptor-like kinase
RLP	Receptor-like protein
RNAi	RNA interference
RNA-Seq	RNA sequencing
SA	Salicylic acid
SCF	Skp1-cullin 1-F-box
SDL	Segregation distortion locus (<i>singular</i>) / loci (<i>plural</i>)
SIM	Simple interval mapping
<i>SLB</i>	<i>STRIGOLACTONE BIOSYNTHESIS</i>
sNMF	Sparse non-negative matrix factorization
SNP	Single nucleotides polymorphism
SSR	Simple sequence repeat
TE	Transposable element
TF	Transcription factor
TFE	<u>T</u> ime to <u>f</u> irst parasite <u>e</u> mergence

Chapter 1

General introduction

1.1 Parasitic plants

Parasitic plants feed on resources from another plant, the host. Plant parasitism has evolved independently at least twelve times in angiosperms (Bromham *et al.*, 2013), and occurs in over 4000 angiosperm species from 19 families (Spallek *et al.*, 2013; Poulin, 2011; Westwood *et al.*, 2010). The degree to which a parasitic plant relies on its host to complete its life cycle varies from genus to genus. Holoparasites, such as members of the genera *Cuscuta* L. (Figure 1.1A), *Orobanche* L. (Figure 1.1B) and *Phelipanche* Pomel (Figure 1.1C), are heterotrophs that have lost their ability to fix carbon through photosynthesis and depend on their host for water, nutrients and photosynthates throughout their life cycle (Westwood *et al.*, 2010). In contrast, hemiparasites, such as members of the genera *Viscum* L. (Figure 1.1D), *Striga* Lour. (Figure 1.1E), *Triphysaria* Fisch. & C.A.Mey. (Figure 1.1F) and *Phtheirospermum* Bunge ex Fisch. & C.A.Mey. (Figure 1.1G), are autotrophic and can photosynthesise to some degree once they emerge above ground (Westwood *et al.*, 2010). Hemiparasites can be further categorised into facultative hemiparasites (including *Triphysaria* and *Phtheirospermum* species; Figure 1.1F & G), which can complete their life cycle independently of a host, but will parasitise a host when available, and obligate hemiparasites (including *Striga* spp.; Figure 1.1E), which need to attach to a host to survive. Parasitic plants can also be categorised according to their sites of attachment. Stem parasites, such as *Cuscuta* spp. (Figure 1.1A) and mistletoes (e.g. *Viscum* L.; Figure 1.1D), attach to the stem of the host. On the contrary, root parasites, such as *Striga*, *Orobanche* and *Tryphysaria* species (Figure 1.1B–C & E–G), attach to the roots of the host.

Despite diverse lineages with varying degrees of parasitism, virtually all parasitic plants possess a unifying feature that is a specialised organ, called a haustorium (*plural* haustoria). Parasitic plants use haustoria to infect and establish vascular connections with the host, which facilitate the transfer of water, inorganic and organic nutrients and other macromolecules, such as mRNA from host to parasite and *vice versa* (Shahid *et al.*, 2018; Yoshida *et al.*, 2016; Nikolov *et al.*, 2014). Some parasite haustoria form direct connections with the xylem of their hosts, for example, *Striga*, *Triphysaria* and *Phtheirospermum* species, or to both the xylem and

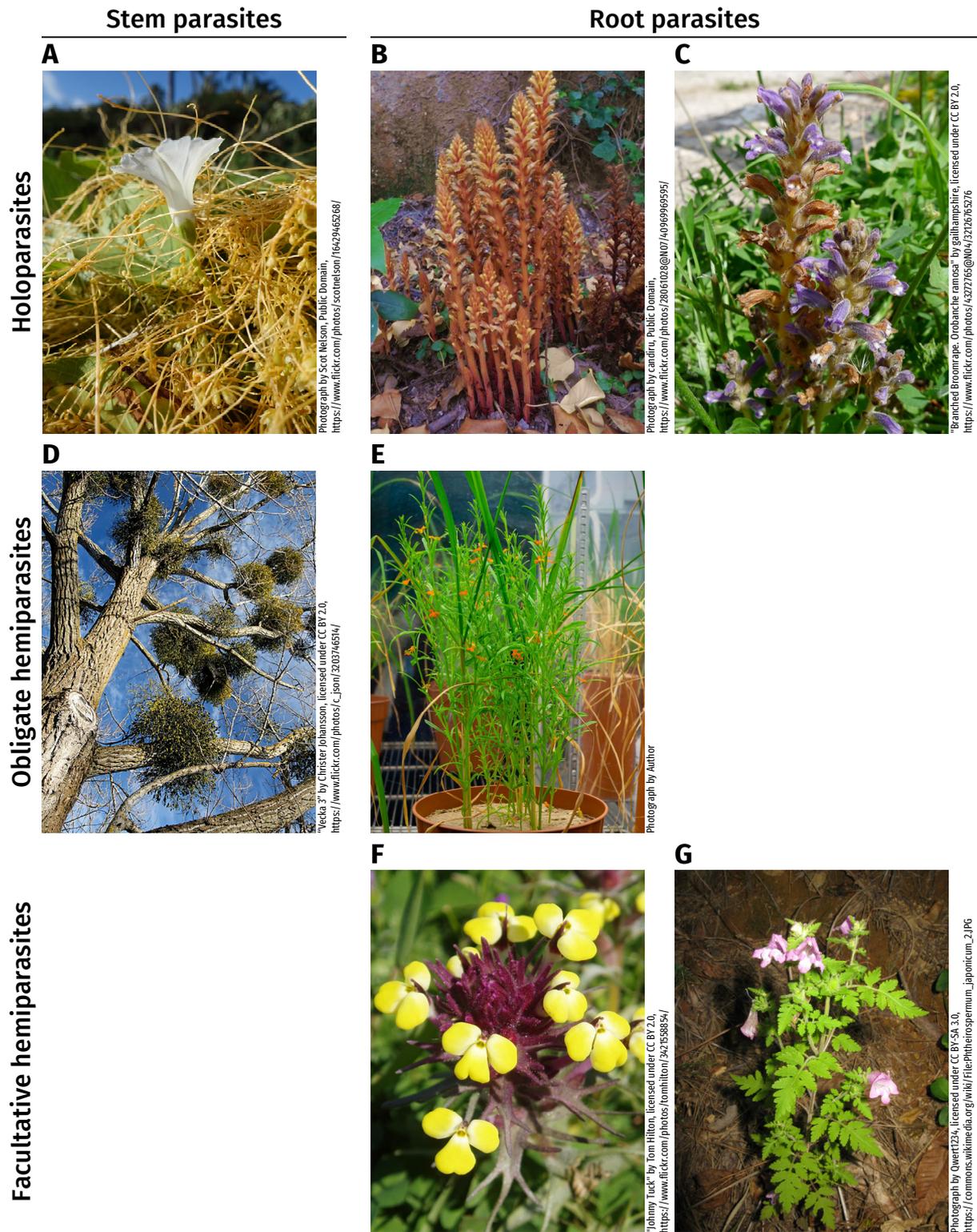


Figure 1.1 Parasitic plants can be categorised by the degree of parasitism and the site of attachment. Holoparasites (**A–C**) cannot photosynthesise and derive nutrients from the host, whilst hemiparasites (**D–G**) retains some ability to photosynthesise. Hemiparasites can be further divided into obligate hemiparasites (**D–E**), which require a host to complete their life cycle, and facultative hemiparasites (**F–G**), which can independently complete their life cycle but will parasitise a host when available. Stem parasites (**A & D**) attach to the stem of the host, whereas root parasites (**B–C & E–G**) attach to the host via the root. **A**, *Cuscuta sandwichiana*; **B**, *Orobanche hederiae*; **C**, *Phelipanche ramosa*; **D**, *Viscum album*; **E**, *Striga asiatica*; **F**, *Triphysaria eriantha*; **G**, *Phtheirospermum japonicum*.

phloem of their hosts, such as in *Cuscuta*, *Orobanche* and *Phelipanche* species. The OROBANCHACEAE Vent. is the most diverse family of parasitic plants, with at least 2110 species from 99 genera (Stevens, 2017; Spallek *et al.*, 2013; Bennett & Mathews, 2006). It is mostly comprised of root-parasitic plant species and is also unique in that its members exhibit a wide range of the degrees of parasitism, from the non-parasitic genus *Lindenbergia* Lehm. to the facultative hemiparasitic genera *Triphysaria* and *Phtheirospermum* and the obligate hemiparasitic genus *Striga* to the holoparasitic genera *Orobanche* and *Phelipanche* (Mutuku *et al.*, 2020; Tsybalyuk & Mosyakin, 2018; Fernández-Aparicio *et al.*, 2016; Spallek *et al.*, 2013; Westwood *et al.*, 2010). *Striga* and *Orobanche* species are globally the most damaging parasitic plants to agriculture and are considered a serious biotic threat to food security (Westwood *et al.*, 2010; Scholes & Press, 2008; Mohamed *et al.*, 2006).

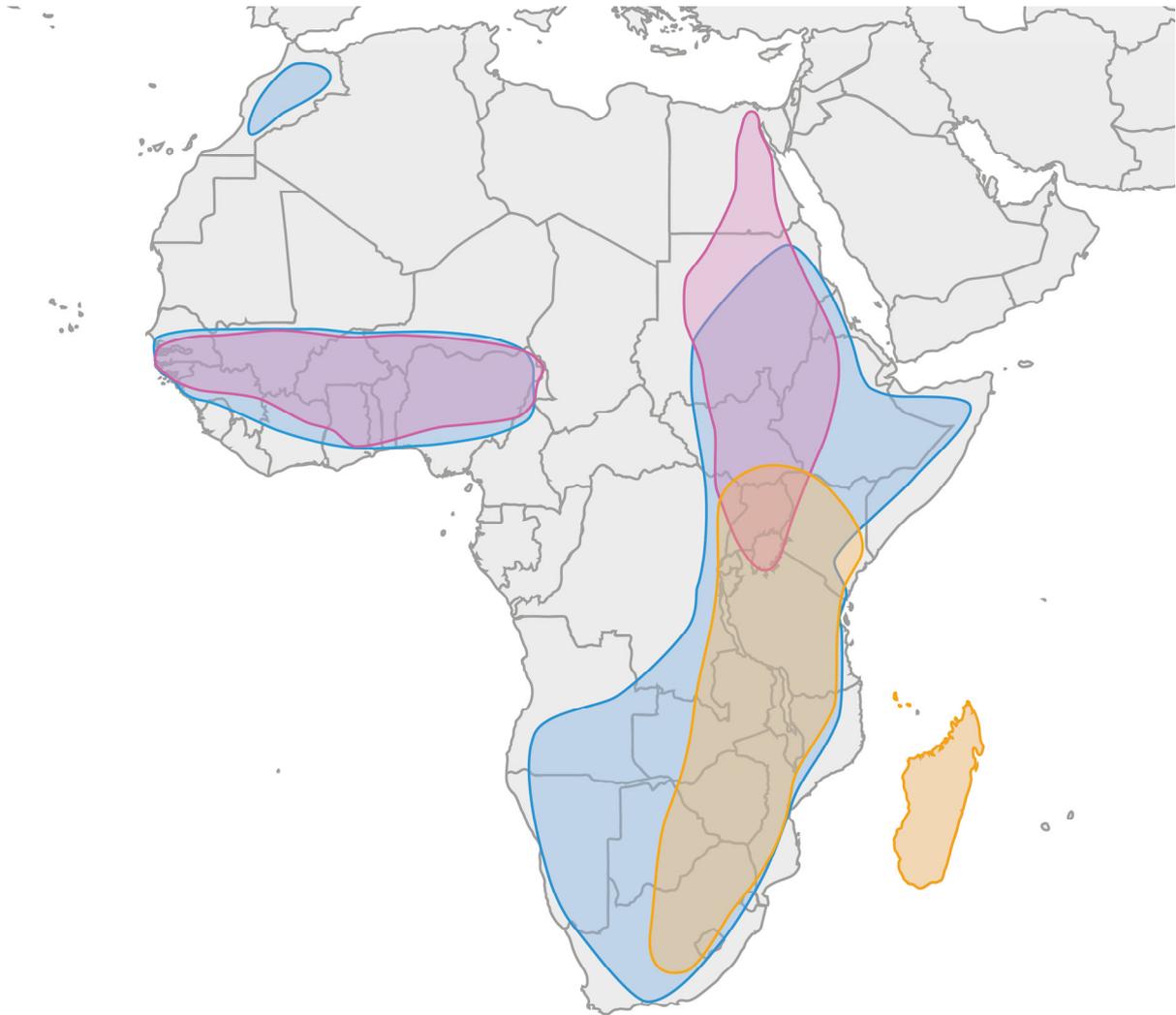
1.2 The genus *Striga*

The genus *Striga* is made up of at least 27 species (Spallek *et al.*, 2013; The Plant List, 2013), which are mainly endemic to Africa (Spallek *et al.*, 2013; Parker, 2012; Mohamed *et al.*, 2006). At least eleven *Striga* species infect crop plants (Mohamed *et al.*, 2001) and they are the most devastating agricultural pests to subsistence farmers in Africa (Scholes & Press, 2008). This is because the severity of *Striga* infection is negatively correlated with the soil fertility (Spallek *et al.*, 2013; Cochrane & Press, 1997) and technologies for soil improvement, such as fertiliser application, are often inaccessible to resource-limited farmers. *Striga* spp. thrive in semi-arid climates, which cover regions frequently used for rain-fed upland agriculture to produce staple cereal and legume crops in Africa. More than 500 000 km² (and growing) areas of arable land—in at least 25 African countries—are infested with one or more species of *Striga* (Spallek *et al.*, 2013; Westwood *et al.*, 2010; Parker, 2009). Farmers report yield losses due to *Striga* infestation, ranging from 20 to 90 %. Sometimes total crop failure occurs, which can result in the abandonment of *Striga*-infested farmland (Spallek *et al.*, 2013; Atera & Itoh, 2011; Scholes & Press, 2008). More than 100 million people are affected by agricultural losses due to *Striga* infestation (Yoder & Scholes, 2010; Scholes & Press, 2008), which is

economically equivalent to a yearly loss of US\$1 billion (Spallek *et al.*, 2013; Waruru, 2013; Parker, 2009; Berner *et al.*, 1995).

Three species, *Striga hermonthica* (Delile) Benth., *S. asiatica* (L.) Kuntze and *S. gesnerioides* (Willd.) Vatke, cause the most damage to crop production in Africa and inflict the largest economic losses (Spallek *et al.*, 2013; Scholes & Press, 2008; Mohamed *et al.*, 2006). *Striga hermonthica* has the largest impact, followed by *S. asiatica* and *S. gesnerioides*, respectively (Parker, 2009). The three *Striga* species can infect a wide range of host species, including economically important cultivated crops. *Striga gesnerioides* infects eudicots, such as cowpea [*Vigna unguiculata* (L.) Walp.] and tobacco (*Nicotiana* L. species), whilst *S. hermonthica* and *S. asiatica* infect monocots such as maize (*Zea mays* L.), sorghum [*Sorghum bicolor* (L.) Moench], upland rice (Asian rice, *Oryza sativa* L. and African rice, *O. glaberrima* Steud.), sugarcane (*Saccharum* L. species), pearl millet [*Cenchrus americanus* (L.) Morrone] and finger millet (*Eleusine coracana* Gaertn.) (Kountche *et al.*, 2016; Spallek *et al.*, 2013; Parker, 2012, 2009; Berner *et al.*, 1995).

The infestation of *Striga* weeds is widespread in Africa (Figure 1.2). *Striga hermonthica* is found in West and East Africa, where it predominately affects maize, sorghum and millets (Spallek *et al.*, 2013; Parker, 2009). *Striga hermonthica* is now also a problem in rain-fed upland rice production as the crop has gained traction from farmers in the regions in recent years (Atera & Itoh, 2011). *Striga asiatica* has the widest global distribution, identified in Africa, Asia, Australia and North America (Spallek *et al.*, 2013). Within Africa, *S. asiatica* is mainly found in Southern Africa, the south of the equator in East Africa, and also on the islands of Madagascar and Comoros (Rodenburg *et al.*, 2016; Spallek *et al.*, 2013) (Figure 1.2). Like *S. hermonthica*, crops affected by *S. asiatica* include maize, sorghum, millets and upland rice (Parker, 2009). In Madagascar, a major upland rice-growing area in Africa, *S. asiatica* is the main *Striga* species parasitising rice. Lastly, *S. gesnerioides* is found in West, East and Southern Africa (Figure 1.2), but causes problems mainly in the production of cowpea in West Africa (Spallek *et al.*, 2013).



 *Striga hermonthica*

 *Striga asiatica*

 *Striga gesnerioides*



Figure 1.2 Distributions of the three *Striga* species that cause the most damage to agriculture in Africa: *Striga hermonthica*, *S. asiatica* and *S. gesnerioides*. The map of *Striga* species distributions was adapted from Spallek et al. (2013).

1.3 Life cycle of *Striga*

Despite the farmers' urgent need for control measures against *Striga* spp., the progress towards finding effective solutions has been slow compared to other plant pathogens (Scholes & Press, 2008). In addition to many documented socio-economic constraints faced by farmers, aspects of the biology of the host–parasite interactions make control extremely difficult. Firstly, the parasite seed bank is large, long-lived and genetically diverse (Parker, 2009; Mohamed *et al.*, 2007). Secondly, the life cycle of the parasite and host are closely linked by the exchange of chemical signals (Clarke *et al.*, 2019; Cui *et al.*, 2018; Yang *et al.*, 2016; Ichihashi *et al.*, 2015; Bouwmeester *et al.*, 2003). Thirdly, *Striga* begins to adversely affect the growth and development of the host very shortly after attachment, thus control of the parasite must take place early during the underground phase of the life cycle before irreversible damage to the host occurs (Clarke *et al.*, 2019; Scholes & Press, 2008).

Understanding the life cycle of *Striga* and the chemical signals exchanged between host and parasite is essential to the design of effective control strategies. Dust-like seeds of *Striga* spp. (approximately 0.1–0.6 mm) go through two dormancy periods (Mohamed *et al.*, 1998; Parker & Riches, 1993) (Figure 1.3a). After maturing on the mother plant and released from the pods into the soil, *Striga* seeds enter a period of after-ripening dormancy for at least 6–12 months (Mohamed *et al.*, 1998; Parker & Riches, 1993). After-ripened seeds then become responsive to external cues for germination. Once *Striga* seeds have received a prolonged period of hydration in elevated temperature of ~ 30 °C for two weeks, they enter the period of wet dormancy, where the perception of root exudates from nearby plants breaks the dormancy and triggers the germination process (Cardoso *et al.*, 2011; Mohamed *et al.*, 1998; Parker & Riches, 1993) (Figure 1.3b). Root exudates of host and non-host species contain compounds known as germination stimulants, the best known of which are strigolactones (Bouwmeester *et al.*, 2019; Zwanenburg *et al.*, 2016; Yoneyama *et al.*, 2010). Strigolactones are sesquiterpene lactones derived from β -carotene (Jia *et al.*, 2018; Al-Babili & Bouwmeester, 2015). Recently, a set of strigolactone receptors have been identified that can recognise

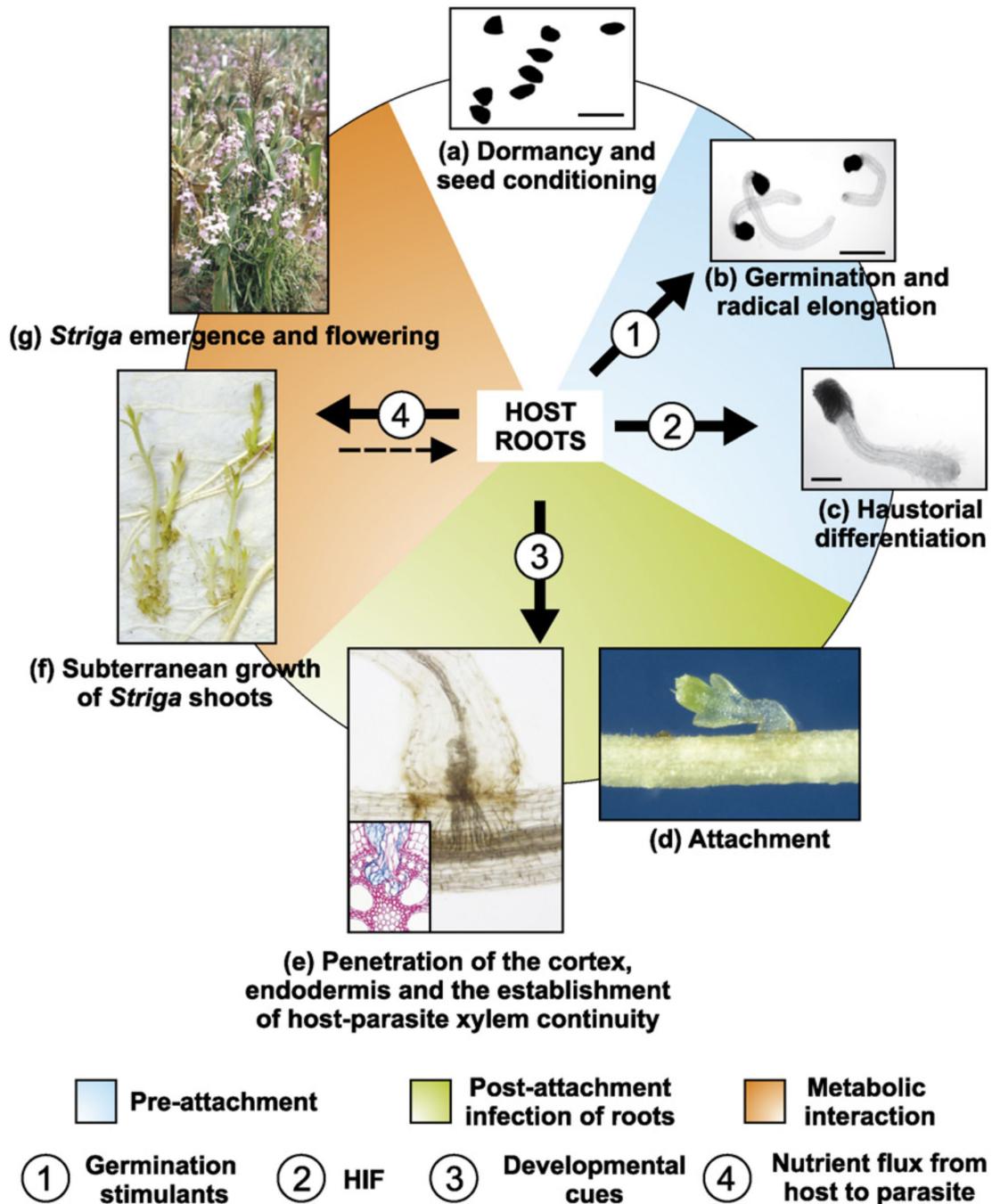


Figure 1.3 Life cycle of *Striga*. (a) *Striga* seeds can lie dormant in the soil up to 20 years before a period of elevated temperature and moisture, called conditioning period, prompts the seeds into wet dormancy. (b) In the wet dormancy, germination of *Striga* seeds are triggered by germination stimulants in host root exudates; the most well-known are strigolactones. (c) After germination, *Striga* radicle grows towards host roots, and begin to differentiate the haustorium once haustorium-inducing factors (HIFs) are detected. (d) The haustorium attaches to the host, and (e) penetrates through the host root cortex and endodermis of the host before forming direct connections to the host xylem. (f) Cotyledons and leaves develop, and stem elongates in the subterranean growth when the parasite completely relies on the host for water and nutrients. (g) *Striga* emerges above ground and begins to perform photosynthesis while still receiving water and nutrients from the host. Four to six weeks after emergence, *Striga* flowers. Each plant can produce up to 50 000–500 000 seeds. Scale bars = 0.25 mm (a–c). The figure was taken from Scholes and Press (2008).

different strigolactones at picomolar (10^{-12} mol L⁻¹) concentrations (Lumba *et al.*, 2017a, b; Toh *et al.*, 2015). The high sensitivity of these receptors ensures that *Striga* seeds only germinate in the proximity of the host to maximise their survival, as microscopic *Striga* seeds need to attach to the host soon after germination since they contain very limited resources. *Striga* seeds are viable in the ground for upwards of 20 years, providing more opportunities for the parasites to encounter potential hosts (Parker & Riches, 1993). Varieties of rice and sorghum that produce a low concentration of strigolactones or produce different types of strigolactones, to which *Striga* spp. are less sensitive, can offer pre-attachment resistance and are used as a part of integrated management against *Striga* (Gobena *et al.*, 2017; Cardoso *et al.*, 2014; Jamil *et al.*, 2011a).

After a *Striga* seed has germinated, an emerging radicle perceives host-derived haustorium-inducing factors (HIFs), such as 2,6-dimethoxy-1,4-benzoquinone (DMBQ), and develop a pre-haustorium, which is characterised by a hemispherical shape covered with haustorial hairs, within several hours after the perception of HIFs (Goyet *et al.*, 2019) (Figure 1.3c). Once the radicle with a pre-haustorium has extended towards and eventually reached the host root, the attachment and haustorium maturation occur (Goyet *et al.*, 2019; Yoshida *et al.*, 2016) (Figure 1.3d). The epidermal cells of the haustorium differentiate into intrusive cells, which separate host cells apart by releasing cell wall degrading enzymes to modify the integrity of host cell walls (Clarke *et al.*, 2019; Yoshida *et al.*, 2016). A comparative transcriptomic study on members of the family OROBANCHACEAE, including *S. hermonthica*, identified recruitment of genes, through gene duplication, from other invasive growth structures, such as pollen tube growth, to function in haustorial penetration (Yang *et al.*, 2015). The maturing haustorium penetrates through the host root cortex and endodermis until it reaches the host vasculature where *Striga* cells differentiate into tracheary cells that connect the parasite xylem to the host xylem vessels (Clarke *et al.*, 2019; Goyet *et al.*, 2019; Dorr, 1997) (Figure 1.3e). Only xylem–xylem connections, without any phloem connections, are formed between *Striga* spp. and their hosts (Dorr, 1997). Some varieties of rice, sorghum, maize and cowpea

can stop the penetration of parasite haustoria in the root cortex or at the endodermis in incompatible interactions, resulting in arrested growth and eventual death of parasites (Mutinda *et al.*, 2018; Mbuvi *et al.*, 2017; Cissoko *et al.*, 2011; Gurney *et al.*, 2006, 2003).

Subterranean *Striga* plants lack access to sunlight and live like a holoparasite, by completely relying on the host for access to water, organic and inorganic nutrients, which are transported from cell to cell or through direct xylem connections (Mohamed *et al.*, 2001; Parker & Riches, 1993) (Figure 1.3f). Fixed carbon was transferred from the host in the form of a sugar alcohol mannitol and free amino acids, such as asparagine and glutamine, whilst nitrogen was transferred as amino acids and nitrates (Pageau *et al.*, 2003; Press *et al.*, 1991). To facilitate the extraction of water and nutrients, *Striga* spp. irreversibly change the morphology and physiology of their host as soon as the vascular connections are established (Scholes & Press, 2008). Hence, any interventions to successfully protect the host from the adverse effects of *Striga* must act before or soon after the attachment of the parasites (Scholes & Press, 2008). In the host, more biomass being diverted into growing larger root systems, which when combined with a low transpiration rate of the host and an elevated transpiration rate of the parasites, results in a larger pool of water and nutrients that stay in host roots being transported to the parasites (Frost *et al.*, 1997). Parasites also reprogramme the physiology and growth of their hosts by changing levels of host plant hormones, such as elevated abscisic acid (ABA) in the host resulting in host's stomatal closure (Frost *et al.*, 1997). The loss in biomass of host plants infected with *Striga* is often not linearly proportional to the gain in parasite biomass (Gurney *et al.*, 1999). Partially, this is because the restriction on host stomatal opening leads to lower photosynthesis of the host, resulting in stunted and chlorotic hosts (Graves *et al.*, 1989). Also, it has been suggested that *Striga* spp. deliver toxins that inhibit the growth of the host (Waweru *et al.*, 2019; Spallek *et al.*, 2013). However, the identity of the toxic compounds is yet to be discovered.

After growing underground for several weeks, depending on the depth of the soil, *Striga* plants emerge above ground; they produce chlorophyll and begin to fix their own carbon via

photosynthesis (Parker & Riches, 1993) (Figure 1.3g). In some cases, the above-ground tissues of the hosts completely die off, leaving only the root systems for the parasite to keep using for water and nutrient absorption. Approximately 4–6 weeks after emergence, *Striga* plants start flowering. *Striga hermonthica* is self-incompatible and requires cross-pollination to set seeds (Safa *et al.*, 1984). This results in high heterozygosity in *S. hermonthica* and highly diverse seed banks (Welsh & Mohamed, 2011). On the other hand, *S. asiatica* and *S. gesnerioides* can outcross but preferentially self-pollinate (Lane *et al.*, 1994; Nickrent & Musselman, 1979). Within four weeks after pollination, a *Striga* plant can produce up to 50 000–500 000 seeds, which are then released to the soil (Berner *et al.*, 1995). High fecundity of *Striga* allows the seed bank to build up rapidly in fields if the parasites are left unmanaged.

1.4 Current controls for the management of *Striga* weeds

Controlling *Striga* weeds is a difficult task because of not only the intertwining life cycle between the parasites and their hosts but also the inaccessibility of underground parasites at the time when any interventions are most effective in protecting against irreversible damage to the hosts. Several practices are currently used by farmers with varying degrees of success. The main *Striga* control used by smallholder farmers in Africa is manual weeding (N'cho *et al.*, 2019; Tippe *et al.*, 2017; Hearne, 2009). Although hand weeding prevents the return of more *Striga* seeds into the seed bank if carried out before parasite flowering, it is inefficient and labour intensive, and it cannot reverse the damage the parasites have already inflicted upon the crop hosts (N'cho *et al.*, 2019; Hearne, 2009). Another way to reduce the number of parasites is to improve soil fertility. As the availability of nitrogen and phosphorus increases, a lower quantity of exudates from host roots is secreted into the rhizosphere and fewer *Striga* seeds germinate (Jamil *et al.*, 2011b). However, fertilisers are not widely available and are too expensive for smallholder farmers, only a third of whom report some fertiliser applications in their fields, preventing extensive adoption of the measure (Mrema *et al.*, 2017; Atera *et al.*, 2012a; Hearne, 2009).

Instead, crop rotation and intercropping can also improve soil fertility and are more affordable. Use of catch crops, such as fodder grass and pasture legumes, which are planted to initiate germination of *Striga* seeds but are then destroyed or harvested before the parasites reproduce, would deplete the *Striga* seed bank after a few consecutive years of application (Teka, 2014). Subsistence farmers are less likely to adopt this strategy as it requires the land not being used for food production for several successive seasons (Mrema *et al.*, 2017; Hearne, 2009). Alternatively, intercropping staple crops with trap crops, such as a fodder legume *Desmodium* spp. can alleviate *Striga* infestation (Kifuko-Koech *et al.*, 2012; Khan *et al.*, 2002). Not only do *Desmodium* spp. trigger suicidal germination of *Striga* seeds and increase soil nitrogen availability, but *Desmodium* spp. also produce compounds inhibiting *Striga* attachment to maize (Khan *et al.*, 2002). However, intercropping also faces similar adoption hurdles as crop rotation; that are farmers' unwillingness to plant non-food crops and the requirement of repeated usage over a few seasons before the change in *Striga* infestation level becomes noticeable (Mrema *et al.*, 2017; Teka, 2014; Hearne, 2009).

Although many herbicides, such as 2,4-D, selectively target eudicot weeds, which *Striga* spp. are, without harming monocot crops, subterranean growth before emergence prevents effective herbicide application (Teka, 2014). The only effective use of herbicide to control *Striga* spp. exists as acetolactate synthase-inhibiting herbicides-coated seeds of herbicide-resistant maize, which are already in use in East Africa (Rodenburg *et al.*, 2010; De Groote *et al.*, 2008; Manyong *et al.*, 2008; Kanampiu *et al.*, 2003). However, herbicide-coated seeds are sensitive to the environmental conditions as inappropriate timing of heavy rainfall may wash away the herbicide before *Striga* seeds are affected (Kanampiu *et al.*, 2003). Although farmers are enthusiastic about the technology, the availability and the recurring cost of herbicide-treated seeds are likely to prevent a wide adoption in smallholder farmers as they tend to save seeds on the farm to cultivate from one season to the next (Manyong *et al.*, 2008).

Striga spp. can also be controlled by the use of resistant and tolerant crops. Resistance is defined as a reduction in parasite load on the host, whilst tolerance is characterised by the

ability of a host variety to be less affected by a given parasite load than another variety of the same species (Rodenburg & Bastiaans, 2011). Genetic resistance is compatible with the current farmer practice as farmers can cultivate food crops continually and save the seeds of resistant/tolerant cultivars for future planting. The use of resistant crops has an extra benefit of depleting *Striga* seed bank (Scholes & Press, 2008).

Several sources of genetic resistance have been reported in sorghum (Mohemed *et al.*, 2018; Mbuvi *et al.*, 2017; Haussmann *et al.*, 2004, 2001b; Rich *et al.*, 2004; Gworgwor, 2003; Showemimo, 2003; Ejeta & Butler, 1993), maize (Gasura *et al.*, 2019; Menkir & Meseke, 2019; Mutinda *et al.*, 2018; Midega *et al.*, 2016; Amusan *et al.*, 2008; Menkir, 2006; Gethi & Smith, 2004; Gurney *et al.*, 2003), rice (Beardon, 2018; Rodenburg *et al.*, 2017, 2015; Samejima *et al.*, 2016; Cissoko *et al.*, 2011; Jamil *et al.*, 2011a; Yoshida & Shirasu, 2009; Kaewchumnong & Price, 2008; Gurney *et al.*, 2006; Harahap *et al.*, 1993), pearl millet (Sattler *et al.*, 2018) and cowpea (Li & Timko, 2009; J. Li *et al.*, 2009; Lane *et al.*, 1993). Compatibility between parasites and host (resulting in susceptible or sensitive hosts in the compatible interactions) depends on several factors, including the infestation level, the environmental conditions and the specificity of host genotypes against parasite strains/ecotypes (Rodenburg *et al.*, 2017; Haussmann *et al.*, 2001b, c). So far none of the resistant crop varieties offers a complete resistance against *Striga* spp., as a few parasites can grow on even highly resistant cultivars (Beardon, 2018; Gurney *et al.*, 2006, 2003). There is a risk of resistance breakdowns as surviving *Striga* individuals receive a strong selection pressure. Furthermore, a vast *Striga* seed bank harbours immense genetic diversity, especially in highly heterozygous *S. hermonthica*, and the seed viability is long-lasting in the soil seed bank (Welsh & Mohamed, 2011; Berner *et al.*, 1995). Emphasis should be made on stacking multiple resistance genes from different sources to attain durable resistance.

In addition to resistance, several sources of tolerance have been identified, using the host's ability to maintain its height or photosynthetic rate under *Striga* infestation, in rice (Rodenburg *et al.*, 2017), sorghum (Rodenburg *et al.*, 2008, 2005) and maize (Pierce *et al.*,

2003; Gurney *et al.*, 2002). Although resistance and tolerance do not often occur in the same host, combining both aspects in crop varieties recommended for farmers will result in high yielding varieties despite being infected by a few parasites (Rodenburg *et al.*, 2017; Atera *et al.*, 2012b; Rodenburg & Bastiaans, 2011).

1.5 Different phenotypes of *Striga* resistance

Resistance against parasitic plants is generally regarded as non-host resistance and host resistance, where plant species that are not parasitised under natural conditions are considered non-hosts (Timko & Scholes, 2013). Host-resistance, which is relevant to crop protection as many crop species are natural hosts of *Striga*, has been reported in various cultivars and landraces for over half a century and involves several mechanisms (Timko & Scholes, 2013; Ejeta, 2007). In order to incorporate various forms of *Striga* resistance into a host genotype by stacking multiple resistance genes together, the genetic loci (and possibly gene identities), the mechanisms and the phenotypes of resistance genes must be identified and characterised first, to ensure compatible and diverse resistance genes are included, which is a key to durable resistance (Dormatey *et al.*, 2020; Mundt, 2014).

The phenotypes of resistance can be categorised according to infection stages at which parasite invasion is stopped by the host, including during the germination of parasite seeds and the induction of parasite haustoria before attachments, during the penetration through host root cortex and endodermis after the attachments and after the formation of host–parasite vascular connections (Figure 1.3) (Timko & Scholes, 2013; Yoshida & Shirasu, 2009; Scholes & Press, 2008). *Striga* seeds' requirement for signals from the host to germinate presents an opportunity to prevent germination of parasite seeds. Most common *Striga* germination stimulants in host root exudate are strigolactones, which are produced in different quantities and compositions among varieties of the same crop species (Bouwmeester *et al.*, 2019, 2007; Yoneyama *et al.*, 2010; Awad *et al.*, 2006). Certain strigolactones, such as 5-deoxystrigol, are more potent in stimulating the germination of *Striga* seeds than others (Mohemed *et al.*,

2018; Yoneyama *et al.*, 2009). Varieties of rice and sorghum that produce a low overall quantity of strigolactones or produce predominantly the low-germination-stimulating strigolactones have been identified and shown to have fewer parasite attachments (Mohemed *et al.*, 2018, 2016; Jamil *et al.*, 2011a; Hausmann *et al.*, 2004; Vogler *et al.*, 1996). After germination, *Striga* spp. also require host-derived HIFs to begin the differentiation of a haustorium. Wild sorghums (*S. bicolor*) that produce low HIFs were identified and used in the mapping of the *Lbf* (low haustorial factor) locus (Grenier *et al.*, 2007; Rich *et al.*, 2004). A low HIF-producing phenotype was also discovered in *Tripsacum dactyloides* (L.) L., a wild relative of maize, resulting in resistance to *S. hermonthica* (Gurney *et al.*, 2003).

As *Striga* penetrates through the host root cortex after attachment, resistant hosts respond by rapid accumulation of cytotoxic phenolic compounds, as seen in *S. hermonthica*–sorghum and *S. asiatica*–non-host plants interactions (Arnaud *et al.*, 1999; Hood *et al.*, 1998). Moreover, the cell walls of host root cells are also modified in response to parasite invasion. Cell walls may exhibit increased lignification, suberisation and callose deposition at the site of attachment, for example, in cowpea–*S. gesnerioides* associations (Botanga & Timko, 2005; Lane, 1996). Some hosts respond to parasite invasion with a hypersensitive response (HR)-like response, which appears as browning, relatively rapidly and eventual death of host cortical cells surrounding the site of attachment (Timko & Scholes, 2013). HR-like responses have also been reported in incompatible *Striga*–host interactions, such as in *S. hermonthica*–rice, *S. hermonthica*–sorghum and *S. gesnerioides*–cowpea interactions (Gurney *et al.*, 2006; Mohamed *et al.*, 2003; Lane *et al.*, 1993).

Before parasites can initiate parasite–host xylem connectivity, they must breach the endodermis, which acts as a physical barrier that substantially prevents the penetration of root parasitic weeds (Timko & Scholes, 2013). In highly resistant hosts, the *Striga* endophyte cannot penetrate through the endodermis and grows around the host root steel, often exiting the host root. This type of resistance has been observed in the interactions between the parasites, *S. hermonthica* and *S. asiatica*, and hosts, including rice (Beardon, 2018; Cissoko *et*

al., 2011; Yoshida & Shirasu, 2009; Gurney *et al.*, 2006), sorghum (Hausmann *et al.*, 2004; Maiti *et al.*, 1984), maize (Mutinda *et al.*, 2018) and its wild relative, *Zea diploperennis* Iltis, Doebley & R.Guzmán (Amusan *et al.*, 2008). In addition, some hosts cannot completely block the penetration of endodermis but can restrict parasites' withdrawal of water and nutrients through limited vascular connections, resulting in slow-growing or arrested growth of parasites, for example, in the *S. hermonthica*–sorghum (Mbuvi *et al.*, 2017; Arnaud *et al.*, 1999), *S. hermonthica*–*Z. diploperennis* (Amusan *et al.*, 2008) and *S. hermonthica*–rice interactions (Cissoko *et al.*, 2011). These parasites with restricted penetration of the endodermis were observed with the deposition of densely stained materials inside the xylem vessels (Cissoko *et al.*, 2011) and poor differentiation of parasite hyaline body, which is an area inside the haustorium used for the acquisition and storage of nutrients received from the host (Yoshida *et al.*, 2016; Amusan *et al.*, 2008; Gurney *et al.*, 2003).

Despite successful host–parasite vascular connections, some hosts can still elicit resistance through abiosis and haustorium disorganisation (Timko & Scholes, 2013; Yoshida & Shirasu, 2009). A transfer of toxic compounds has been suggested to be responsible for *T. dactyloides*'s ability to kill off the parasites despite no obvious blockage in the xylem–xylem connections (Gurney *et al.*, 2003). *T. dactyloides* also has the haustorium-inhibitive phenotype, in which *S. hermonthica* attached to it could not form a subsequent secondary haustorium on normally susceptible maize (Gurney *et al.*, 2003). The sizes of parasites are also smaller on some hosts, such as the intermediately resistant rice cultivar Kasalath (Gurney *et al.*, 2006) and the maize variety KSTP 94 (Mutinda *et al.*, 2018), compared to susceptible hosts in spite of apparent successful xylem–xylem connections, suggesting some degrees of parasite–host physiological incompatibility (Timko & Scholes, 2013). In addition, disorganisation of haustoria, where vascular differentiation inside parasite tubercle is absent, was reported in incompatible *S. gesnerioides*–cowpea and wild legumes interactions (Botanga & Timko, 2005).

1.6 Dissection of the genetic basis of *Striga* resistance

Natural variations in the levels of resistance existing in host varieties and landraces have been used by geneticists to dissect the genetic basis of *Striga* resistance, such as the identities and the heritability of genes or gene clusters that underly resistance phenotypes (Timko & Scholes, 2013). Generally, resistance against *Striga* spp. is polygenic that includes several resistance genes with major and minor effects and has large genotype \times environment interactions in members of the family POACEAE Barnhart, which encompasses all cereal crop species (Timko & Scholes, 2013; Scholes & Press, 2008). On the other hand, resistance against *S. gesnerioides* is largely monogenic, which is reflected in the apparent race structures of the parasite (Timko & Scholes, 2013; J. Li *et al.*, 2009). Over the years, several techniques, including linkage mapping, association mapping and segregation analyses in F₁ (first filial generation), F₂ and BC₁ (first backcross generation) populations, have been employed to derive the inheritance patterns, locations and identities of resistance genes against *Striga* spp.

Resistance against *S. gesnerioides* in cowpea is by far the most understood. Observations of cowpea cultivars that are resistant to *S. gesnerioides* from one location being susceptible to *S. gesnerioides* from other geographically distant locations prompted a suggestion that distinct races exist between *S. gesnerioides* from different geographic locations (Timko & Scholes, 2013). Owing to its reproduction strategy, preferentially inbred *S. gesnerioides* has been found to have a clear race structure, comprised of seven known races with differential abilities to infect a test panel of cowpea cultivars (J. Li *et al.*, 2009; Lane *et al.*, 1994). The position of resistance gene, specific to each *S. gesnerioides* race, was mapped to two linkage groups of cowpea (Timko *et al.*, 2007). A resistance gene, *RSG3-301*, was the first to be cloned from a resistant cowpea cultivar B301, using a positional cloning approach (Li & Timko, 2009). *RSG3-301* encodes a membrane-bound protein that contains coiled-coil (CC), nucleotide-binding (NB) and leucine-rich repeat (LRR) domains, which is similar to plant resistance (R) proteins that are involved in pathogen recognition (Li & Timko, 2009).

In sorghum, the segregation analysis of low germination stimulant phenotype in F₁, F₂ and BC₁ mapping populations, derived from crosses between the resistant sorghum variety SRN39 and susceptible genotypes, revealed that the resistance phenotype is controlled by a single recessive nuclear gene (Vogler *et al.*, 1996). The position of the *LGS1* (*LOW GERMINATION STIMULANT 1*) gene, responsible for the resistance phenotype in SRN39, was later mapped using a population of recombinant inbred lines (RILs) derived from a cross between SRN39 and the susceptible landrace Shanqui Red (Satish *et al.*, 2012). It was later found that the *LGS1* allele from SRN39 contains a loss-of-function mutation that switches the dominant strigolactone in sorghum root exudate from the highly germination-stimulating 5-deoxystrigol to the less potent orobanchol (Gobena *et al.*, 2017). In contrast, Haussmann *et al.* (2001a) found that the low germination stimulant phenotypes in three other sorghum genotypes are controlled by a major recessive gene with additional several minor genes. A subsequent analysis of two sets of a RIL population, derived from a cross between the low germination stimulant producing genotype IS9830 and the susceptible genotype E36-1, revealed nine and eleven quantitative trait loci (QTL), explaining ~ 80 % of the phenotypic variance (Haussmann *et al.*, 2004). Among these QTL, the QTL with the highest association significance was located at the same position as the *LGS1* locus. In the same study, Haussmann *et al.* (2004) also identified nine and eleven QTL underlying mechanical resistance that explained ~ 80 % of the phenotypic variance, using two sets of a RIL population derived from a cross between the resistant genotype N13 and the susceptible genotype E36-1. In this case, resistance was a quantitative trait, controlled by many genes.

In another study, the low production of HIF, which was observed in the wild sorghum (*S. bicolor*) PQ434 in an *in vitro* assay, was found to be elicited by a single dominant nuclear locus, *Lbf* (Grenier *et al.*, 2007). A mapping approach using two mapping populations of F₂-derived F₃ (F_{2:3}), originated from PQ434 × Shanqui Red cross was used in the study (Grenier *et al.*, 2007). In a different study, F₁, F₂ and BC₁ mapping populations derived from crosses between another two wild sorghum genotypes (CK32 and KP33, possessing strong HRs) and

two sorghum cultivars (having no HR) were used to identify two nuclear loci for HR to *Striga* (*Hrs1* and *Hrs2*) (Mohamed *et al.*, 2010). The mechanism of endodermal resistance and the identity of underlying genes have yet to be identified in sorghum.

In maize, Amusan (2011) used an F₂ mapping population that derived from a cross between the susceptible genotype 5057 and the resistant genotype ZD05 — which is derived from a backcross of maize and its wild relative *Z. diploperennis* (Amusan *et al.*, 2008)— to identify two major QTL (each explaining 27 and 28% of the phenotypic variance) for post-attachment resistance against *S. hermonthica*. In the resistant host, *Striga* endophyte often failed to penetrate the host endodermis; those that breached the endodermis form limited connections to host xylems and had poorly developed hyaline body.

In rice, a wide range of variations in the levels of both pre- and post-attachment resistance exist in both Asian and African cultivated rice species. For pre-attachment resistance, a major QTL on chromosome 1 for *S. hermonthica*-resistance that explains 18% of the variance in the number of emerged parasites was identified using an *O. sativa* mapping population that comprises of RILs derived from a cross between the resistant cultivar Bala and the susceptible cultivar Azucena (Kaewchumnong & Price, 2008). A subsequent QTL analysis using the same mapping population found that the position of the resistance QTL co-located with a major QTL for strigolactone biosynthesis (Cardoso *et al.*, 2014). The Bala allele of the QTL contains a deletion of two cytochrome P450 genes, *STRIGOLACTONE BIOSYNTHESIS 1* (*SLB1*) and *SLB2*, that are orthologs of the *Arabidopsis thaliana* (L.) Heynh. *MAX1* gene, which is involved in strigolactone biosynthesis, resulting in a reduction in the quantity of the germination-stimulating 5-deoxystrigol produced, in root exudate (Cardoso *et al.*, 2014).

For post-attachment resistance, several mapping populations have been heavily utilised to unravel the genetic basis of resistance against *S. hermonthica*. An *O. sativa* population of a BC₁F₁₄ backcross inbred lines (BILs), derived from the Nipponbare/Kasalath//Nipponbare cross between the resistant cultivar Nipponbare and the more susceptible cultivar Kasalath,

revealed seven QTL on chromosomes 1, 4, 5, 6, 7, 8 and 12, explaining a total of 31 % of the phenotypic variance (Gurney *et al.*, 2006). All QTL, except that on chromosome 4, had the resistance alleles coming from Nipponbare background (Gurney *et al.*, 2006). Large allelic substitution values, compared to parental scores, and the bimodal distribution of BIL phenotypic scores indicated that the resistance is controlled by few genes of major effects (Gurney *et al.*, 2006). In a follow-up study, Swarbrick *et al.* (2009) verified that the Kasalath-derived resistance allele of the QTL on chromosome 4 acts independently of the genetic backgrounds. This was demonstrated using another *O. sativa* population of BC₁F₇ BILs, derived from the Koshihikari/Kasalath//Koshihikari cross between the intermediately resistant cultivar Kasalath and the more susceptible cultivar Koshihikari. In the study, one genome-wide significant QTL, which explained 16 % of the phenotypic variance, was identified on chromosome 4 with an overlapping location to that in the previous study (Swarbrick *et al.*, 2009).

A QTL of major effect for post-attachment resistance against *S. hermonthica* has recently been identified using an *O. sativa* population of RILs by Beardon (2018). The population was derived from a cross between the resistant cultivar IR64 and the susceptible cultivar Azucena. A single major QTL, which was derived from IR64, was detected on chromosome 12 and explained 68 % of the phenotypic variance. This QTL on chromosome 12 has been mapped twice, first by Gurney *et al.* (2006) and second in a study involved a BC₄F₅ BIL population, derived from a backcross between two *O. sativa* cultivars, Nipponbare and Koshihikari (Nipponbare/Koshihikari//Koshihikari) (Scholes *et al.*, unpublished data). The Nipponbare × Koshihikari BIL population also yielded only a single QTL on chromosome 12, which accounted for 55 % of the phenotypic variance (Scholes *et al.*, unpublished data). The QTL spans over the 5.7–6.7 Mbp positions on the Nipponbare reference genome and contains 131 and 76 predicted genes on Nipponbare and IR64 QTL, respectively (Beardon, 2018). Attempts to narrow down the span of the QTL met with difficulty as there was a lack of recombination breakpoints within the QTL region (Beardon, 2018). From analysing the

sequences of genes within the QTL, it was found that the QTL also contains a cluster of resistance genes encoding receptor-like proteins (RLP), which were annotated as orthologs of *Verticillium* wilt resistance (*R*) genes in tomato (*Solanum lycopersicum* L.) (Beardon, 2018).

Recently, two genome-wide association studies (GWASes) have been carried out to identify resistance loci against *S. hermonthica*. Adewale *et al.* (2020) performed a GWAS on early maturing white tropical maize inbred lines, using *Striga* damage scores in the field to infer host resistance, and identified two loci significantly associated with a reduction in *Striga* damage, each explaining 29 and 42 % of the phenotypic variance. The candidate genes at the two loci were the *ZmCCD1* gene, which encodes a maize carotenoid cleavage dioxygenase and was associated with the colonisation of arbuscular mycorrhizal fungi that could limit the germination of *Striga* seeds (Sun *et al.*, 2008), and the *amt5* gene, which encodes an ammonium transporter protein. The authors of the study hypothesised that the expression of *amt5* changes host nitrogen status which in turn affects defence against *S. hermonthica* parasitism, since an application of nitrogen fertilisers reduces the emergence of parasites (Adewale *et al.*, 2020).

Another GWAS for *S. hermonthica*-resistance was carried out on sorghum. Kavuluko *et al.* (2020) performed a GWAS on a diverse panel of sorghum genotypes, using the number, the mean length and the biomass of parasites in a soil-free system to determine the level of *Striga*-resistance. Many candidate genes at several loci that were significantly associated with the resistance were identified. These included genes involved in secondary metabolite transport [e.g. pleiotropic drug resistance (PDR) family of ATP-binding cassette (ABC) transporter gene], cell wall metabolism (e.g. fasciclin-like arabinogalactan and xylanase inhibitor genes) and the regulation of defence response (e.g. ethylene-responsive transcription factor and disease resistance protein genes) (Kavuluko *et al.*, 2020).

1.7 Expanding rice production in Africa is threatened by the infestation of *Striga* weeds.

Rice is one of the most important staple crops in developing countries, especially in Africa where it is grown in 38 countries to meet a growing demand, which has increased noticeably in recent years due to population growth, the rise of income and a shift in consumer preference (Seck *et al.*, 2012; Balasubramanian *et al.*, 2007). One of the two species of cultivated rice, *O. glaberrima*, has been domesticated and grown for consumption in Africa for centuries (Cubry *et al.*, 2018; Nwanze *et al.*, 2006). However, in recent years *O. sativa* has gained more popularity in Africa as both a staple crop and a cash crop because it yields higher than locally adapted *O. glaberrima* (Cubry *et al.*, 2018; Rodenburg *et al.*, 2010). It was estimated that 40 % of the rice consumed in Africa comes from elsewhere as domestic production cannot keep up with the fast-growing demand (Seck *et al.*, 2010). One of the major biotic constraints to rice production in Africa is the infestation of *Striga* spp., occurring in at least 31 upland rice-growing countries (Rodenburg *et al.*, 2016) (Figure 1.4).

The infestation of one or more species of *Striga* has been conservatively estimated to affect the 8870 km² rain-fed upland rice-growing area and cause 488 000 tonnes of rice yield loss (equivalent to 293 000 tonnes milled rice), costing US\$117 million annually (Rodenburg *et al.*, 2016; N'cho *et al.*, 2014). Efforts to develop rice varieties to overcome both biotic and abiotic constraints in rice production have resulted in varieties recommended for farmers, such as *Striga*-resistant New Rice for Africa (NERICA) varieties that were developed by crossing high yielding *O. sativa* with more biotically and abiotically resilient *O. glaberrima* (Rodenburg *et al.*, 2015; Diagne *et al.*, 2013; Cissoko *et al.*, 2011; Jamil *et al.*, 2011a).

Most *Striga*-resistance work on rice has been carried out on interactions with *S. hermonthica*, leaving a gap in knowledge regarding rice interactions with *S. asiatica*, which is widespread in rice-growing regions in Southern Africa and the southern half of East Africa, including Comoros and Madagascar (Figure 1.2). More diverse sources for resistance against both

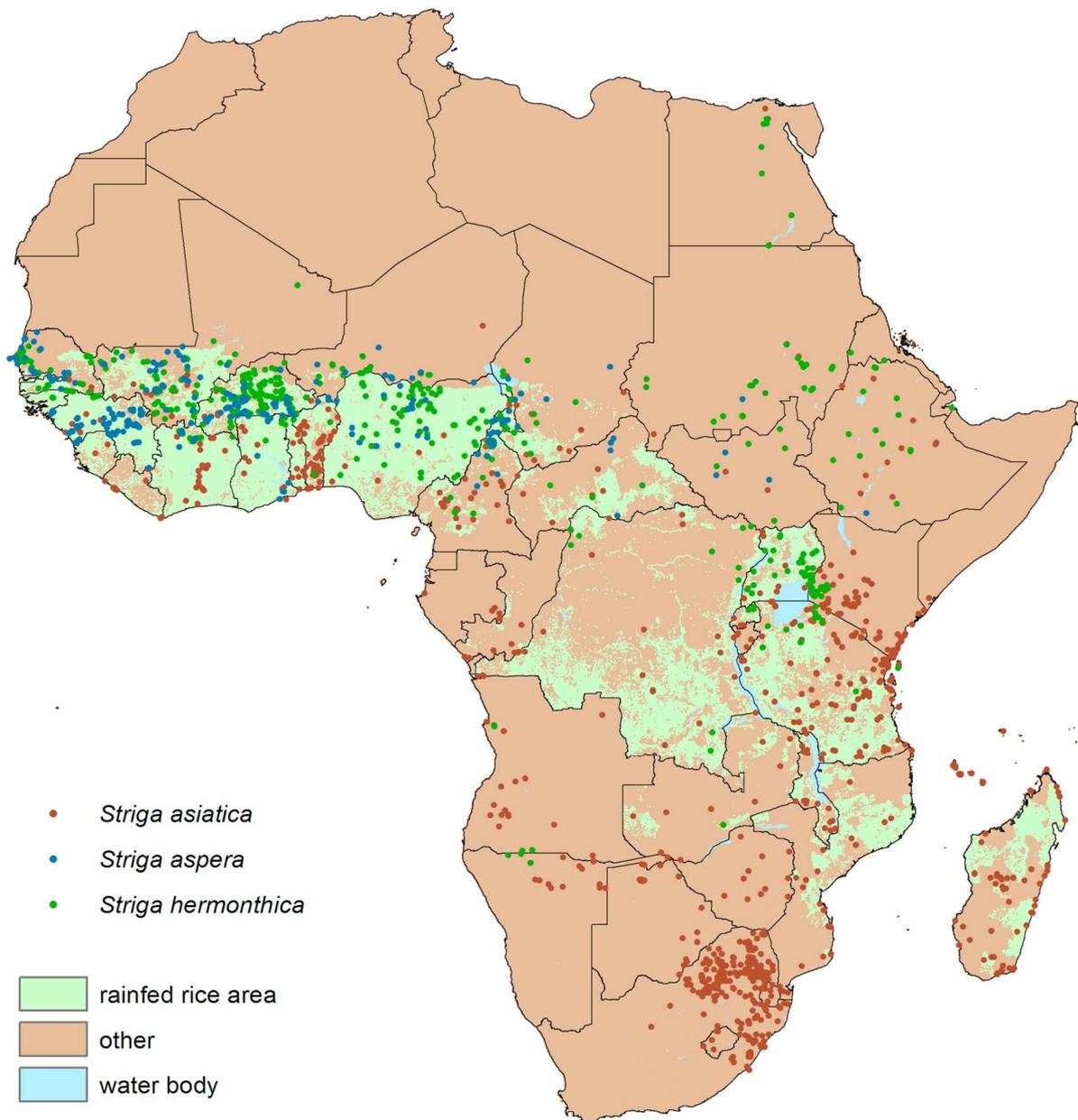


Figure 1.4 Distribution of the collection sites of herbarium samples of *Striga* species and rain-fed rice-producing areas in Africa. The figure was adapted from Rodenburg *et al.* (2016).

S. asiatica and *S. hermonthica* are still needed in anticipation of resistance breakdown and breeding of durable resistance through pyramiding of multiple resistance genes that elicit different resistance mechanisms.

1.8 Aims of Thesis

The aims of this thesis are (i) to identify and characterise resistance phenotypes in rice against *S. asiatica* and to identify genetic loci and candidate genes underlying the resistance

phenotype, using quantitative genetic approaches, and (ii) to perform a GWAS on an existing dataset of the levels of resistance/susceptibility in an *O. glaberrima* diversity panel to *S. hermonthica*, to identify loci underlying resistance. Specific aim(s) for each chapter include:

- Chapter 2: (i) To select a suitable RIL population by phenotyping the parental genotypes of three RIL populations to select one where the two parental genotypes exhibit different levels of resistance/susceptibility to *S. asiatica* in order to map the genetic basis of the phenotype, and (ii) to characterise the phenotype of resistance/susceptibility of the chosen parental genotypes at a macroscopic and microscopic level.
- Chapter 3: To map QTL that underlie the *S. asiatica*-resistance phenotype, which was identified and characterised in Chapter 2, using several different QTL mapping approaches on a RIL population selected in Chapter 2.
- Chapter 4: To identify possible mechanisms and candidate genes within the regions of the QTL underlying the resistance to *S. asiatica*, which were mapped in Chapter 3, by comparing the genomes of the parental genotypes of the RIL population.
- Chapter 5: To identify loci that are associated with resistance to *S. hermonthica* in rice by performing a GWAS, using a panel of diverse *O. glaberrima* accessions, whose resistance/susceptibility phenotypes were available from a phenotypic screen done prior to the start of the PhD programme.

Chapter 2

**Selection of a rice
recombinant inbred line (RIL) population
for mapping resistance to *Striga asiatica*:**

**Characterisation of
the resistance/susceptibility phenotype of
the parents of three RIL populations**

2.1 Introduction

Rice is the most rapidly expanding staple crop in Africa (Muthayya *et al.*, 2014; Balasubramanian *et al.*, 2007). From 1961 to 2013, the production and the consumption of rice in Africa have swiftly increased by 7.5 and 9.0 times, respectively (International Rice Research Institute, 2013). The demand for rice in the region is also projected to keep rising from rapid population growth and growing consumer preference for rice as a result of quick urbanisation (Van Oort *et al.*, 2015). However, over 40 % of rice consumed in Africa is imported because the production cannot keep up with the demand, and this supply–demand gap is also getting wider (Demont, 2013; Seck *et al.*, 2010). One of the major constraints imposed onto rice production in Africa is the parasitic weed *Striga*, which is widespread in rain-fed upland rice-producing areas. The two main species of *Striga* that affect rice production are *S. hermonthica* (in West, Central Africa and East Africa) and *S. asiatica* (in Southern Africa and the southern half of East Africa, including Comoros and Madagascar) (Parker, 2013). Annually, *Striga* infestation results in a loss of 293 000 tonnes of milled rice, which costs over US\$100 million to the economy (Rodenburg *et al.*, 2016).

Striga infestation disproportionately affects resource-limited subsistent farmers. Current *Striga* control measures, which include manual weeding, application of fertilisers, crop rotation, intercropping and the use of herbicide coated seeds, are either ineffective, unaffordable, not widely available or not widely adopted (Mrema *et al.*, 2017; Atera *et al.*, 2012b; Hearne, 2009) (see Section 1.4). The use of resistant crop varieties has the potential to be feasible and widely adopted by impoverished smallholder farmers. Improved rice varieties that contain desirable traits, such as high yielding and resistance to biotic/abiotic stresses, have a real impact on the livelihood of farmers. For example, adopters of NERICA rice varieties, which have been developed by crossing high-yielding *O. sativa* subsp. *japonica* rice varieties with the more stress-tolerant and the resistant *O. glaberrima* rice variety CG14 (Jones *et al.*, 1997a, b), saw their per-capita income more than double, from US\$25 to US\$58 (Arouna *et al.*, 2017).

Since both *S. hermonthica* and *S. asiatica* irreversibly damage the host shortly after the parasites successfully form xylem continuity with the host (while they are underground), effective *Striga* resistance should aim to prevent parasite attachment (pre-attachment attachment) or terminate the growth of parasites soon after attachment (post-attachment resistance) (Scholes & Press, 2008).

Most *Striga* resistance work in rice has been carried out on the resistance to *S. hermonthica*. Pre-attachment resistance is associated with low production of *Striga* germination stimulants, such as strigolactones, in host root exudates (Cardoso *et al.*, 2014; Jamil *et al.*, 2011a). Post-attachment resistance phenotypes are characterised by the inability of parasites to penetrate the endodermis of the host (Beardon, 2018; Cissoko *et al.*, 2011; Gurney *et al.*, 2006) or ineffective xylem–xylem connections between the host and the parasites after their formation (Cissoko *et al.*, 2011; Yoshida & Shirasu, 2009). (See Section 1.5 for more details.)

The expression of resistance (in incompatible interactions between the host and parasites) is a function of both the resistance genes of the host and the species and the genetic makeup of the parasites (Rodenburg *et al.*, 2017; Scholes & Press, 2008). From field and laboratory studies, rice varieties with resistance to *S. hermonthica* are not always resistant to *S. asiatica* and *vice versa* as demonstrated by Rodenburg *et al.* (2017, 2015). There is a need for knowledge on *S. asiatica*-specific post-attachment resistance in rice.

RIL populations have been used successfully in the discovery of QTL for both pre- and post-attachment resistance to *S. hermonthica* in rice and other cereal crops (examples in Section 1.6). A RIL population have an advantage for QTL mapping in that it is considered an immortal population, which allows genetically identical replicates of the genotypes to be shared and screened many times in different environments. Also, it has accumulated a large number of recombination breakpoints from repeated selfing, which helps refine the resolution of QTL detection (Ferne & Keurentjes, 2018; Mauricio, 2001). In addition, a RIL population, which segregates multiple QTL simultaneously, enables detection of possible

epistatic interactions between QTL (Fernie & Keurentjes, 2018). A rice nested-association mapping (NAM) population, comprising of ten RIL sub-populations that are derived through crossing of a common parental genotype (*O. sativa* subsp. *indica*) with ten diversity donors [*O. sativa* subsp. *japonica* (subgroup *tropical japonica*) and *O. sativa* subsp. *indica*], has been developed. Each of these RIL populations contains ~200 RILs that had been genotyped by sequencing (GBS), yielding dense genetic markers (Fragoso *et al.*, 2017). This rice NAM population (and its RIL sub-populations) provides an excellent resource for discovery of novel *S. asiatica* resistance genes as it contains a wide range of genetic diversity, and would allow a subsequent dissection of the genetic basis underlying the phenotype through QTL mapping.

2.1.1 Aims and objectives

The aims of this study were to (i) phenotype the parental genotypes of three RIL populations (part of the rice NAM population) to select a RIL population where the two parental genotypes exhibit different levels of resistance/susceptibility to *S. asiatica*, in order to map the genetic basis of the phenotype, and (ii) characterise the phenotype of resistance/susceptibility of the chosen parental genotypes at a macroscopic and microscopic level. The specific objectives were to:

- i) Phenotype the resistance/susceptibility of four rice parental genotypes by growing them in rhizotrons for quantification of resistance by measuring the biomass, cumulative length and number of *S. asiatica* individuals on host root systems.
- ii) Perform a detailed analysis of the growth of *S. asiatica* individuals on the two rice parental genotypes of the chosen RIL population by measuring changes in biomass, length and number of parasites over time.
- iii) Characterise the infection process of *S. asiatica* on the parental genotypes of the chosen RIL population to determine whether there was a difference in the speed at which the parasites penetrate the host root and form xylem connections with the host.

- iv) To determine the number and timing of the emergence of *S. asiatica* individuals on the two rice parental genotypes of the chosen RIL population, when grown in pots containing *S. asiatica* seeds.

2.2 Methodology

2.2.1 Plant materials

An accession of *S. asiatica* seeds was collected from a local population of *S. asiatica* individuals parasitising sorghum plants in farmers' fields in Ethiopia in 1993. The *S. asiatica* seeds have been bulked subsequently in Sheffield on a highly susceptible rice genotype, IAC165 [*O. sativa* subsp. *japonica* (subgroup *tropical japonica*)]. The seeds of IAC165 were sourced from the National Institute of Agrobiological Sciences (NIAS), Ibaraki, Japan.

Four parental genotypes of the rice NAM population—developed at the International Center for Tropical Agriculture (CIAT), Cali, Colombia (Fragoso *et al.*, 2017)—consisted of the common parent, IR64 (*O. sativa* subsp. *indica*), and three diversity donors: CT8556-37-2-3-1-M (*O. sativa* subsp. *indica*), and CT10037-56-6-M-M-1 and ITA164 [*O. sativa* subsp. *japonica* (subgroup *tropical japonica*)].

2.2.2 Growth and infection of rice genotypes with *Striga asiatica*

Rice seeds were germinated between two sheets of damp filter paper (grade 0858; GE Healthcare UK Ltd, Buckinghamshire, UK) placed between two blocks of horticultural-grade mineral wool (Rockwool; Grodan B.V., Roermond, the Netherlands) in propagators. The seeds were incubated at 30 °C for 2 d, after which the propagators were moved to a controlled environment growth chamber with a day/night temperature of 28/25 °C, relative humidity of 60 % and a 12 h photoperiod (photon flux density of 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at plant height). At 7 d after germination, seedlings of a similar size were transferred into rhizotrons, as described by Cissoko *et al.* (2011). A rhizotron is a root observation chamber, consisting of a 25 cm \times 25 cm \times 2 cm Perspex base, filled with Rockwool, onto which a 100 μm nylon

mesh (Plastok Associates Ltd, Wirral, UK) is placed. There are openings at the top and bottom of the rhizotron for shoot growth and drainage, respectively. Seedlings were placed on the mesh so that most of the roots grew down the mesh. The Perspex lid of the rhizotron was then placed onto the base allowing observation of root growth and access to the root system. The rhizotron was covered with aluminium foil to prevent the light from reaching the roots. Each rhizotron received 10 ml of 40 % Long Ashton solution containing 2 mol m^{-3} ammonium nitrate (Hewitt, 1966) four times a day. Eight days after being transferred into the rhizotrons, each seedling was inoculated with $\sim 10 \text{ mg}$ of conditioned and germinated *S. asiatica* seeds using paintbrushes (Cissoko *et al.*, 2011). The *S. asiatica* seeds had been surface-sterilised in 10 % (v/v) commercial bleach (CLE0300; Scientific Laboratory Supplies Ltd, Nottingham, UK). Approximately 40 mg of *S. asiatica* seeds were then conditioned on glass-fibre filter paper (GF/A grade; GE Healthcare UK Ltd, Buckinghamshire, UK) in a sealed 9 cm Petri dish (SLS2002; Scientific Laboratory Supplies Ltd, Nottingham, UK) at $30 \text{ }^{\circ}\text{C}$ for 13 d. To ensure synchronous attachments, each Petri dish of *S. asiatica* seeds received 3 ml of $0.1 \text{ } \mu\text{g L}^{-1}$ aqueous solution of GR24, an artificial germination stimulant, 16 h before the inoculation of *S. asiatica* seeds.

2.2.3 Quantification of resistance/susceptibility of the four parental genotypes of three RIL populations

Five replicate plants of each NAM parental genotypes were inoculated with *S. asiatica*. At 28 d after inoculation (DAI), a high-resolution [1200 dpi (dot per inch)] scan of the whole root system of every rice plant was taken using a Canon CanoScan 9000F Mark II scanner. *Striga asiatica* plants were then removed from each root system and placed in Petri dishes. Each Petri dish was placed on a fluorescent lightbox and the *S. asiatica* individuals were photographed, using a Sony ILCE-6000 digital camera with a Sony SEL30M35 lens. The number, cumulative length and mean length of *S. asiatica* individuals per host were then quantified from the photographs using ImageJ version 1.51j8 (Rasband, 2016). When quantifying the cumulative length, mean length and number of parasites per root system,

only parasites larger than 2.5 mm (i.e. successful attachments) were counted. The parasites were then dried at 50 °C for at least 3 d to obtain the biomass (dry weight) per host plant. The overall differences in the biomass, cumulative length, mean length and number of parasites between the four rice genotypes were statistically tested using one-way analysis of variance (ANOVA), followed by a post hoc Tukey's honest significance test, in the R software environment version 3.6.3 (R Core Team, 2017).

2.2.4 Characterisation of the development of *Striga asiatica* on IR64 and CT8556-37-2-3-1-M during the early phase of the life cycle

In order to determine whether there was a difference in the rate at which *S. asiatica* penetrated the host root and formed parasite–host xylem connections, 20 rhizotrons of IR64 and CT8556-37-2-3-1-M plants were established and inoculated with pre-germinated *S. asiatica* seeds. *Striga asiatica* attachments from five replicates of each rice genotype were collected at 2, 3, 5 and 8 DAI. For each rice replicate, the host root system with *S. asiatica* attachments were cut into ~ 5 cm pieces and placed in 71.4 % (m/m) aqueous solution of chloral hydrate (Sigma-Aldric Co, St Louis, Missouri, US) for at least 14 d to clear the root system. The root segments were then stained for the presence of lignin in a mixture of phloroglucinol–HCl at 2:1 ratio of 3 % (m/v) solution of phloroglucinol (Sigma Chemical Co, St. Louis, Missouri, US) dissolved in absolute ethanol solution (Fisher Scientific UK Ltd, Loughborough, UK) to 32 % hydrochloric acid (VWR International SAS, Fontenay-sous-Bois, France), for at least 2 min. While the root segments were in the staining agent, the attachments were photographed using a Leica M165 FC stereomicroscope. For each rice replicate, ~ 120 *S. asiatica* individuals were classified into different developmental stages.

Six developmental stages of attached parasites were defined. Before the parasite attachment, *S. asiatica* seeds were triggered to germinate with the artificial germination stimulant GR24. A radicle emerged from the seed coat and grew towards the host root (Figure 2.1A).

DEVELOPMENTAL STAGE 1—Haustorial attachment to the host root (Figure 2.1Bi–ii): When the *S. asiatica* radicle reached the host root, it differentiated in response to HIFs, to form the attachment and penetration organ, the haustorium, which was covered with numerous haustorial hairs that attached it to the epidermis of the host aiding the penetration of the parasite (Figure 2.1Bi). Figure 2.1Bii shows an example of *S. asiatica* individual with an elongated haustorium.

DEVELOPMENTAL STAGE 2—Penetration through the host root cortex (Figure 2.1Ci–ii): Once a haustorium had attached to the epidermis of the host, parasite intrusive cells began to penetrate through the host root cortex. As the parasite intrusive cells penetrated through the host root cortex, lignin was deposited around the site of attachment, which is seen as a ring of pink cells in Figure 2.1Cii. Development Stage 2 covers the period from parasite penetration of the host root cortex to just prior to the formation of host–parasite xylem–xylem connections.

DEVELOPMENTAL STAGE 3—Breaching the endodermis of the host and formation of xylem–xylem continuity with the host (Figure 2.1Di–ii): The parasite entered Stage 3 of development upon breaking the endodermis of the host and the establishment of xylem–xylem connections between host and parasites. Pink staining shows lignin associated with the host–parasite vascular connection(s) as well as the parasite xylem vessels in the parasite shoot (Figure 2.1Di). Figure 2.1Dii shows a parasite with its cotyledons beginning to expand inside the seed coat.

DEVELOPMENTAL STAGE 4—Emergence of the cotyledons from the seed coat: Figure 2.1Ei–ii show the emerged cotyledons. The parasite vascular system (stained pink) was visible throughout the parasite shoot and cotyledons (Figure 2.1Ei) and at the xylem–xylem bridge (Figure 2.1Eii).

DEVELOPMENTAL STAGE 5—Emergence of the first pair of leaves: The first pair of leaves have distinct midveins (Figure 2.1F).

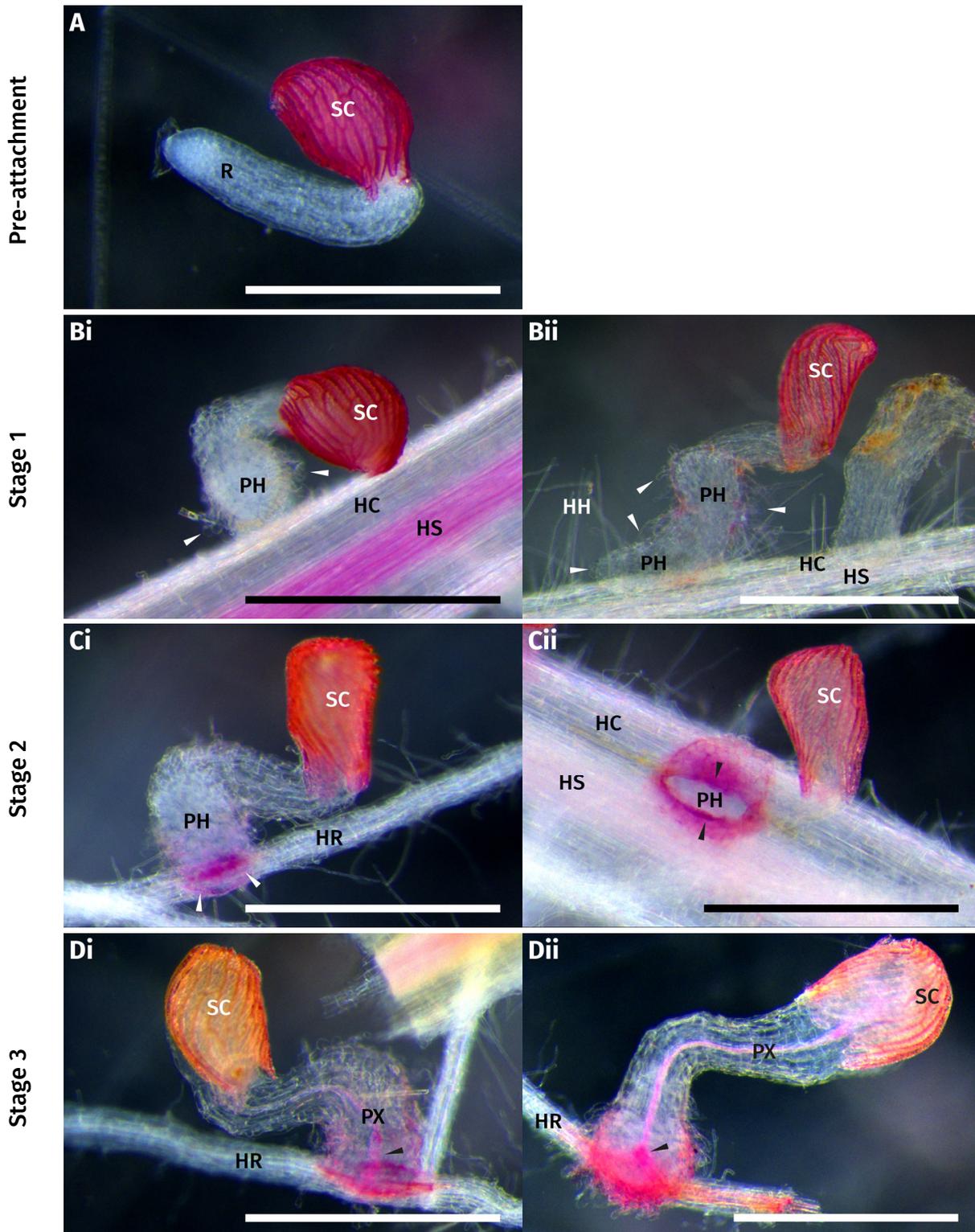


Figure 2.1 Developmental stages of *Striga asiatica* individuals between 2 and 8 d after inoculation. *Striga asiatica* seed and attachments were destained and had lignin deposition stained pink with chloral hydrate and hydrochloric acid. **A**, Germinated *S. asiatica* seed before attachment. **Bi–ii**, Stage 1 parasites—Haustorial attachment to the host root—with arrowheads pointing at haustorial hairs. **Ci–ii**, Stage 2 parasites—Penetration through the cortex—with arrowheads indicating lignin depositions at the sites of attachment. **Di–ii**, Stage 3 parasites—Breaching the endodermis of the host and formation of xylem-xylem continuity with the host—with arrowheads pointing at the xylem-xylem bridge(s). **Ei–ii**, Stage 4 parasites—Emergence of the cotyledons from the seed coat. **F**, Stage 5

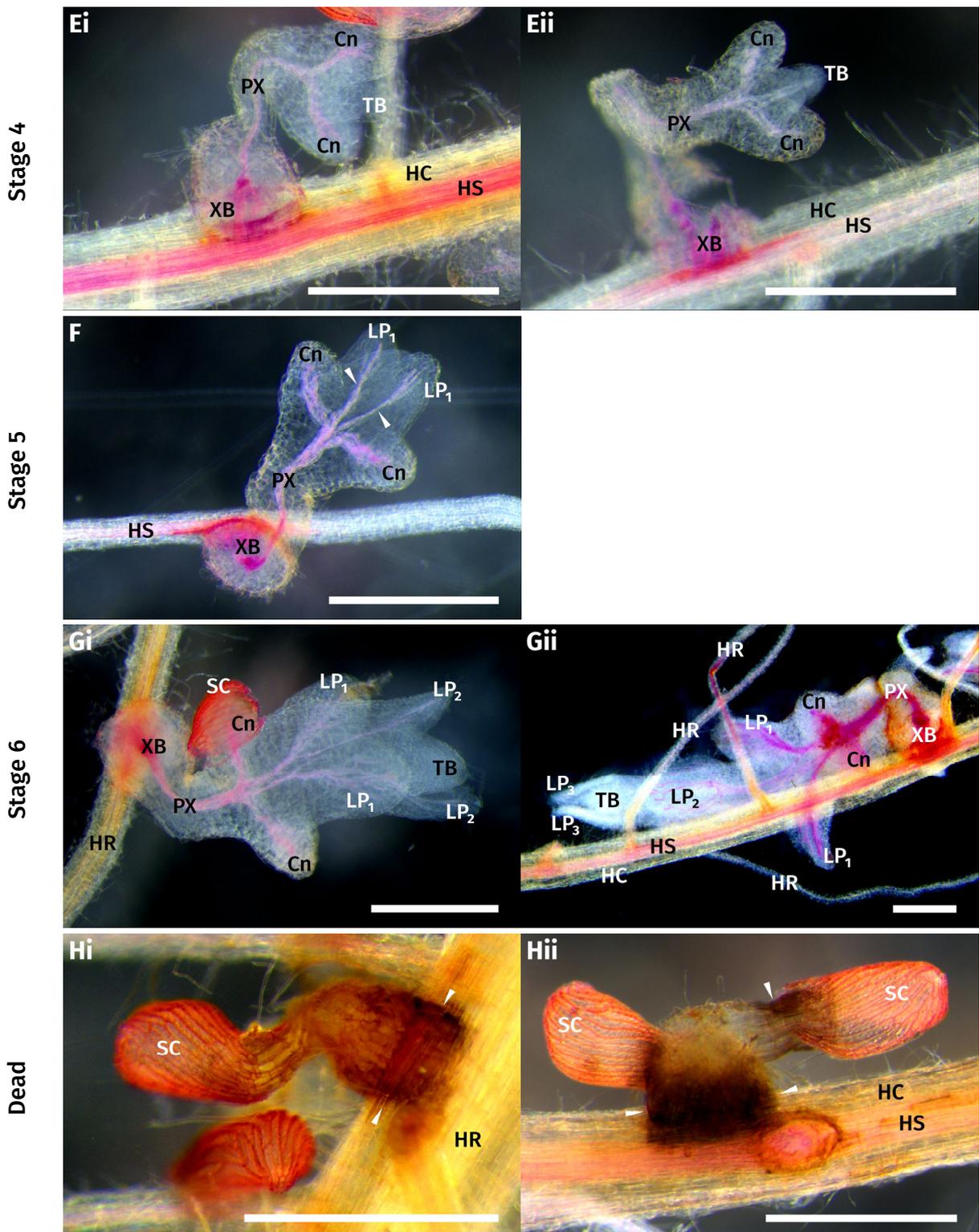


Figure 2.1 (continued)

parasite—Emergence of the first pair of leaves—with arrowheads highlighting the midveins of the first leaf pair. **Gi–ii**, Stage 6 parasites—Emergence of the second pair of leaves. **Hi–ii**, Dead parasites with arrowheads pointing at necrotic tissues at the sites of attachment and on parasites. SC, seed coat; R, parasite radicle; PH, parasite haustorium; HC, host root cortex; HS, host stele; HH, host root hair; HR, host root; PX, parasite xylem; XB, xylem-xylem bridge; Cn, cotyledon; TB, parasite terminal bud; LP_{*i*}, *i*th leaf pair. Scale bars = 0.5 mm.

DEVELOPMENTAL STAGE 6 — Emergence of the second pair of leaves (Figure 2.1Gi–ii).

Finally, some parasites died shortly after attaching to the host due to a strong host resistance response (Figure 2.1Hi–ii). The host resistance response was characterised by brown (Figure 2.1Hi) or black (Figure 2.1Hii) necrotic tissues at the site of attachment of the parasite. The parasite tissue also exhibited extensive necrosis (Figure 2.1Hii).

The proportion of the parasites at each developmental stage on a host at each time point was calculated over the total number of attached parasites at each time point. The statistical difference between the proportions of the parasites at each developmental stage at each time point from five replicates of the two rice genotypes was determined by Mann–Whitney’s U test as implemented in R software environment version 3.6.3 (R Core Team, 2017).

2.2.5 Characterisation of changes in biomass, length and number of *Striga asiatica* plants from 12 to 42 d after inoculation

Thirty replicates of IR64 and CT8556-37-2-3-1-M plants were set up in rhizotrons and inoculated with pre-germinated *S. asiatica* seeds (Section 2.2.2). The root systems of five replicates of each rice genotype were harvested at 12, 16, 21, 28, 35 and 42 DAI. The root systems were scanned, and parasites larger than ~ 0.5 mm were harvested and processed as described in Section 2.2.3. When quantifying the mean length and number of parasites per root system only parasites larger than 2.5 mm (i.e. successful attachments) were counted at 21, 28 and 35 DAI and 5.0 mm at 42 DAI. The statistical differences between the biomass, mean length and number of parasites from the two rice genotypes at the different time points were tested using two-way ANOVA. The average relative growth rates (RGR) of parasites for each rice genotype were calculated from the parasite biomass means, in Equation 2.1.

$$RGR_{t_{i+1}} = \frac{\ln m_{i+1} - \ln m_i}{t_{i+1} - t_i} \quad (2.1)$$

where t_i is the time of the first biomass measurement m_i , t_{i+1} is the time of the second biomass measurement m_{i+1} and $RGR_{t_{i+1}}$ is the average RGR between t_i and t_{i+1} .

The leaf pairs (nodes) of parasites at 16 and 28 DAI were counted on the harvested parasites from three replicates of IR64 and CT8556-37-2-3-1-M. The difference between the numbers of leaf pairs at each time point was assessed using Mann–Whitney’s *U* test with Holm–Bonferroni method to adjust *P*-values for multiple comparisons. The statistical tests were carried out in the R software environment version 3.6.3 (R Core Team, 2017).

2.2.6 Characterisation of the emergence of *Striga asiatica* plants growing on IR64 and CT8556-37-2-3-1-M in pots

To determine whether there was a difference in the number and timing of emergence of the parasites growing on the two rice genotypes, 20 pots (10 cm × 10 cm × 19 cm) containing *S. asiatica* seeds were established. The base of each pot was lined with 2 cm–thick Rockwool. After the pots were filled with washed river sand up to 7 cm from the top, 40 mg *S. asiatica* seeds were mixed with ~ 15 g of sand and evenly spread in the centre of the pot, leaving ~ 1 cm gap from each side. The pots with *S. asiatica* seeds were then filled with more sand, up to 2 cm from the top, and watered with tap water to saturation. The pots were placed in a growth chamber under the same environmental conditions as described in Section 2.2.2. Pots were watered to keep the sand moist when needed. Ten days later, a rice seed, which had been germinated on damp filter paper (grade 0858; GE Healthcare UK Ltd, Buckinghamshire, UK) 2 d prior, was planted in each pot at the depth of 1 cm. Five, eight and seven pots were planted with IAC165, IR64 and CT8556-37-2-3-1-M seeds, respectively. The pots were watered thoroughly after planting and then left for a week. At 7 d after planting (DAP), each pot received 100 ml of tap water to keep the top layer of sand moist. From 11 DAP onwards, each pot received 15 ml of 40 % Long Ashton solution containing 2 mol m⁻³ ammonium nitrate (Hewitt, 1966) four times a day.

The number of emerged parasites in each pot was recorded daily from 35 to 84 DAP. At 84 DAP, representative pots of each rice genotype were photographed. The above-ground parts of *S. asiatica* individuals were then harvested to obtain the fresh weight per host plant.

Three root systems of each rice genotype were carefully washed and photographed. Samples of underground *S. asiatica* plants on the washed root systems were placed in a Petri dish. Each Petri dish was placed on a fluorescence lightbox and photographed. All photography was conducted using the Sony ILCE-6000 digital camera with a Sony SELP18105G lens. The area under *S. asiatica* number progress curve (ASNPC; Haussmann *et al.*, 2000) for each pot were calculated using function `audpc` in package `agricolae` version 1.3-3. The differences in the ASNPC values and the above-ground fresh weights between the three rice genotypes were statistically tested using Mann–Whitney’s *U* test with Holm–Bonferroni method to adjust *P*-values for multiple comparisons. All calculations and statistical tests were carried out in the R software environment version 3.6.3 (R Core Team, 2017)

2.3 Results

2.3.1 The four parental genotypes of three rice RIL populations exhibited different levels of susceptibility to *Striga asiatica*.

In order to select a RIL population from an existing rice NAM population, the common parent (IR64) and three diversity donors (CT8556-37-2-3-1-M, CT10037-56-6-M-M-1 and ITA164) were infected with an accession of *S. asiatica* from Ethiopia to determine their levels of resistance/susceptibility by quantifying the biomass (Figure 2.2A), the cumulative length (Figure 2.2B), the number (Figure 2.2C) and the mean length of parasites (Figure 2.2D) per host root system at 28 DAI. The four rice genotypes exhibited significantly different levels of resistance/susceptibility to this accession of *S. asiatica* in all four measurements (one-way ANOVA — parasite biomass: $F = 10.65$, $df = 3, 16$, $P < 0.001$; parasite cumulative length: $F = 15.04$, $df = 3, 16$, $P < 0.001$; number of parasites: $F = 7.80$, $df = 3, 16$, $P < 0.01$; parasite mean length: $F = 9.89$, $df = 3, 16$, $P < 0.001$). When each diversity donor was compared to IR64, only CT8556-37-2-3-1-M was significantly less susceptible than IR64 in all four measurements (Tukey’s honest significance test: $P < 0.05$) (Figure 2.2A–D),

whilst CT10037-56-6-M-M-1 and ITA164 were as susceptible to *S. asiatica* as IR64 in all four measurements (Tukey's honest significance test: $P \geq 0.05$) (Figure 2.2A–D).

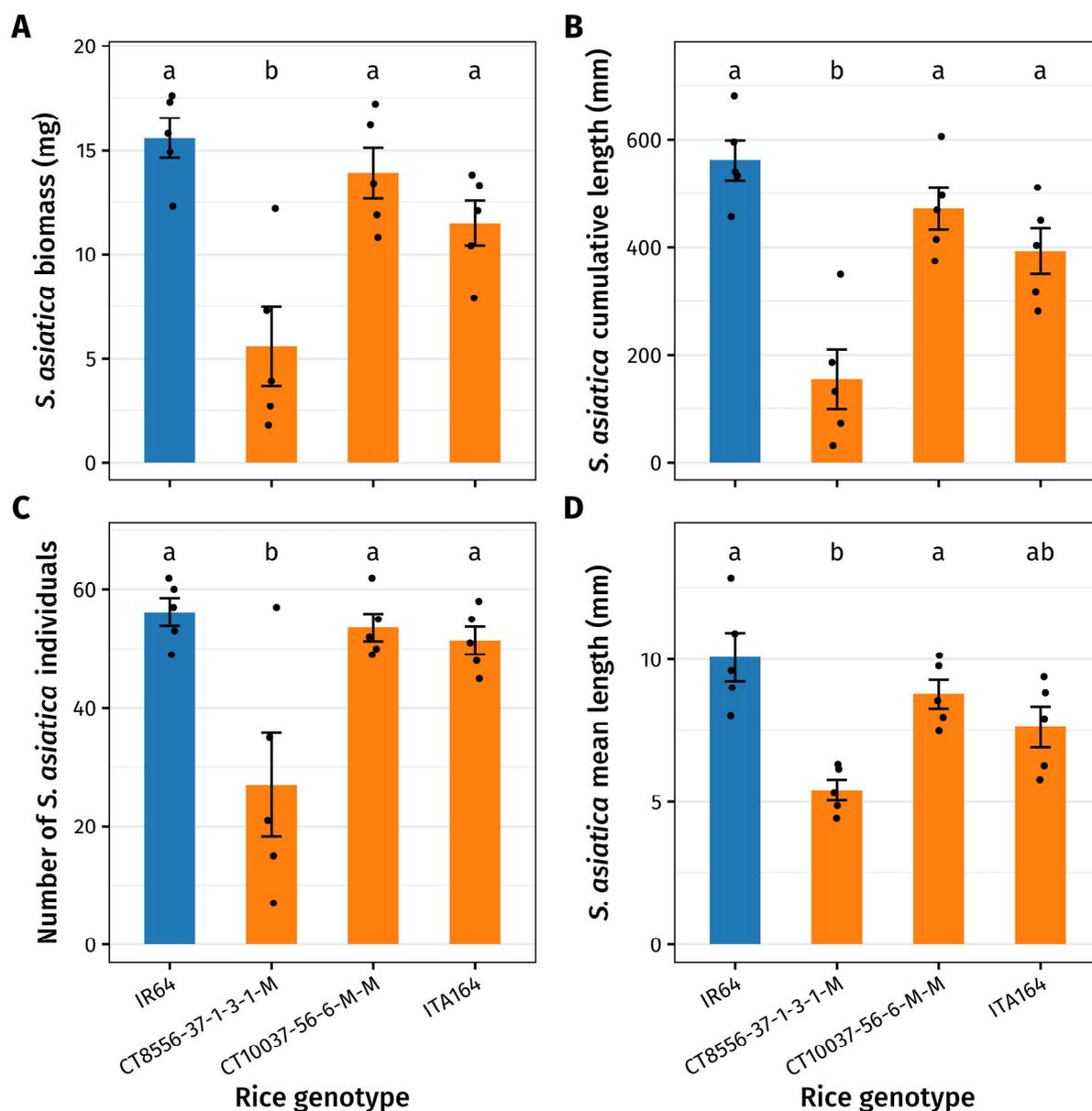


Figure 2.2 Susceptibility/resistance of the common parental line (IR64) and three selected diversity donors (CT8556-37-2-3-1-M, CT10037-56-6-M-M-1 and ITA164) of a nested association mapping population. The biomass (dry weight) (A), the cumulative length (B), the number (C) and the mean length (D) of *Striga asiatica* individuals per rice host at 28 d after inoculation of germinated seeds of an accession of *S. asiatica* from Ethiopia. Each bar shows the mean with the standard error of the mean ($n = 5$). There was a significant effect of rice genotype on the *S. asiatica* biomass (one-way ANOVA: $P < 0.001$), cumulative length (one-way ANOVA: $P < 0.001$), number (one-way ANOVA: $P < 0.01$) and mean length of parasites (one-way ANOVA: $P < 0.001$). Bars that share the same letter are not significantly different (Tukey's honest significance test: $P \geq 0.05$)

On average, each CT8556-37-2-3-1-M host plant supported less than half the number of parasites on each IR64 plant, with 35.8 % of the parasite biomass (Figure 2.2A & C). The mean length of an individual parasite from CT8556-37-2-3-1-M was 53.7 % of that of parasites on IR64 (Figure 2.2D). The cumulative length of parasites from CT8556-37-2-3-1-M was only 27.5 % of the cumulative length of parasites from IR64 (Figure 2.2C). Thus, the phenotypes of CT8556-37-2-3-1-M and IR64 were characterised in detail to examine possible mechanisms that underlie the differences in their susceptibility to *S. asiatica*.

2.3.2 Why are *Striga asiatica* plants smaller on CT8556-37-2-3-1-M compared to IR64?

Two hypotheses to explain why *S. asiatica* individuals on CT8556-37-2-3-1-M were smaller than those on IR64 were investigated. Firstly, the smaller parasites on CT8556-37-2-3-1-M may be due to slower penetration of the *S. asiatica* endophyte through the host root cortex, endodermis and the formation of xylem connections with the host (Hypothesis 1), or alternatively slower growth of parasites after they had formed vascular connections (Hypothesis 2).

To address Hypothesis 1, the infection process of parasites on IR64 and CT8556-37-2-3-1-M was compared from 2 to 8 DAI. This was done by counting the number of parasites at each developmental stage (Section 2.2.4) throughout the early stages of the parasite life cycle (at 2, 3, 5 and 8 DAI), and then comparing the proportion of parasites at each developmental stage at each time point. There was no significant difference between the proportions of parasites at different developmental stages on CT8556-37-2-3-1-M and IR64, at any time points over the course of the experiment (Mann–Whitney’s *U* test: $P > 0.05$) (Figure 2.3). The results of the statistical tests are shown in Table 2.1.

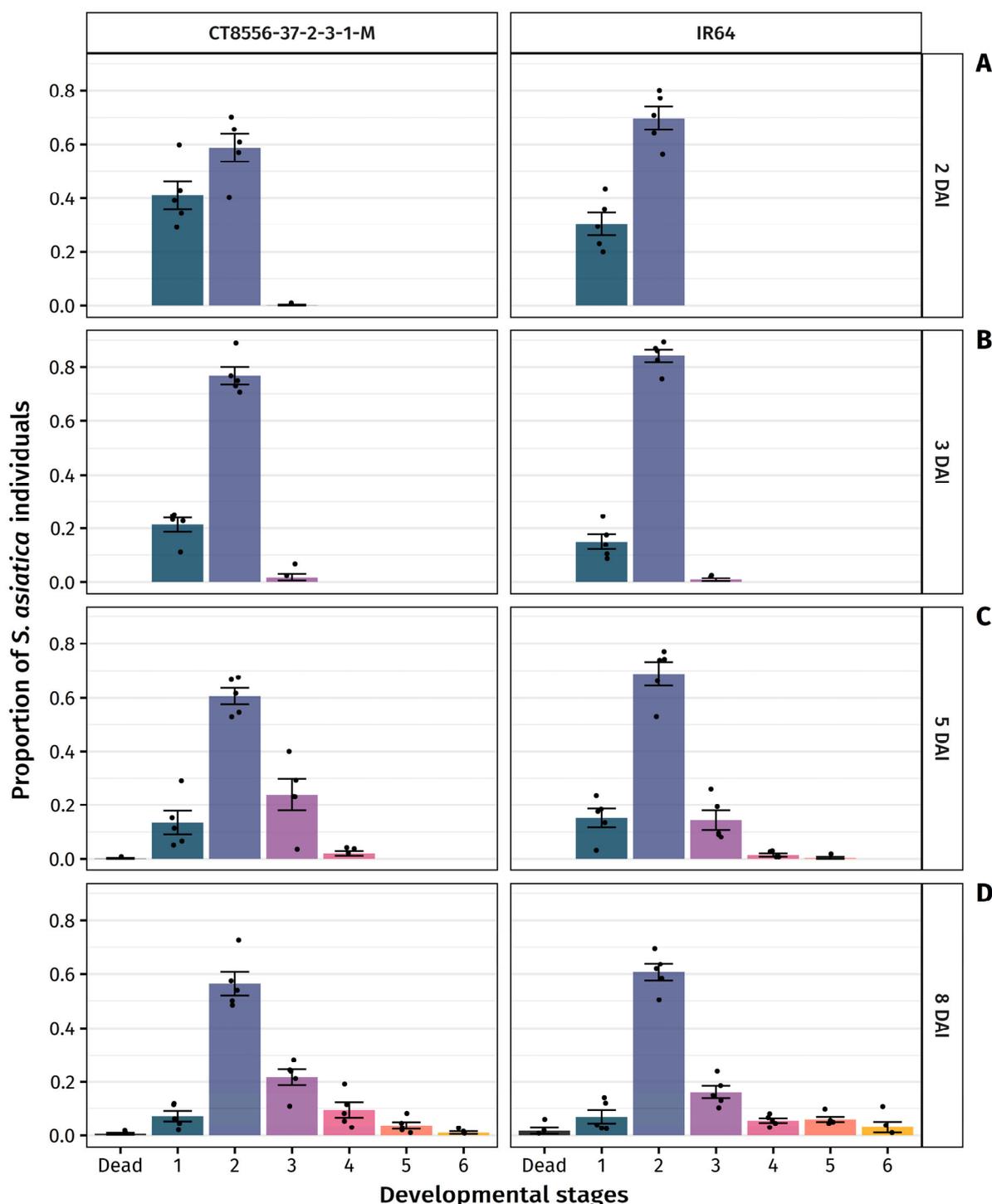


Figure 2.3 Distributions of *Striga asiatica* individuals in different developmental stages between 2 and 8 d after inoculation (DAI). At each time point (A, 2 DAI; B, 3 DAI; C, 5 DAI; D, 8 DAI), a sample of *S. asiatica* individuals ($\bar{N} = 117$; $86 \leq N \leq 153$) from a different replicate of rice host were categorised into: Stage 1, parasites having a haustorium; Stage 2, parasite's haustorium penetrating host root cortex; Stage 3, host-parasite xylem bridge(s) formed; Stage 4, cotyledons emerged from the seed coat; Stage 5, first pair of leaves emerged, with midveins; Stage 6, parasites with more than one pairs of leaves; Dead parasites with necrosis. The proportions of parasites in different developmental stages within a replicate (black dots) were used to calculate the mean and standard error of the mean ($n = 5$), represented by bars and error bars. Bars of the same colour correspond to parasites in the same stage of development.

Table 2.1 Results of the Mann–Witney’s *U* tests on the differences between the proportions of parasites at different developmental stages on the two rice genotypes, CT8556-37-2-3-1-M and IR64, at 2, 3, 5 and 8 d after inoculation (DAI).

DAI	Developmental stages	<i>U</i> statistics	<i>P</i> -value
2	1	18	0.310
	2	6	0.220
3	1	20	0.151
	2	5	0.151
	3	13	0.906
5	1	10	0.690
	2	6	0.222
	3	18	0.310
	4	14	0.833
8	1	12	> 0.999
	2	7	0.310
	3	19	0.222
	4	17	0.421
	5	5	0.151
	6	10	0.666
	Dead	9	0.504

Five replicate hosts for each rice genotype.

2.3.3 *Striga asiatica* plants grew more slowly on CT8556-37-2-3-1-M than on IR64 from 8 d after inoculation.

The previous section established that the lower biomass, length and number of *S. asiatica* individuals on CT8556-37-2-3-1-M was not caused by a delay in the formation of xylem–xylem connections between the parasites and the host. The difference in the time the parasites took to develop while growing on the two rice genotypes appeared after the formation of vascular connections. To examine whether this difference translated into the differences in the biomass, the cumulative length and the average size of *S. asiatica* individuals seen in Figure 2.2, a detailed analysis of the growth of parasites on CT8556-37-2-3-1-M and IR64, from 12 to 42 DAI, was carried out.

As the biomass of *S. asiatica* increased throughout the time course, the parasite biomass increased significantly faster on IR64 compared to parasites on CT8556-37-2-3-1-M (two-way ANOVA: $F_{\text{Rice genotypes} \times \text{Time}} = 10.04$, $df = 5, 47$, $P < 0.001$) (Figure 2.4A). Overall, the

biomass of parasites was significantly greater on IR64 compared to CT8556-37-2-3-1-M (two-way ANOVA: $F_{\text{Rice genotypes}} = 55.59$, $df = 1, 47$, $P < 0.001$) (Figure 2.4A). As expected, the relative growth rate (RGR)—the relative rates of biomass accumulation—of parasites showed that the parasites from both rice genotypes grew more slowly with time, as they became larger (Figure 2.4B). The RGR of the parasites on CT8556-37-2-3-1-M was lower than the RGR of those on IR64 (Figure 2.4B).

As *S. asiatica* individuals grew longer over the time course, the mean length of parasites increased significantly faster on IR64 than on CT8556-37-2-3-1-M (two-way ANOVA: $F_{\text{Rice genotypes} \times \text{Time}} = 37.71$, $df = 5, 47$, $P < 0.001$) (Figure 2.4C). On average, parasites growing on CT8556-37-2-3-1-M were significantly shorter than those growing on IR64 (two-way ANOVA: $F_{\text{Rice genotypes}} = 208.58$, $df = 1, 47$, $P < 0.001$) (Figure 2.4C).

When the distributions of the lengths of parasite individuals were considered, greater proportions of *S. asiatica* individuals growing on CT8556-37-2-3-1-M were in smaller size classes, compared to IR64. At 12 DAI, most parasites on CT8556-37-2-3-1-M were 1–2.5 mm long, compared to 2.5–5 mm in length on IR64 (Figure 2.4D & 2.5). More than 20 % of parasites on CT8556-37-2-3-1-M were shorter than 1 mm, compared to less than 5 % of those on IR64 (Figure 2.4D). At 16 DAI, 79.1 % and 64.8 % of parasites on CT8556-37-2-3-1-M and IR64, respectively, were shorter than 5 mm (Figure 2.4D & 2.5). The proportion of parasites in the smallest size class (< 1 mm) almost diminished from both rice genotypes (Figure 2.4D). At 21 and 28 DAI, most of the parasites on CT8556-37-2-3-1-M were still shorter than 5 mm, whilst those on IR64 were mostly longer than 5 mm and 10 mm, respectively (Figure 2.4D & 2.5). By 35 DAI, most parasites on CT8556-37-2-3-1-M were longer than 10 mm, whilst most from IR64 were longer than 25 mm (Figure 2.4D & 2.5). At the end of the experiment (42 DAI), 44.6 % of parasites on IR64 were longer than 50 mm, whilst only 8.8 % of those on CT8556-37-2-3-1-M were in these size classes (Figure 2.4D & 2.5). Most of the parasites on CT8556-37-2-3-1-M were < 25 mm (Figure 2.4D & 2.5).

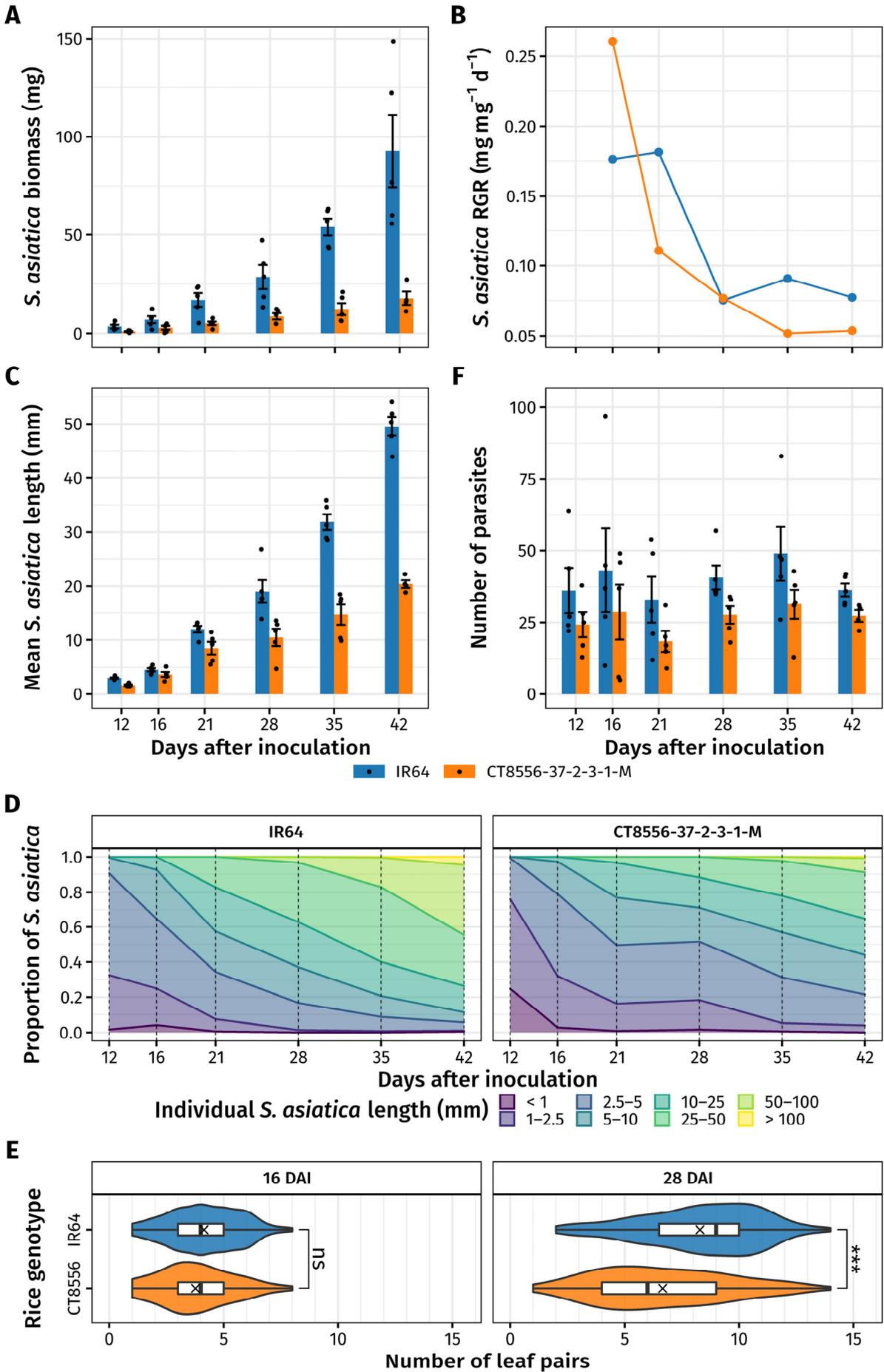


Figure 2.4 Growth of *Striga asiatica* individuals between 12 and 42 d after inoculation (DAI). Measurements of the biomass, the mean length and the number of parasites were taken on *S. asiatica* individuals at 12, 16, 21, 28, 35 and 42 DAI. **A**, Biomass (dry weight) of parasites per host plant. **B**, Relative rate of biomass accumulation (relative growth rate, RGR) was calculated from the means of parasite biomass per host plant. **C**, Mean length of parasites. **D**, Distribution of the length of parasites. The proportions of parasites were calculated from the number of parasites across five rice host replicates ($\bar{N} = 175$; $117 \leq N \leq 252$). **E**, Numbers of leaf pairs (nodes) of each parasite at 16 and 28 DAI are represented in violin plots with box plots. Crosses indicate the mean. Significance levels of differences were determined by Mann–Whitney’s *U* test with Holm–Bonferroni method: ns, non-significance; ***, adjusted $P < 0.001$. **F**, Number of parasites per host plant. Bar and error bars (**A**, **C** & **F**) represent the mean and the standard error of the mean ($n = 4$ to 5), respectively. Bars, lines, dots and shaded areas in **A–C** and **E–F** are coloured according to the rice hosts: IR64 in blue and CT8556-37-2-3-1-M in orange.

In summary, Figure 2.4D shows a much slower progression of parasites through size classes on CT8556-37-2-3-1-M compared to parasites on IR64, which is reflected by the more gentle gradients of the lines that separate size classes of parasites on CT8556-37-2-3-1-M compared to those on IR64.

To examine whether the size difference of parasites growing on the two rice genotypes was a result of differential stem elongation or differences in the developments of internodes, the numbers of leaf pairs of parasites were counted at 16 and 28 DAI (Figure 2.4E). At 16 DAI, the numbers of leaf pairs of parasites on CT8556-37-2-3-1-M and IR64 were not significantly different (Mann–Whitney’s test: $U = 4283$, $n_{\text{IR64}} = 134$, $n_{\text{CT8556}} = 56$, adjusted $P = 0.119$) (Figure 2.4E), suggesting the length difference was due to the internode growth. However, at 28 DAI parasites had significantly fewer pairs of leaves on CT8556-37-2-3-1-M than on IR64 (the medians were 6 and 9 pairs of leaves, respectively) (Mann–Whitney’s test: $U = 10\,861$, $n_{\text{IR64}} = 139$, $n_{\text{CT8556}} = 117$, adjusted $P < 0.001$) (Figure 2.4E), showing that CT8556-37-2-3-1-M affected both the growth and development of parasites at the later time point.

Finally, the number of parasites that successfully infected the host did not significantly change over the time course (two-way ANOVA: $F_{\text{Time}} = 0.95$, $df = 5, 47$, $P = 0.455$) (Figure 2.4F). There were significantly fewer parasites growing on CT8556-37-2-3-1-M than on IR64 (two-way ANOVA: $F_{\text{Rice genotype}} = 4.66$, $df = 1, 47$, $P = 0.036$).

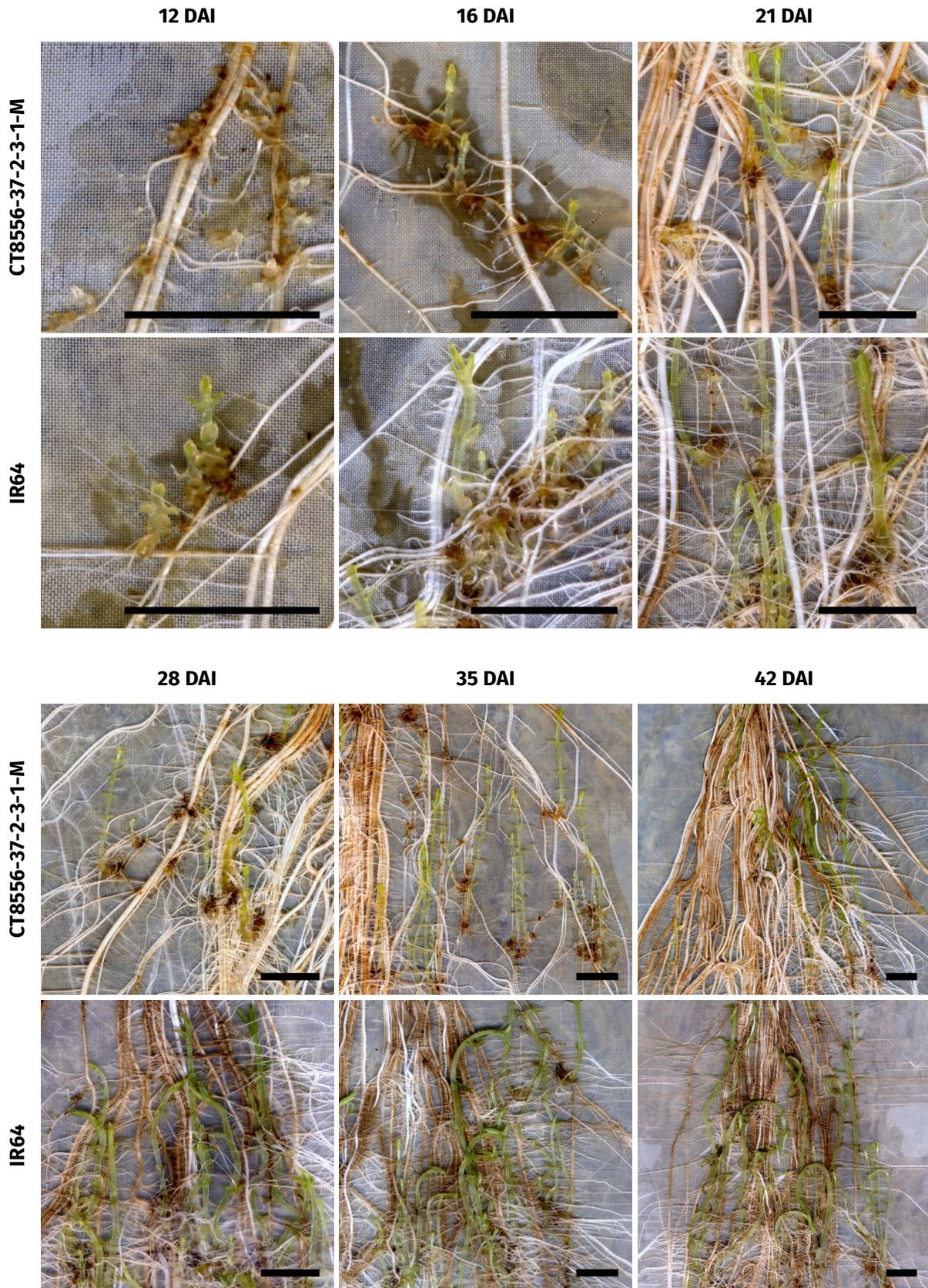


Figure 2.5 Images of *Striga asiatica* plants on the root systems of CT8556-37-2-3-1-M and IR64 from 12 to 42 d after inoculation (DAI). The different rice hosts grown in rhizotron systems were scanned at 12, 16, 21, 28, 35 and 42 DAI before the parasites were removed for counting and measurements of their length and biomass. Scale bars = 10 mm.

2.3.4 Fewer *Striga asiatica* plants emerged more slowly when growing on CT8556-37-2-3-1-M compared to IR64.

The previous section showed that the difference in the time the parasites took to develop after the formation of vascular connections on the two rice genotypes gave rise to the differences in the biomass and length of *S. asiatica* individuals in rhizotron systems. To examine whether the differences in the growth and development of parasites in rhizotron systems translated into differences in the emergence of parasites above ground, a pot-based experiment was set up to compare the number, biomass and time of emergence of parasites on CT8556-37-2-3-1-M, IR64 and, as a comparison, the very susceptible rice genotype IAC165.

Striga asiatica individuals first emerged on IAC165 at 36 DAP, followed by IR64 at 47 DAP and CT8556-37-2-3-1-M at 58 DAP (Figure 2.6A). By the end of the experiment (84 DAP), on average IAC165 pots contained six *S. asiatica* plants, whilst IR64 pots and CT8556-37-2-3-1-M pots contained four and one parasites, respectively (Figure 2.6A; Figure 2.7Ai–iii,

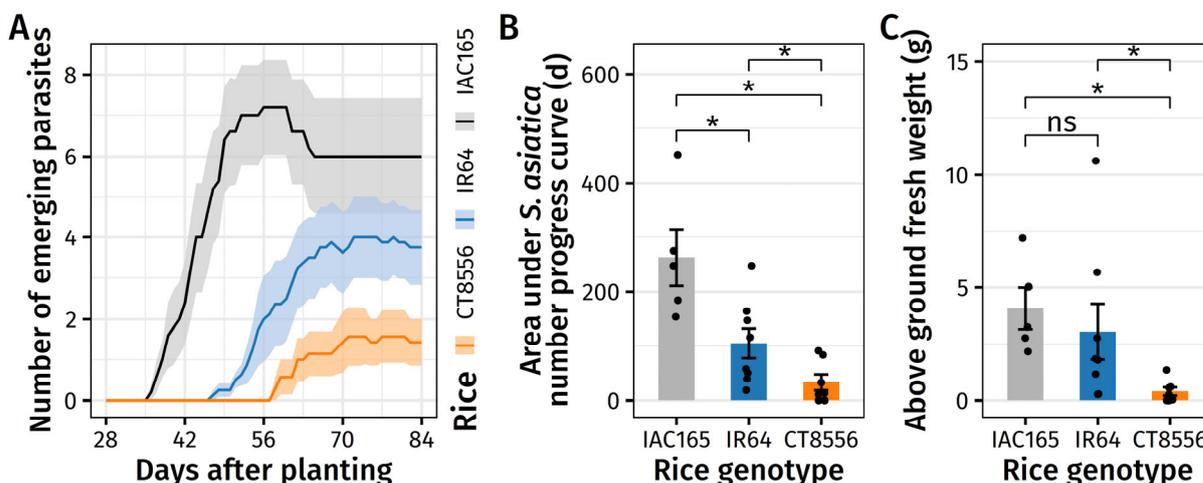


Figure 2.6 Above ground growth of *Striga asiatica* plants on IAC165, IR64 and CT8556-37-2-3-1-M rice genotypes in pots. **A**, Emergence of *S. asiatica* plants is shown in *S. asiatica* number progress curve. The lines and shaded areas represent the mean and the standard error of the mean of the number of parasites that have emerged above ground per rice host (in each pot) on each day (IAC165: $n = 5$; IR64: $n = 8$; CT8556-37-2-3-1-M: $n = 7$). **B**, Area under the *S. asiatica* number progress curve for each rice host. **C**, Cumulative fresh weight of the above-ground part of *S. asiatica* plants per rice host at 84 days after planting. Colours in each plot correspond to the rice genotypes: black/grey, IAC165; blue, IR64; orange, CT8556-37-2-3-1-M. Significance levels of the differences in **B** and **C** were evaluated using Mann–Whitney’s U test with Holm–Bonferroni method for multiple comparisons: ns, non-significance; *, adjusted $P < 0.05$.

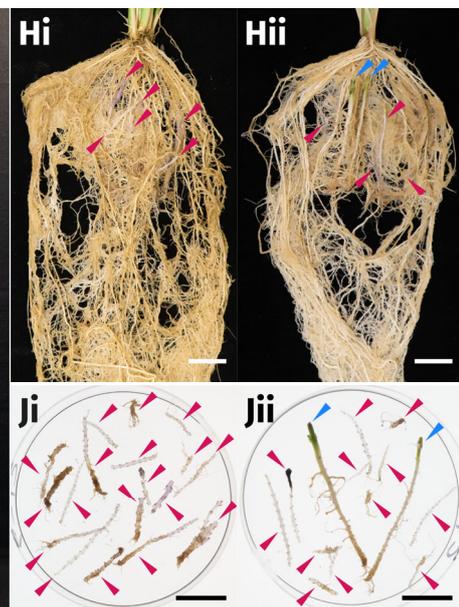
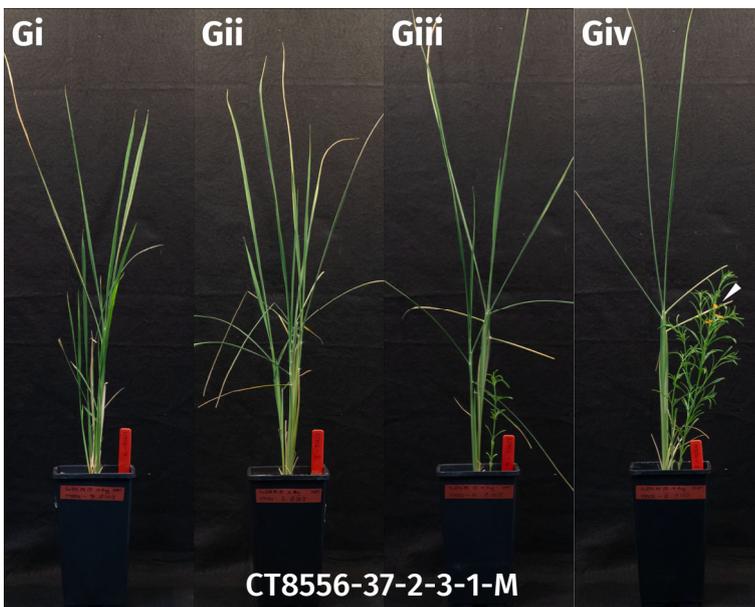
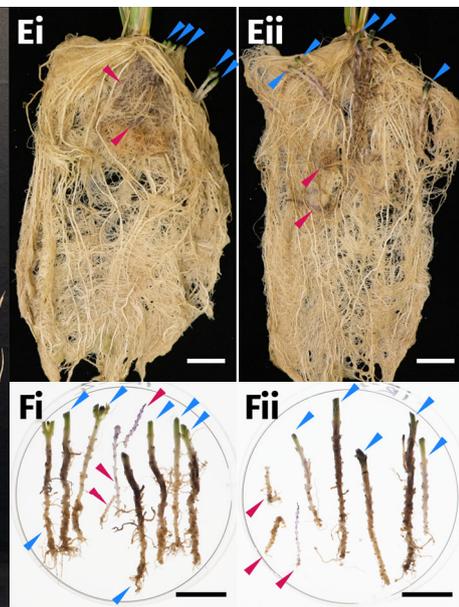
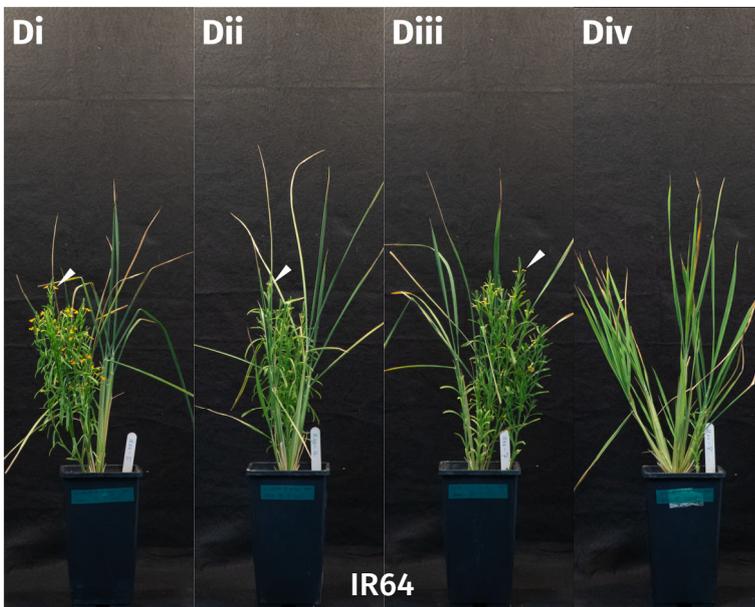
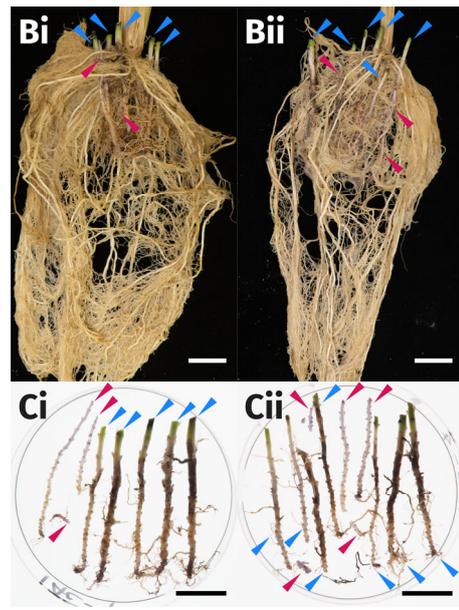
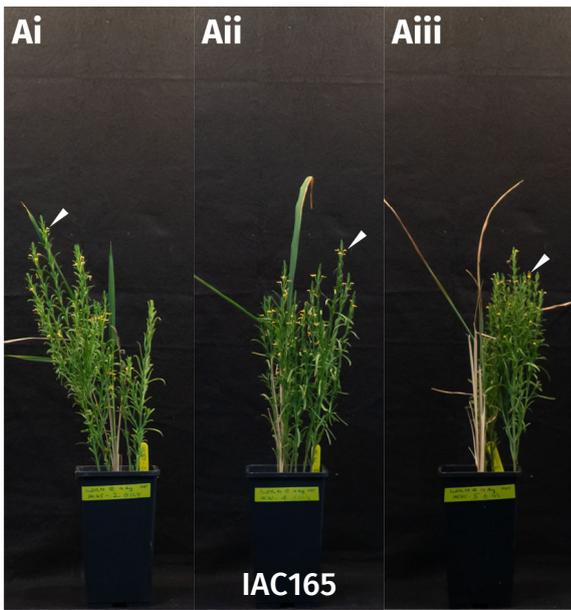


Figure 2.7 *Striga asiatica* plants parasitising on IAC165, IR64 and CT8556-37-2-3-1-M rice genotypes in pots. Ai-iii, Di-iv & Gi-iv, Representative pots of *S. asiatica* plants growing on IAC165, IR64 and CT8556-37-2-3-1-M, respectively. White arrowheads point at example flowers. All photographs of the pots are on the same scale. The dimension of the inner side of the top of the pots is 10 cm × 10 cm. **Bi-ii, Ei-ii & Hi-ii**, Representative photographs of the root systems of IAC165, IR64 and CT8556-37-2-3-1-M, respectively. **Ci-ii, Fi-ii & Ji-ii**, Representative photographs of underground parts of *S. asiatica* individuals growing on IAC165, IR64 and CT8556-37-2-3-1-M, respectively. Blue arrowheads indicate emerged parasites. Pink arrowheads point at parasites that attached to the host roots but did not emerge. Scale bars = 2 cm.

Di-iv & Gi-iv). The ASNPC of IAC165 was significantly greater than those of IR64 (Mann-Whitney's U test: $U = 36.5$, $n_{\text{IAC165}} = 5$, $n_{\text{IR64}} = 8$, adjusted $P = 0.038$) and CT8556-37-2-3-1-M (Mann-Whitney's U test: $U = 35.0$, $n_{\text{IAC165}} = 5$, $n_{\text{CT8556}} = 7$, adjusted $P = 0.017$) (Figure 2.6B). The ASNPC for parasites growing on IR64 was also significantly greater than that for parasites growing on CT8556-37-2-3-1-M (Mann-Whitney's U test: $U = 47.0$, $n_{\text{IR64}} = 8$, $n_{\text{CT8556}} = 7$, adjusted $P = 0.032$) (Figure 2.6B).

When the fresh weights of above ground parasites were compared, the parasites growing in the IR64 and the IAC165 pots did not differ significantly in weights (Mann-Whitney's U test: $U = 29$, $n_{\text{IAC165}} = 5$, $n_{\text{IR64}} = 8$, adjusted $P = 0.222$) (Figure 2.6C). However, the above-ground biomass of parasites on CT8556-37-2-3-1-M was significantly lower than those growing on either IAC165 (Mann-Whitney's U test: $U = 35$, $n_{\text{IAC165}} = 5$, $n_{\text{CT8556}} = 7$, adjusted $P = 0.017$) or IR64 (Mann-Whitney's U test: $U = 49$, $n_{\text{IR64}} = 8$, $n_{\text{CT8556}} = 7$, adjusted $P = 0.018$) (Figure 2.6C), showing that the smaller size and slower growth of parasites on CT8556-37-2-3-1-M significantly reduce the number of parasites that could complete their life cycle.

At the end of the experiment, all IAC165 and IR64 pots had at least one emerged *S. asiatica* plant, whilst two out of seven pots of CT8556-37-2-3-1-M had no emerged parasites at all (Figure 2.7Di-ii). Out of ten parasites growing on CT8556-37-2-3-1-M, only one flowered at the end of the experiment (Figure 2.7Giv), compared to 16 out of 30 parasites on IR64 that started flowering more than two weeks before the one flowering parasite on CT8556-37-2-3-

1-M. After the above-ground tissues were harvested, representative pots were emptied and the root systems were washed to see whether there were underground parasites that failed to emerge (Figure 2.7Bi-ii, Ei-ii & Hi-ii). On IAC165 and IR64, most parasites that had successfully parasitised the host emerged, and only a few underground parasites failed to emerge (Figure 2.7Bi-ii, Ci-ii, Ei & Fi). In contrast, some parasites attach to the roots of CT8556-37-2-3-1-M but failed to emerge by the end of the experiment (Figure 2.7Hi-ii & Ji-ii), when most parasites on IAC165 and IR64 had emerged and had been flowering for several weeks (Figure 2.7Ai-iii & Di-iii).

2.4 Discussion

This chapter shows that out of the four parental rice genotypes of RIL populations, which were phenotyped for post-attachment resistance/susceptibility to *S. asiatica*, CT8556-37-2-3-1-M was less susceptible than the common parent, IR64, whilst two other diversity donors were as susceptible as IR64. The *S. asiatica*-resistance seen in CT8556-37-2-3-1-M had an interesting phenotype that consisted of two components, (i) the number of parasites growing on CT8556-37-2-3-1-M was significantly lower than on IR64 and (ii) parasites that attached to the host grew significantly more slowly on CT8556-37-2-3-1-M than on IR64. Also, the parasites growing on CT8556-37-2-3-1-M hosts were less likely to emerge above ground than those on IR64, and the few that emerged did so much later and were much smaller than those on IR64. The parasites on CT8556-37-2-3-1-M also flowered later and were unlikely to produce as many seeds as parasites on IR64.

2.4.1 The number of parasites was lower on CT8556-37-2-3-1-M than on IR64.

Even though there was no difference in the time taken for parasites on CT8556-37-2-3-1-M to form vascular connections with their host, the number of successful attachments was significantly lower on CT8556-37-2-3-1-M than on IR64. In the rhizotron experiment, roots of the two rice genotypes were inoculated with pre-germinated *S. asiatica* seeds, thus

differences in the exudation of germination stimulants could not explain the difference in the number of attached parasites. Therefore, the higher proportion of unsuccessful parasites on CT8556-37-2-3-1-M was likely to result from a resistance response that occurred early in the life cycle of the parasite.

A fast resistance response was seen against some parasites, as necrosis of the host root surrounding the site of attachment was visible. The necrosis began to appear on both rice genotypes 5–8 DAI (Figure 2.1Hi–ii), but more parasites died on CT8556-37-2-3-1-M than IR64, reflected in a lower number of parasites at 12 DAI (Figure 2.4). The timing of the necrosis was similar to the HR to *Striga* spp. in sorghum and cowpea (Mohamed *et al.*, 2003). In sorghum, necrotic lesions around the attachment sites of *S. asiatica* were observed in the sorghum cultivar Framida from 3 DAI and increased in number until at least 12 DAI (Mohamed *et al.*, 2003). In cowpea, *S. gesnerioides* elicited an HR on the cowpea varieties 58-57 and B301, 3–4 d after the parasite began to penetrate the host roots, which was approximately 4–7 DAI (Lane *et al.*, 1994).

2.4.2 Parasites grew more slowly on CT8556-37-2-3-1-M compared to IR64

In addition to fewer parasites on CT8556-37-2-3-1-M, a detailed analysis of the growth of *S. asiatica* individuals revealed that parasites grew significantly more slowly — with a lower relative growth rate — on CT8556-37-2-3-1-M than on IR64. This is the first report of a slow growth phenotype of *S. asiatica* identified in rice.

One hypothesis to explain the slower growth phenotype of *S. asiatica* is that there was a delay in the formation of xylem–xylem connections between *S. asiatica* and CT8556-37-2-3-1-M compared to IR64. This could delay access to host nutrients and thus delay parasite growth and development. Cissoko *et al.* (2011) found that a small number of *S. hermonthica* overcame the blockage at the endodermis of the highly resistant rice cultivar, NERICA10, and formed only a few connections to the host xylem vessels. These parasites also took longer

to breach the endodermis of resistance cultivars, compared to those on more susceptible hosts. They remained small and were associated with the accumulation of dense unknown staining materials at the host–parasite interface and the occlusion of the xylem connections (Cissoko *et al.*, 2011). Mbuvi *et al.* (2017) also reported a similar phenotype in wild sorghum (WSE-1) infected with *S. hermonthica*. They showed that a small fraction of *S. hermonthica* individuals sometimes evaded host endodermal resistance and established limited connections to the resistant WSE-1 plants. The phenotype was also associated with small endophyte and poor differentiation of the parasite hyaline body (Mbuvi *et al.*, 2017).

To investigate this hypothesis, the microscopic study of early phases of infection of parasites on CT8556-37-2-3-1-M and IR64 showed that *S. asiatica* established xylem–xylem continuity (Stage 3 in Figure 2.1Bi–ii) with a similar timing at 3 DAI (Figure 2.3). However, with the technique employed in the current study, a difference in the number of xylem–xylem connections or their permeability between the two hosts could not be realised and ruled out from explaining the slower growth of *S. asiatica* on CT8556-37-2-3-1-M. It is possible that the slower growth of *S. asiatica* on CT8556-37-2-3-1-M resulted from impeded nutrient uptake from fewer xylem–xylem connections. Three-dimensional reconstruction of *Striga* attachments from a series of thin sections is a suitable technique for quantifying the number of vascular connections (Masumoto *et al.*, 2020).

In addition to potential differences in the formation of xylem–xylem connections between host and parasites, resistance can arise after the formation of vascular connections via physiological incompatibility between the parasite and the host (Joel *et al.*, 2007). For example, the xylem sap of resistant hosts may contain metabolites that are less suitable for the parasites to assimilate, resulting in sub-optimal growth. This has been demonstrated indirectly in several studies that grew a generalist root parasitic plant species, such as species of *Rhinanthus* L., with a wide range of host species, resulting in very different rates of the growth of parasites (Matthies, 2017; Rowntree *et al.*, 2014; Hautier *et al.*, 2010; Seel & Press, 1993). The flux of nutrients from the host to parasites may be reduced in resistant hosts by a change

in water potential gradient at the haustorial interface. This can be achieved by an accumulation of host-produced osmotic solutes or insensitivity of the host to changes in the level of plant hormones caused by the parasite that would normally drive the extraction of water and nutrients from the host, such as the plant hormone ABA (Fujioka *et al.*, 2019a; Frost *et al.*, 1997). The composition of plant hormones in the xylem sap of resistant hosts may be detrimental to the growth and development of the parasites. And lastly, the resistant hosts may produce compounds in xylem sap that are growth-inhibitive or toxic to the parasite. Examples of such cytotoxic compounds include phenolic compounds, such as those produced by sorghum under *S. hermonthica* infection (Arnaud *et al.*, 1999), and phytoalexins, which are associated with resistance to the root parasite *Orobanche crenata* Forssk. in pea (*Pisum sativum* L.) and *Medicago truncatula* Gaertn. (Pérez-de-Luque *et al.*, 2009; Lozano-Baena *et al.*, 2007).

Several cases in which the growth and development of *S. hermonthica* were repressed by physiological incompatibility after the formation of xylem–xylem connections had been identified in host species. From studying longitudinal sections of parasite attachment on susceptible and resistant hosts, Gurney *et al.* (2003) attributed the arrested growth and development of *S. hermonthica* after the formation of vascular connections with a wild relative of maize, *T. dactyloides*, to the impairment of haustorial development and poor differentiation of hyaline body because no occlusion was observed in the vascular connections. Gurney *et al.* (2003) also suggested that toxic compound(s) produced by the host were responsible for the ability of *T. dactyloides* to kill off *S. hermonthica* in spite of no blockage in the xylem–xylem connections. The same study also provided evidence that *T. dactyloides* may produce compounds that inhibit haustorial development. *Striga hermonthica* individuals that attached to *T. dactyloides* via a secondary haustorium on one of the parasite lateral roots could not form a subsequent secondary haustorium on another parasite lateral root on normally susceptible maize (Gurney *et al.*, 2003). The presence of similar but less potent compound(s) that inhibit the growth of *S. asiatica* in CT8556-37-

2-3-1-M xylem sap may explain the slower growth of parasites on CT8556-37-2-3-1-M, compared to IR64.

2.4.3 How useful is this slower growth phenotype of *Striga asiatica* for *Striga* control?

The slower growth of *S. asiatica* on CT8556-37-2-3-1-M than on IR64 resulted in a delay in the parasite emergence and a reduction in the number of emerged parasites, which led to lower above-ground parasite biomass (Figure 2.6). Parasites from CT8556-37-2-3-1-M were likely to produce fewer seeds than those on IR64 since they accumulate significantly lower biomass, emerged and flowered later. This is directly beneficial to farmers in *Striga* infested areas as CT8556-37-2-3-1-M decreases parasite fitness and reduces the replenishment of *Striga* seeds back into the soil seed bank.

Although the root exudate of CT8556-37-2-3-1-M was not characterised for its ability to trigger the germination of *Striga* seeds, which is a mechanism of pre-attachment resistance (Mohemed *et al.*, 2018; Jamil *et al.*, 2011a), several *S. asiatica* individuals that failed to emerge were observed underground after the sand was washed away from the host root systems showing that CT8556-37-2-3-1-M root exudate could germinate *S. asiatica* seeds and initiate parasite attachment, and the reduced emergence was likely to be a result of slower growth of parasites, rather than a form of pre-attachment resistance.

In its current form, the slower growth phenotype of *S. asiatica* may not seem attractive to farmers and plant breeding community because CT8556-37-2-3-1-M still allows the parasites to attach and form vascular connections for a long period. The damage *Striga* plants cause to host growth and yield is likely to be significant since *Striga* spp. inflict damages to the host soon after attachment (Scholes & Press, 2008). However, a recent study that compared the levels of resistance/susceptibility in two maize inbred lines identified a similar slow growth phenotype of *S. hermonthica* individuals growing on the resistant maize genotype TZSTRI108, also known as ZD05 (Amusan *et al.*, 2008; Menkir, 2006), compared to the

susceptible genotype 5057 (Unachukwu *et al.*, 2020). TZSTRI108 is derived from the wild maize relative *Z. diploperennis* (Amusan *et al.*, 2008; Menkir, 2006). TZSTRI108 was more vigorous than 5057 under *S. hermonthica* infection and showed a reduction in the damage caused by the parasites (Unachukwu *et al.*, 2020). When tested under field conditions at two locations in Nigeria for two years, TZSTRI108 produced a higher yield than the susceptible genotype 5057, showing the value of the slow growth resistance in crop protection (Unachukwu *et al.*, 2020).

Understanding the mechanism by which the growth of *S. asiatica* individuals was slowed down would be useful as knowledge of the underlying mechanisms causing the slower growth of the parasites could pave a way to further reduce the growth of parasites and possibly completely prevent above ground emergence. The trait is suitable to be used in conjunction with other *Striga* control measures. Also understanding the mechanism behind the phenotype would be interesting in terms of further the understanding of the complex interactions between parasitic plants and their hosts. Identification of genes responsible for the phenotype is the key to unlock the underlying mechanism of the host to control the parasite. In the next chapter, a population of RILs derived from a cross between CT8556-37-2-3-1-M and IR64 will be used to identify the genetic region(s) associated with lower biomass and number of *S. asiatica* individuals growing on CT8556-37-2-3-1-M.

Chapter 3

Identification of quantitative trait loci underlying resistance to *Striga asiatica* in rice

3.1 Introduction

In Chapter 2, a difference in the levels of resistance/susceptibility to *S. asiatica* between two rice (*O. sativa* subsp. *indica*) genotypes, IR64 and CT8556-37-2-3-1-M, was identified and characterised. CT8556-37-2-3-1-M caused significant reductions in the biomass, the size and the number of parasites per host in the rhizotron system, as well as a significant delay in the emergence and a significant decrease in the number of emerged parasites in pots, compared to IR64. The identification of QTL that underlie the resistance would allow plant breeders to integrate the resistance into the breeding of resistant rice varieties, used for *Striga* control. Also, knowledge of the genetic basis of the resistance is the first step towards the discovery of genes that control the resistance and an understanding of the mechanism by which the growth and development of the parasites were influenced by the host. The RIL population derived from a cross between IR64 and CT8556-37-2-3-1-M can be used to identify QTL underlying the resistance phenotype.

In a QTL mapping study, genetic markers, such as amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR), diversity arrays technology (DArT) and single nucleotide polymorphism (SNP) markers, are regressed on the phenotypic traits of the genotypes in a segregating population, such as F₂, double haploid (DH), BIL and RIL populations, to identify their associations (Collard *et al.*, 2005; Falconer & Mackay, 1996). Many QTL for resistance to *Striga* spp. have been identified in sorghum, maize and rice (see Section 1.6), and some have led to the discovery of genes that govern the resistance phenotypes. In sorghum, a single major-effect locus for low germination stimulant production (*LGS*) was mapped in a population of 354 RILs, derived from a cross between two sorghum genotypes, SRN39 (resistant) and Shanqui Red (susceptible), with 358 DArT and SSR markers (Satish *et al.*, 2012). The maximum germination distance of *S. asiatica* and *S. hermonthica* seeds — a proxy for the measurement of the germination stimulant activity — was used as the input phenotypic data (Satish *et al.*, 2012). A single major QTL was identified on the SBI-05 chromosome, with a logarithm of the odds (LOD) score

of 82, which explained 70 % of the phenotypic variance in the mapping population (Satish *et al.*, 2012). After further fine mapping, using additional SSR markers identified by bulked segregant analysis (Satish *et al.*, 2012), Gobena *et al.* (2017) compared the genomic sequences of five resistant sorghum genotypes at the *LGS* locus and identified that the low germination stimulant phenotype was a result of deletions of or a loss-of-function mutation in the *Sobic.005G213600* gene, which was involved in strigolactone biosynthesis.

In rice, several QTL have been mapped for both pre- and post-attachment resistance to *S. hermonthica* (Table 3.1). Multiple QTL for resistance to an accession of *S. hermonthica* from Mali were discovered, using a population of 115 F₆ RILs, derived from a cross between the resistant genotype Bala (*O. sativa* subsp. *indica*) and the susceptible genotype Azucena [*O. sativa* subsp. *japonica* (subgroup *tropical japonica*)] (Kaewchumnong & Price, 2008). The RILs were grown in pots containing *S. hermonthica* seeds. Several traits associated with *S. hermonthica* resistance were measured, including the time to first parasite emergence (TFE), and the parasite biomass (dry weight) and the number of parasites at 11 weeks after planting. In total, two QTL for the parasite biomass (on chromosomes 1 and 6), one for the number of parasites (on chromosome 1) and five for the TFE (on chromosomes 1, 2, 11 and 12) were identified with LOD scores ranging from 3.6 to 7.9 and each explaining between 9.8 and 18.4 % of the phenotypic variance for different traits (Kaewchumnong & Price, 2008) (Table 3.1).

The rice cultivar Azucena exudes high quantities of strigolactones which trigger the germination of *S. hermonthica* seeds, whereas the cultivar Bala is a low germination stimulant producer. Cardoso *et al.* (2014) used the same Bala × Azucena RIL population to map a major QTL on chromosome 1 for the production of strigolactones, *qSTRIGOLACTONE BIOSYNTHESIS 1.1* (*qSLB1.1*). This QTL was the same as the QTL identified by Kaewchumnong & Price (2008) (*qShNI.1*, *qShB1.1* and *qShTFE1.1* in Table 3.1). By comparing the genomic sequences — within the *qSLB1.1* region — of the two rice genotypes, Cardoso *et al.* (2014) discovered that there was a deletion of two cytochrome P450 genes,

Table 3.1 Quantitative trait loci (QTL) for resistance to *Striga hermonthica* identified in rice.

Study (year)	Mapping population		<i>Striga</i> source	Chr.	Name*	Trait	Physical position [†] (Mbp)	LOD score	PVE (%)	Source of resistance allele
Kaewchumnong & Price (2008)	Bala × Azucena	RILs	Mali	1	<i>qShB1.1</i>	Parasite biomass	31.4	4.9	11.0	Bala
					<i>qShN1.1</i>	Number of parasites	31.4	7.6	18.4	Bala
					<i>qShTFE1.1</i>	Time to emergence	29.4	5.8	14.4	Azucena
				2	<i>qShTFE2.1</i>	Time to emergence	—	3.9	12.9	Bala
					<i>qShTFE2.2</i>	Time to emergence	29.6–30.3	4.0	9.8	Bala
				6	<i>qShB6.1</i>	Parasite biomass	22.9	4.8	11.2	Bala
				11	<i>qShTFE11.1</i>	Time to emergence	2.0	3.6	10.0	Azucena
				12	<i>qShTFE12.1</i>	Time to emergence	21.3	4.9	12.3	Bala
Cardoso <i>et al.</i> (2014)	Bala × Azucena	RILs	Not reported	1	<i>qSLB1.1</i>	Strigolactone biosynthesis	29.0 [‡]	—	—	Bala
Gurney <i>et al.</i> (2006)	Nipponbare × Kasalath	BILs	Kibos, Kenya	1	<i>qShR1.1</i>	Parasite success rate	31.4	3.9	1.8	Nipponbare
				4	<i>qShR4.1</i>	Parasite success rate	8.2	14.4	7.6	Kasalath
				5	<i>qShR5.1</i>	Parasite success rate	22.6–24.5	4.3	1.9	Nipponbare
				6	<i>qShR6.1</i>	Parasite success rate	29.2	9.8	4.2	Nipponbare
				7	<i>qShR7.1</i>	Parasite success rate	20.3	12.5	5.5	Nipponbare
				8	<i>qShR8.1</i>	Parasite success rate	24.2	4.7	2.1	Nipponbare
				12	<i>qShR12.1</i>	Parasite success rate	7.5	13.7	7.4	Nipponbare
Swarbrick <i>et al.</i> (2009)	Kasalath × Koshihikari	BILs	Kibos, Kenya	3	<i>qShR3.1</i>	Parasite success rate	16.2	2.6	7.0	Kasalath
				4	<i>qShR4.1</i>	Parasite success rate	10.4	6.0	16.2	Kasalath
				10	<i>qShR10.1</i>	Parasite success rate	9.6	2.6	7.1	Koshihikari
Scholes <i>et al.</i> , (unpublished)	Nipponbare × Koshihikari	BILs	Kibos, Kenya	12	<i>qShR12.1</i>	Parasite biomass	5.7–6.7 [‡]	23	55	Nipponbare
Beardon (2018)	IR64 × Azucena	RILs	Kibos, Kenya	12	<i>qShR12.1</i>	Parasite biomass	5.7–6.7 [‡]	14.1	67.9	IR64

* Name of the QTL created for reference in this Thesis; [†] Physical position of the nearest marker on the Nipponbare reference genome in Gramene database (Youens-Clark *et al.*, 2011); [‡] Physical position of the QTL on the Nipponbare reference genome, reported in the study; Chr., chromosome; PVE, percentage of phenotypic variance explained.

SLB1 and *SLB2* in the Bala genome. These genes are involved in the biosynthetic pathway of strigolactones and their deletion was responsible for the reduction in the germination stimulant activity in the rice genotype Bala.

Several QTL for post-attachment resistance to *S. hermonthica* have also been identified (Table 3.1). Seven QTL were identified for post-attachment resistance to *S. hermonthica* from Kibos in Kenya, by phenotyping 98 BILs derived from a cross between the resistant *O. sativa* subsp. *japonica* (subgroup *temperate japonica*) cv. Nipponbare and the susceptible *O. sativa* subsp. *indica* (subgroup *aus*) cv. Kasalath in a rhizotron system (Gurney *et al.*, 2006). Of the seven QTL that each explained from 1.8 to 7.6% of the phenotypic variance in the mapping population, one QTL (*qShR4.1*) had the Kasalath allele conferring the resistance (Table 3.1), whilst all the other QTL derived resistant alleles from Nipponbare. This *qShR4.1* was later remapped by phenotyping 182 BILs derived from another cross between Kasalath (more resistant) and the *O. sativa* subsp. *japonica* (subgroup *temperate japonica*) cv. Koshihikari (more susceptible), backcrossed to Koshihikari, for resistance to the same *S. hermonthica* accession (Swarbrick *et al.*, 2009). The new results showed that *qShR4.1* explained 16% of the phenotypic variance in this new population and it conferred resistant independently of the genetic background (Swarbrick *et al.*, 2009) (Table 3.1).

In the study by Gurney *et al.* (2006), the QTL which explained the highest percentage of phenotypic variance for post-attachment resistance was on chromosome 12 (*qShR12.1*). This QTL was identified again in two separate studies (J. D. Scholes, personal communication; Beardon, 2018). The first study aimed to resolve the locations and effects of the Nipponbare QTL previously identified, without the influence of the Kasalath allele at *qShR4.1*, using a new BIL population derived from Nipponbare/Koshihikari//Koshihikari (J. D. Scholes, personal communication). Only one highly significant QTL was identified at the same locus as *qShR12.1*, with the LOD score of 23 and explaining 55% of phenotypic variance (Table 3.1; Scholes *et al.*, unpublished results). The second study by Beardon (2018) involved a two-step strategy where 64 randomly selected RILs, derived from a cross between the resistant rice

cultivar IR64 and the susceptible cultivar Azucena, were initially phenotyped for resistance to the same accession of *S. hermonthica* from Kibos, Kenya. As this RIL population had high-density SNP markers, a preliminary QTL analysis was carried out and revealed one major-effect QTL on chromosome 12. Then to narrow the QTL region further, 20 more RILs that had recombination breakpoints within the QTL region were screened and the results were pooled with those from the initial screen for a subsequent QTL analysis which increased the LOD score of the QTL on chromosome 12 (Table 3.1).

The phenotype of resistance characterised in Chapter 2 consisted of a slow growth phenotype of *S. asiatica* attachment on CT8556-37-2-3-1-M compare to IR64, which is likely to be a complex polygenic trait that is influenced by many genes of minor effects. The polygenic nature of the trait needs to be considered when carrying out a QTL analysis. Over the years, many methods for QTL mapping have been developed with increasing complexity and sophistication, from marker regression that cannot account for loci within gaps between markers to the mapping of multiple QTL, each interacting with one another (Collard *et al.*, 2005; Broman, 2001). In marker regression (MR) (Soller *et al.*, 1976) and simple interval mapping (SIM) (Lander & Botstein, 1989), the position of a hypothetical QTL was tested at each marker or locus, respectively, one at a time, with the assumption that only a single QTL is present in the genome. SIM, which requires the construction of a linkage map beforehand, use the information on the linkage between markers to infer the likelihood of the presence of QTL between markers to improve the estimates of QTL effect sizes since it incorporates the effects of recombination and allows the inclusion of individuals with missing marker data via considering the status of flanking makers (Broman, 2001). These advantages of SIM are very useful when the number of genetic markers is limited. Although SIM assumes the presence of one QTL, it can be used to detect multiple QTL, especially when QTL are located on different chromosomes, which are independently segregated. Even though using SIM to detect multiple QTL would not allow any detection of epistatic interactions and has a lower detection power than more sophisticated methods, which require building and selecting

models that include multiple QTL, SIM provides a baseline result with a higher certainty in the QTL it detects, as no model selection process is involved.

Examples of QTL detection methods that integrate the presence of multiple QTL into the inferred sets of QTL and their possible interactions (referred to as QTL models) includes composite interval mapping (CIM), inclusive CIM (ICIM), multiple interval mapping (MIM) and multiple QTL mapping (MQM). The widely used CIM (Zeng, 1994, 1993; Jansen & Stam, 1994; Jansen, 1993) uses a subset of markers as cofactors (covariates), which represent other QTL in the model while performing interval mapping. CIM increases the power and the precision of QTL detection by controlling the effects of linked QTL via the covariates and reducing the residual variation. However, the selection of cofactors is critical as the QTL detection power decreases if too many or too few markers are selected and the uncertainty of the selected cofactors is not considered when the subsequent interval mapping is performed (Broman, 2001). A newer method, ICIM (Li *et al.*, 2008, 2007) improved the estimates of the QTL effects in the CIM algorithm by selecting the cofactors only once through stepwise regression, which simultaneously includes all marker information. In a different approach, MIM (Kao *et al.*, 1999) extended the method used in CIM to include multiple QTL and their possible interactions in the QTL model with multiple regression. Since searching the full model space, which involves evaluating all possible QTL at any locations, is not feasible, MIM uses stepwise regression to select QTL to include in the final model. Lastly, MQM (Arends *et al.*, 2010) selects important markers using multiple regression and backward elimination of unnecessary markers and uses the remaining markers as cofactors to carry out interval mapping. MQM is superior to CIM in that it fixes the residual variance from both the full model and the reduced model, in which the QTL under test is removed from the full model, whereas CIM estimates the residual error variance for the reduced model that will absorb the QTL variance, resulting in a lower test statistic (Jansen, 2007). As a result, a larger number of cofactors can be used in MQM, compared to CIM, without the problem of overfitting (Arends *et al.*, 2010).

These multiple methods of QTL detection can be used in the mapping of QTL on any biparental mapping populations; one of which is a RIL population. Although a RIL population takes longer to develop than, for example, F₂ population, it can be screened on multiple occasions in multiple environments as it is considered an immortal population because the alleles at all loci are virtually fixed in the F₇ generation (with theoretical 1.6 % heterozygosity) which is commonly used in the development of a RIL population. Even though a RIL population contains some heterozygous loci, which are rendered as missing markers in many QTL mapping programs (they only accept homozygous markers), a RIL population is still better than purely homozygous DH population because there are more opportunities for recombination events to happen as the generation advances through the population development, unlike DH population, which only goes through one round of meiosis. More recombinations allow for better resolution of QTL detection.

The RIL population, derived from IR64 × CT8556-37-2-3-1-M, is a part of a rice NAM population and consists of 190 RILs (Fragoso *et al.*, 2017). The population was genotyped by sequencing, in which the genomes of IR64, CT8556-37-2-3-1-M and the RILs were sequenced and mapped onto the Nipponbare genome (Kawahara *et al.*, 2013), which is the standard rice reference genome. Sequence variants were called and imputed between the sequence reads using the positions on the Nipponbare reference genome, yielding 5 126 982 variants in total, which were reduced to essential 6504 SNP markers that mark all the recombination sites in the population.

3.1.1 Aim and objectives

This chapter aimed to map QTL (using different QTL mapping methods) that underlie the *S. asiatica*-resistance phenotype of CT8556-37-2-3-1-M, using the RIL population derived from IR64 × CT8556-37-2-3-1-M. Specific objectives include:

- i) Phenotyping the RIL population using a two-step phenotyping strategy, in which a third of the RIL population (randomly selected) will be phenotyped first, followed by

a QTL analysis to assess if any QTL could be detected. If QTL are detected, additional RILs that contain recombination breakpoints around the QTL will then be phenotyped in a second screen. Data from the first and second screens will then be combined and used in a second QTL analysis. If no QTL are detected in the first screen, all remaining RILs will be phenotyped in the second screen and all phenotypic data will be used in the second QTL analysis.

- ii) Preparing the phenotypic data from the phenotypic screen for QTL analyses. This involves removing escapes from the dataset, using transformation on data (if necessary), calculating repeatability and calculating conditional means using general linear mixed-effect models, including the validation of model assumptions.
- iii) Constructing a linkage map of the RIL population using the SNP markers information, provided by Dr Mathias Lorieux.
- iv) Performing QTL analyses, using SIM, ICIM, MIM and MQM, to compare the results from the multiple QTL mapping approaches.

3.2 Methodology

3.2.1 Plant materials

The same batch of *S. asiatica* seeds from the Ethiopia accession used in Chapter 2 was used again in this study.

A rice mapping population, consisting of 190 RILs, derived from a cross between two *O. sativa* subsp. *indica* rice genotypes (IR64 and CT8556-37-2-3-1-M), was used in this study. The RIL population was a part of a NAM population (Fragoso *et al.*, 2017), which was developed at CIAT, Colombia. The RILs were produced by crossing the parental genotypes to produce ~ 300 F₁ hybrids. The F₁ plants were then selfed to create a segregating population until the F₇ generation was reached, through single-seed descent. Due to environmental conditions and partial sterility, only 190 RILs were produced. Out of these, 184 RILs were genotyped by sequencing (Fragoso *et al.*, 2017), yielding 6504 SNP markers, marking all

recombination breakpoints in the population. Six RILs with high heterozygosity were excluded from the study, leaving 178 RILs.

3.2.2 Phenotypic screen, data preparation and statistical analyses

The growth and infection of rice with *S. asiatica* was performed according to the protocol described in Section 2.2.2. The biomass (dry weight) of parasites harvested from the root system of each host was recorded at 28 DAI after the root system was scanned at a high-resolution (1200 dpi). A workflow of experimental design, data preparation and statistical analyses prior to QTL analyses is shown in Figure 3.1.

3.2.2.1 Phenotypic scoring of the recombinant inbred line population using a two-step phenotypic scoring strategy

A two-step phenotypic scoring strategy (Beardon, 2018) was employed to reduce the time required to score the RIL population for resistance to *S. asiatica*. The two-step strategy consisted of an initial phenotypic screen of a small number of randomly selected rice genotypes, followed by a QTL analysis to determine whether any QTL could be detected. If QTL were detected, more RILs from the population with recombination breakpoints around the positions of the QTL, could be selected and phenotyped in a second screen (together with some RILs from the initial set of randomly selected RILs). This strategy allows the confidence interval of the QTL to be narrowed down and increases the significance of the putative QTL without having to screen the whole population, though it is less sensitive to QTL of small effects. If no significant QTL is detected in the first screen, all remaining genotypes within the RIL population will need to be phenotyped in the second screen to see whether any small-effect QTL can be detected. The two-step phenotypic scoring strategy was particularly useful for screening for a post-attachment resistance against *S. asiatica* because the growth and infection of *S. asiatica* are time-consuming and cumbersome. In this study, 57 of the 178 RILs were screened for *S. asiatica* resistance in the first screen (Screen 1).

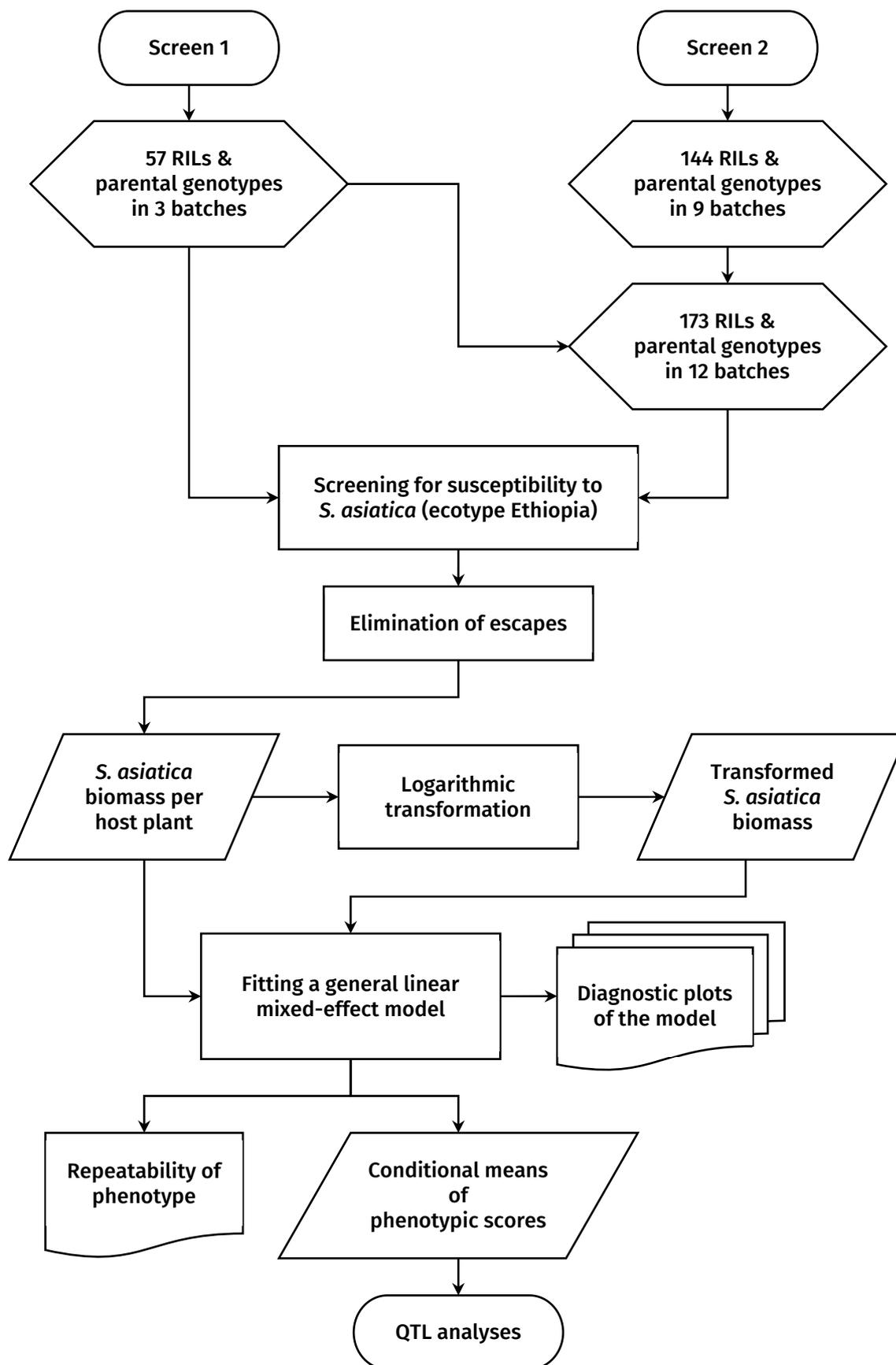


Figure 3.1 Flowchart of experimental design, data preparation and statistical analyses prior to quantitative trait locus (QTL) analyses. Statistical and QTL analyses were carried out on both data from Screen 1 and the combined data from Screen 1 and Screen 2.

From previous work, e.g. Beardon (2018) and Cissoko (2012), four replicates of each rice genotype in a mapping population were required to robustly score the resistance/susceptibility level to *Striga*. In order to maintain a uniform and synchronous inoculation of *S. asiatica* across replicate hosts, the maximum number of hosts to be inoculated at the same time was limited to approximately 80 plants, hence Screen 1 was carried out in three batches. For each RIL, one replicate was placed in each of the three batches, whilst the fourth replicate was randomly assigned to one of the three batches. This experimental design enabled a statistical comparison between batches to assess the inter-batch variation. Each batch also contained four replicates of each of the two parental genotypes. Batch sizes ranged from 78 to 86 plants (Supplementary Table S3.1).

Following the QTL analysis of the data from Screen 1, Screen 2 was designed to include all available RILs to increase the population size, hence higher detection power of QTL. To make use of the data from Screen 1, twelve RILs along with the parental genotypes were rescored in Screen 2 to allow the normalisation of any differences in infection level between the two screens, when the two datasets were combined. The twelve RILs were selected to cover the whole range of susceptibility in Screen 1. Screen 2 comprised of 144 RILs with four replicates per RIL. Screen 2 was divided into nine batches, each containing the parental genotypes (four replicates each) and 64 RILs. The four replicates of each RIL were randomly assigned to the nine batches with the condition that there was ≤ 1 replicate of each RIL in each batch. Also, the replicates of the twelve RILs being rescored were assigned across the nine batches in a way that each batch had similar number of resistant and susceptible RILs. The placement of replicates in Screen 2 can be found in Supplementary Table S3.2.

3.2.2.2 Elimination of escapes

Prior to data analyses, replicate plants with aberrant root architecture were removed from the dataset. If there were fewer than three replicates per RIL after the exclusion, the RIL was omitted from the dataset. An example of a replicate that was removed from RIL number 176

is shown in Figure 3.2. The plant root system of replicate D had not grown and thus this replicate was removed (Figure 3.2D). In total, 58 plants from ten genotypes in Screen 1 and 110 plants from 54 genotypes from Screen 2 were removed.

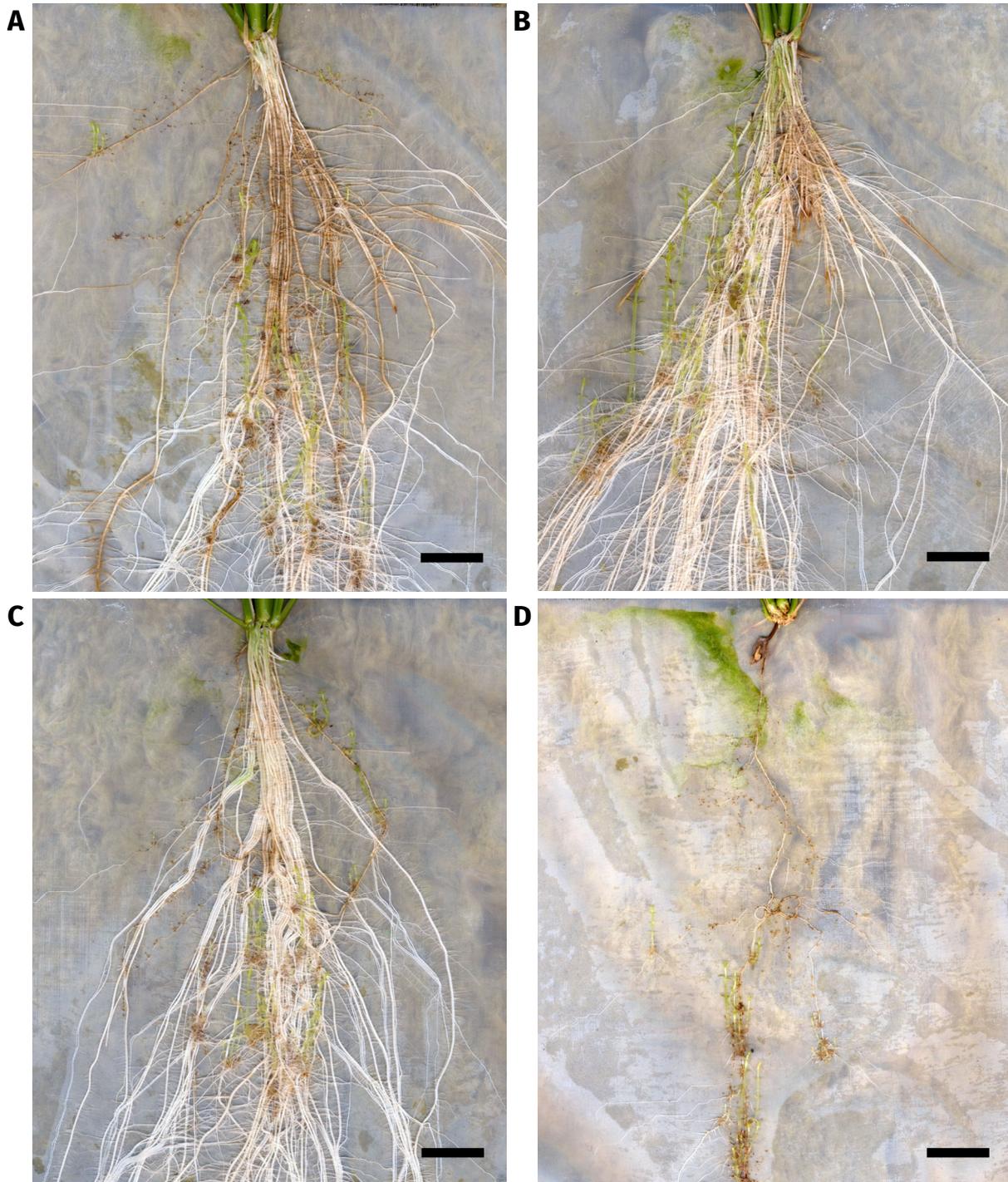


Figure 3.2 Scans of rhizotrons of the rice recombinant inbred line (RIL) number 176 before the *Striga asiatica* plants were harvested at 28 d after inoculation. Four rhizotrons contained four replicates of the same RIL (A–D). Compared to other replicates, replicate D, which had an undeveloped root system (smaller root system), was excluded from the dataset before data analyses. Scale bars = 1 cm.

3.2.2.3 Selection of the trait to represent the phenotype

The level of resistance or susceptibility to *Striga* spp. in rice has been quantified by the proportion of successful parasites (parasite success rate), the number of parasites per host plant, the biomass of parasites per host plant (Table 3.1) and the cumulative length of parasites per host plant (Beardon, 2018; Cissoko *et al.*, 2011). The length and the biomass of *Striga* per host plant have the advantage of reflecting a more realistic susceptibility level when numerous, small unsuccessful *Striga* plants are found on the host. In Screen 1, there was a very strong positive correlation between the biomass and the cumulative length of *S. asiatica* plants per rice host ($r = 0.911$, $n = 185$, $P < 0.001$) (Figure 3.3A), thus analysing both traits for QTL would likely yield the same set of QTL. Also in Screen 1, there was a strong positive correlation between the biomass and the number of parasites per rice host ($r = 0.618$, $n = 185$, $P < 0.001$) (Figure 3.3B), thus some of the QTL identified using the parasite biomass would likely underlie the difference in the number of parasites as well. Therefore, the biomass of *S. asiatica* per host plant was chosen for all further analyses as this trait was easiest to measure and would speed up screening time if the whole RIL population had to be screened.

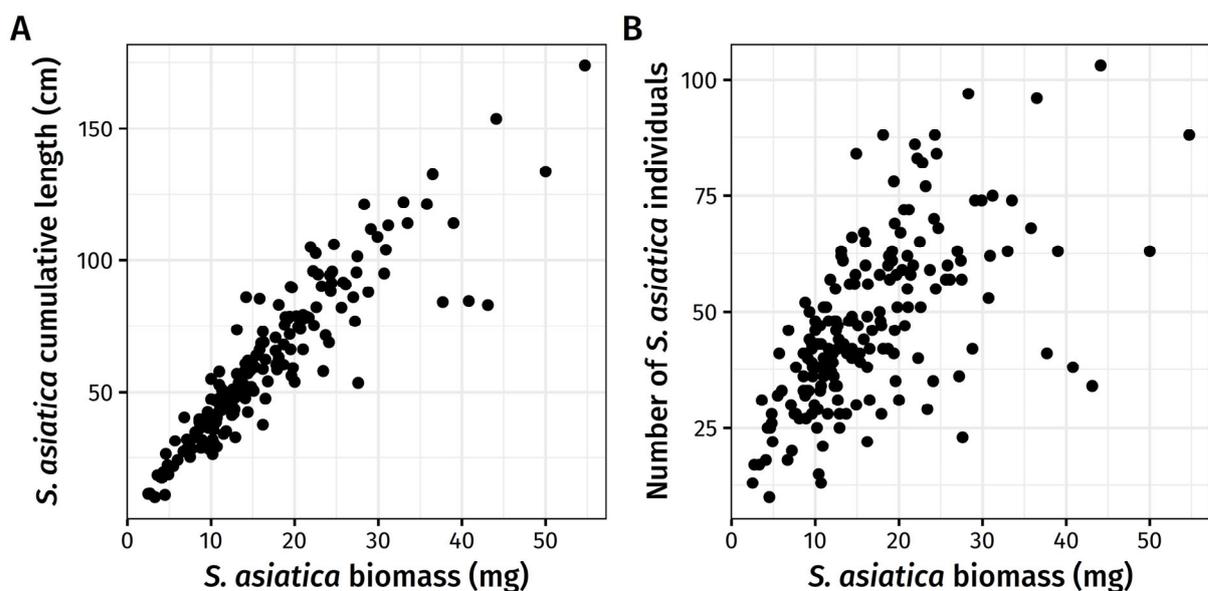


Figure 3.3 Relationships between the measurements of traits related to *Striga asiatica*-resistance in Screen 1. A, Relationship between the biomass and the cumulative length of *S. asiatica* per host plant. B, Relationship between the biomass and the number of *S. asiatica* individuals per host plant.

3.2.2.4 Assessment of the phenotypic consistency of the genotypes between batches

To assess the consistency of the phenotype scored by carrying out the screen in batches, the repeatability (the intra-class correlation) of genotypes and batches were calculated by partitioning overall variance into the inter-genotype variance (genetic variations) and the inter-batch variance, respectively. The repeatability and its confidence interval and significance level were calculated using the `rptR` package (Stoffel *et al.*, 2017) in the R software environment version 3.6.0 (R Core Team, 2017). To separate the genetic and the non-genetic components of variance, genotypes and batches were fitted as a pair of crossed random effects with the intercept (population mean) fitted as a fixed effect in a generalised linear mixed-effects model (GLMM), shown in the `lme4` notation (Bates *et al.*, 2015) in Equation 3.1. The confidence interval of repeatability was calculated using 10 000 parametric bootstrap iterations. The significance level of repeatability was tested using a permutation test with 10 000 permutations.

$$\text{Phenotypic score} \sim 1 + (1 \mid \text{Genotype}) + (1 \mid \text{Batch}) \quad (3.1)$$

In this chapter, GLMMs with a normal error distribution and an identity link function, also called general linear mixed-effect models (LMMs), were used. The assumptions of an LMM for gaussian data, which was offered in the `rptR` package, include normal distribution and homoscedasticity of data (Stoffel *et al.*, 2017). Preliminary data exploration showed that the *S. asiatica* biomass per host was right-skewed. Therefore, to ensure that there was no violation of the model assumptions, logarithmic transformation with the common logarithm, $x' = \log_{10}(x)$, was applied to the data. To assess whether the assumptions of the model were followed, diagnostic plots and Shapiro–Wilk’s normality test—as implemented in R software version 3.6.0 (R Core Team, 2017)—were used. Diagnostic plots consisted of the frequency histogram of the residuals, the residuals versus fitted values, the scale–location and the normal quantile–quantile (Q–Q) plots.

3.2.2.5 Did the biomass of parasites meet the assumptions of LMM?

In Screen 1, the distribution of untransformed biomass of *S. asiatica* per host plant was significantly different from a normal distribution (Shapiro–Wilk test: $W = 0.910$, $n = 185$, $P < 0.001$). On the other hand, there was no significant difference between the distribution of log-transformed biomass of *S. asiatica* per host plant and normal distribution (Shapiro–Wilk test: $W = 0.990$, $n = 185$, $P = 0.242$). This was supported by the diagnostic plots (Figure 3.4). The frequency histogram of residuals of the log-transformed data showed more normally distributed residuals compare to the untransformed data (Figure 3.4A–B). The quantile values of the experimental data in the normal Q–Q plot matched the theoretical quantile line of a normal distribution more closely in the log-transformed data (Figure 3.4C–D). Log-transformation also resulted in the elimination of systematic increase in residual values as the biomass of *S. asiatica* per host plant increased (Figure 3.4E–F). Lastly, the assumption of homoscedasticity was met after the data were log-transformed. The residual variance of the untransformed data increased with the fitted value (Figure 3.4G), whilst the variance of the log-transformed data remained relatively constant (Figure 3.4H). Therefore, subsequent statistical and QTL analyses were carried out on the log-transformed *S. asiatica* biomass for Screen 1.

In the combined dataset (Screen 1 and 2), Shapiro–Wilk’s test of normality showed that the distributions of both untransformed and log-transformed biomass of *S. asiatica* per host plant were significantly different from a normal distribution (untransformed data: $W = 0.897$, $n = 785$, $P < 0.001$; log-transformed data: $W = 0.994$, $n = 785$, $P = 0.003$). However, the significant results were likely caused by the test being too sensitive to minute deviations from normality, due to the large sample size. This hypothesis was confirmed after the diagnostic plots were considered (Figure 3.5). When the untransformed and the log-transformed data were compared, the same pattern observed in Screen 1 emerged. The log-transformed data better conformed to both assumptions of normality and homoscedasticity; thus, the log-transformed *S. asiatica* biomass plant was selected for all further analyses.

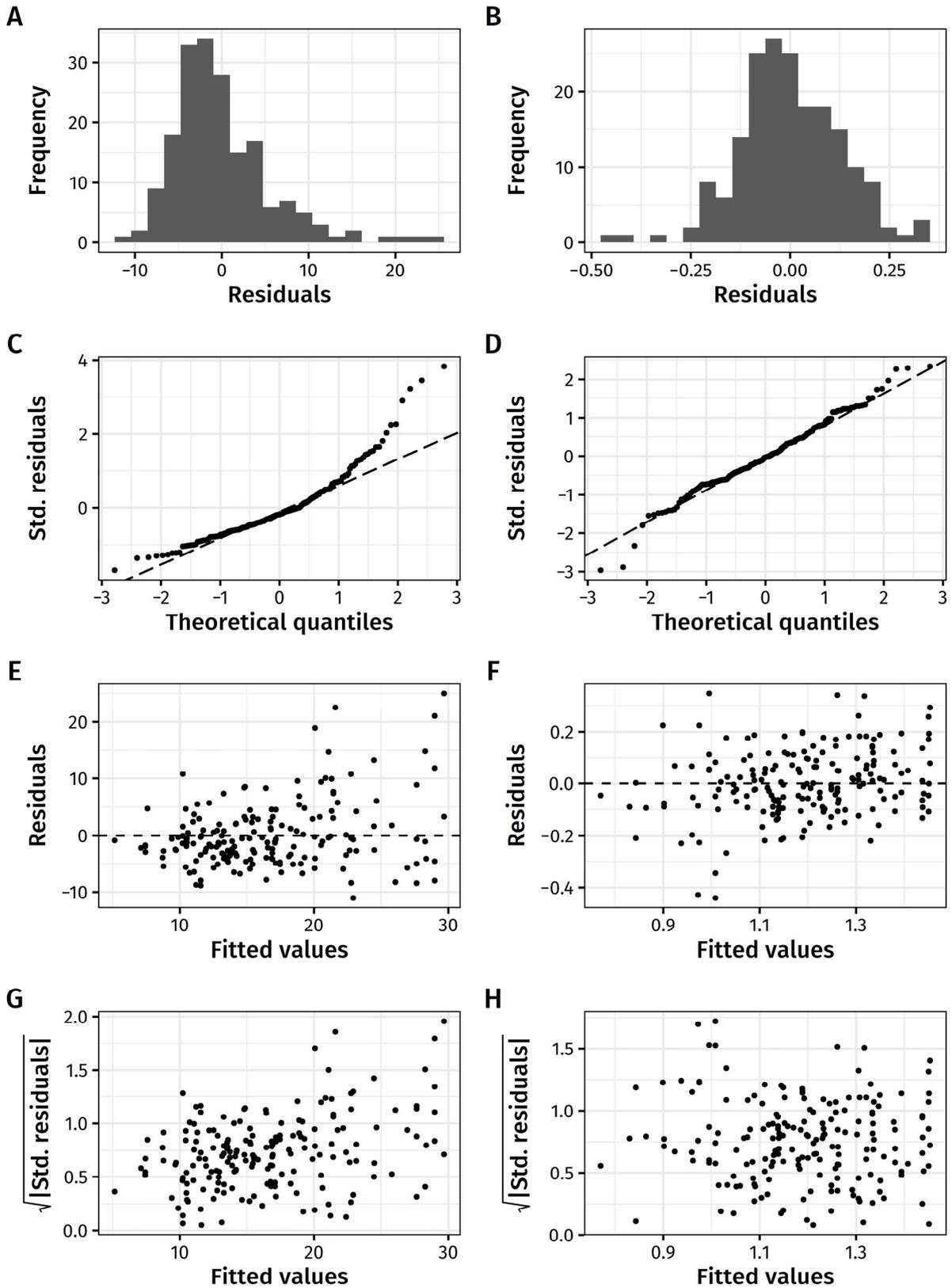


Figure 3.4 Diagnostic plots of the untransformed and log-transformed biomass of *Striga asiatica* per host from Screen 1 after the data was fitted to a linear mixed-effect model. The untransformed data are shown in **A, C, E** and **G**. The transformed data are shown in **B, D, F** and **H**. Diagnostic plots include the frequency histogram of residuals (**A–B**), the normal quantile–quantile plot (**C–D**), the residuals versus fitted values plots (**E–F**) and the scale–location plots (**G–H**). Std. = standardised.

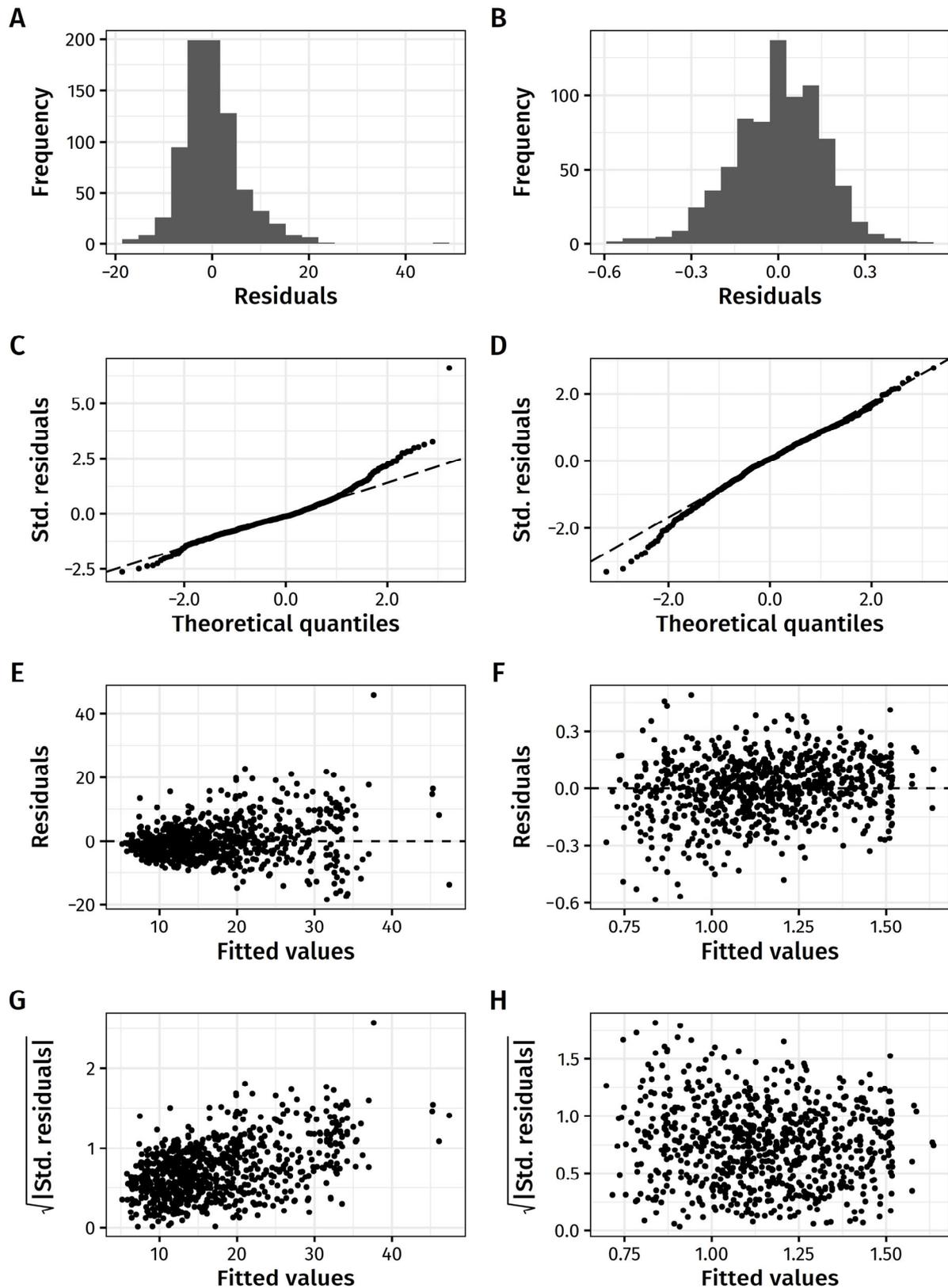


Figure 3.5 Diagnostic plots of the untransformed and log-transformed biomass of *Striga asiatica* per host from Screen 1 and Screen 2 after the data was fitted to a linear mixed-effect model. Diagnostic plots of the untransformed data are shown in A, C, E and G. Diagnostic plots of the transformed data are shown in B, D, F and H. A–B, Frequency histogram of residuals. C–D, Normal quantile–quantile plot. E–F, Residuals versus fitted values plots. G–H, Scale–location plots. Std. = standardised.

3.2.2.6 Phenotypic inputs for QTL analysis from the LMM

The conditional means of the phenotypic scores of different genotypes, which came from the best linear unbiased predictors (BLUPs) that were normalised for the inter-batch variations in the LMM (Equation 3.1), were used as phenotypic inputs in subsequent QTL analyses. Package `lme4` (version 1.1-21) was used to calculate the conditional means via the `ranef` function (Bates *et al.*, 2015).

3.2.3 QTL analyses

Different QTL analyses were used to detect QTL underlying *S. asiatica*-resistance in the RIL population. Phenotypic data from Screen 1 were analysed via SIM for QTL detection, whilst the combined phenotypic data from Screen 1 and 2 were analysed using SIM, ICIM, MIM and MQM. Although SIM can be used to detect multiple loci that are associated with a trait, the underlying QTL model of SIM assumes the presence of single QTL in the genome and does not consider any epistatic interactions between loci. More advanced approaches, such as ICIM, MIM and MQM, were designed to take epistatic interactions into account and control for the loci with major effects, to improve the detection power of loci with smaller effects. Thus, the data in the combined dataset were analysed using MQM (approach 1 and 2), ICIM and MIM. For each mapping method, the significance threshold of LOD scores was computed at the genome-wide significance levels, $\alpha = 0.01$ and $\alpha = 0.05$, using a permutation test (Hyndman & Fan, 1996; Churchill & Doerge, 1994) with 10 000 permutations. The support intervals of the QTL were calculated using a 1.5–LOD score drop either side of the peak.

3.2.3.1 Linkage map construction

Marker information, which was provided by Dr Mathias Lorieux, had been imputed using LB-Impute (Fragoso *et al.*, 2016) and BP-Impute (Fragoso *et al.*, 2017). The RILs had been genotyped by sequencing, in which the sequence reads of the RILs were mapped onto the Nipponbare reference genome (IRGSP-1.0) (Fragoso *et al.*, 2017; Kawahara *et al.*, 2013). To

create genetic markers, SNPs were called from the resulting alignment and named after the position on the IRGSP-1.0 chromosomes. Approximately 2.0 % of marker information was missing (Figure 3.6A) because heterozygous markers were treated as missing data by the QTL mapping software (Broman *et al.*, 2003). The software **MapDisto** version 2.0 beta 106 (Heffelfinger *et al.*, 2017; Lorieux, 2012) was used to construct a linkage map with the Kosambi mapping function (Kosambi, 1944). The resulting linkage map had a total length of 1550.5 cM. The mean and the maximum gap between adjacent markers were 0.44 cM and 16.0 cM, respectively (Figure 3.6B).

The segregation distortion markers were detected using Pearson's χ^2 goodness-of-fit tests to measure the deviation from the 1:1 allelic ratio of AA:BB at each marker. The *P*-value was adjusted for multiple testing, using the sequential Bonferroni method (Reflinur *et al.*, 2014; Rice, 1989). There were significant segregation distortions markers on chromosomes 4, 7 and 8 (Figure 3.6C).

3.2.3.2 Simple interval mapping (SIM)

A QTL search via SIM was carried out in the R/**qt1** software version 1.44-9 (Broman *et al.*, 2003). Using function **scanone** of R/**qt1**, SIM was performed with extended Haley-Knott regression method (Feenstra *et al.*, 2006) after the conditional genotype probabilities were calculated every 1 cM, when that was a gap wider than 1 cM between markers, using the **calc.genoprob** function with the Kosambi map function and default settings.

When SIM was performed on the combined dataset and there was a significant secondary LOD score peak near the main peak, the validity of the secondary peak was assessed by fake-QTL evaluation (Lorieux, 2018), where the observed LOD score and the expected LOD scores, $E(Z_M)$, calculated from two sets of recombination fractions were analysed. If the secondary LOD score peak arises from spurious recombinations between loci of the main peak and the secondary peak, the observed LOD curve should follow the expected LOD scores $E_p(Z_M)$, calculated from the actual pairwise recombination fraction between the

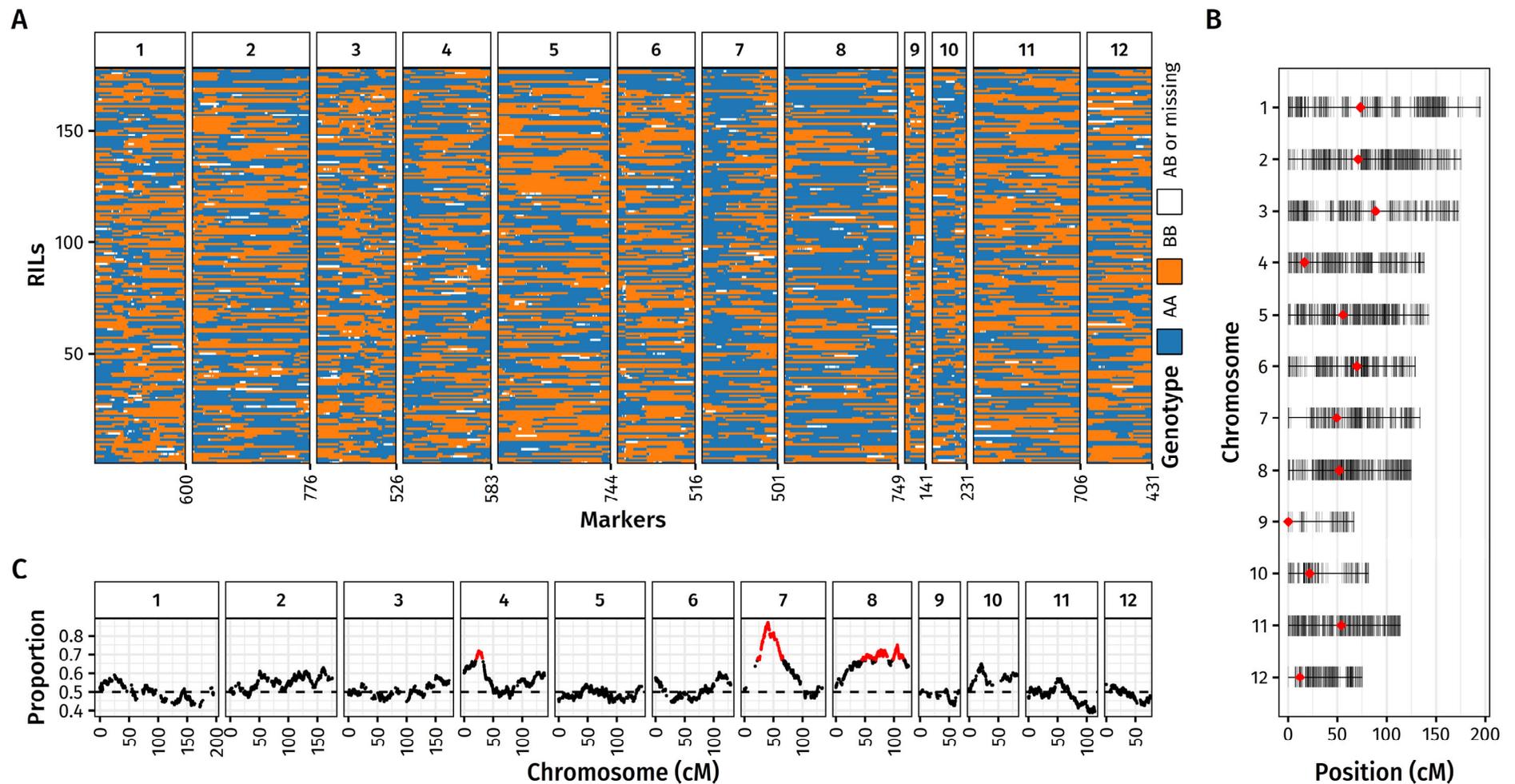


Figure 3.6 Genotypic data of the recombinant inbred line (RIL) population derived from CT8556-37-2-3-1-M \times IR64 for interval mapping. Alleles of 6504 SNP markers (**A**) are shown in colours. Blue = homozygous IR64 allele; orange = homozygous CT8556-37-2-3-1-M allele; white = heterozygous markers or missing data. Linkage map (**B**) shows genetic positions of the SNP markers with the centromeres marked by red diamonds. The segregation distortion (**C**) is shown by the proportion of IR64 alleles (AA) in each marker from 178 RILs. Red markers were significantly distorted at the genome-wide $\alpha = 0.05$ level with sequential Bonferroni correction.

marker at the global peak of LOD scores and other markers (\hat{r}_p), and appear greater than the expected LOD scores $E_d(Z_M)$, calculated from the recombination fractions that are derived from the genetic distance (\hat{r}_d). The \hat{r}_p was calculated in R/qtl (version 1.44-9) using the `est.rf` function. The \hat{r}_d was calculated from the inverted Kosambi's mapping function (Vinod, 2011), as seen in Equation 3.2.

$$\hat{r}_d = \frac{1}{2} \tanh(2d) \quad (3.2)$$

where \tanh is the hyperbolic tangent function and d is the genetic distance in Morgan (100 cM).

3.2.3.3 Inclusive composite interval mapping (ICIM)

A QTL analysis by ICIM (Li *et al.*, 2007) was carried out in the QTL IciMapping software version 4.1.0 (Meng *et al.*, 2015). The mapping was based on the ICIM of QTL with additive effects ('ICIM-ADD' option) every 0.1 cM along the length of each chromosome. The P -value threshold in the stepwise regression for the removal of markers as covariates from the model was set at 0.02 (Jansen, 1994) by setting the 'PIN' setting to 0.01.

3.2.3.4 Multiple interval mapping (MIM)

MIM, proposed by Kao *et al.* (1999), was carried out as implemented in the QGene software version 4.4.0 (Joehanes & Nelson, 2008) with default settings. The selection and usage of cofactors were done automatically by the software.

3.2.3.5 Multiple QTL mapping (MQM)

MQM was proposed by Jansen (2007, 1993) and implemented in the software R/qtl using two different approaches. The first approach (MQM-1) followed a framework set by Sen and Churchill (2001) in which a QTL model comprising of multiple QTL and their epistatic interactions is constructed by analysing the results from two-dimensional scans for pairs of loci using the `scantwo` function (Manichaikul *et al.*, 2009; Broman *et al.*, 2003). In the

two-dimensional scan, four models exploring epistatic interactions (Equation 3.3–3.6) were built for each pair of loci in 1 cM × 1 cM grids, using Haley–Knott regression method (Haley & Knott, 1992).

$$\text{Full model: } y = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \gamma(Q_1 \times Q_2) + \varepsilon \quad (3.3)$$

$$\text{Additive model: } y = \beta_0 + \beta_1 Q_1 + \beta_2 Q_2 + \varepsilon \quad (3.4)$$

$$\text{Single-QTL model: } y = \beta_0 + \beta_1 Q_1 + \varepsilon \quad (3.5)$$

$$\text{Null model: } y = \beta_0 + \varepsilon \quad (3.6)$$

where y is the phenotype, β_i denotes the regression coefficient associated with the i^{th} term, Q_i is the genotype of the i^{th} QTL, γ is the interaction between the QTL pair and ε is a random normal variable.

The four models were compared in five ways of comparisons — full model *versus* null model (Z_{full}), full model *versus* single-QTL model (Z_{full-1}), full model *versus* additive model ($Z_{interaction}$), additive model *versus* null model ($Z_{additive}$) and additive model *versus* single-QTL model ($Z_{additive-1}$). Each model comparison produced a LOD score. The significance levels of the LOD scores were calculated in a permutation test (10 000 permutations). Because the two-dimensional scan was used as an exploratory tool to find possible additional QTL and interactions to include in an initial QTL model before their validity was formally tested again when the final QTL model was fitted, a less stringent significance threshold of $\alpha = 0.1$ was used. When the initial QTL model was constructed, a pair of QTL with epistasis interactions were included when Z_{full} and either Z_{full-1} or $Z_{interaction}$ were significant, and a pair of QTL without epistatic interactions were added when $Z_{additive}$ and $Z_{additive-1}$ were significant. The initial QTL model was then used as the starting point of a model selection process with a stepwise search algorithm, using the `stepwiseqtl` function (Manichaikul *et al.*, 2009; Broman *et al.*, 2003). The penalties for the stepwise search algorithm were set at $\alpha = 0.01$ and were calculated from the 10 000 permutations of the LOD scores of the two-dimensional scan. The output of this process is the final QTL model of MQM-1.

In the second approach (MQM-2), a model of multiple QTL and their interactions with cofactors was automatically fitted by the `mqmscan` function (Arends *et al.*, 2010). A standard `mqmscan` routine involves augmentation of missing genotypes prior to the model building step. However, the genotypic data from Section 3.2.3.1 contained too many missing genotypes for `mqmaugment` function to process. Therefore, the missing genotypes were imputed with the `argmax` function in R/`qtl`. Also, the `mqmscan` function is not designed to process as many as 6504 genetic markers. Only a subset of 1221 markers was included in MQM-2 (Figure 3.7A). The process of collapsing the genetic map involved selecting markers which were closest to every whole number of cM position from 0 cM to the end of the chromosome. Furthermore, the `mqmscan` function requires a set of initial cofactors that will be subjected to backward elimination process in the model fitting step, but different sets of initial cofactors result in very different MQM results, hence the set of initial cofactors requires a careful selection. In this study, a set of 100 cofactors was chosen on the probability that the cofactors would most likely remain in the final MQM-2 model after the backward elimination steps (Figure 3.7B). The probability for each marker being selected was calculated by running 10 000 iterations of MQM process. In each iteration, a set of initial cofactors were randomly selected by the `mqmautocofactors` function with a different seed for the pseudo-random number generator. The 100 cofactors with the highest probability, given that there were no other markers with a higher probability within 5 cM distance, were selected. The number of initial markers ($k = 100$) was chosen within the theoretical limit of $k < (n - 20 = 153)$ (D. Arends, personal communication). This set of 100 cofactors, which were located on all twelve chromosomes (Figure 3.7C), was used as the initial set of cofactors in the final MQM process via the `mqmscan` function. The settings for `mqmscan` function included picking the threshold for the inclusion of markers as cofactors to be at 0.02 (Jansen, 1994, 2007), moving the position under test to every 1 cM along the length of each chromosome and dropping cofactors when they were within 25 cM of the position under test (default settings).

3.3 Results

3.3.1 Phenotypes of RILs from Screen 1

The biomass of *S. asiatica* on most RILs was between that of the more susceptible parental genotype, IR64, and that of the more resistant parental genotype, CT8556-37-2-3-1-M. However, some genotypes were more susceptible and some more resistant than the parental genotypes, indicative of transgressive segregation (Figure 3.8). RIL number 107 and 253 had *S. asiatica* biomass higher than IR64, whilst RIL number 144, 69, 83 and 6 had *S. asiatica* biomass lower than CT8556-37-2-3-1-M. The biomass of *S. asiatica* of all RILs had a positive-skewed unimodal distribution where the mode was situated between the means of the two parental genotypes, much closer to CT8556-37-2-3-1-M. The distribution of the trait showed that the resistance was highly quantitative and thus likely to be polygenic in nature.

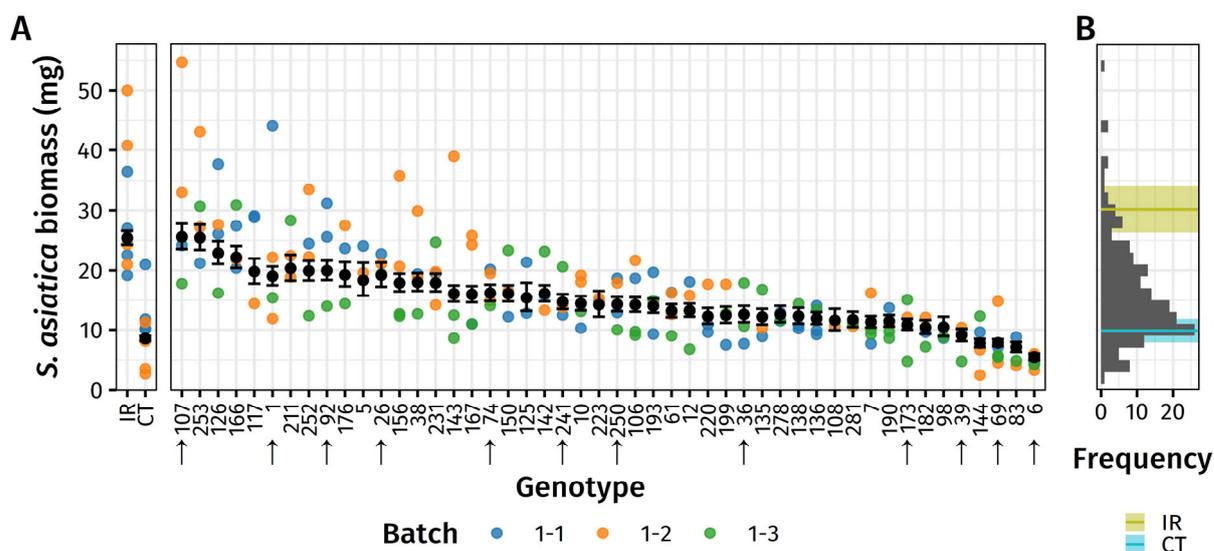


Figure 3.8 Biomass of *Striga asiatica* per rice plant on 48 randomly selected recombinant inbred lines (RILs), derived from CT8556-37-2-3-1-M × IR64, in Screen 1. (A) The left panel shows the biomass (dry weight) of *S. asiatica* per host plant on the parental genotypes. In the right panel, the RILs were ranked from the most susceptible to the most resistant. Differently coloured symbols represent *S. asiatica* biomass measured on RIL replicates in different batches (blue = Batch 1-1; orange = Batch 1-2; green = Batch 1-3). Each black symbol represents the back-transformed conditional mean of *S. asiatica* biomass for each genotype from the linear mixed-model (error bar = mean ± standard error). Genotype names represent the line number in the CT22591-3 RIL population from CIAT. RILs marked with arrows (↑) were rescored in Screen 2. (B) The frequency histogram of the *S. asiatica* biomass of the RILs was plotted against the mean (horizontal line) and the standard error (shaded area) of the parental genotypes. CT = CT8556-37-2-3-1-M (teal); IR = IR64 (lime).

To determine whether carrying out the phenotyping screen in a series of batches affected the phenotypic scores, a repeatability analysis was carried out (Table 3.2). The genetic variation between RILs (the genotype) significantly explained 49.8% of the overall variance of *S. asiatica* biomass. This is larger than the inter-batch variation, which contributed only 3.9% to the total variance (although this was significant). Therefore, the conditional mean of each RIL, which took the inter-batch variation into account, was used as the phenotypic input in subsequent QTL analyses.

Table 3.2 Summary of repeatability of the phenotypic scores from Screen 1.

Trait	Factor	Repeatability	95% confidence interval*	P-value†
<i>S. asiatica</i> biomass	Genotype	0.498	[0.329, 0.631]	< 0.001
	Batch	0.039	[0.000, 0.149]	0.005

* 95% confidence intervals were calculated from parametric bootstrap (10 000 iterations).

† P-values were calculated from a permutation test (10 000 permutations).

3.3.2 Simple interval mapping (SIM) of phenotypic data from Screen 1 revealed one significant QTL on chromosome 1.

There was one significant QTL (genome-wide significance, $P < 0.05$) for resistance to *S. asiatica* on chromosome 1, using the biomass of parasites as the phenotypic score (Figure 3.9). The QTL had a LOD score of 4.08 (genome-wide significance, $P = 0.01$) at the 143.8 cM position. There were also other non-significant LOD peaks, such as a second peak on chromosome 1 and a peak on chromosomes 3 and 11 (Figure 3.9). These non-significant peaks could potentially contribute to the control of resistance to *S. asiatica* but failed to be detected due to small sample size (a small number of RILs). Therefore, Screen 2 was designed to include all available RILs to increase the detection power of QTL.

3.3.3 Phenotypes of RILs in the combined dataset

The *S. asiatica* biomass of the RILs again ranged from more susceptible than IR64 to more resistant than CT8556-37-2-3-1-M (Figure 3.10), with the majority of RILs exhibiting the levels of resistance/susceptibility in between those of the two parental genotypes. Two RILs

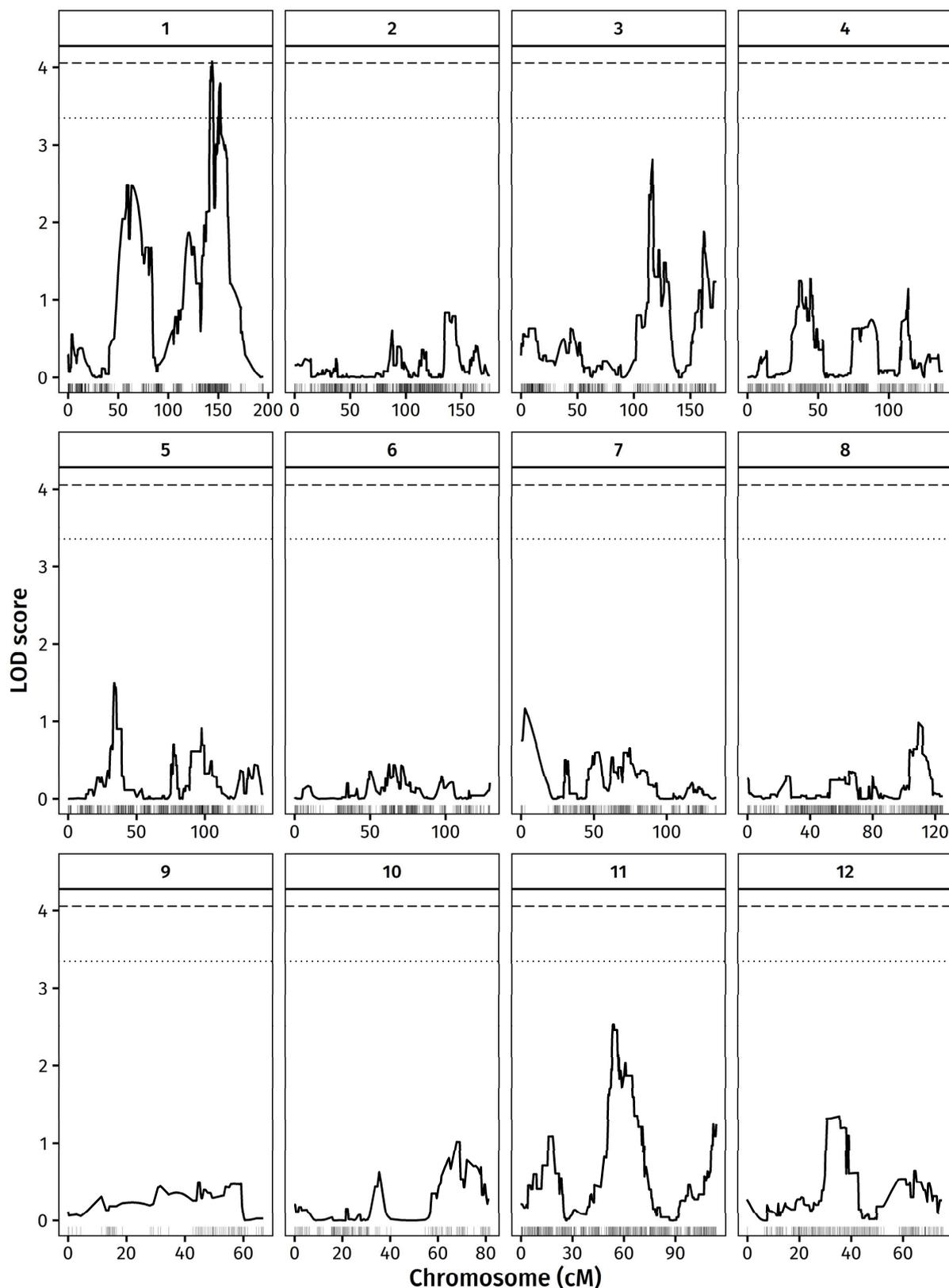


Figure 3.9 Simple interval mapping (SIM) of quantitative trait loci (QTL) for *Striga asiatica*-resistance in Screen 1. The logarithm of the odds ratios (LOD) scores were generated by a one-dimensional scan in the R/qtl software, using extended Haley–Knott regression method. Genome-wide significance levels (dotted line at $\alpha = 0.05$ and dashed line at $\alpha = 0.01$) were calculated using a permutation test (10 000 permutations). Vertical lines at the bottom of the plot mark positions of markers.

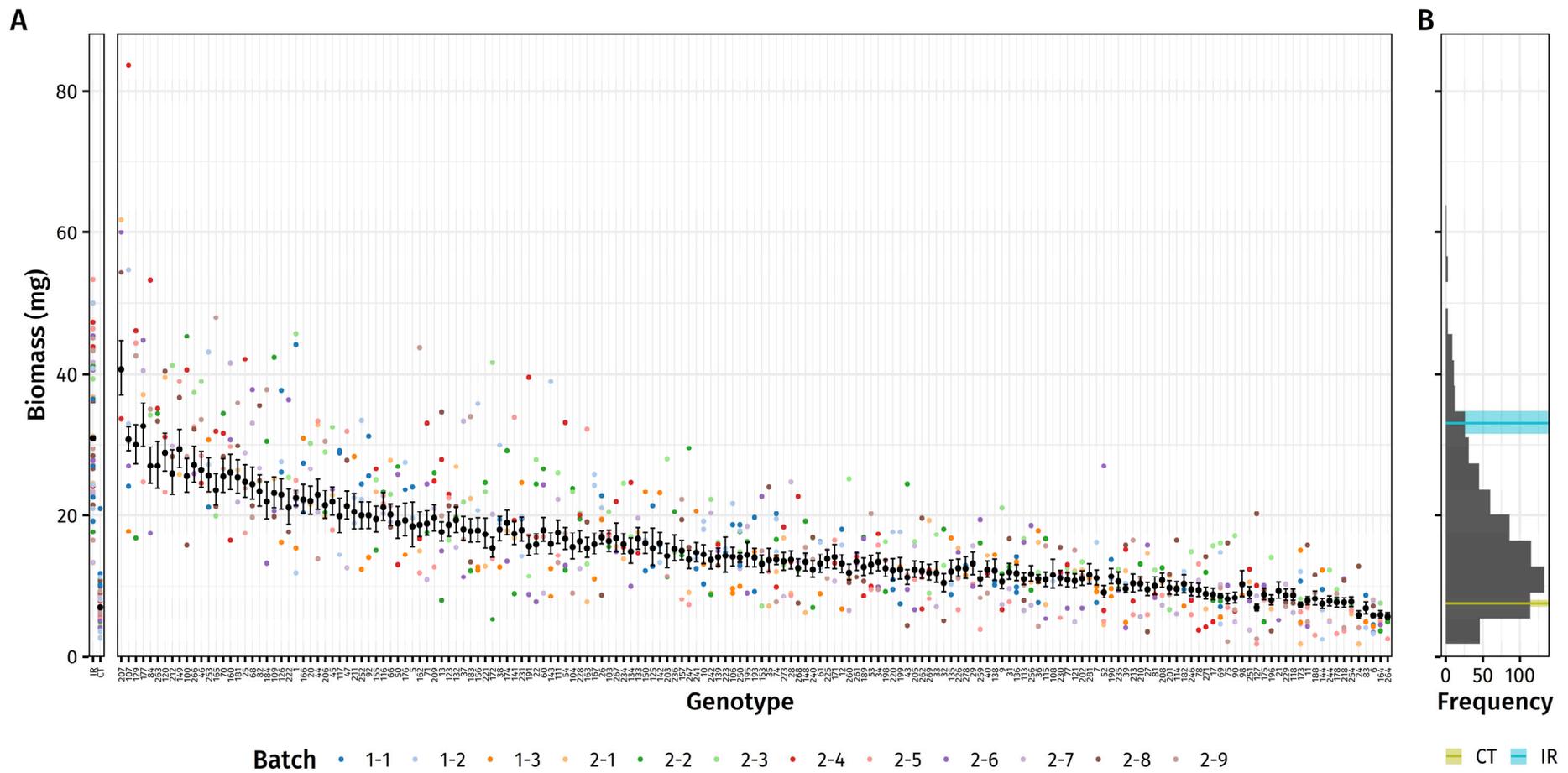


Figure 3.10 The biomass of *Striga asiatica* per host plant of a population of recombinant inbred lines (RILs) derived from CT8556-37-2-3-1-M \times IR64. (A), The left panel shows the measurements of the parental genotypes. In the right panel, the RILs were ranked from the most susceptible to the most resistance. Differently coloured symbols represent measurements from different batches in which host plants were grown. Black symbols represent the back-transformed conditional means of the random effects of genotypes from the linear mixed-effect model (error bar = mean \pm standard error). Genotype names are the line number in CIAT's CT22591-3 population. The frequency histogram (B) of the *S. asiatica* biomass of the RILs was plotted against the mean (horizontal line) and the standard error (shaded area) of the parental genotypes. CT = CT8556-37-2-3-1-M (lime); IR = IR64 (teal).

(number 207 and 177) had a higher *S. asiatica* biomass than IR64, whilst six RILs (number 24, 83, 6, 164 and 264) had a lower *S. asiatica* biomass than CT8556-37-2-3-1-M. The distribution of *S. asiatica* biomass of the RILs in the combined dataset had the same distribution as that of the data from Screen 1, with one mode located between the means of the parental genotypes. The repeatability analysis showed that genotype significantly explained 52.0% of the overall variance, whilst the inter-batch variation significantly accounted for only 3.0% of the overall variance (Table 3.3). Therefore, the phenotypic input for subsequent QTL analyses was the conditional means of the RILs, which account for the inter-batch variation.

Table 3.3 Summary of repeatability of the phenotypic scores from the combined dataset

Trait	Factor	Repeatability	95% confidence interval *	P-value †
<i>S. asiatica</i> biomass	Genotype	0.520	[0.440, 0.589]	< 0.001
	Batch	0.030	[0.005, 0.068]	< 0.001

* 95% confidence intervals were calculated from parametric bootstrap (10 000 iterations).

† P-values were calculated from a permutation test (10 000 permutations).

3.3.4 QTL analyses of combined dataset detected many QTL underlying resistance to *Striga asiatica*.

The combined dataset was first analysed using SIM. Three significant main LOD score peaks ($P < 0.01$), at 142.0 cM on chromosome 1 (*qSaB1.1*; QTL for *Striga asiatica* biomass on chromosome 1, number 1), 90.0 cM on chromosome 5 (*qSaB5.1*) and 62.4 cM on chromosome 11 (*qSaB11.1*), were detected via SIM (Figure 3.11). Additionally, there was one significant secondary peak ($P < 0.01$) on chromosome 1 at 124.0 cM (*qSaB1.2*) (marked with an arrow in Figure 3.11). For all the QTL, the allele associated with *S. asiatica*-resistance came from CT8556-37-2-3-1-M.

The evaluation of *qSaB1.2* was conducted by taking spurious linkage into account. The expected LOD curves did not show any indication of spurious linkage between the locus of the secondary peak and the locus of the primary peak on chromosome 1 as the $E_p(Z_M)$ curve was consistently below the $E_d(Z_M)$ curve. When the observed LOD curve was compared to

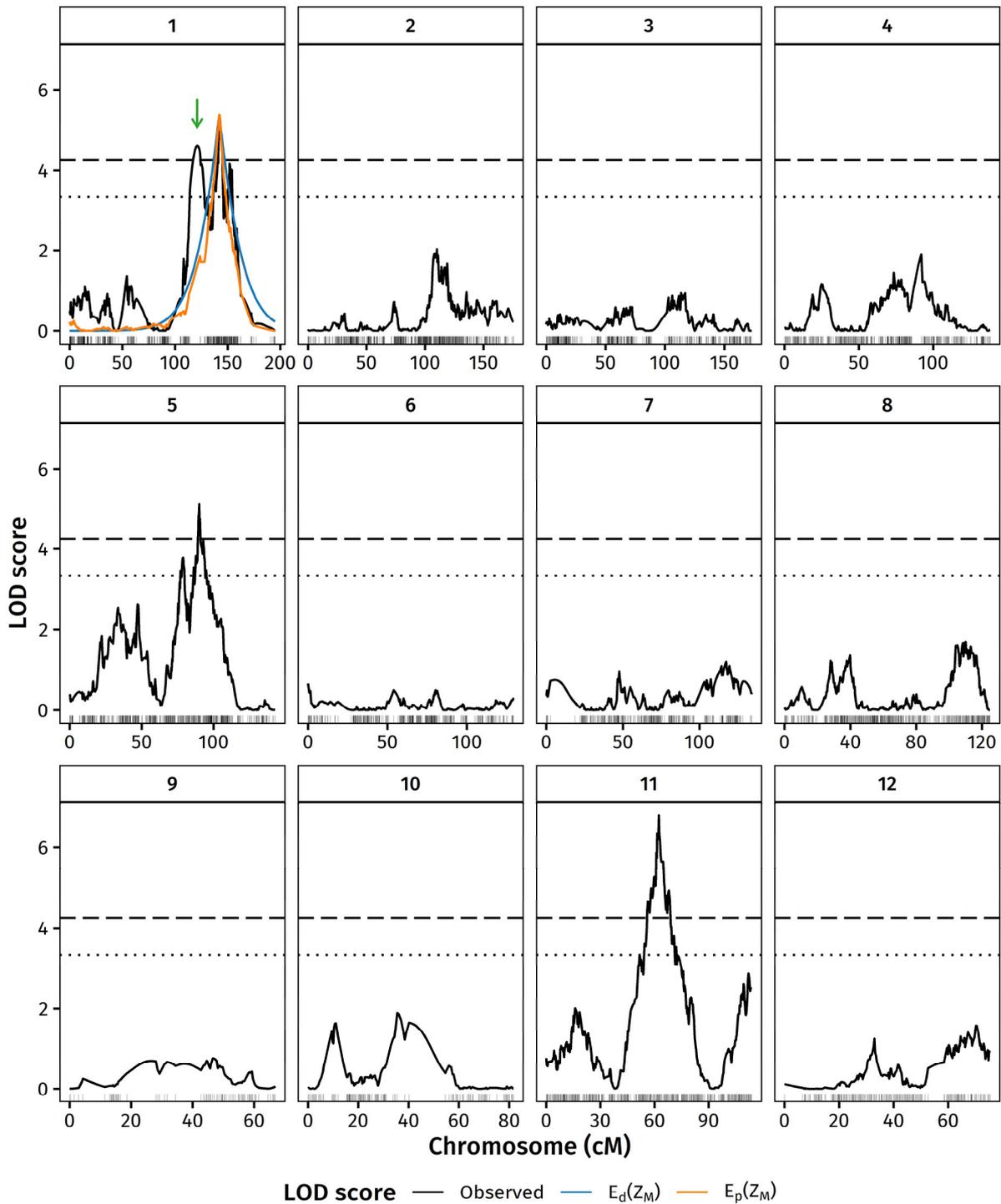


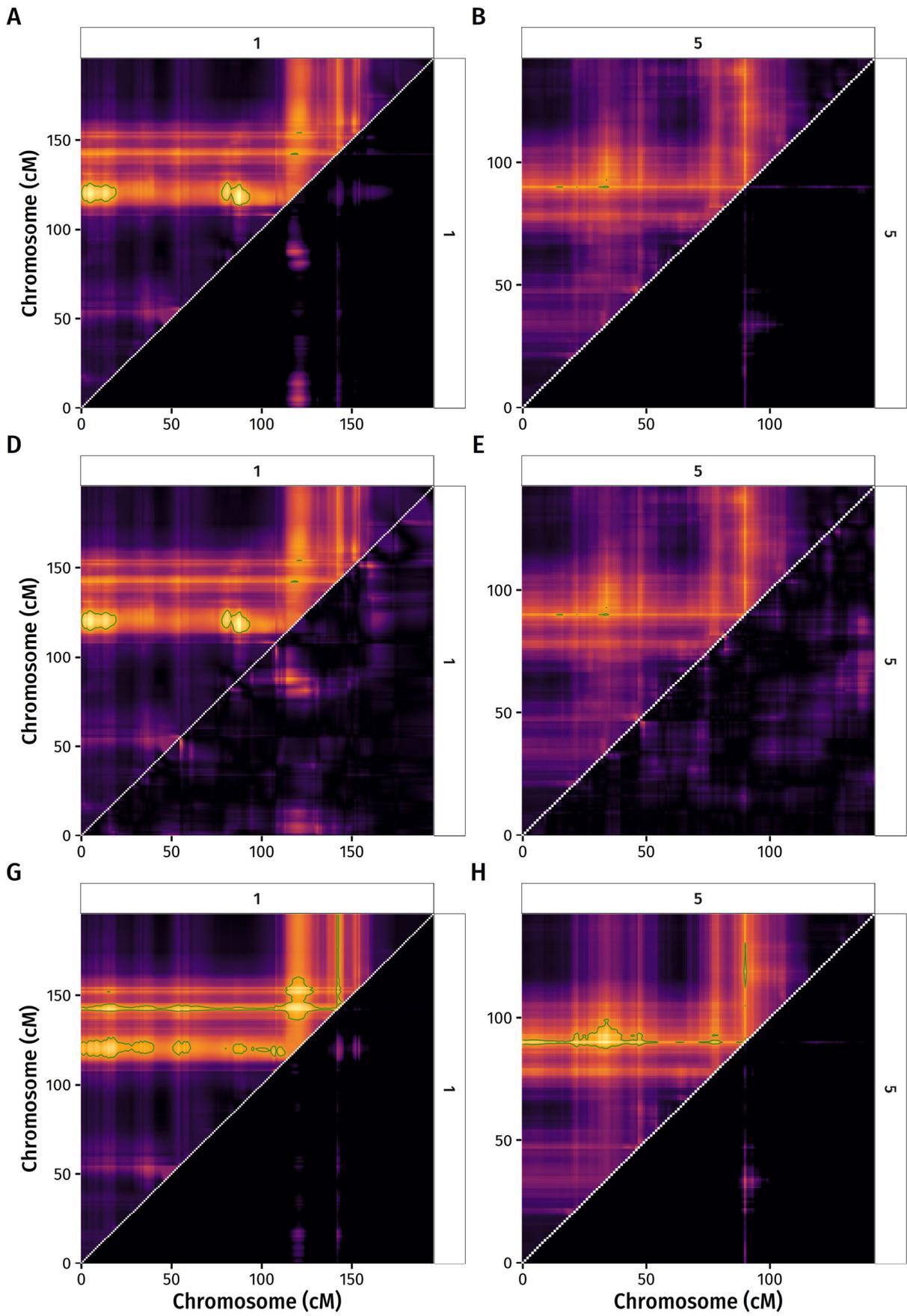
Figure 3.11 Simple interval mapping (SIM) of quantitative trait loci (QTL) for *Striga asiatica*-resistance in the combined the dataset. The logarithm of the odds ratios (LOD) scores were generated by a one-dimensional scan in the R/qt l software, using extended Haley–Knott regression method (black solid line). Genome-wide significance levels (dotted line at $\alpha = 0.05$ and dashed line at $\alpha = 0.01$) were calculated using a permutation test (10 000 permutations). Vertical lines below x-axis mark positions of markers. Green arrow marks a secondary LOD score peak under test. The expected LOD score $E_d(Z_M)$ in blue was calculated from the recombination fractions derived from the genetic distance (\hat{r}_d). The expected LOD scores $E_p(Z_M)$ in orange was calculated from the observed pairwise recombination fractions (\hat{r}_p) between markers and the marker at the LOD peak.

the expected LOD curves, the observed LOD curve did not follow the shape of expected LOD curves at the secondary peak locus, indicating that these observed secondary LOD peak did not arise from unexpected recombination and was likely to be a genuine QTL peak. In conclusion, four significant QTL for resistance to *S. asiatica* containing CT8556-37-2-3-1-M genes were identified by SIM.

In addition to SIM, the data were analysed using MQM (approaches 1 and 2), ICIM and MIM to take into account the interactions between QTL and to improve the detection power for QTL with smaller effects. The first MQM approach (MQM-1) involved building a QTL model using a two-dimensional scan. A two-dimensional scan consists of two steps. First SIM was performed to locate the primary QTL (first-dimensional scan). A second scan (on the second dimension), was performed to determine whether a secondary QTL was present. The presence/absence of epistatic interactions was also evaluated in the second scan.

Within each chromosome, MQM-1 suggested that there was weak evidence for a secondary QTL on chromosome 11 at 108 cM ($P = 0.051$), which had no significant interaction with *qSaB11.1* ($P > 0.1$) (Figure 3.12). MQM-1 also showed that there was only one QTL on chromosomes 1 and 5 (*qSaB1.1* and *qSaB5.1*) because any additions of a secondary QTL on either of the two chromosomes to the QTL model did not significantly improve the model fitting ($P > 0.1$), whether epistatic interactions were allowed or not (Figure 3.12). For the addition of *qSaB1.2* given the presence of *qSaB1.1*, the LOD scores of model comparisons were: $Z_{full} = 6.26, P = 0.165$; $Z_{full-1} = 0.92, P > 0.999$; $Z_{interaction} < 0.01, P > 0.999$; $Z_{Additive} = 6.25, P = 0.017$; and $Z_{Additive-1} = 0.92, P > 0.999$.

To assess whether there were any epistatic interactions among *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, pairs of loci on different chromosomes were subjected to a two-dimensional scan. There was stronger evidence for an additive model that lacks epistatic interactions between the three main QTL ($P_{additive} < 0.01$) than that for a model where the three main QTL interact ($P_{interactive} < 0.05$; Table 3.4).



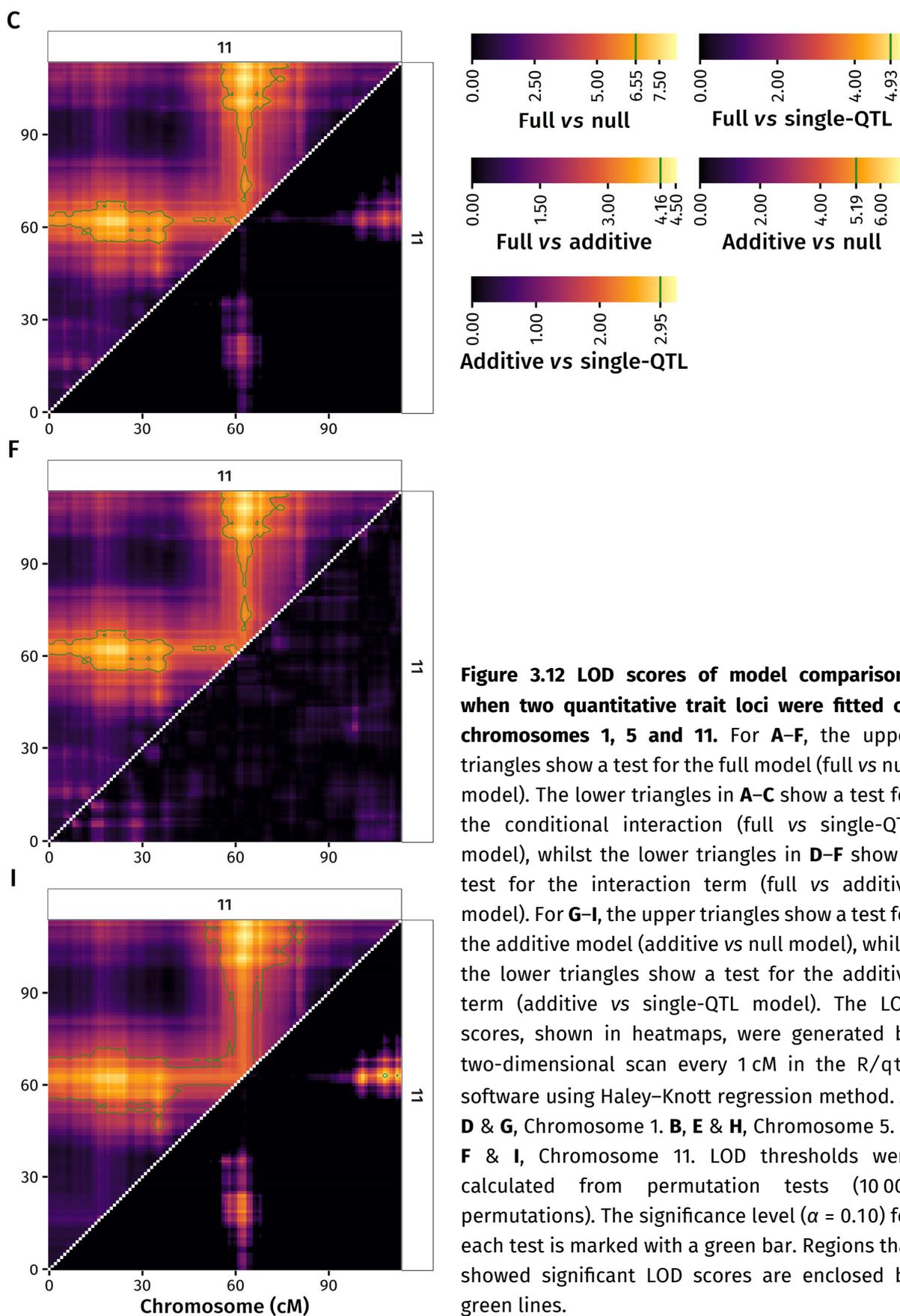


Table 3.4 Summary of all significant results from the two-dimensional scan, whether epistatic interactions between loci were significant or not, using MQM-1.

Locus A		Locus B		LOD score from comparisons of models				
Chr.	Position (cM)	Chr.	Position (cM)	Full vs null	Full vs single	Full vs additive	Additive vs null	Additive vs single
1	142	5	90	10.74 ^{***}	5.40 [*]	0.86 ^{ns}	9.88 ^{***}	4.54 ^{**}
1	121	11	62	11.97 ^{***}	5.72 [*]	0.32 ^{ns}	11.65 ^{***}	5.40 ^{***}
5	90	11	63	11.40 ^{***}	5.15 [†]	0.16 ^{ns}	11.24 ^{***}	4.99 ^{**}
11	63	11	108	9.55 ^{***}	3.30 ^{ns}	0.11 ^{ns}	9.44 ^{***}	3.19 [†]
11	19	12	31	7.22 [*]	0.97 ^{ns}	4.83 [*]	2.40 ^{ns}	< 0.01 ^{ns}

LOD scores in red indicate a case when the presence of QTL pair was statistically significant ($P < 0.05$). Significance levels of LOD score differences were calculated by permutation test (10 000 permutations). ^{***} $P < 0.001$; ^{**} $P < 0.01$; ^{*} $P < 0.05$; [†] $P < 0.1$; ^{ns} non-significance ($P \geq 0.1$). Chr., chromosome.

Besides, the inclusion of a pair of QTL, one at 19 cM on chromosome 11 and another at 31 cM on chromosome 12, with an epistatic interaction significantly improved the model ($P < 0.05$; Table 3.4). In this epistatic interaction, the resistance to *S. asiatica* was associated with the CT8556-37-2-3-1-M allele at the 31 cM position on chromosome 12 when the IR64 allele is present at the 19 cM locus on chromosome 11 (Figure 3.13).

In the model selection step, six QTL and one epistatic interaction were fitted to an initial model (Table 3.5). A stepwise selection algorithm was then applied to the model. The final QTL model from MQM-1 had four QTL at 142.4 cM on chromosome 1, at 78.0 cM on chromosome 5, at 62.4 and 109.4 cM on chromosome 11 (Table 3.5). The epistatic

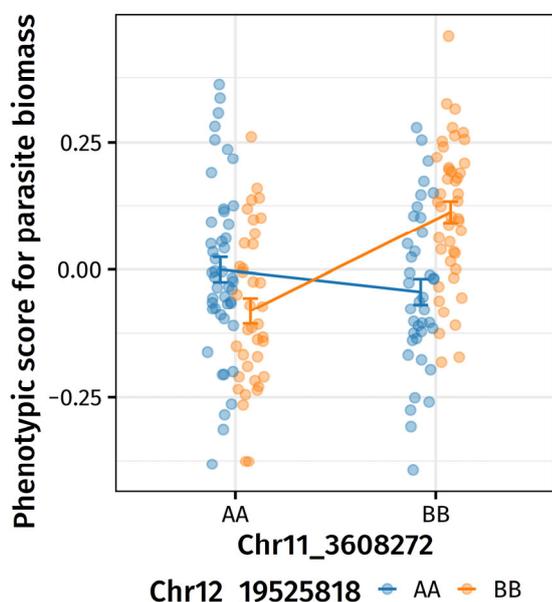


Figure 3.13 Epistatic interaction between two loci on chromosome 11 at 19 cM (Chr11_3608272 maker) and chromosome 12 at 31 cM (Chr12_19525818 marker).

The phenotypic score is the conditional mean of log-transformed *S. asiatica* biomass of each RIL from the general linear mixed-effect model. The more negative the score, the higher the resistance level. Each dot represents the phenotypic score each RIL with corresponding genotypes. Alleles of Chr11_3608272 are shown on the x-axis. Different colours show different alleles of Chr12_19525818. Colour lines connect mean phenotypic scores. Error bars represent one standard error. The A alleles (blue) came from IR64, whilst the B alleles (orange) came from CT8556-37-2-3-1-M.

interaction was discarded because the main effects of the two loci (Q5 and Q6 in Table 3.5) were non-significant ($P > 0.01$). All except the secondary QTL on chromosome 11 (at 109.4 cM) had CT8556-37-2-3-1-M alleles associated with *S. asiatica* resistance.

Table 3.5 Locations of putative quantitative trait loci (QTL) identified from the two-dimensional scan in MQM-1, and the results of QTL model selection procedure using a stepwise search algorithm.

Putative QTL	Chromosome	Initial location (cM)	Was the putative QTL kept by a stepwise search algorithm?	Final location (cM)
Q1	1	142	Yes	142.0
Q2	5	90	Yes	78.0
Q3	11	62	Yes	62.4
Q4	11	108	Yes	109.4
Q5	11	19	No	—
Q6	12	31	No	—
Q5 × Q6	—	—	No	—

Another MQM approach (MQM-2) involves the automation of model building process using backward elimination of an initial set of cofactors. Within the final MQM-2 QTL model, a total of 25 markers were used as cofactors. MQM-2 identified more than ten LOD score peaks that were above the threshold of the genome-wide significance level of 0.01 (Figure 3.14). Not all these peaks were likely real QTL because the observed LOD scores were very high given the repeatability of the genotypes and the population size. Although the number of cofactors was within the theoretical limit of 26, too many cofactors could have inflated the LOD score. Strangely, the permutation test did not produce a higher significance threshold. Four peaks were clearly above the threshold and were most likely to be real QTL. These included one peak on chromosome 1 (at 142 cM, with a shoulder peak at 124 cM), one peak on chromosome 5 (at 93 cM, with a shoulder peak at 78 cM) and two peaks on chromosome 11 (at 60 and 100 cM). Out of these four QTL, only one at 100 cM on chromosome 11 had the *S. asiatica*-resistance allele coming from IR64. The rest had CT8556-37-2-3-1-M alleles associated with the resistance.

In addition to MQM, another two independent QTL mapping methods, ICIM and MIM, were also performed on the combined dataset of Screen 1 and 2 to explore the possibility of

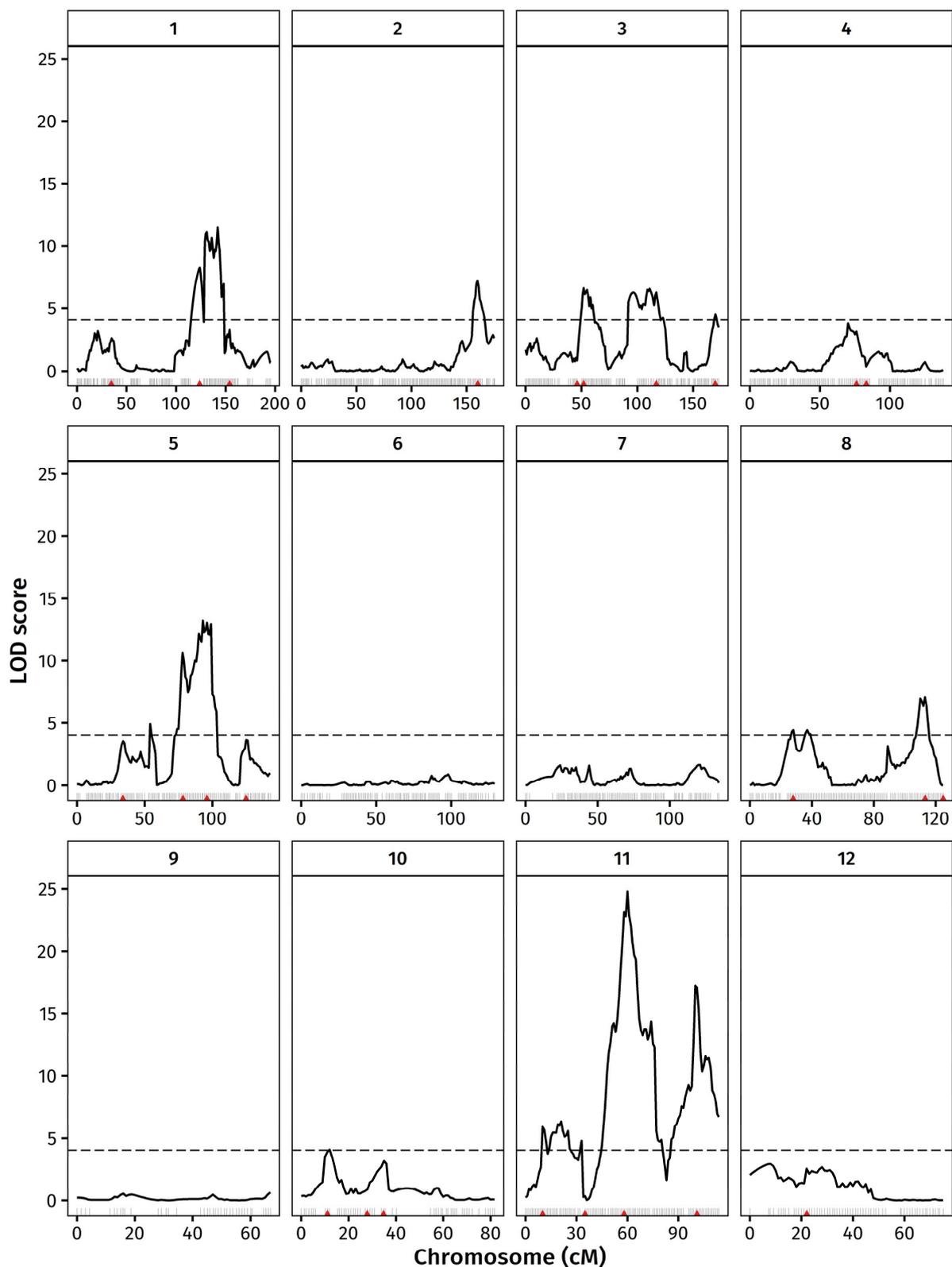


Figure 3.14 Multiple QTL mapping (MQM) of quantitative trait loci (QTL) for *Striga asiatica*-resistance in the combined dataset. The logarithm of the odds ratios (LOD) scores were generated by the `mqmscan` function in the R/`qtl` software (solid line). Genome-wide significance level at $\alpha = 0.01$ (dashed line) was calculated using a permutation test (10 000 permutations). Vertical lines below the x-axis show the positions of SNP markers. Red triangles mark the positions of 27 cofactors used in the final QTL model.

detecting additional QTL of smaller effects. ICIM was designed to take epistatic interactions into account and control for the loci with major effects, to improve the detection power of loci with smaller effects. Four significant QTL at 142.4 cM on chromosome 1, at 78.0 cM on chromosome 5 and at 62.4 and 109.4 cM on chromosome 11 ($P < 0.01$) were detected by ICIM (Figure 3.15). CT8556-37-2-3-1-M contributed the alleles associated with *S. asiatica* resistance for all except one QTL at 109.4 cM on chromosome 11, where IR64 contributed the allele for *S. asiatica*-resistance. There was also one suggestive QTL ($0.01 \leq P < 0.05$) at 15.1 cM on chromosome 11 where IR64 contributed the allele associated with *S. asiatica* resistance. In the ICIM model, five markers at these QTL were fitted as covariates (cofactors).

MIM employs a QTL model that assumes the presence of multiple QTL, which is appropriate in this study as *S. asiatica*-resistance was polygenic as seen from various QTL mapping methods. Unexpectedly, MIM revealed only one significant QTL ($P < 0.01$) on chromosome 11 at 62.4 cM (Figure 3.16). There was one suggestive QTL on chromosome 1 at 144.0 cM position ($0.01 \leq P < 0.05$). The alleles associated with *S. asiatica*-resistance came from CT8556-37-2-3-1-M in both cases. A total of three cofactors were fitted in the model on chromosomes 1, 5 and 11 (Figure 3.16).

Table 3.6 summarises all QTL detected using different mapping methods. Different QTL mapping methods consistently revealed three QTL with the largest effect sizes on chromosomes 1, 5 and 11 (*qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively), one shoulder peak on chromosome 1 (*qSaB1.2*), and one QTL with a smaller effect on chromosome 11 (*qSaB11.2*) (Table 3.6). The locations of *qSaB1.1* and *qSaB11.1* from different mapping methods were in good agreements, at 142.0 and 62.4 cM, respectively. There was a disagreement on the position of *qSaB5.1*. SIM placed this QTL at 90.0 cM, whilst ICIM and MQM-1 positioned it at 78.0 cM. MQM-2 predicted the main peak at 93 cM and a shoulder peak at 78 cM. The QTL from SIM had a larger effect size than those from other methods. *qSaB5.1* and *qSaB5.2* were designated to the QTL at 90 cM and 78 cM, respectively.

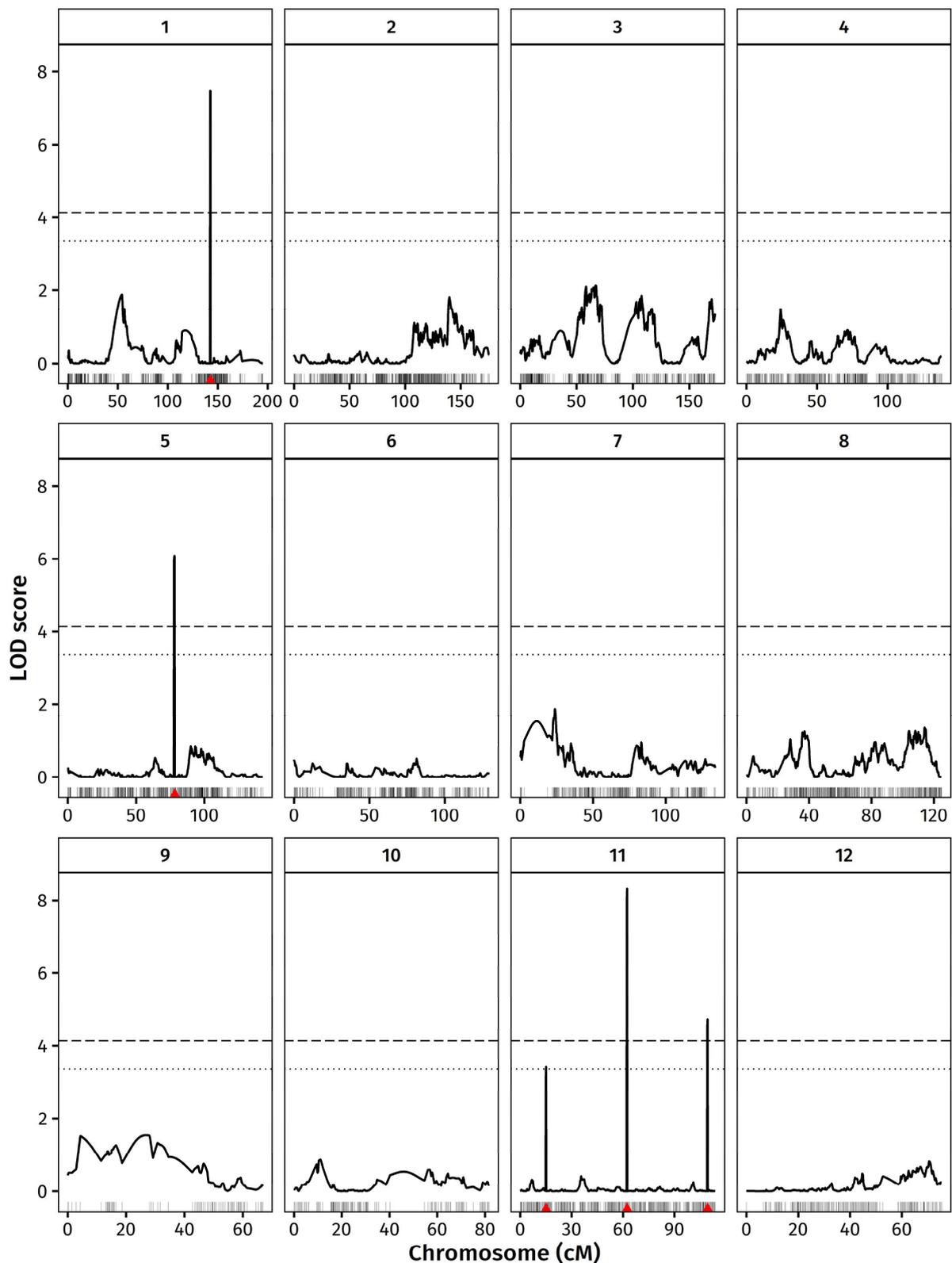


Figure 3.15 Inclusive composite interval mapping (ICIM) of quantitative trait loci (QTL) for *Striga asiatica*-resistance in the combined dataset. The logarithm of the odds ratios (LOD) scores were generated by ICIM with additive effects in the QTL IciMapping software (solid line). Genome-wide significance levels (dotted line at $\alpha = 0.05$ and dashed line at $\alpha = 0.01$) were calculated using a permutation test (10 000 permutations). Vertical lines below the x-axis show the positions of SNP markers. Red triangles mark the positions of five cofactors in the QTL model.

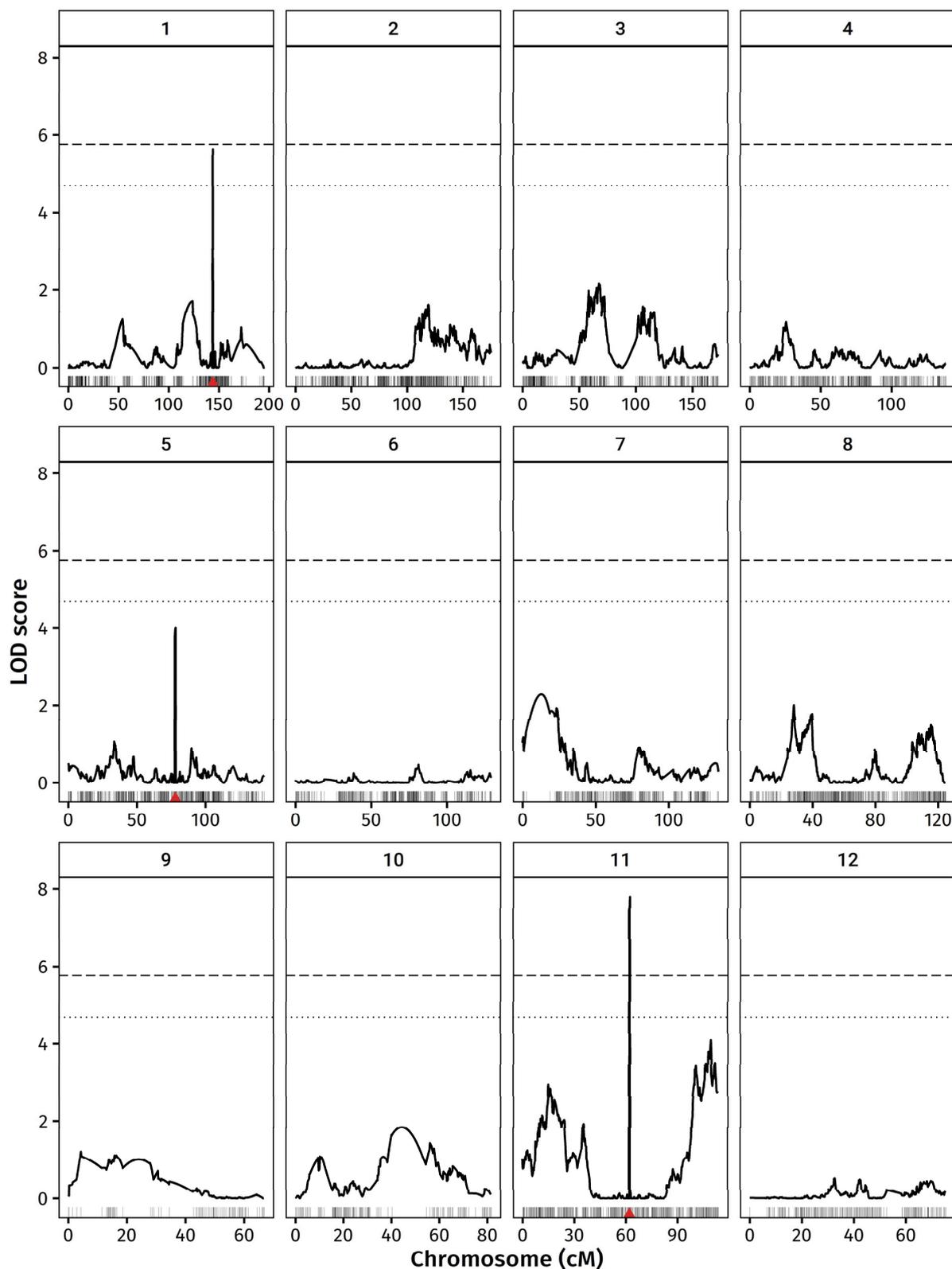


Figure 3.16 Multiple interval mapping (MIM) of quantitative trait loci (QTL) for *Striga asiatica*-resistance in the combined dataset. The logarithm of the odds ratios (LOD) scores were generated by single-trait MIM in the QGene software (solid line). Genome-wide significance levels (dotted line at $\alpha = 0.05$ and dashed line at $\alpha = 0.01$) were calculated using a permutation test (10 000 permutations). Vertical lines below the x-axis show the positions of SNP markers. Red triangles mark the positions of three cofactors in the QTL model.

Table 3.6 Summary of results from different methods of mapping quantitative trait loci (QTL) for the slow growth resistance to *Striga asiatica*.

QTL	Method	Chr.	Position (cM)	LOD	P-value*	Additive effect (× pheno. SD) [†]	PVE [‡]	LOD support interval (cM) [§]	Left marker	Peak marker	Right marker
qSaB1.1	SIM	1	142.0	5.44	< 0.001	+0.062 (+0.33)	13.5	[139.8, 154.0]	35 098 668	35 217 650	38 082 498
	ICIM	1	142.4	7.49	< 0.001	+0.057 (+0.30)	12.4	[142.4, 142.7]	35 373 206	35 373 206	35 424 814
	MQM-1	1	142.0	6.88	< 0.01	+0.057 (+0.30)	11.3	[137.0, 145.0]	34 372 239	35 217 650	35 802 756
	MQM-2	1	142.0	11.48	< 0.001	—	—	[129.0, 143.1]	32 261 997	35 237 157	35 478 304
qSaB1.2	SIM	1	121.0	4.63	0.004	+0.058 (+0.30)	11.6	[114.0, 128.0]	28 757 754	32 071 374	32 157 735
	MQM-2	1	124.0	8.26	< 0.001	—	—	[120.0, 125.0]	28 945 062	32 071 374	32 103 038
qSaB5.1	SIM	5	90.0	5.12	0.001	+0.060 (+0.32)	12.7	[87.6, 93.7]	21 276 428	22 190 668	22 653 177
	MQM-2	5	93.0	13.26	< 0.001	—	—	[90.0, 99.0]	22 182 010	22 522 049	23 404 664
qSaB5.2	ICIM	5	78.0	6.09	< 0.001	+0.051 (+0.27)	10.0	[77.7, 78.2]	20 130 013	20 146 589	20 146 589
	MQM-1	5	78.0	5.82	< 0.01	+0.052 (+0.28)	9.4	[76.4, 91.1]	19 951 939	20 146 589	22 282 726
	MQM-2	5	78.0	10.66	< 0.001	—	—	[78.0, 79.1]	20 130 013	20 146 589	20 362 194
qSaB11.1	SIM	11	62.4	6.79	< 0.001	+0.069 (+0.36)	16.5	[61.0, 65.0]	17 930 831	18 455 653	19 281 255
	ICIM	11	62.4	8.31	< 0.001	+0.061 (+0.32)	13.9	[62.2, 62.4]	18 452 762	18 455 653	18 455 653
	MIM	11	62.4	7.80	< 0.01	+0.063 (+0.33)	17.5	[62.1, 62.4]	18 361 595	18 684 231	18 684 231
	MQM-1	11	62.4	8.08	< 0.01	+0.063 (+0.33)	13.5	[61.0, 68.5]	17 930 831	18 455 653	19 937 299
	MQM-2	11	61.0	24.80	< 0.001	—	—	[60.0, 60.1]	17 907 712	17 924 485	17 924 485
qSaB11.2	ICIM	11	109.4	4.73	0.003	-0.046 (-0.24)	7.5	[109.2, 109.4]	27 674 570	27 677 288	27 677 288
	MQM-1	11	109.4	4.78	< 0.01	-0.048 (-0.25)	7.6	[100.3, 113.9]	24 800 811	27 677 288	29 008 614
	MQM-2	11	101.0	17.23	< 0.001	—	—	[100.0, 101.1]	24 798 240	24 800 811	24 827 997

*P-value for genome-wide significance level was calculated by permutation test (10 000 permutations). [†]A positive additive effect means the CT8556-37-2-3-1-M allele is associated with *S. asiatica* resistance, whilst a negative additive effect means the IR64 allele is associated with the resistance. The effect size was measured in standard deviations where the phenotypic standard deviation between different genotypes was 0.1897. [‡]Percentage variance explained (PVE) of a QTL in each mapping method. [§]LOD support interval was the interval that supports a 1.5-LOD score decrease from the QTL peak. Chr., chromosome; qSaBi.j = jth QTL on the ith chromosome of rice for *Striga asiatica* biomass (dry weight). SIM, simple interval mapping. ICIM, inclusive composite interval mapping. MQM, multiple QTL mapping. MIM, multiple interval mapping.

3.4 Discussion

This chapter aimed to establish the genetic basis of the resistance to *S. asiatica* on CT8556-37-2-3-1-M that was characterised in Chapter 2. A QTL mapping approach using a RIL population derived from a cross between two parental genotypes (IR64 × CT8556-37-2-3-1-M) that exhibited different levels of resistance/susceptibility when infected with *S. asiatica* was taken. Using the conditional means of the log-transformed parasite biomass, the resistance to *S. asiatica* was found to be polygenic, consistent with the continuous variation, with the unimodal distribution, of the phenotypes of the RILs. Six significant QTL for the resistance were discovered, using several QTL mapping methods (Table 3.6). The QTL, which acted additively, each explained 7.5–17.5 % of the phenotypic variance in the mapping population. All except one QTL (*qSaB11.2*) had the alleles for resistance originating from CT8556-37-2-3-1-M. Combinations of the favourable IR64 allele (at *qSaB11.2*) and other CT8556-37-2-3-1-M alleles could explain the transgressive segregation seen in some RILs, such as RIL number 6 and 264, where their parasite biomass was lower than that of CT8556-37-2-3-1-M.

3.4.1 How do these QTL for *Striga asiatica*-resistance compare to QTL for *Striga*-resistance from other studies?

This study was the first to map genetic loci associated with post-attachment resistance to *S. asiatica* in rice. Most work published so far on genetic mapping of both pre- and post-attachment resistance has been on resistance to *S. hermonthica*. The resistance to *S. asiatica* in this study was identified to be quantitative and polygenic, similar to other resistance against *S. hermonthica* in sorghum and rice. In sorghum, Hausmann *et al.* (2004) identified nine and eleven QTL for resistance to *S. hermonthica* by screening two RIL populations in field trials at five different locations, in Mali and Kenya, across two planting seasons. The resistance phenotype in that study was likely a combination of pre- and post-attachment resistance, though the major locus underlying the resistance was the locus for low germination stimulant

production (*LGS*). Significant QTL × environmental effects were also identified (Hausmann *et al.*, 2004).

In rice, several QTL studies for resistance to *S. hermonthica*, carried out on various BIL and RIL populations using survival, biomass, number and emergence of *S. hermonthica* (Table 3.1), revealed that *S. hermonthica*-resistance in rice often is polygenic and controlled by several QTL with large and small effects. Some of the QTL for the biomass of *S. asiatica* identified in the current study were in the vicinity of the QTL for post-attachment resistance from previous studies on *S. hermonthica*. The physical locations of *qSaB1.2* (at 28.8–32.2 Mbp) and *qSaB5.1* (at 21.3–22.7 Mbp) respectively coincided with the positions of *qShR1.1* (at 31.4 Mbp; Table 3.1) and *qShR5.1* (at 22.6–24.5 Mbp; Table 3.1), which were the minor QTL for *S. hermonthica* survival ratio from Gurney *et al.* (2006). In their study, the resistance phenotype involved the inability of the parasites to penetrate the endodermis of the root of the host, resulting in the death of parasite as they could not access host water and nutrients through xylem connections (Gurney *et al.*, 2006). This was different from the phenotype in the current study in which parasites could form vascular connections with the host but grow more slowly afterwards on CT8556-37-2-3-1-M. However, some of these QTL for post-attachment resistance to *S. hermonthica* and the QTL in the current study could be controlled by the same set of genes.

The physical location of *qSaB1.2* (28.9–32.1 Mbp) identified in the current study also coincided with that of a major QTL for *S. hermonthica*-resistance in rice (*qShB1.1*, *qShN1.1* and *qShTFE1.1*; Table 3.1), later named *qSLB1.1*, at 29.3 Mbp (Cardoso *et al.*, 2014; Kaewchumnong & Price, 2008). The resistance allele of *qSLB1.1* from the rice genotype Bala contained a deletion of two genes involved in the biosynthesis of a strigolactone (a germination stimulant), resulting in lowered germination of parasites, hence decreasing the number of parasites (Cardoso *et al.*, 2014). It is possible that common genes were involved in both *qSLB1.1* and *qSaB1.2*, mapped here. Strigolactones can act as not only a germination stimulant but also a plant growth regulator (Omoarelojie *et al.*, 2019; Zwanenburg *et al.*,

2016), which can influence host susceptibility to *Striga* through cell wall remodelling (Louden, 2017). Louden (2017) found that rice mutants that had their strigolactone biosynthesis pathway disrupted became more susceptible to *S. hermonthica*, especially at the early stages of infection in which the parasite penetrates through the root cortex of the rice mutants at a faster rate than in that of the wild type hosts.

3.4.2 Differences and similarities between QTL mapping methods

The QTL mapping methods used in this study revealed different numbers of QTL in their respective QTL models: four QTL from SIM together with the fake-QTL evaluation; four QTL and one suggestive peak from ICIM; one QTL and one suggestive peak from MIM; four QTL from MQM-1; and 15 QTL (four were highly significant), with complex interactions among themselves, from MQM-2.

The differences in the number of QTL detected from different methods were expected, especially in a polygenic trait, because the epistatic interactions and residual error variance are incorporated differently from method to method (Jansen, 2007). In SIM, which was the simplest QTL mapping method in this study, loci were tested for the association to the trait independently without considering the interference from other loci (Lander & Botstein, 1989). The lack of interaction modelling in the SIM QTL model explained why SIM detected fewer QTL than the more sophisticated methods such as ICIM, MIM and MQM. For example, on chromosome 11 only *qSaB11.1* was detected by SIM, whilst other methods also identified an additional QTL (*qSaB11.2*) and a suggestive peak on the same chromosome. In these more complex mapping methods, the large effect of *qSaB11.1* was controlled by fitting the markers nearby as a covariate in the QTL models. Another advantage of incorporating interactions between loci in QTL model is an increase in detection power of QTL with marginal effects on other chromosomes (Jansen, 2007; Zeng, 1993). This was reflected in the 15 QTL detected by MQM-2, benefitting from 25 cofactors fitted. Many of these QTL were only present in the MQM-2 QTL model. Although as many as 25 cofactors were included in

the MQM-2 QTL model, much higher than would be seen in CIM or ICIM given the number of RILs ($n = 173$), they were in line with the recommended theoretical maximum number of cofactors for MQM at $2 \times \sqrt{n} = 2 \times \sqrt{173} \approx 26$ (Jansen, 2007).

The results on the secondary QTL on chromosome 1 (*qSaB1.2*) from the fake-QTL evaluation contradicted those from the two-dimensional scan in MQM-1. The fake-QTL evaluation concluded that the presence of a *qSaB1.2* was not a product of spurious linkage between markers (Figure 3.11), whilst the two-dimensional scan showed that the inclusion of *qSaB1.2* did not significantly improve the QTL model (Figure 3.12A, D & G). However, *qSaB1.2* also appeared in MQM-2 model as a shoulder peak of *qSaB1.1*.

ICIM and MIM have the advantage of focusing the LOD score peak down to specific loci because when the outside-of-interval markers, which are linked to the peak markers, are under a statistical test they are conditioned for by the peak marker and their statistical significance diminishes (Li *et al.*, 2007; Zeng, 1993, 1994). This was reflected in sizes of the 1.5-LOD support intervals, which were smaller than 1 cM in these two methods. The slim LOD support intervals would narrow down the number of genes within the regions. However, since the location of QTL from these mapping methods depend on the choice of cofactors, the selection of wrong cofactors may result in missing the causal genes from the QTL.

Despite the differences in the number of QTL identified, the QTL mapping methods yielded consistent positions of the QTL commonly identified by the different methods. Two QTL (*qSaB1.1* and *qSaB11.1*) were present in all QTL models (though only one QTL was significant in the MIM results), giving confidence that these QTL were genuine rather than statistical artefacts. Two QTL were identified on chromosome 5 (*qSaB5.1* and *qSaB5.2*) by different methods. Both were credible since *qSaB5.1* identified by SIM was unbiased as covariate selection and model search were not involved in the process of QTL model testing, whilst *qSaB5.2* mapped by ICIM and MQM took the effects of other QTL into account when the location of the QTL was estimated.

Different QTL methods estimated different magnitudes of the QTL effects. *qSaB11.1*, which had the largest additive effects in all mapping methods, explained 13.5–17.5 % of the variance of the RIL phenotypes, followed by *qSaB1.1* at 11.3–13.5 %, *qSaB5.1* at 12.7 %, *qSaB5.2* at 9.4–10.0 % and *qSaB11.2* at 7.5–7.6 %. The additive effects of these QTL were also compared to the phenotypic standard deviation within a genotype to assess the QTL magnitude (Gurney *et al.*, 2006; Falconer & Mackay, 1996). The estimate effects of all QTL identified were less than 0.36 times the phenotypic standard deviation, indicating that the underlying genes were likely to be minor genes with significant effects.

3.4.3 What might have affected the QTL analyses?

Segregation distortion loci (SDL) are characterised by a deviation of allelic frequencies from the expected Mendelian segregation ratio, which is 1:1 for a RIL population derived from a cross in this study. Whilst segregation distortion is always disadvantageous to the detection power of QTL with dominant effects, Xu (2008) found that segregation distortion can be either beneficial or detrimental to the power to detect QTL for their additive effects. Two theoretical studies of SDL with all possible combination of allelic frequencies under an assumption that SDL occurs randomly under a uniform distribution found that the effects of SDL on the detection power of QTL depend on the degree of dominance of the QTL (L. Zhang *et al.*, 2010; Xu, 2008). On the other hand, when the dominance effects of QTL cannot be estimated, such as in a RIL population which were used in this study, the power of detecting additive effects is highest when segregation distortion is absent (L. Zhang *et al.*, 2010). In other words, SDL is always detrimental to a QTL detection on a RIL population. From the genetic map data (Figure 3.6C), there were several loci with severe segregation distortion on chromosomes 7, 8 and 4, ranked from the most to the least distorted. As a result, an apparent absence of any QTL on chromosome 7 could arise from a loss in detection power. Incorporation of the segregation distortion information in a more sophisticated QTL mapping method could alleviate such problem (Xu, 2008), though currently unavailable in QTL mapping software.

The resistance to *S. asiatica* was discovered to be associated with several QTL. There were a few issues that complicated the process of dissecting the genetic basis of the complex trait. Firstly, type I error (false positive) rate of QTL analyses increases with an increase in the number of true QTL, an increase in the genome size, a decrease in the mapping population size and a decrease in the heritability of the trait (Bernardo, 2004). Bernardo (2004) performed a study in a simulated F₂ maize population with varying numbers of QTL and trait heritability and proposed that the genome-wide significance threshold should be lowered from the conventional $\alpha < 0.05$ to limit the false discovery rate (FDR), though a more stringent threshold comes a cost of a reduction in the detection power of QTL. Therefore, the threshold of 0.01 was chosen for this study to limit type I error without sacrificing too much detection power. Secondly, the effect size of QTL tends to be inflated, especially in a QTL study that involves a small mapping population and several true QTL, in what is known as the Beavis effect (Beavis, 1994). The phenotypic variation explained (PVE) by the QTL (also denoted as R^2) reported in this study was likely inflated, given the size of the RIL population and the number of QTL detected (Schön *et al.*, 2004; Beavis, 1998, 1994).

Lastly, a complex trait is polygenic and likely controlled by multiple QTL of small effects. A QTL of a smaller effect has a broad LOD support interval that contains many genes, making the process of identifying the candidate genes underlying the QTL difficult. The 1.5–LOD support intervals of the QTL with the largest effect size in this study—*qSaB11.1*—as estimated by SIM and MQM, which did not artificially narrow the QTL regions, were larger than 3 cM, an equivalence of ~ 800 kbp on the physical map and potentially containing hundreds of genes. A large QTL region on the physical map could be a result of several factors. First, a smaller effect size of a QTL directly translates to a lower LOD score of the peak and a gentler drop of the LOD curve around the peak. A larger linkage region could translate to a larger physical interval. This factor was inevitable as it was the nature of the trait. Second, a small mapping population will result in smaller test statistics, hence lower LOD scores. The RIL population used in this study was of a typical size for a QTL study in rice, 150–240 RILs

(Wu *et al.*, 2013). Third, a lack of recombination events around the QTL peak induces a lower resolution to separate loci apart. Lack of recombination will result in a small LOD support interval covering a large region on the physical map. A lack of recombination could arise from dissimilarity between the parental genomes, small population size and a lack of opportunity for chromosomal crossovers to happen or a low local recombination rate. Some genetic regions harbour more recombination events than others, whilst some regions such as the centromeric regions are void of them. For example, the *qSaB11.1* was situated at 17.9–18.7 Mbp, near the centromere of chromosome 11 (at 12.0 Mbp) (Fragoso *et al.*, 2017), so it was likely that this QTL region might not narrow further despite increasing the size of the RIL population.

3.4.4 Identification of candidate genes underlying the resistance to *Striga asiatica*

Generally, the positions of QTL for resistance to pests and diseases are valuable to the breeding of resistant varieties. In its current form, the resistance to *S. asiatica* is associated with slower growth of parasites on the resistant host compared to those on the susceptible host, which then results in fewer emerged parasites and slower parasite emergence. These traits are desirable as they are likely to decrease the fecundity of the parasite, reducing the number of seeds in infested fields. However, the rice hosts were not completely protected as a proportion of parasites could still successfully attach to the host and potentially affect the growth and yield of the host. There have been studies looking at the size and the effects of *S. asiatica* on host yield; a reduction in parasite biomass per host is not linearly related to an increase in the biomass or the yield of the hosts (Rodenburg *et al.*, 2017; Cissoko *et al.*, 2011). A recent study by Unachukwu *et al.* (2020) compared the susceptibility/resistance of two maize inbred lines and identified a slow growth phenotype to *S. hermonthica* on the more resistant maize genotype (TZSTRI108), compared to the susceptible genotype 5057, similar to the that observed in this study. In addition to the slow growth of parasite individuals on TZSTRI108, there were also fewer attachments of parasites. When the growth of the two

maize genotypes was compared with and without *S. hermonthica* infection, TZSTRI108 grew more vigorously than the very susceptible 5057 when compared to the uninfected controls. Also, the maize genotypes were tested in the field at two locations in Nigeria over two consecutive seasons. TZSTRI108 had a much higher yield than the more susceptible genotype 5057, showing the value of different types of resistance to farmers. Unachukwu *et al.* (2020) concluded that the maize genotype TZSTRI108 exhibited several forms of resistance including fewer *S. hermonthica* attachments and slower growth of the parasites and that understanding the genetic basis of these traits would facilitate breeding programmes that incorporate different types of resistance into farmer preferred varieties.

Similarly, understanding the mechanism by which the growth of *S. asiatica* individuals was slowed down on CT8556-37-2-3-1-M would be useful for breeding programmes. Knowledge of the underlying mechanisms causing the slower growth of the parasites could pave a way to further reduce their growth and possibly completely prevent the above-ground emergence of parasites. The trait in an improved form would then be suitable to be used in conjunction with other *Striga* control measures. Also understanding the mechanism behind the phenotype would be interesting in terms of further understanding the complex interactions between host and parasite.

Identification of genes responsible for the phenotype is the key to unlocking the underlying mechanism of host resistance to the parasite. As discussed previously, the QTL regions identified in this study potentially contain hundreds of genes and ideally QTL need to be narrowed down further to identify the candidate genes. Fine mapping can be done by developing a very large heterozygous inbred family for each QTL by selfing RILs that segregate for markers within the QTL regions, introducing new recombination breakpoints within the QTL region in one or two generations (Tuinstra *et al.*, 1997). More recombination will result in a smaller physical interval for each QTL region. Creating a mapping population is not a trivial task as it requires a lot of resources to grow a large collection of plants and the expertise in crossing the plants. Instead, the available parental genomes can be compared to

identify the presence and absence of the genes within the QTL regions or their variations. These genetic differences can be used to produce a candidate list of genes responsible for the difference in phenotypes. This approach is investigated in Chapter 4.

Chapter 4

Identification of candidate genes underlying resistance in rice to *Striga asiatica*

4.1 Introduction

The linkage mapping study in Chapter 3 revealed that the resistance to *S. asiatica* parasites in the rice genotype CT8556-37-2-3-1-M was a quantitative resistance trait that was controlled by at least six QTL on chromosomes 1, 5 and 11. These QTL contributed in an additive manner to the lower number of parasites and slow growth phenotype of the parasite. Of these QTL, three with the largest effect sizes (*qSaB1.1*, *qSaB5.1* and *qSaB11.1*) were consistently detected using different QTL mapping methods and were chosen for further investigation to identify underlying candidate genes. The genes within these three QTL regions of CT8556-37-2-3-1-M and IR64 must be evaluated and compared to identify possible candidate genes. This will pave the way for future functional validation of genes underlying the resistance phenotype. The identification of the causal genes of this unique slow growth resistance to *S. asiatica* in rice will help elucidate mechanisms by which host plants defence themselves from parasitic weeds and add to the understanding of complex interactions between parasitic plants and their hosts.

Previous studies on the molecular mechanisms of post-attachment resistance to the root parasitic plants *S. gesnerioides* (Li & Timko, 2009), *S. hermonthica* (Beardon, 2018), *Orobancha cumana* Wallr. (Duriez *et al.*, 2019) and the shoot parasite *Cuscuta reflexa* Roxb. (Hegenauer *et al.*, 2016) have identified the importance of *R* genes in the perception of parasites and elicitation of the defence response. In the case of the *S. gesnerioides*–cowpea interaction, a single gene (*RSG3-301*), which encodes a protein with a CC interaction domain at the N terminus, and an NB site and an LRR domain at the C terminus, was found to interact in a race-specific manner against Race 3 of *S. gesnerioides*, eliciting an HR at the site of attachment of the parasite (Li & Timko, 2009). In *S. hermonthica*–rice and *C. reflexa*–tomato interactions, RLP resistance genes, which encode cell surface immune receptors, are thought to perceive the parasite, triggering a defence response (Beardon, 2018; Hegenauer *et al.*, 2016). Finally, Duriez *et al.* (2019) used a map-based cloning approach to identify the resistance gene *HaOr7*, which encodes an LRR receptor-like kinase (RLK), in roots of

resistant varieties of sunflower (*Helianthus* L. species). Recognition of the parasite by HaOr7 prevented the parasite from forming connections with the vascular system of the host.

Other studies that mapped the underlying genetic basis of resistance to *S. hermonthica* have also shown that some forms of resistance are polygenically inherited as described in Section 3.1. Still, few candidate genes underlying polygenic resistance to *Striga* have been identified at present. However, in other pathosystems, genes or QTL underlying quantitative resistance were found to have diverse functions, not restricted to only the detection of foreign invaders by major *R* genes (Pilet-Nayel *et al.*, 2017; Poland *et al.*, 2009). Examples of the physiological/biochemical processes and genes that are involved in quantitative resistance mechanisms include those controlling aspects of morphology and development of the host. For example, loci controlling stomatal density and openness were associated with resistance to bacterial diseases (Melotto *et al.*, 2017, 2006). Plants with low stomata density were less susceptible to bacteria, fungal and oomycete pathogens as stomatal pores provide an opening in the leaf that many foliar pathogens can enter to access inner leaf tissues (Tateda *et al.*, 2019; Jordá *et al.*, 2016). Quantitative resistance can also be controlled by signal transduction through transporters. For example, the *Lr34* gene, which encodes an ABC transporter, confers slow-rusting resistance to fungal pathogens in wheat (Ellis *et al.*, 2014; Krattinger *et al.*, 2009). The substrate of LR34 was found to be ABA, and the fungal resistance was a result of the redistribution of ABA in the leaf (Krattinger *et al.*, 2019). Another way in which quantitative resistance is controlled is through the metabolism of defence hormones. For instance, the *STV11* gene, which encodes a sulfotransferase (OsSOT1), confers resistance to *Rice stripe tenuivirus* in rice (Q. Wang *et al.*, 2014). OsSOT1 converts salicylic acid (SA) into sulfonated SA. This leads to an accumulation of sulfonated SA, which inhibits the replication of the virus (Q. Wang *et al.*, 2014). Finally, some quantitative resistance QTL were associated with traits such as the production of antimicrobial secondary metabolites and cell wall reinforcement. In *A. thaliana*, the induction of the production of camalexin, a major antimicrobial phytoalexin, was associated with resistance to the clubroot protist

Plasmodiophora brassicae (Lemarié *et al.*, 2015). In potato, quantitative resistance to the oomycete *Phytophthora infestans* — late blight causing pathogen — was associated with the thickening of host cell walls from the deposition of secondary compounds such as hydroxycinnamic acid amides and flavonoids (Yogendra *et al.*, 2015).

In Chapter 2 and 3, a few hypotheses were proposed to explain the resistance phenotype of CT8556-37-2-3-1-M to *S. asiatica*. Firstly, it was hypothesised that a difference in the physiology of the two rice genotypes might have led to the difference in the rate of growth of *S. asiatica* individuals. For example, the acquisition of water, organic and mineral nutrients may have been limited either by differences in physical barriers in the host root cortex or xylem connections between the host and the parasite, exemplified by the deposition of the accumulation of dense staining materials in *S. hermonthica*-resistance in rice and sorghum hosts (Mbuvi *et al.*, 2017; Cissoko *et al.*, 2011). Alternatively, differences in the availability of essential nutrients, such as nitrate, ammonium and amino acids in the host xylem, which are the main source of carbon and nitrogen for *Striga* (Pageau *et al.*, 2003), may have differed between the rice genotypes. The flux of water and nutrients from the host to the parasite could also differ between rice genotypes due to differences in water potential gradient at the host–parasite interface due to differences in host-derived osmotic solutes (Joel *et al.*, 2007) or to differences in hormone regulation, such as ABA in the host. Infection of rice and maize by *Striga* elevates the concentration of ABA, leading to stomatal closure (Fujioka *et al.*, 2019b; Frost *et al.*, 1997). This in turn reduces the transpiration of the host, which drives the flow of xylem sap into the parasite (Fujioka *et al.*, 2019a). Secondly, it was hypothesised that the growth of the parasite may be affected by the presence of toxins or growth-inhibitor compounds produced by CT8556-37-2-3-1-M. An example of the delivery of unknown toxic chemical(s) was seen in the incompatible interaction between *S. hermonthica* and *T. dactyloides* (a wild relative of maize) (Gurney *et al.*, 2003). Finally, the difference in the number of attached parasites may reflect a difference in basal resistance mechanism in the two rice genotypes.

4.1.1 Aim and objectives

This chapter aimed to identify possible mechanisms and candidate genes, within the regions of the three major QTL previously mapped on chromosomes 1, 5 and 11, that underlie the difference in the susceptibility of IR64 and CT8556-37-2-3-1-M rice genotypes to *S. asiatica* parasites. The specific objectives were to:

- i) Identify the regions corresponding to the three major QTL on the IR64 genome, which had been assembled and was available, by aligning SNP marker sequences to the IR64 genome sequence, as the positions of the SNP markers used for QTL mapping were for the Nipponbare reference genome.
- ii) Identify the genes within the QTL regions of IR64 genome sequence by performing gene prediction analyses on the chromosomal sequences of IR64 and comparing resulting gene models with transcriptome data to check whether there were any expressed genes missing from the gene models.
- iii) Annotate the predicted genes of IR64 within the QTL regions by comparing the translated sequences of the genes with peptide sequence databases.
- iv) Assemble CT8556-37-2-3-1-M read sequences, which were obtained from Illumina sequencing, into contigs and align the contigs onto the IR64 genome sequence to obtain the CT8556-37-2-3-1-M sequences within the QTL regions.
- v) Perform gene prediction and gene annotation analyses on CT8556-37-2-3-1-M contigs that aligned within the IR64 QTL regions to identify IR64-equivalent genes of CT8556-37-2-3-1-M.
- vi) Identify structural differences between IR64 and CT8556-37-2-3-1-M genome sequences by mapping CT8556-37-2-3-1-M read sequences onto IR64 genome sequence and compare the positions and annotations of the genes within the QTL regions of the two rice genotypes.

4.2 Methodology

4.2.1 Genome sequences

The genome sequence of the rice genotype IR64 (*O. sativa* subsp. *indica*), that has been assembled to the chromosomal level, was provided by Dr Mathias Lorieux [CIAT, Cali, Colombia and Institut de Research pour le Développement (IRD), Montpellier, France]. This IR64 sequence combined the IR64 reference sequence published by Zhou *et al.* (2020) with additional PacBio long-read and Illumina short-read sequences. The genome sequence of CT8556-37-2-3-1-M rice genotype (*O. sativa* subsp. *indica*) is currently being sequenced by collaborators of Dr Mathias Lorieux and was expected to be available for use in this project in early spring of 2020. However, due to coronavirus disease 2019 (Covid-19), the genome sequencing is behind schedule and was not available for use in this project. Nevertheless, some Illumina sequences of this rice genotype were provided by Dr Mathias Lorieux. The CT8556-37-2-3-1-M sequences comprised of two libraries of paired-end Illumina short-read sequences with the same insert size of ~ 200 bp. Each library was sequenced at ~ 30 × depth of coverage of reads. These short-read sequences had already been cleaned and had their adapter sequences trimmed. The ‘Os-Nipponbare-Reference-IRGSP-1.0’ (IRGSP-1.0) reference genome sequence of the Nipponbare rice genotype [*O. sativa* subsp. *japonica* (subgroup *temperate japonica*)] (Kawahara *et al.*, 2013) was downloaded from the website of the Rice Annotation Project Database (RAP-DB) project (<https://rapdb.dna.affrc.go.jp/download/irgsp1.html>).

4.2.2 Conversion of positions of markers within the QTL regions from Nipponbare coordinates to IR64 coordinates

The SNP markers used in the mapping of the QTL for *S. asiatica* resistance (Chapter 3) had been called by mapping IR64 and CT8556-37-2-3-1-M short-read sequences onto the Nipponbare IRGSP-1.0 rice reference genome sequence. This means that the SNP positions of the mapped QTL referred to the positions of Nipponbare SNPs. The location of SNP markers and the QTL on the IR64 genome sequence were needed for downstream analyses.

To achieve this, firstly the sequences of chromosomes 1, 5 and 11 of Nipponbare and IR64 were aligned to identify any structural differences (chromosome rearrangements) between the sequences of the two rice genotypes using the **Progressive Mauve** program within the multiple genome aligner **Mauve** version 20150226 build 10, with the default parameters (Darling *et al.*, 2004).

Overall, there were few chromosomal rearrangements on chromosomes 1 and 5 (Figure 4.1A–B). In contrast, there were numerous inversions and translocations on chromosome 11 (Figure 4.1C). Within the three QTL regions, there were four minor chromosome rearrangements. Within *qSaB1.1* on chromosome 1, there was a small sequence of 4.8 kbp that was translocated from position 37.1 Mbp on Nipponbare sequence to position 37.0 Mbp on IR64 sequence (Figure 4.1A). There was an inversion of a short sequence of 6.3 kbp within *qSaB5.1* on chromosome 5 where the Nipponbare sequence at 22.5 Mbp was inverted on IR64 at 23.9 Mbp (Figure 4.1B). There were two minor translocations of sequence around within the *qSaB11.1* region on chromosome 11 (Figure 4.1C). A sequence of 4.7 kbp starting at 18.4 Mbp on the Nipponbare sequence was translocated out of the QTL region to position 23.9 Mbp on the IR64 sequence. In addition, another sequence of 4.1 kbp at 0.8 Mbp on the Nipponbare sequence was translocated into the QTL region at position 20.7 Mbp on the IR64 sequence. However, none of the SNP markers within the QTL regions were located within these sequence rearrangements. This allowed a straightforward conversion of sequence coordinates as the marker orders were identical on both rice genotypes.

Secondly, to convert the marker positions on the Nipponbare genome sequence to the IR64 genome sequence, 5 kbp of Nipponbare sequence around each SNP marker (marker \pm 2500 bp) were used to search for homologous sequence on the IR64 genome sequence using the nucleotide–nucleotide BLAST (**blastn**) program version 2.9.0+ (Zhang *et al.*, 2000), with the default parameters and penalties. In total, the positions of 185 SNP markers within the three QTL regions were successfully converted from Nipponbare coordinates to IR64

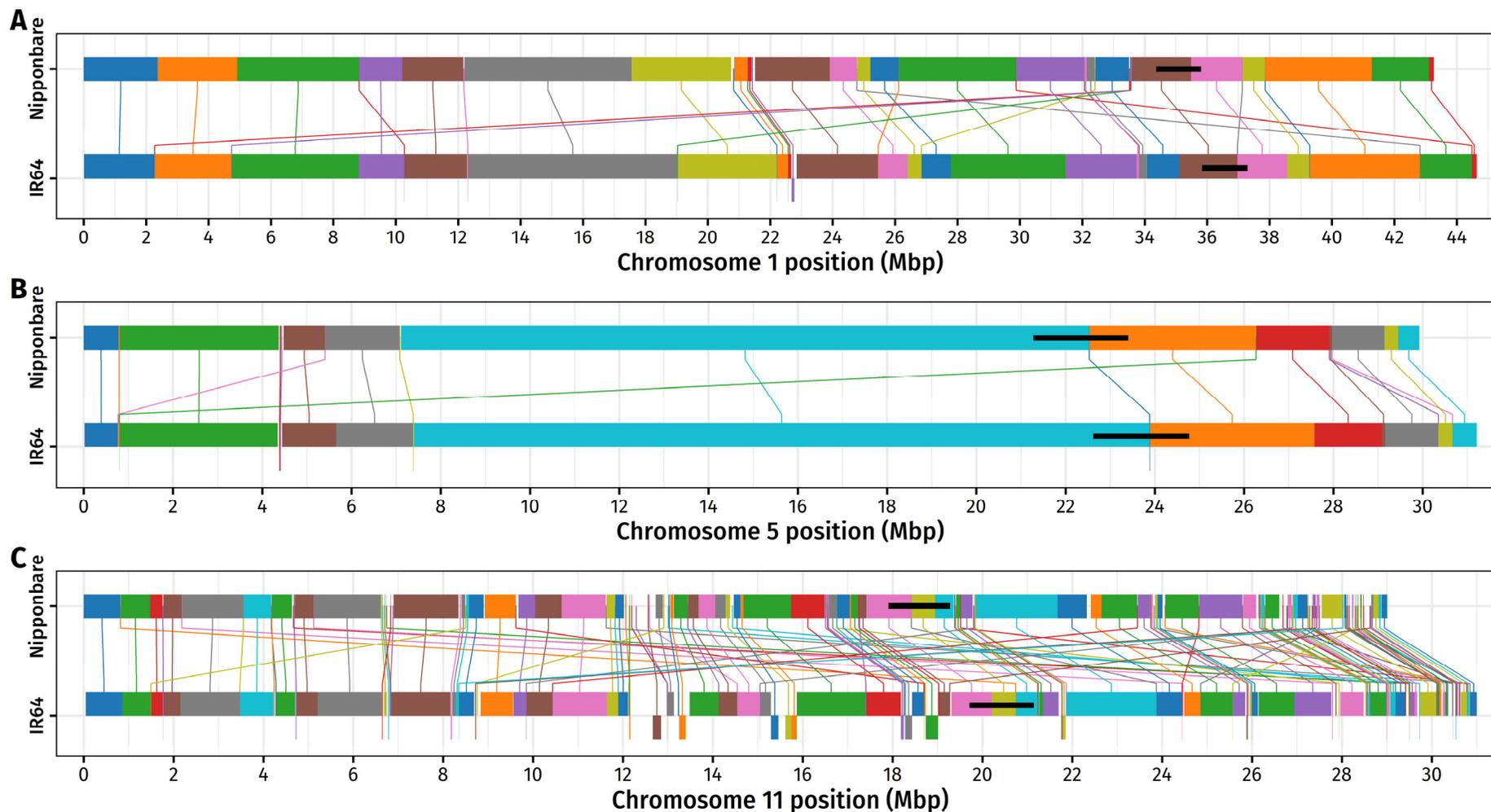


Figure 4.1 Alignments of chromosomes 1, 5 and 11 of the rice genotypes, Nipponbare and IR64. A pair of blocks of the same colour between Nipponbare and IR64 sequences, which are connected by a line of the same colour, contain homologous sequence. IR64 blocks that are above the central IR64 tick marks contain IR64 sequences with the same orientation as the homologous Nipponbare sequences. IR64 blocks that are below the central IR64 tick marks contain IR64 sequences in the reverse orientation of the homologous Nipponbare sequences (inversion). Black horizontal bars mark the positions of the three major QTL for the slow growth resistance phenotype of *Striga asiatica* in rice: *qSaB1.1* (A), *qSaB5.1* (B) and *qSaB11.1* (C).

coordinates. Only one SNP marker on chromosome 1 did not match the sequence of IR64 and was removed from the list of markers. The positions of the start, the end and the peak, and the length of each QTL are presented in Table 4.1.

Table 4.1 Positions of the three major quantitative trait loci (QTL) for the slow growth resistance phenotype of *Striga asiatica* on the genome sequences of the rice genotypes, Nipponbare and IR64.

QTL	Position (bp)					
	Nipponbare			IR64		
	Region	Peak	Length	Region	Peak	Length
qSaB1.1	34 372 239–	35 217 650–	1 430 518	35 839 502–	36 740 520–	1 455 600
	35 802 756	35 237 157		37 295 101	36 760 068	
qSaB5.1	21 276 428–	22 190 668	2 128 237	22 619 144–	23 547 071	2 151 091
	23 404 664			24 770 234		
qSaB11.1	17 907 712–	18 455 653–	1 373 544	19 733 593–	20 243 940–	1 411 210
	19 281 255	18 684 231		21 144 802	20 491 413	

4.2.3 Prediction and annotation of IR64 genes

Whole-genome gene prediction and annotation of the IR64 genome sequence had been carried out by a collaborator of Dr Mathias Lorieux using the **EuGene** pipeline (Foissac *et al.*, 2008). To increase the confidence of the **EuGene** gene predictions, three additional gene prediction algorithms were performed, and their outputs were compared with the **EuGene** gene models. First, the online *ab initio* gene predictor **FGenesH** version 2.6 (Softberry Inc., 2016; Solovyev *et al.*, 2006), with the ‘*O. sativa indica (long-grain rice)*’ specific parameters, was used to predict genes on the whole sequences of chromosomes 1, 5 and 11 of IR64. Second, the gene prediction software **AUGUSTUS** version 3.2.2 (Stanke *et al.*, 2008) was used — in the *ab initio* mode — to make predictions on the whole sequences of chromosomes 1, 5 and 11 of IR64. The parameters specifically trained to predict genes of *Oryza brachyantha* A.Chev. & Roehr. ‘**rice**’ were used as the species was the most closely related to *O. sativa* on the list of trained parameters. Third, the *ab initio* gene prediction program **GeneMark-ES** version 4.38 (Lomsadze *et al.*, 2005) was used, with the default settings for self-training algorithm, to predict genes on the whole sequences of chromosomes 1, 5 and 11 of IR64.

The use of different gene prediction algorithms can result in the prediction of different gene sequences, start and stop sites or positions of exons. In order to determine whether predicted genes are likely to be real, it is often informative to map transcriptome data to a genome sequence. Data from an RNA sequencing (RNA-Seq) study of the effect of *S. hermonthica* on gene expression in roots of IR64 at different stages of infection (2, 4 and 8 DAI), compared to uninfected controls, was available in the Scholes laboratory. These data were used to validate the presence of genes predicted by the four gene-prediction algorithms. It must be remembered, however, that the absence of gene expression data for some genes does not mean they are not real, as they may not be expressed under the experimental conditions used.

The Illumina short-read sequences had been trimmed and cleaned before mapping onto the combined sequences of IR64 genome and *S. hermonthica* genome. The use of the combined rice-*S. hermonthica* genome was to separate the proportion of rice mRNA transcripts from those of *S. hermonthica*, as the harvested samples were a mix of tissues from both species. The resulting alignment **BAM** file was provided by David P. Rapley (Scholes laboratory). Only the primary alignment location of each read in the **BAM** file was considered. This was achieved by filtering out non-primary alignment locations using the ‘`view -F 0x100`’ command in **SAMtools** version 1.9 (H. Li *et al.*, 2009). After the reads were sorted by the ‘`sort`’ command in **SAMtools**, the depth of coverage of the mapped reads at every position within the three QTL regions was extracted by the ‘`depth -m 0`’ command in **SAMtools**. The depth of coverage from each tissue sample was normalised to the sequencing depth (the total read count for each sample), giving the read counts per one million reads (CPMs), before the mean CPMs across the 24 sequenced tissue samples was calculated and compared to the gene models from gene prediction programs.

To annotate the functions of the predicted genes, the translated sequences of the predicted genes were searched against three annotation databases using the protein-protein BLAST (`blastp`) program version 2.9.0+ with the default parameters and penalties. The first database was the 2019-09-09 release of the RAP-DB annotations of the IRGSP-1.0

Nipponbare rice genes that had been verified with cDNA and EST alignments. The database was downloaded from the RAP-DB website (<https://rapdb.dna.affrc.go.jp/download/archive/irgsp1/>). The second database was the Michigan State University (MSU) Rice Genome Annotation Project annotation database version 7.0 of the IRGSP-1.0 Nipponbare rice genes that had been verified with cDNA and EST alignments. The MSU database was download from http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/. The third database searched was the ‘**Non-redundant protein sequence (nr)**’ database (version 2020-10-12) downloaded from the National Center for Biotechnology Information (NCBI) website (<https://ftp.ncbi.nlm.nih.gov/blast/db/>). The **nr** database comprised of the non-redundant GenBank CDS translations, the Protein Data Bank, the Swiss-Prot, the Protein Information Resource and the Protein Reference Foundation, excluding the environmental samples from whole-genome shotgun projects, databases. A **blastp** hit was considered when E-value < 10^{-10} , query coverage > 50 % and percentage identity > 50. For each query, the hit with the highest bit score was reported.

The translated sequences of the predicted genes were also searched for **SUPERFAMILY** annotations (Wilson *et al.*, 2009), which are structural classifications of proteins at the superfamily level (Andreeva *et al.*, 2014), and presence of known protein domains based on the **Pfam** database (El-Gebali *et al.*, 2019), using the **InterProScan** program version 5.47-82.0 (Jones *et al.*, 2014), with **SUPERFAMILY** version 1.75 and **Pfam** version 33.1. The associated gene ontology (GO) terms (Carbon *et al.*, 2019; Ashburner *et al.*, 2000) of resulting **SUPERFAMILY** and **Pfam** annotations were also reported.

4.2.4 Assembling CT8556-37-2-3-1-M reads using reference-guided *de novo* assembly

As the genome sequence of CT8556-37-2-3-1-M was not available for use in this project, as explained in Section 4.2.1, existing Illumina read sequences were assembled to give continuous sequence using a reference-guided *de novo* assembly approach, adapted from

Lischer & Shimizu (2017). First, the two libraries of CT8556-37-2-3-1-M Illumina reads were combined and mapped to the IR64 genome sequence — used as the reference sequence — using `Novoalign` version 3.09.02 (Novocraft, 2019), with the options ‘`-r Random`’ and ‘`-i +- 200-400`,’ which place reads with multiple alignment locations randomly at one of the locations and set the expected size of proper pair fragments to 200–400 bp. Next, the reads in the resulting alignment file had the non-primary alignments filtered out (`samtools view -F 0x100`) and then sorted (`samtools sort`) by `SAMtools` version 1.9 (H. Li *et al.*, 2009). Duplicated reads that resulted from the PCR step in the process of library preparation were removed using the `MarkDuplicates` tool from the `Picard` toolkit version 1.129 (Broad Institute, 2019). Local realignment of reads around indels was performed with `RealignerTargetCreator` and `IndelRealigner` tools in the Genome Analysis Toolkit (`GATK`) version 3.4-46 (Van der Auwera *et al.*, 2013).

Second, for each chromosome, the mapped reads and the reads that did not map to any chromosomes were extracted from the alignment file, using the ‘`view -F 0x4`’ command for the mapped reads and the ‘`view -f 132 -F 8`’, ‘`view -f 68 -F 8`’ and ‘`view -f 13`’ commands for the unmapped reads in `SAMtools`. For a given chromosome, the extracted reads were then *de novo* assembled using the `ABYSS` assembler version 2.1.5 (Jackman *et al.*, 2017), with a Bloom filter and a de Bruijn graph at several k-mer sizes (Table 4.2). Five k-mer sizes — the optimal, two smaller than optimal and two larger than optimal — and their corresponding minimum k-mer count thresholds were calculated by `KmerGenie` version 1.7051 (Chikhi & Medvedev, 2014). The use of multiple k-mer sizes in five assemblies for each chromosome was to minimise the loss of information due to either the lack of specificity in k-mers with shorter k-mer lengths leading to ambiguity in the reconstruction of contig sequences, or the inability to overlap k-mers with longer k-mer length leading to disjointed sequences and a higher number of small contigs (Lischer & Shimizu, 2017). Contigs that were shorter than 200 bp were discarded.

Table 4.2 k-mer lengths and their corresponding minimum k-mer count thresholds used in *de novo* assemblies of CT8556-37-2-3-1-M reads with a Bloom filter and a de Bruijn graph.

Chromosome	Optimal k		Shorter than optimal k				Longer than optimal k			
	k	kc	k	kc	k	kc	k	kc	k	kc
1	67	4	57	5	41	6	71	4	81	3
5	71	4	61	5	51	6	77	4	81	3
11	63	5	51	6	41	7	71	4	81	3

k, k-mer length; kc, corresponding minimum k-mer count threshold.

Third, the redundancy from multiple assemblies was eliminated by assembling the contigs from the five assemblies of each chromosome using the comparative genome assembler, **AMOScmp** version 3.1.0 (Treangen *et al.*, 2011) with the IR64 sequence as a reference. The alignment of the contigs to the reference was carried out by the **nucmer** program from the **MUMmer** package version 3.9.4alpha (Kurtz *et al.*, 2004). The default parameters of **AMOScmp** were used, except in two instances where different values were picked as outlined in Lischer & Shimizu (2017). The maximum ignorable trim length was set to 1000 bp (`-t 1000`) in the ‘`casm-layout`’ step where **AMOScmp** positioned redundant contigs onto the reference sequence based on the alignment information from **nucmer**, and the minimum overlap bases was set to 10 bp (`-o 10`) in the ‘`make-consensus`’ step where consensus sequences were built from redundant contigs.

The IR64 sequence-guided *de novo* assembly of short Illumina reads of CT8556-37-2-3-1-M yielded more than 5000 contigs for each of chromosomes 1, 5 and 11 (Table 4.3). These contigs could not be scaffolded further because of the lack of longer read sequences from sequencing technologies such as PacBio sequencing (Rhoads & Au, 2015) or Nanopore sequencing (Bowden *et al.*, 2019). Thus, the gaps between contigs could not be closed.

Table 4.3 Summary statistics of the IR64-guided assembly of CT8556-37-2-3-1-M contigs (only contigs longer than 200 bp are considered).

Chromosome	CT8556-37-2-3-1-M contigs					IR64 genome sequence size (bp)
	Number of contigs	N50	L50	Mean contig size (bp)	Total contig size (bp)	
1	7311	818	13 311	5006	36 602 393	44 659 248
5	6361	750	9154	3653	23 234 134	31 259 799
11	5091	541	10 370	3765	19 169 982	31 117 384

4.2.5 Prediction and annotation of CT8556-37-2-3-1-M genes

The CT8556-37-2-3-1-M contigs, which were separately assembled from the CT8556-37-2-3-1-M reads mapped to either chromosome 1, 5 or 11 of IR64, were then aligned to the corresponding IR64 chromosome using the **Progressive Mauve** program within the **Mauve** aligner version 20150226 build 10 (Darling *et al.*, 2004), with the default parameters. Within the QTL regions, 159, 482 and 245 contigs aligned to the IR64 sequences of *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively. As some of these contigs were shorter than a typical rice gene and too short for a gene prediction algorithm to run effectively, a cut-off length for contigs was set to 2442 bp, which is the median length of rice genes from the RAP-DB database, leaving totals of 87, 214 and 108 contigs within *qSaB1.1*, *qSaB5.1* and *qSaB11.1* regions, respectively.

To focus prediction and annotation of genes on CT8556-37-2-3-1-M contigs to primarily at the peak of the three major QTL, contigs within 2 cM genetic distance from the peak of each QTL (36 619 686–37 103 937 bp positions on chromosome 1, 23 029 535–23 651 819 bp positions on chromosome 5 and 19 736 416–21 124 051 bp positions on chromosome 11) were selected for gene prediction analysis. The 2 cM genetic distance was chosen because it has been reported that most causal genes are discovered within this distance from the position of the QTL peak (Price, 2006). In total, 29, 54 and 106 contigs that were longer than the cut-off length aligned within the 2 cM distance from the peak of *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively. These contig sequences were analysed using the **FGenesh** gene predictor as it gave the best results out of the four predictors used on IR64. The gene prediction analysis and subsequent annotation for the CT8556-37-2-3-1-M contigs were carried out in the same manner as the IR64 sequences, as described in Section 4.2.3.

4.3 Results

4.3.1 Locating the physical positions of three major QTL on the IR64 genome sequence

Three major QTL for the slow growth resistance phenotype of *S. asiatica* that were identified in Chapter 3 were located on the IR64 genome sequence (Table 4.4). First, *qSaB1.1*, which spanned 8.72 cM on the linkage map, was 1.46 Mbp wide and located at 35.84–37.30 Mbp position on chromosome 1, with two markers with the maximum LOD score (peak markers) situated at 36.74 and 36.76 Mbp positions (Figure 4.2A; Table 4.4). The 2 cM boundaries of *qSaB1.1* were at positions 36.62 and 37.10 Mbp (Figure 4.2A; Table 4.4). Second, *qSaB5.1*, which spanned 11.89 cM, was 2.15 Mbp wide and was located at 22.62–24.77 Mbp on chromosome 5 (Figure 4.2C; Table 4.4). The peak marker of *qSaB5.1* was at the 23.55 Mbp position, whilst the left and right flanking markers 2 cM from the peak marker were located at 23.03 and 23.65 Mbp positions, respectively (Figure 4.2C; Table 4.4). Finally, *qSaB11.1* spanned the 5.52 cM and was 4.08 Mbp wide, positioned at 19.73 Mbp to 21.14 Mbp on chromosome 11 (Figure 4.2E; Table 4.4). *qSaB11.1*'s two peak markers were situated at 20.24 and 20.49 Mbp positions (Figure 4.2E; Table 4.4). The 2 cM boundaries of *qSaB11.1* were almost at the start position and the end position of the QTL, at 19.73 and 21.12 Mbp, respectively (Figure 4.2E; Table 4.4).

Table 4.4 Locations, peak positions, the 2 cM distance from the peaks and lengths of three major quantitative trait loci (QTL) for the slow growth resistance phenotype of *Striga asiatica* on the IR64 genome sequence.

QTL	Chromosome	Whole QTL			2 cM distance	
		Boundaries	Length (bp)	Peak	Boundaries	Length (bp)
<i>qSaB1.1</i>	1	35 839 502– 37 295 101	1 455 600	36 740 520– 36 760 068	36 619 686– 37 103 937	484 252
<i>qSaB5.1</i>	5	22 619 144– 24 770 234	2 151 091	23 547 071	23 029 535– 23 651 819	622 285
<i>qSaB11.1</i>	11	19 733 593– 21 144 802	1 411 210	20 243 940– 20 491 413	19 736 416– 21 124 051	1 387 636

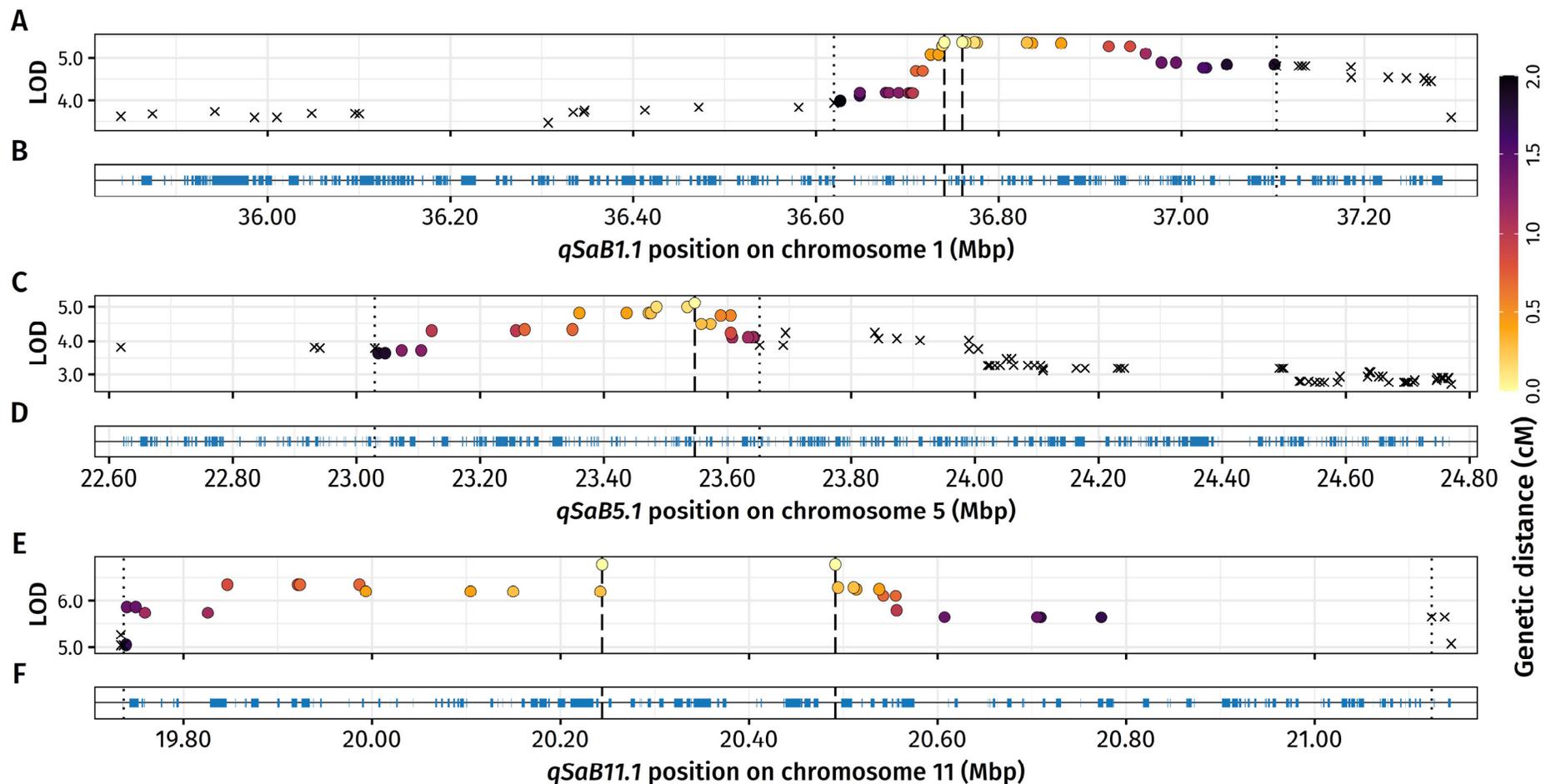


Figure 4.2 Physical locations of the logarithm of the odds ratio (LOD) support intervals of *Striga asiatica*-resistance quantitative trait loci (QTL) on the IR64 genome sequence. LOD scores of *qSaB1.1* (A), *qSaB5.1* (C) and *qSaB11.1* (E) from the simple interval mapping were plotted against the physical locations of the SNP markers (circles and crosses) on chromosomes 1, 5 and 11, respectively. Dashed lines mark the positions of the markers with the highest LOD scores (peak markers). SNP markers are shown as circles, which are bounded by vertical dotted lines when they are within 2 cM genetic distance from the peak markers. The colour of the marker symbols (yellow to black) corresponds to the genetic distance from the nearest peak marker (0–2 cM). Markers over 2 cM from the closest peak marker are shown as crosses. B, D & F, Genes predicted by the EuGene pipeline within the QTL regions are shown as blue rectangles.

The **EuGene** pipeline predicted 215, 157 and 138 genes within the region of *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively (Figure 4.2B, D & F). Of these, 78, 76 and 136 genes were situated within the interval of 2 cM genetic distance from either side of the nearest peak marker of *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively. The 2 cM distance from the peak of *qSaB11.1* (1388 kbp) was more than double the size of those of *qSaB1.1* (484 kbp) and *qSaB5.1* (622 kbp); therefore, more genes were predicted within the 2 cM distance of *qSaB11.1*. This occurred because the population-specific local rate of recombination was much lower at *qSaB11.1* than at the other two QTL. This could also be seen from the two tightly linked peak markers of *qSaB11.1*, which were 247 kbp apart and contained 16 **EuGene**-predicted genes but shared the same cM position (Figure 4.2E–F).

4.3.2 Gene models and transcriptome of IR64

IR64 genome sequences of chromosomes 1, 5 and 11 were analysed with three gene-prediction algorithms: **FGenesH**; **Augustus**; and **GeneMark**. The results from these gene prediction programs were compared to those from **EuGene**. Overall, the four programs showed good agreement on the start and stop sites of many exons (Figure 4.3A, 4.4A & 4.5A). **FGenesH**, **Augustus** and **GeneMark** predicted more genes within all three QTL regions than **EuGene** (Table 4.5), with **FGenesH** giving the highest number of genes, followed by **GeneMark** and **Augustus**, respectively. There were some sections of the IR64 genome sequence where **EuGene** failed to predict the presence of genes, contradicting the predictions from the other three programs, for example, at positions 37.058–37.069 Mbp on

Table 4.5 Number of genes within the whole quantitative trait locus (QTL) regions and the 2 cM distance from the QTL peaks of three major *Striga asiatica*-resistance QTL on chromosomes 1, 5 and 11 of rice, predicted by four gene prediction programs, **EuGene**, **FGenesH**, **Augustus** and **GeneMark**.

QTL	Whole QTL region				2 cM dsitance from QTL peaks			
	EuGene	FGenesH	Augustus	GeneMark	EuGene	FGenesH	Augustus	GeneMark
<i>qSaB1.1</i>	215	342	230	246	79	123	78	92
<i>qSaB5.1</i>	286	559	346	386	76	156	89	117
<i>qSaB11.1</i>	138	318	175	208	135	312	172	206
Total	639	1221	751	840	290	592	339	415

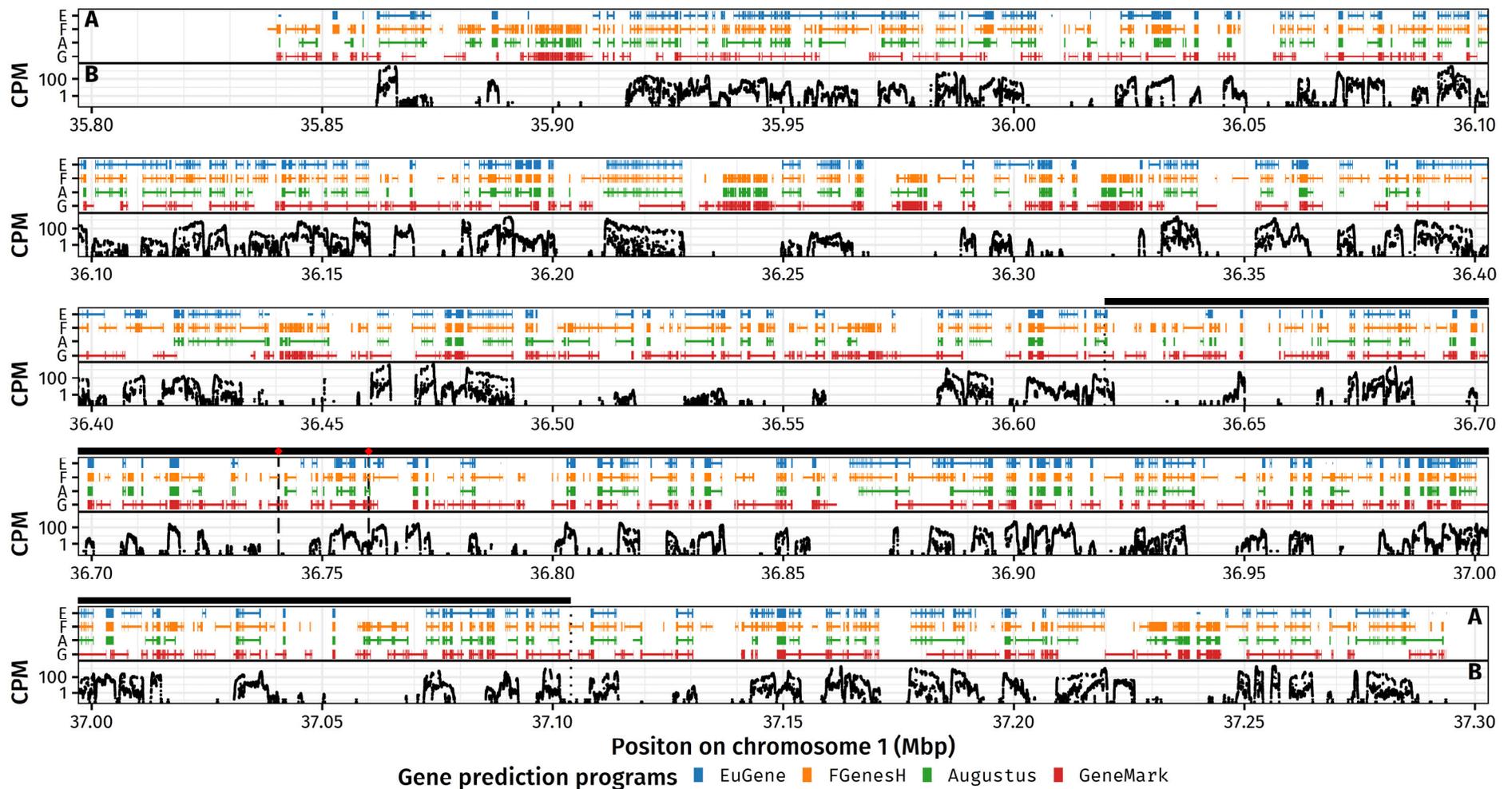
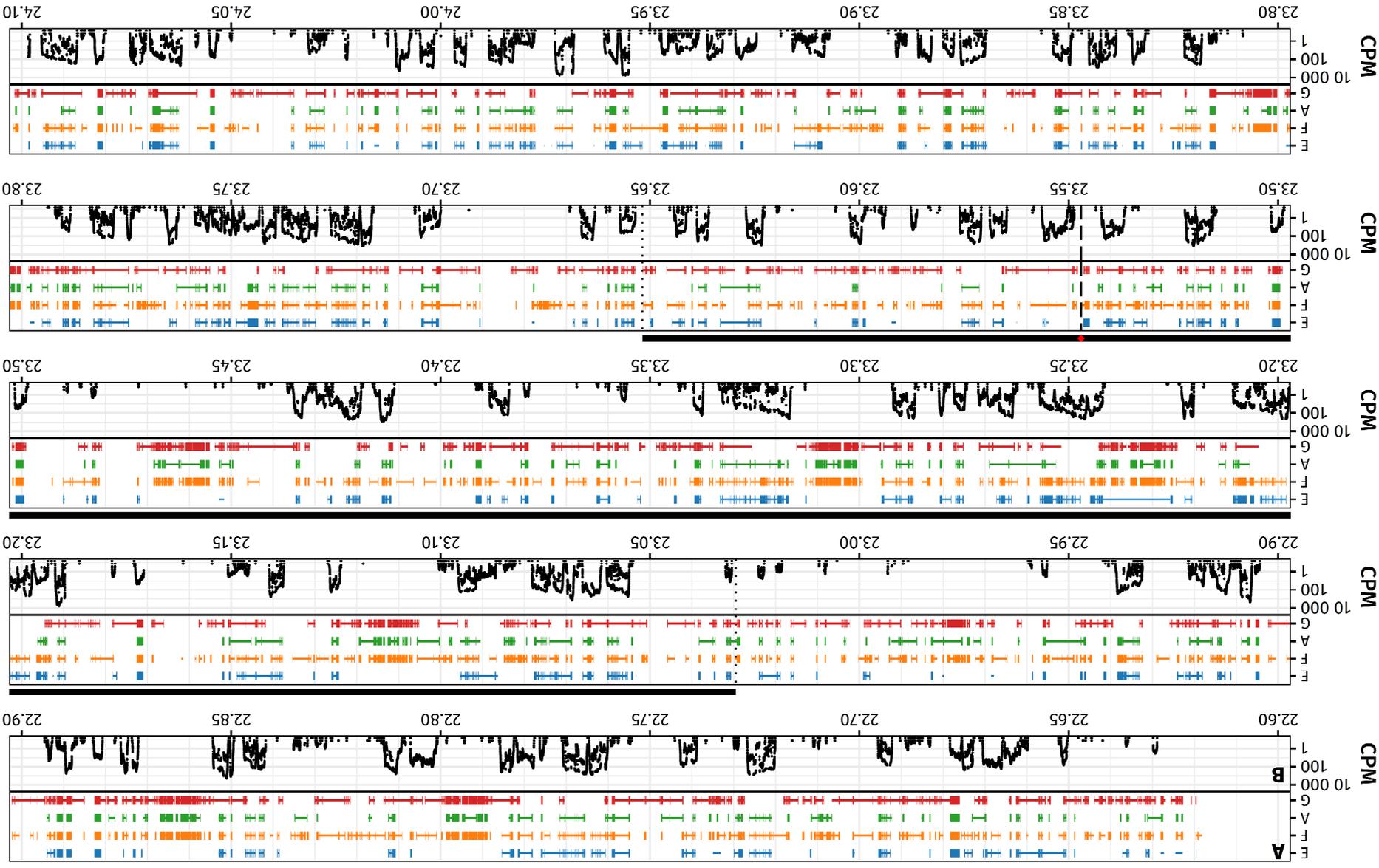


Figure 4.3 Genes predicted by different gene prediction programs within the *Striga asiatica*-resistance quantitative trait locus (QTL), *qSaB1.1*, on the IR64 genome. Gene models (A) were produced using four gene prediction programs: EuGene (blue); FGenesH (orange); Augustus (green); and GeneMark (red). Exons (coloured rectangles) of the same gene are connected by a horizontal bar. The expression of IR64 transcripts (B) is shown as the count per one million base pairs (CPM) of cDNA reads on a log-scale. Horizontal black bar (36.620–37.104 Mbp) marks the 2 cM distance from the QTL peak. Two red diamonds (36.741 and 36.760 Mbp) mark the positions of markers at the QTL peak.



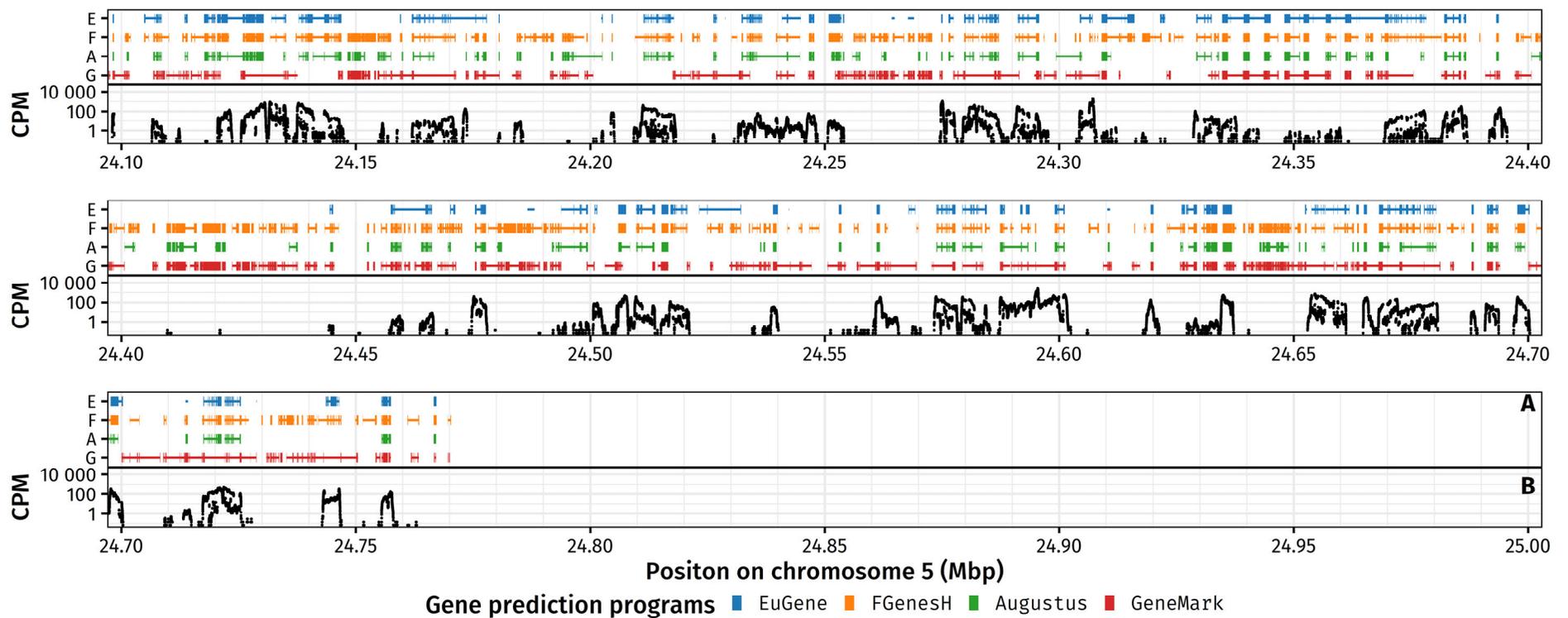


Figure 4.4 Genes predicted by different gene prediction programs within the *Striga asiatica*-resistance quantitative trait locus (QTL), *qSaB5.1*, on the IR64 genome. Gene models (A) were produced using four gene prediction programs: EuGene (blue); FGenesH (orange); Augustus (green); and GeneMark (red). Exons (coloured rectangles) of the same gene are connected by a horizontal bar. The expression of IR64 transcripts (B) is shown as the count per one million base pairs (CPM) of cDNA reads on a log-scale. Horizontal black bar (23.030–23.652 Mbp) marks the 2 cM distance from the QTL peak. Red diamond (23.547 Mbp) mark the positions of markers at the QTL peak.

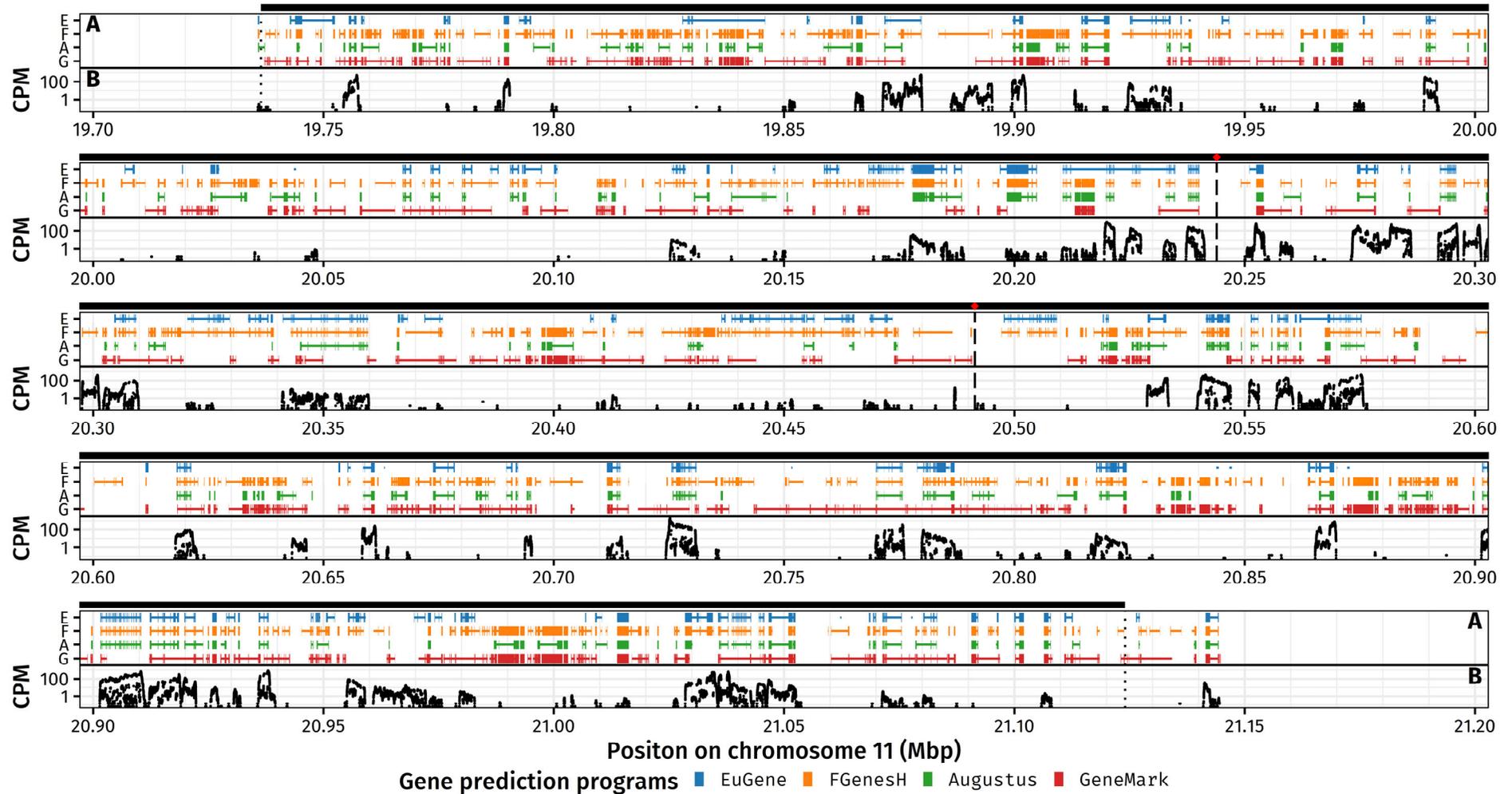


Figure 4.5 Genes predicted by different gene prediction programs within the *Striga asiatica*-resistance quantitative trait locus (QTL), *qSaB11.1*, on the IR64 genome. Gene models (A) were produced using four gene prediction programs: EuGene (blue); FGenesH (orange); Augustus (green); and GeneMark (red). Exons (coloured rectangles) of the same gene are connected by a horizontal bar. The expression of IR64 transcripts (B) is shown as the count per one million base pairs (CPM) of cDNA reads on a log-scale. Horizontal black bar (19.736–21.124 Mbp) marks the 2 cM distance from the QTL peak. Two red diamonds (20.244 and 20.491 Mbp) mark the positions of markers at the QTL peak.

chromosome 1 (Figure 4.3A), positions 23.097–23.123 Mbp on chromosome 5 and positions 20.213–20.217 Mbp and 20.987–21.003 Mbp on chromosome 11.

Transcriptomic data from an RNA-Seq analysis of changes in gene expression in roots of IR64 following infection with *S. hermonthica* were also used to verify the accuracy of the gene models. Since the gene expression data were from an analysis of the effect of *S. hermonthica* on gene expression at three time points after infection of IR64 roots, they could only be used to confirm that a gene was real (where the mapped cDNAs showed the expression of the gene), but not to determine whether a gene prediction was false (in the absence of expression).

In general, the locations of the predicted genes matched the location of mapped reads (expression profile) of IR64 genes (Figure 4.3B, 4.4B & 4.5B). The gene expression data revealed that several genes predicted by **FGenesH**, **Augustus** and **GeneMark**, but missed by **EuGene**, were real. For instance, predicted genes that were expressed according to the transcriptome data at 36.708–36.709 Mbp and 36.722–36.724 Mbp in *qSaB1.1* (Figure 4.3), at 23.580–23.585 Mbp and 23.587 Mbp in *qSaB5.1* (Figure 4.4) and at 20.298–20.301 Mbp and 20.694–20.695 Mbp in *qSaB11.1* (Figure 4.5) were predicted by **FGenesH** or **Augustus**, but were absent from the **EuGene** gene predictions. When **FGenesH**, **Augustus** and **GeneMark** were compared, the **FGenesH** gene prediction matched the expression pattern more closely than the gene models from the other two programs. **Augustus** did not predict many genes that had strong evidence of expression, for example, at 37.008 Mbp on chromosome 1, 23.200 Mbp on chromosome 5 and 20.225 Mbp on chromosome 11. **GeneMark** tended to predict genes over long stretches of sequence (with a mean length of 3.6 kbp) with more exons (on average nine exons per gene), compared to other prediction algorithms (2.2–2.6 kbp in length with three to four exons on average). Therefore, the **FGenesH** gene model was selected for further analyses, and **FGenesH** was chosen to predict genes on CT8556-37-2-3-1-M sequences.

4.3.3 Mapping of the reads and assembled contigs of CT8556-37-2-3-1-M onto the IR64 genome sequence

Figure 4.6–4.8 show the positions of IR64 genes predicted by FGenesH (Figure 4.6A, 4.7A & 4.8A), the depth of coverage of CT8556-37-2-3-1-M reads (Figure 4.6B, 4.7B & 4.8B) and the assembled CT8556-37-2-3-1-M contigs that were mapped onto the IR64 genome (Figure 4.6C, 4.7C & 4.8C). The mean depth of coverage of CT8556-37-2-3-1-M reads across the genome was 57.31 times (Figure 4.6B, 4.7B & 4.8B). In several areas, the depth of coverage of CT8556-37-2-3-1-M reads deviated from the mean and became highly variable, for example, at positions 36.63–36.64 Mbp in *qSaB1.1* (Figure 4.6B), positions 23.10–23.12 Mbp in *qSaB5.1* (Figure 4.7B), and positions 19.90–19.92 Mbp and 20.21–20.22 Mbp in *qSaB11.1* (Figure 4.8B). Fluctuation in the depth of coverage of CT8556-37-2-3-1-M reads was likely to be caused by the ambiguity associating with the alignment of repetitive sequences between CT8556-37-2-3-1-M reads and IR64 genome. The occurrences of varying CT8556-37-2-3-1-M read depths also coincided with predicted IR64 genes annotated as transposable elements (TEs). (The details of genes annotations will be given in the next section and Supplementary Table S4.1–S4.3). Also, no assembled CT8556-37-2-3-1-M contigs aligned to any of these areas of variable depth of coverage of CT8556-37-2-3-1-M reads (Figure 4.6B, 4.7B & 4.8B).

In addition, the depth of coverage of CT8556-37-2-3-1-M reads was extremely high at two loci, position 37.071 Mbp on chromosome 1 with 797 times depth (14 times the average) and position 20.496 Mbp on chromosome 11 with 23 649 times depth (413 times the average). This could indicate a sequencing bias or an error of the sequence aligner. However, the sequences at the two loci were moderately more repetitive than surrounding sequences (Supplementary Figure S4.1).

Furthermore, within the three whole regions of the major QTL, only one region between positions 20.11 and 20.17 Mbp of chromosome 11 — adjacent to the peak of *qSaB11.1* — had almost no CT8556-37-2-3-1-M reads mapped onto it (Figure 4.8B). A lack of CT8556-

37-2-3-1-M reads mapped here indicated that there may be a deletion of the CT8556-37-2-3-1-M sequence on chromosome 11, with reference to the IR64 genome.

4.3.4 Predicted genes within 2 cM distance from QTL peaks

Within 2 cM distance from the peak markers of *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, FGenesH predicted 123, 156 and 313 genes on the IR64 sequences, respectively (Table 4.6). Within the same regions, 27, 97 and 57 CT8556-37-2-3-1-M contigs that were longer than 2442 bp aligned to the *qSaB1.1*, *qSaB5.1* and *qSaB11.1* sequences of IR64, respectively (Table 4.6). Fewer genes were predicted on these CT8556-37-2-3-1-M contigs, compared to the number of genes predicted on IR64 genome (Table 4.6). However, the real number of genes within these CT8556-37-2-3-1-M regions of 2 cM distance from the QTL peaks was probably different for two reasons. One, genes on contigs shorter than 2442 bp were not included. Two, a gene that spanned more than one contig was counted multiple times. Also, approximately 12 % of these CT8556-37-2-3-1-M genes were partial, due to the proximity to the end of contigs, i.e. they did not contain a stop codon (Table 4.6).

Table 4.6 Number of genes predicted by FGenesH on the sequences of the IR64 genome and CT8556-37-2-3-1-M contigs (longer than 2442 bp) within 2 cM distance from the peaks of the three major *Striga asiatica*-resistance QTL.

QTL	Number of IR64 genes	CT8556-37-2-3-1-M			Total number of genes
		Number of Contigs	Number of complete genes	Number of partial genes	
<i>qSaB1.1</i>	123	27	93	7	100
<i>qSaB5.1</i>	156	97	110	4	114
<i>qSaB11.1</i>	313	53	207	12	219
Total	592	177	410	23	433

The details of the predicted genes, from both IR64 and CT8556-37-2-3-1-M within *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, can be found in Supplementary Table S4.1–4.3, which contain gene identification (ID), CDS start and stop sites, DNA strand orientation, number of exons, number of amino acids, annotation, BLAST hit and its statistics, SUPERFAMILY and Pfam domain hits, associated GO terms and percentage identity between IR64 and CT8556-37-2-3-1-M genes.

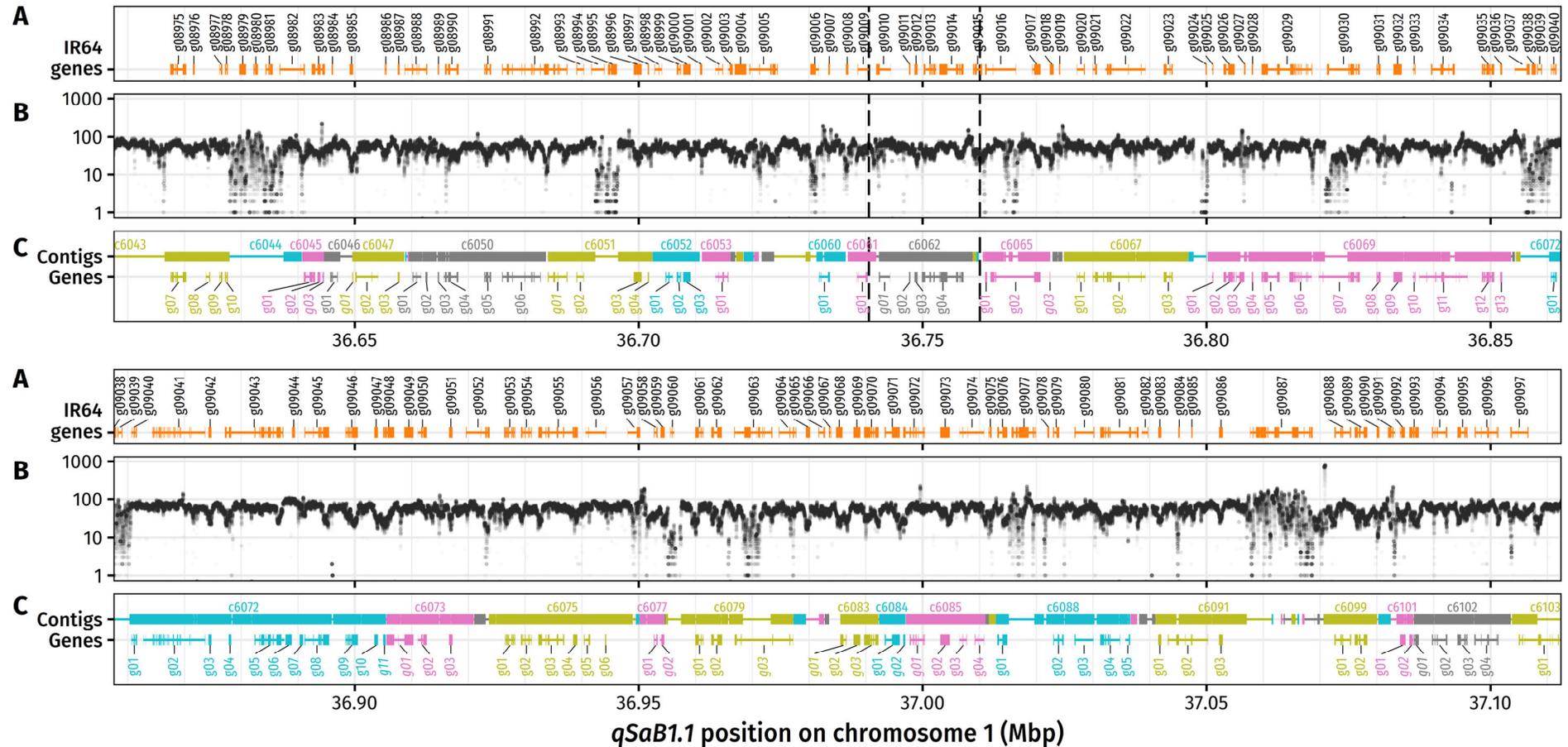


Figure 4.6 Comparison of predicted genes on the IR64 genome sequence and CT8556-37-2-3-1-M contigs within 2 cM from the peak of the *Striga asiatica*-resistance quantitative trait locus (QTL), *qSaB1.1*. (A) Predicted genes on the IR64 genome sequence. Exons of the same gene are connected by a horizontal bar. Gene IDs are labelled. (B) Depth of coverage of CT8556-37-2-3-1-M reads that mapped onto the IR64 genome sequence. (C) Contigs and predicted genes of CT8556-37-2-3-1-M that were aligned to the IR64 genome sequence. Blocks of CT8556-37-2-3-1-M contigs show loci that CT8556-37-2-3-1-M contigs and the IR64 sequence aligned (homologous sequences). Contigs longer than 2442 bp were labelled. For the predicted genes, exons of the same gene are connected by a horizontal bar. Predicted genes share the same colour with their corresponding contigs. Gene IDs are labelled. Gene IDs of truncated genes are in italic. The vertical dashed lines (36.741 and 36.760 Mbp) mark the boundaries of the QTL peak.

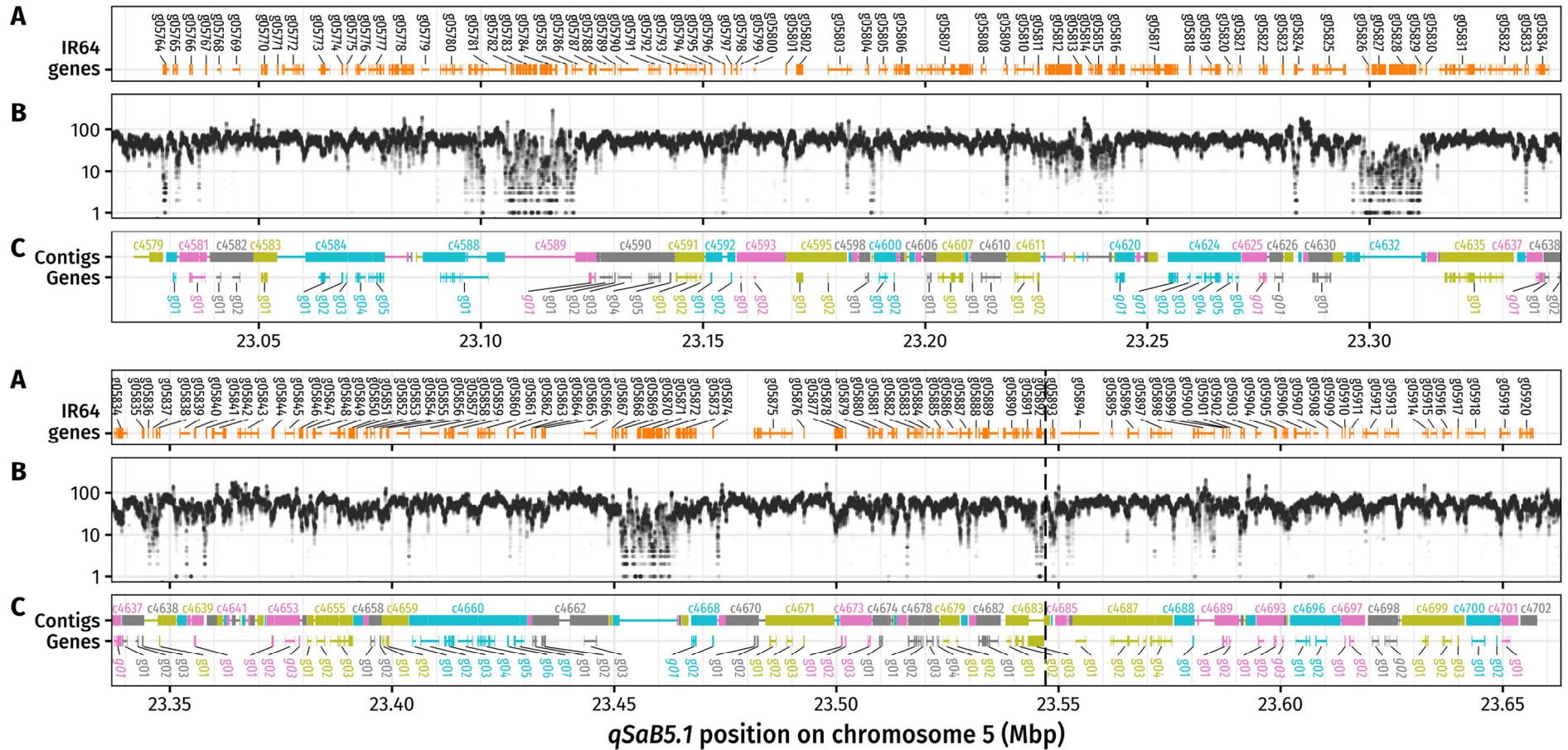
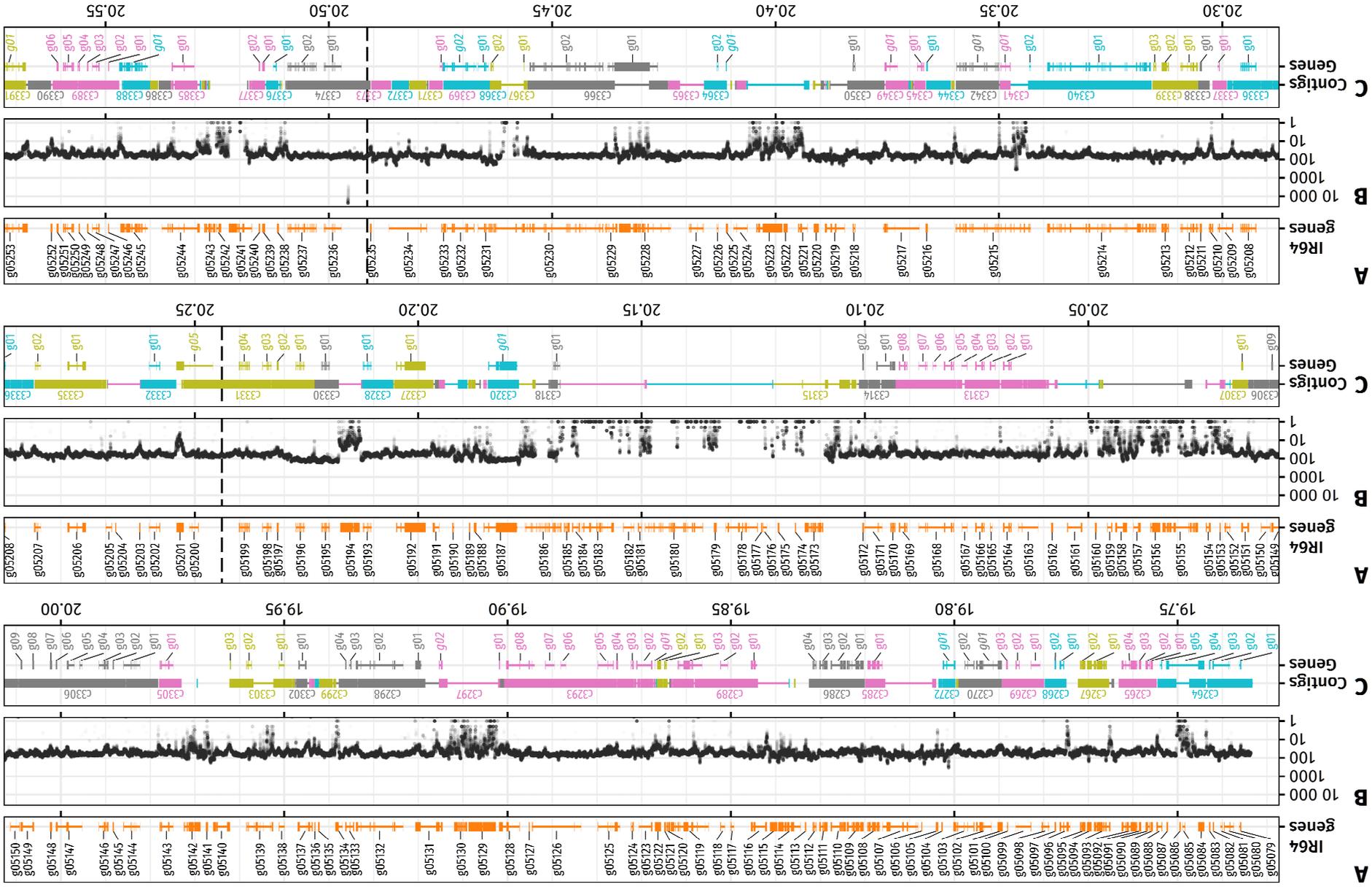


Figure 4.7 Comparison of predicted genes on the IR64 genome sequence and CT8556-37-2-3-1-M contigs within 2 cM from the peak of the *Striga asiatica*-resistance quantitative trait locus (QTL), *qSaB5.1*. (A) Predicted genes on the IR64 genome sequence. Exons of the same gene are connected by a horizontal bar. Gene IDs are labelled. (B) Depth of coverage of CT8556-37-2-3-1-M reads that mapped onto the IR64 genome sequence. (C) Contigs and predicted genes of CT8556-37-2-3-1-M that were aligned to the IR64 genome sequence. Blocks of CT8556-37-2-3-1-M contigs show loci that CT8556-37-2-3-1-M contigs and the IR64 sequence aligned (homologous sequences). Contigs longer than 2442 bp were labelled. For the predicted genes, exons of the same gene are connected by a horizontal bar. Predicted genes share the same colour with their corresponding contigs. Gene IDs are labelled. Gene IDs of truncated genes are in italic. The vertical dashed line (23.547 Mbp) marks the QTL peak.



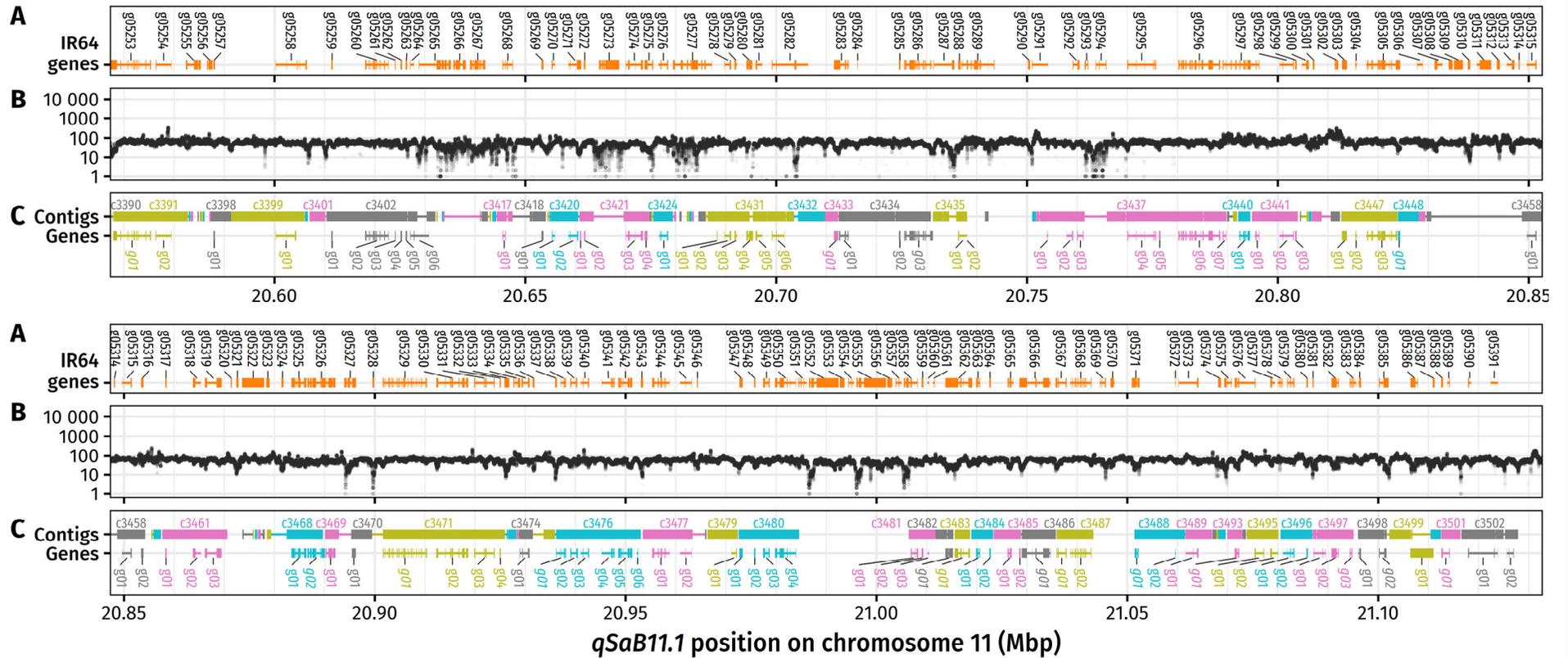


Figure 4.8 Comparison of predicted genes on the IR64 genome sequence and CT8556-37-2-3-1-M contigs within 2 cM from the peak of the *Striga asiatica*-resistance quantitative trait locus (QTL), *qSaB11.1*. (A) Predicted genes on the IR64 genome sequence. Exons of the same gene are connected by a horizontal bar. Gene IDs are labelled. (B) Depth of coverage of CT8556-37-2-3-1-M reads that mapped onto the IR64 genome sequence. (C) Contigs and predicted genes of CT8556-37-2-3-1-M that were aligned to the IR64 genome sequence. Blocks of CT8556-37-2-3-1-M contigs show loci that CT8556-37-2-3-1-M contigs and the IR64 sequence aligned (homologous sequences). Contigs longer than 2442 bp were labelled. For the predicted genes, exons of the same gene are connected by a horizontal bar. Predicted genes share the same colour with their corresponding contigs. Gene IDs are labelled. Gene IDs of truncated genes are in italic. The vertical dashed lines (20.244 and 20.491 Mbp) mark the boundaries of the QTL peak.

The predicted genes encoding proteins with similar functions were assigned the same functional group. The predicted genes in Supplementary Table S4.1–S4.3 are presented in the physical order they appear on the chromosomes, except for predicted genes whose amino acid sequences did not match any entries in the protein databases used or genes that lack any known protein domains. These genes, which are unlikely to be real, are shown at the bottom of each table and are marked as ‘unknown.’ These unknown genes with no annotation will not be discussed further.

In total, 21 (21 %), 23 (20 %) and 33 genes (15 % of all CT8556-37-2-3-1-M genes) were only predicted within the 2 cM distance from the peak of CT8556-37-2-3-1-M *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively, and were not present in the IR64 genome. Of these CT8556-37-2-3-1-M unique genes, three were annotated as domain-containing proteins (which were a protein containing F-box domain, a protein in UDP-Glycosyltransferase/glycogen phosphorylase family and a leucine-rich repeat domain-containing protein), four as TEs, 16 as hypothetical proteins and 59 had no annotation (Supplementary Table S4.1–S4.3).

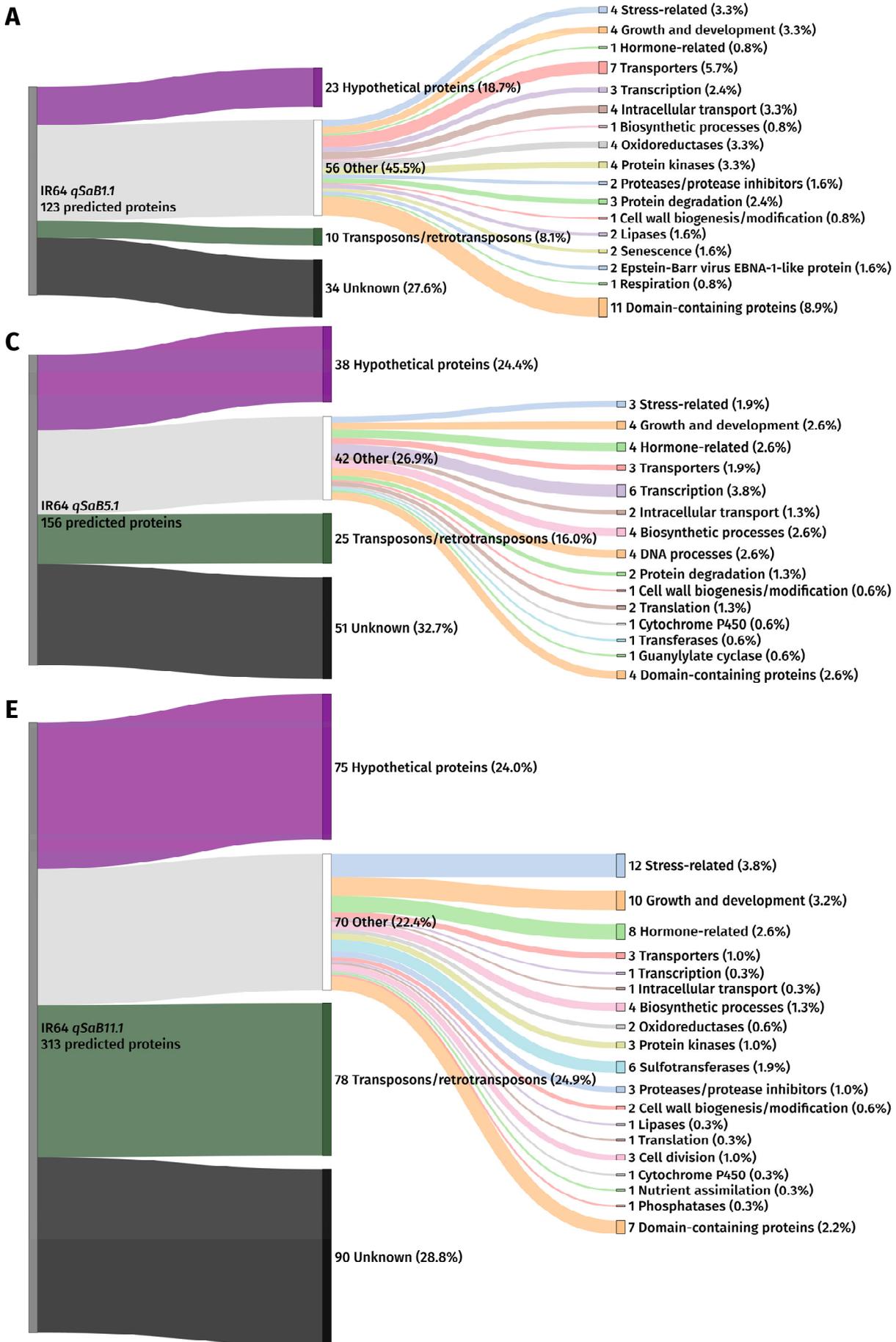
Also, 75, 86 and 170 genes were predicted within the 2 cM intervals of *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively, on both IR64 genome and CT8556-37-2-3-1-M contigs. Of these, 4, 3 and 13 genes from *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively, spanned multiple CT8556-37-2-3-1-M contigs (Supplementary Table S4.1–S4.3).

Table 4.7 and Figure 4.9 show genes categorised into different functional groups. Most of the genes predicted on both IR64 genome and CT8556-37-2-3-1-M contigs encoded similar proteins. The mean percentage identity between the amino acid sequences of IR64 and CT8556-37-2-3-1-M proteins was 79 %. Approximately 40 % of the protein pairs were more than 98 % identical over the full length of their amino acid sequences (Supplementary Table S4.1–S4.3). The differences in the amino acid sequences in the rest of genes predicted on both rice genotypes, likely came from a combination of real differences in the genome sequences of IR64 and CT8556-37-2-3-1-M and the errors associated with predicting genes on short

contig sequences, for instance, incomplete genes at the end of contigs. For example, the genes IR64_01g09030 and CT8556_01c6069g07 from *qSaB1.1*, which were both annotated as transmembrane amino acid transporter protein, shared only 71 % of their amino acid sequences, or the genes IR64_05g05825 and CT8556_05c4630g01 from *qSaB5.1*, which were annotated as AUX1-like protein, shared only 73 % of their amino acid sequences (Table 4.7; Supplementary Table S4.1–S4.2).

Figure 4.9 summarises the number of predicted proteins, within the 2 cM distance from the peak of the QTL, categorised into groups of annotated functions. In general, similar functional groups of genes were found within all three QTL regions (Figure 4.9; Table 4.7). All three QTL regions contained genes encoding proteins involved in growth and development, transmembrane transport via transporters, hormone-mediated signalling pathways, cell wall biogenesis and modification, intracellular transport, biotic/abiotic stress responses and transcription factors (Figure 4.9; Table 4.7).

Within the 2 cM distance from the peak of *qSaB1.1* on chromosome 1, the numbers of proteins in different functional groups were similar between the two rice genotypes (Figure 4.9A–B). Examples of annotated proteins included transmembrane transporter proteins, including two amino acid transporters, two ammonium transporters, two heavy metal transporter/detoxification proteins and a nodulin-like domain-containing protein. The gene encoding the ABA receptor was also present in *qSaB1.1*, as well as genes encoding proteins involved in growth and development, such as protein G1-like7 and a lateral organ boundary domain (LBD)-containing protein, which regulate the formation of constitutive aerenchyma and lateral roots. There were also genes encoding transcription factors (TFs), for example, OVATE FAMILY PROTEIN 2 (OFP2), WRKY24 and a basic helix-loop-helix (bHLH) TF controlling axillary meristem formation and genes involved in biotic/abiotic stress responses, such as those encoding a laccase precursor, defensin DEFL4 and selT-like protein. Finally, there were genes encoding proteins involved in intracellular transport and vesicle trafficking, such as VHS2 and exocyst complex protein EXO70, oxidoreductases, such as glutaredoxin



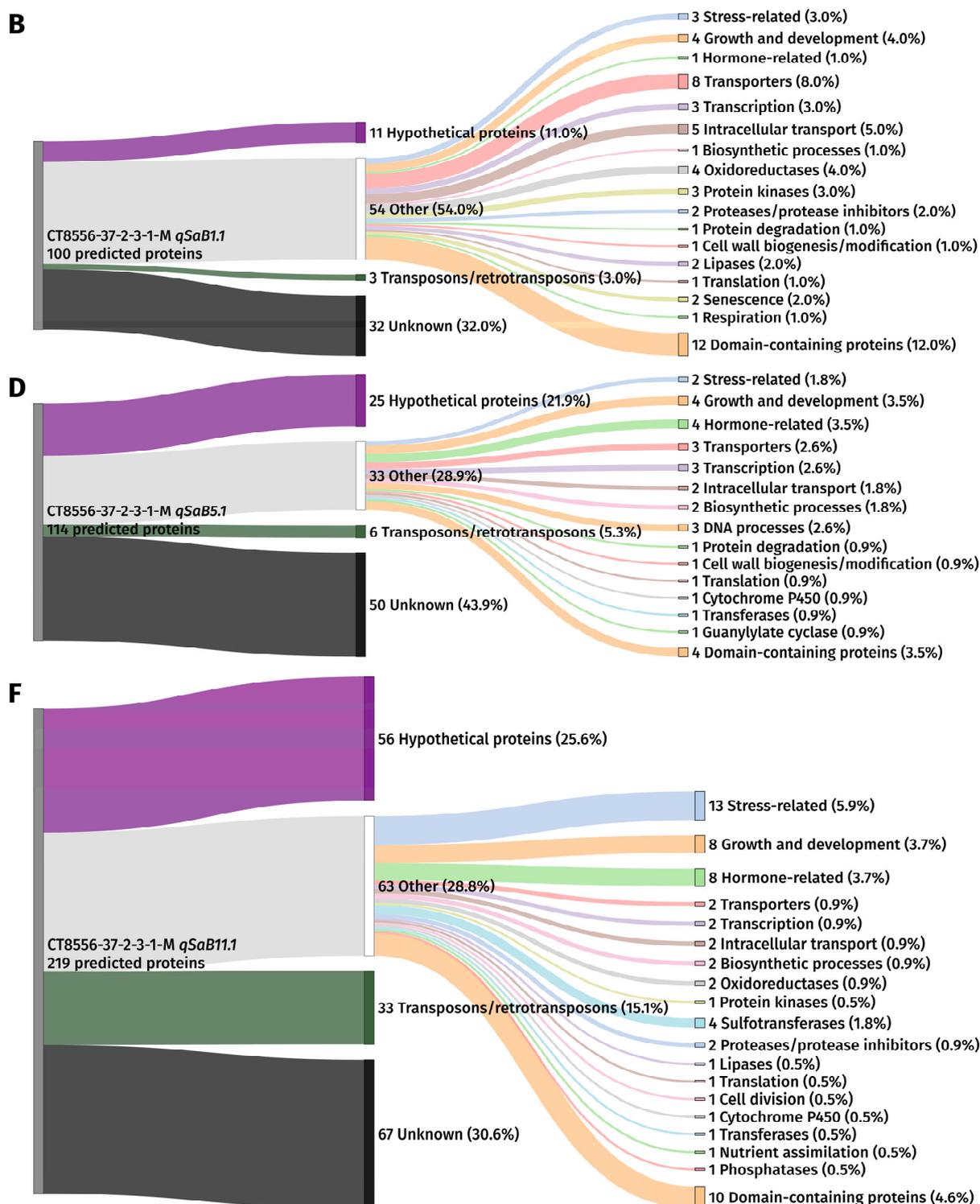


Figure 4.9 Functional categorisation of proteins encoded by genes predicted within the 2 cM distance from peak markers of three major *Striga asiatica*-resistance quantitative trait loci (QTL), *qSaB1.1*, *qSaB5.1* and *qSaB11.1*. Proteins were grouped according to the annotations obtained by comparing predicted protein sequences to rice and NCBI protein databases. If no match was found, SUPERFAMILY and Pfam domains were used to infer functions. 'Unknown' refers to proteins with no database matches and contain no SUPERFAMILY and Pfam domains. Numbers of proteins (in front of group names) correspond to bands' width. Percentage of proteins from the QTL within each group is in parentheses. **A–C**, IR64 proteins from *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively. **D–F**, CT8556-37-2-3-1-M proteins from *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, respectively.

Table 4.7 Annotations of genes within the 2 cM distance from peak markers of three major *Striga asiatica*-resistance quantitative trait loci (QTL), *qSaB1.1*, *qSaB5.1* and *qSaB11.1*, categorised into functional groups. Percent identity of amino acid sequences of IR64 and CT8556-37-2-3-1-M genes; Os ind, *Oryza sativa* subsp. *indica*; Os jap., *O. sativa* subsp. *japonica*; « PF... », Pfam hits; « SSF... », SUPERFAMILY hits.

QTL	IR64		CT8556-37-2-3-1-M		Percent identity
	Gene ID	Annotation	Gene ID	Annotation	
Nutrient assimilation					
<i>qSaB11.1</i>	11g05332	NADH/NADPH-dependent nitrate reductase	11c3471g04	NADH/NADPH-dependent nitrate reductase	100.0
Transporters					
<i>qSaB1.1</i>	01g09026	Nodulin-like domain containing protein.	01c6069g02	Nodulin-like domain containing protein.	94.8
	01g09030	transmembrane amino acid transporter protein, putative, expressed	01c6069g07	transmembrane amino acid transporter protein, putative, expressed	70.6
	01g09031	Amino acid transporter, transmembrane domain containing protein.	01c6069g08	Amino acid transporter, transmembrane domain containing protein.	99.0
	01g09033	Heavy metal transport/detoxification protein domain containing protein.	01c6069g10	Heavy metal transport/detoxification protein domain containing protein.	100.0
	01g09070	Iron(III)-deoxymugineic acid transporter, Translocation of iron in reproductive organs and phloem in joints	01c6083g03	Iron(III)-deoxymugineic acid transporter, Translocation of iron in reproductive organs and phloem in joints	97.1
	01g09089	ammonium transporter protein, putative, expressed	01c6099g02	ammonium transporter protein, putative, expressed	100.0
	01g09093	Similar to Ammonium transporter.	01c6101g02	Similar to Ammonium transporter.	44.1
			01c6102g01	Similar to Ammonium transporter.	52.3
<i>qSaB5.1</i>	05g05791	solute carrier family 35 member F2, putative, expressed	05c4590g04	solute carrier family 35 member F2, putative, expressed	100.0
	05g05793	« PF06027: Solute carrier family 35 »	05c4591g01	PREDICTED: solute carrier family 35 member F1-like isoform X1 [<i>Oryza brachyantha</i>]	91.9
	05g05882	transporter, major facilitator family, putative, expressed	05c4678g01	Major facilitator superfamily protein.	79.2
<i>qSaB11.1</i>	11g05121	Similar to ATPase, coupled to transmembrane movement of substances.	11c3291g01	Similar to ATPase, coupled to transmembrane movement of substances.	65.9
	11g05146	Sugar transporter, TAL effector-mediated susceptibility to bacterial pathogen	11c3306g03	Sugar transporter, TAL effector-mediated susceptibility to bacterial pathogen	99.7
	11g05242	root cap protein 1-like [Os jap.]			
Hormone related					
<i>qSaB1.1</i>	01g09051	ABA receptor 8 [Os ind.]	01c6073g03	abscisic acid receptor PYL4 [Os jap.]	95.7
<i>qSaB5.1</i>	05g05825	Similar to AUX1-like protein.	05c4630g01	Similar to AUX1-like protein.	72.7
	05g05860	periplasmic beta-glucosidase precursor, putative, expressed	05c4660g05	periplasmic beta-glucosidase precursor, putative, expressed	96.1
	05g05890	putative auxin-independent growth promoter [Os jap.]	05c4683g02	putative auxin-independent growth promoter [Os jap.]	32.3
	05g05914	Indole-3-acetic acid (IAA)-amido synthetase, Disease resistance, Abiotic stress tolerance	05c4699g01	putative indole-3-acetic acid-amido synthetase GH3.1 [<i>Prunus yedoensis</i> var. <i>nudiflora</i>]	52.6
<i>qSaB11.1</i>	11g05131	DELLA protein RHT-1 [Os jap.]	11c3297g02	gibberellin response modulator protein, putative, expressed	23.0
			11c3298g01	gibberellin response modulator protein, putative, expressed	50.3
	11g05195	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	11c3330g01	Sorghum bicolor leucine-rich repeat-containing extracellular glycoprotein precursor.	99.6
	11g05196	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	11c3331g01	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	100.0
	11g05198	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	11c3331g03	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	100.0
	11g05206	Transport inhibitor response 1 (TIR1)-like protein, Target of siR109944 (small interfering RNA)	11c3335g01	Transport inhibitor response 1 (TIR1)-like protein, Target of siR109944 (small interfering RNA)	99.6
	11g05286	putative auxin response factor 1 [Os ind.]	11c3434g03	RecName: Full=Auxin response factor 23; AltName: Full=OsARF1 [Os ind.]	96.6
	11g05370	GH3 auxin-responsive promoter family protein [Os jap.]			
	11g05371	Indole-3-acetic acid (IAA)-amido synthetase	11c3488g01	Indole-3-acetic acid (IAA)-amido synthetase, Plant architecture establishment, Drought tolerance	47.6
Growth and development					
<i>qSaB1.1</i>	01g09022	Lateral organ boundary domain (LBD)-containing protein, LBD transcription factor, Regulation of constitutive aerenchyma and lateral root formation	01c6067g02	« PF07859: alpha/beta hydrolase fold PF03195: Lateral organ boundaries (LOB) domain »	57.2
	01g09043	Domain of unknown function DUF399 domain containing protein. « PF11891: Protein RETICULATA-related PF04187: Haem-binding uptake, Tiki superfamily, ChaN »	01c6072g05	uncharacterized protein LOC4327520 [Os jap.] « PF11891: Protein RETICULATA-related PF04187: Haem-binding uptake, Tiki superfamily, ChaN »	66.5
	01g09059	protein G1-like7 [Os jap.]	01c6077g02	protein G1-like7 [Os jap.]	44.3

Table 4.7 (continued)

QTL	IR64		CT8556-37-2-3-1-M		Percent identity	
	Gene ID	Annotation	Gene ID	Annotation		
<i>qSaB5.1</i>	01g09061	thioredoxin-like fold domain-containing protein MRL7 homolog, chloroplastic [Os jap.]	01c6079g01	Thioredoxin-like fold domain containing protein.	89.6	
	05g05765	No apical meristem (NAM) protein domain containing protein.	05c4580g01	No apical meristem (NAM) protein domain containing protein.	54.8	
	05g05774	Similar to Ferredoxin VI, chloroplast precursor (Fd VI).	05c4584g02	Similar to Ferredoxin VI, chloroplast precursor (Fd VI).	100.0	
	05g05777	Similar to Plastid division protein ftsZ1 precursor.	05c4584g05	Similar to Plastid division protein ftsZ1 precursor.	100.0	
	05g05877	UPF0503 protein At3g09070, chloroplastic [Os jap.]	05c4673g01	PREDICTED: UPF0503 protein At3g09070, chloroplastic-like [Oryza brachyantha]	19.7	
	<i>qSaB11.1</i>	11g05164	No apical meristem (NAM) protein domain containing protein.	11c3313g01	No apical meristem (NAM) protein domain containing protein.	97.0
		11g05166	Similar to No apical meristem protein, expressed.	11c3313g03	Similar to No apical meristem protein, expressed.	96.8
		11g05168	no apical meristem protein, putative, expressed	11c3313g05	no apical meristem protein, putative, expressed	60.0
		11g05169	NAM protein [Os jap.]	11c3313g07	Hypothetical protein « SSF101941: NAC domain »	11.7
		11g05193	Small leucine rich repeat (LRR) protein, Cell elongation activity, Regulation of lamina inclination and grain size	11c3313g08	no apical meristem protein, putative, expressed	96.6
		11g05278	TPD1 protein homolog 1-like [Os jap.]	11c3328g01	Small leucine rich repeat (LRR) protein, Cell elongation activity, Regulation of lamina inclination and grain size	99.0
		11g05331	N-alpha-acetyltransferase MAK3 [Os jap.]	11c3431g02	TPD1 protein homolog 1-like [Os jap.]	100.0
		11g05337	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplastic [Os jap.]	11c3471g03	acetyltransferase, GNAT family, putative, expressed	100.0
		11g05338	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplastic [Panicum hallii]	11c3476g01	Similar to predicted protein. « PF12046: Cofactor assembly of complex C subunit B »	71.9
		11g05366	NEF1, putative, expressed	11c3486g01	hypothetical protein OsJ_34095 [Os jap.]	85.6
<u>Cell wall biogenesis/modification</u>						
<i>qSaB1.1</i>	01g08991	Alpha-expansin OsEXPA2.	01c6050g05	Alpha-expansin OsEXPA2.	76.5	
<i>qSaB5.1</i>	05g05835	Similar to H0212B02.6 protein.	05c4638g03	Similar to H0212B02.6 protein.	67.7	
<i>qSaB11.1</i>	11g05128	Omega-hydroxypalmitate O-feruloyl transferase	11c3293g08	Similar to transferase.	33.8	
	11g05375	« PF03254: Xyloglucan fucosyltransferase »				
<u>Abiotic/biotic stress-related</u>						
<i>qSaB1.1</i>	01g09046	Putative laccase precursor, Abiotic stress response	01c6072g09	Putative laccase precursor, Abiotic stress response	99.8	
	01g09067	DEFL4 - Defensin and Defensin-like DEFL family, expressed				
	01g09088	BAG family protein, Control of innate immunity and broad-spectrum disease resistance	01c6099g01	BAG family protein, Control of innate immunity and broad-spectrum disease resistance	99.6	
<i>qSaB5.1</i>	01g09096	Similar to selT-like protein.	01c6102g04	Similar to selT-like protein.	93.3	
	05g05858	Component of the SCF E3 ubiquitin ligase complex, Jasmonate-regulated defense responses, Regulation of leaf senescence	05c4660g03	Component of the SCF E3 ubiquitin ligase complex, Jasmonate-regulated defense responses, Regulation of leaf senescence	100.0	
	05g05872	plant disease resistance polyprotein-like [Os jap.]				
<i>qSaB11.1</i>	05g05916	universal stress protein domain containing protein, putative, expressed	05c4699g02	universal stress protein domain containing protein, putative, expressed	100.0	
	11g05122	IQ calmodulin-binding and BAG domain containing protein, putative, expressed	11c3291g02	IQ calmodulin-binding and BAG domain containing protein, putative, expressed	62.1	
			11c3293g01	IQ calmodulin-binding and BAG domain containing protein, putative, expressed	5.8	
	11g05187	Resistance to Rice stripe virus (RSV), Protection of plant growth from heat stress, (Nipponbare: RSV-susceptible)	11c3320g01	Resistance to Rice stripe virus (RSV), Protection of plant growth from heat stress, (Nipponbare: RSV-susceptible)	92.2	
	11g05189	STV11 [Os jap.]				
	11g05192	STV11-pa1 protein [Os ind.]	11c3327g01	STV11-pa1 protein [Os ind.]	99.9	
	11g05296	stripe rust resistance protein Yr10, putative, expressed	11c3437g06	stripe rust resistance protein Yr10, putative, expressed	91.2	
	11g05305	Similar to Jacalin-like lectin domain containing protein, expressed.	11c3447g03	Similar to Jacalin-like lectin domain containing protein, expressed.	80.3	
			11c3448g01	jacalin-like lectin domain containing protein, expressed	14.9	
	11g05312	Similar to Calcineurin B-like protein.				
11g05330	N-rich protein, putative, expressed	11c3471g02	N-rich protein, putative, expressed	100.0		
11g05376	Bisdemethoxycurcumin synthase	11c3493g01	chalcone synthase, putative, expressed	85.9		

Table 4.7 (continued)

QTL	IR64		CT8556-37-2-3-1-M		Percent identity
	Gene ID	Annotation	Gene ID	Annotation	
			11c3495g01	hypothetical protein EJB05_27085, partial [Eragrostis curvula] « PF00195: Chalcone and stilbene synthases, N-terminal domain »	13.2
	11g05382	Polyketide synthase, type III domain containing protein.	11c3497g02	Polyketide synthase, type III domain containing protein.	100.0
	11g05385	chalcone and stilbene synthases, putative, expressed	11c3498g02	chalcone and stilbene synthases, putative, expressed	47.9
	11g05386	bisdemethoxycurcumin synthase-like [Os jap.]	11c3499g01	bisdemethoxycurcumin synthase-like [Os jap.]	83.9
	Sulfotransferases				
<i>qSaB11.1</i>	11g05089	sulfotransferase domain containing protein, expressed	11c3265g01	Sulfotransferase family protein.	100.0
	11g05090	sulfotransferase domain containing protein, expressed	11c3265g02	Sulfotransferase family protein.	99.6
	11g05091	Flavonol sulfotransferase, putative [Os jap.]	11c3265g03	Flavonol sulfotransferase, putative [Os jap.]	100.0
	11g05097	sulfotransferase domain containing protein, expressed	11c3268g02	sulfotransferase domain containing protein, expressed	99.2
	11g05098	Sulfotransferase domain containing protein [Os jap.]			
	11g05101	sulfotransferase domain containing protein, expressed			
	Cytochrome P450				
<i>qSaB5.1</i>	05g05802	Cytochrome P450 family protein.	05c4595g01	Cytochrome P450 family protein.	99.8
<i>qSaB11.1</i>	11g05319	Similar to Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIA1) (Obtusifoliol 14-alpha demethylase) (Fragment).	11c3461g03	Similar to Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450-LIA1) (Obtusifoliol 14-alpha demethylase) (Fragment).	100.0
	Proteases/protease inhibitors				
<i>qSaB1.1</i>	01g08987	LTPL16 - Protease inhibitor/seed storage/LTP family protein precursor, expressed	01c6047g03	LTPL16 - Protease inhibitor/seed storage/LTP family protein precursor, expressed	100.0
	01g09041	ulp1 protease family, C-terminal catalytic domain containing protein, expressed	01c6072g02	ulp1 protease family, C-terminal catalytic domain containing protein, expressed	78.4
<i>qSaB11.1</i>	11g05259	ICE-like protease p20 domain containing protein, putative, expressed	11c3402g01	ICE-like protease p20 domain containing protein, putative, expressed	97.6
	11g05260	Peptidase S10, serine carboxypeptidase family protein.	11c3402g02	OsSCP63 - Putative Serine Carboxypeptidase homologue, expressed	99.6
	11g05368	Similar to aminopeptidase.	11c3487g02	hypothetical protein Osl_36330 [Os ind.] « PF12576: Protein of unknown function (DUF3754) »	100.0
	Transcription				
<i>qSaB1.1</i>	01g08996	transcription repressor OFP2 [Os jap.]	01c6051g03	transcription repressor OFP2 [Os jap.]	98.3
	01g09035	RecName: Full=WRKY transcription factor WRKY24; Short=OsWRKY24 [Os ind.]	01c6069g12	WRKY24, expressed	90.9
	01g09086	Basic helix-loop-helix (bHLH) transcription factor, Axillary meristem formation	01c6091g03	Basic helix-loop-helix (bHLH) transcription factor, Axillary meristem formation	99.1
<i>qSaB5.1</i>	05g05780	transcription factor TGAL5 isoform X6 [Os jap.]	05c4588g01	transcription factor TGAL5 isoform X6 [Os jap.]	95.9
	05g05787	Similar to T6J4.5 protein (WIP6 protein).	05c4589g01	hypothetical protein Osl_20137 [Os ind.] « SSF57667: beta-beta-alpha zinc fingers »	82.9
	05g05831	mediator of RNA polymerase II transcription subunit 13 [Os jap.]	05c4635g01	expressed protein « PF18296: MID domain of medPIWI »	77.5
	05g05832	hypothetical protein Osl_20152 [Os ind.] « PF11597: Mediator complex subunit 13 N-terminal »			13.2
	05g05846	ethylene-responsive transcription factor ERF094 [Os jap.]	05c4653g03	expressed protein	26.9
	05g05865	MYB family transcription factor, putative, expressed	05c4662g02	MYB family transcription factor, putative, expressed	98.4
<i>qSaB11.1</i>	11g05283	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation	11c3433g01	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation	48.4
			11c3434g01	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation	42.1
	Intracellular transport				
<i>qSaB1.1</i>	01g09029	Similar to VHS2 protein (Fragment).	01c6069g05	EP1-like glycoprotein 3 [Os jap.]	44.9
	01g09047	prenylated rab acceptor, putative, expressed	01c6069g06	VHS and GAT domain containing protein, expressed	49.2
	01g09048	Exo70 exocyst complex subunit family protein.	01c6072g10	prenylated rab acceptor, putative, expressed	99.1
	01g09049	Exo70 exocyst complex subunit family protein.	01c6072g11	Exo70 exocyst complex subunit family protein.	20.0
<i>qSaB5.1</i>	05g05772	sec23/Sec24 zinc finger family protein, expressed	01c6073g01	Exocyst complex component EXO70B1 [Dichantheium oligosanthes]	53.3

Table 4.7 (continued)

QTL	IR64		CT8556-37-2-3-1-M		Percent identity
	Gene ID	Annotation	Gene ID	Annotation	
<i>qSaB11.1</i>	05g05776	Syntaxin 6, N-terminal domain containing protein.	05c4584g04	Syntaxin 6, N-terminal domain containing protein.	100.0
	05g05897	Similar to Phosphatidic acid phosphatase-like protein.	05c4687g04	Similar to Phosphatidic acid phosphatase-like protein.	99.5
	11g05199	BR01 domain domain containing protein.	11c3331g04	BR01 domain domain containing protein.	93.9
	11g05247	Similar to Acyl carrier protein, chloroplast precursor (ACP) (ACP05) (Clone 29C08).	11c3389g02	Similar to Acyl carrier protein, chloroplast precursor (ACP) (ACP05) (Clone 29C08).	98.7
<u>Oxidoreductases</u>					
<i>qSaB1.1</i>	01g08992	« SSF51197: Clavamate synthase-like »	01c6050g06	hypothetical protein Osl_04250 [Os ind.] « PF01283: Ribosomal protein S26e »	37.1
	01g09065	glutaredoxin, putative, expressed	01c6051g01	Hypothetical conserved gene. « SSF51197: Clavamate synthase-like »	25.5
	01g09072	Pyridine nucleotide-disulphide oxidoreductase, NAD-binding region domain containing protein.	01c6084g02	NADH-ubiquinone oxidoreductase, mitochondrial precursor, putative, expressed	26.7
	01g09076	2OG-Fe(II) oxygenase domain containing protein.	01c6085g01	Pyridine nucleotide-disulphide oxidoreductase, NAD-binding region domain containing protein.	68.8
<i>qSaB11.1</i>	11g05271	sex determination protein tasselseed-2, putative, expressed	01c6088g01	2OG-Fe(II) oxygenase domain containing protein.	99.8
	11g05369	Similar to Rubredoxin 1 (Rd-1).	11c3420g02	sex determination protein tasselseed-2, putative, expressed	34.5
			11c3421g01	sex determination protein tasselseed-2, putative, expressed	24.4
<u>DNA processes</u>					
<i>qSaB5.1</i>	05g05807	5-methylcytosine DNA glycosylase/lyase	05c4606g01	5-methylcytosine DNA glycosylase/lyase	9.7
	05g05817	hhH-GPD superfamily base excision DNA repair protein, expressed	05c4607g01	5-methylcytosine DNA glycosylase/lyase	44.7
	05g05917	Os05g0453800 [Os jap.] (GINS complex, Psf1 component family protein)	05c4624g01	hypothetical protein Osl_20146 [Os ind.]	31.5
	05g05918	Similar to DNA replication complex GINS protein PSF1.	05c4699g03	Os05g0453800 [Os jap.] (GINS complex, Psf1 component family protein)	98.2
			05c4700g01	—	56.4
<u>Protein kinases</u>					
<i>qSaB1.1</i>	01g08975	Protein kinase, core domain containing protein.	01c6043g07	Protein kinase, core domain containing protein.	85.3
	01g09014	hypothetical protein Osl_04261 [Os ind.] « SSF56112: Protein kinase-like (PK-like) »	01c6062g04	hypothetical protein Osl_04261 [Os ind.] « SSF56112: Protein kinase-like (PK-like) »	64.5
	01g09091	protein kinase-like [Os jap.]			
	01g09094	probable phytol kinase 2, chloroplastic [Os jap.]	01c6102g02	probable phytol kinase 2, chloroplastic [Os jap.]	98.4
<i>qSaB11.1</i>	11g05183	Protein kinase-like domain containing protein.	11c3366g01	transposon protein, putative, Mutator sub-class [Os jap.]	29.8
	11g05229	« SSF56112: Protein kinase-like (PK-like) »	11c3482g01	hypothetical protein Osl_36325 [Os ind.]	39.1
	11g05362	hypothetical protein Osl_36325 [Os ind.]		« SSF51110: alpha-D-mannose-specific plant lectins »	
		« PF00069: Protein kinase domain PF01453: D-mannose binding lectin »	11c3483g01	hypothetical protein Osl_36325 [Os ind.] « SSF56112: Protein kinase-like (PK-like) »	35.8
<u>Protein degradation</u>					
<i>qSaB1.1</i>	01g09001	probable E3 ubiquitin-protein ligase ATL44 [Os jap.]			
	01g09004	U-box E3ubiquitin ligase, Positive regulation of cold stress response			
	01g09054	gamma-interferon-responsive lysosomal thiol protein [Os jap.]	01c6075g02	gamma-interferon-responsive lysosomal thiol protein [Os jap.]	79.3
<i>qSaB11.1</i>	05g05819	Similar to Arginyl-tRNA--protein transferase 1 (EC 2.3.2.8) (R-transferase 1) (Arginyltransferase 1) (Arginine-tRNA--protein transferase 1). Splice isoform ATE1-2.	05c4624g04	Similar to Arginyl-tRNA--protein transferase 1 (EC 2.3.2.8) (R-transferase 1) (Arginyltransferase 1) (Arginine-tRNA--protein transferase 1). Splice isoform ATE1-2.	79.5
	05g05821	Similar to Arginyl-tRNA--protein transferase 1 (EC 2.3.2.8) (R-transferase 1) (Arginyltransferase 1) (Arginine-tRNA--protein transferase 1). Splice isoform ATE1-2.			
<u>Biosynthetic processes</u>					
<i>qSaB1.1</i>	01g09053	dihydroflavonol-4-reductase, putative, expressed	01c6075g01	dihydroflavonol-4-reductase, putative, expressed	88.0
<i>qSaB5.1</i>	05g05773	glycosyl hydrolases family 17, putative, expressed	05c4584g01	Glycoside hydrolase, family 17 protein.	100.0
	05g05822	Starch-binding domain-containing protein, Starch biosynthesis	05c4625g01	Starch-binding domain-containing protein, Starch biosynthesis	88.2
	05g05842	Similar to glycerol-3-phosphate acyltransferase 8.			

Table 4.7 (continued)

QTL	IR64		CT8556-37-2-3-1-M		Percent identity
	Gene ID	Annotation	Gene ID	Annotation	
<i>qSaB11.1</i>	05g05897	Similar to Phosphatidic acid phosphatase-like protein.	05c4687g04	Similar to Phosphatidic acid phosphatase-like protein.	99.5
	11g05208	serine palmitoyltransferase 2, putative, expressed	11c3336g01	Similar to Serine palmitoyltransferase (Fragment).	100.0
	11g05244	XY4 protein (Fragment).	11c3385g01	uncharacterized protein LOC117930913 [Vitis riparia] « SSF53254: Phosphoglycerate mutase-like »	16.7
	11g05247	Similar to Acyl carrier protein, chloroplast precursor (ACP) (ACP05) (Clone 29C08).	11c3389g02	Similar to Acyl carrier protein, chloroplast precursor (ACP) (ACP05) (Clone 29C08).	98.7
	11g05329	lysosomal alpha-mannosidase precursor, putative, expressed	11c3471g01	lysosomal alpha-mannosidase precursor, putative, expressed	98.5
<u>Lipases</u>					
<i>qSaB1.1</i>	01g09050	GDSL-like lipase/acylhydrolase, putative, expressed	01c6073g02	Lipase, GDSL domain containing protein.	99.7
	01g09095	GDSL-like lipase/acylhydrolase, putative, expressed	01c6102g03	GDSL-like lipase/acylhydrolase, putative, expressed	95.5
<i>qSaB11.1</i>	11g05250	GDSL-like lipase/acylhydrolase, putative, expressed	11c3389g05	GDSL-like lipase/acylhydrolase, putative, expressed	99.7
<u>Translation</u>					
<i>qSaB1.1</i>	01g08992	« SSF51197: Clavaminate synthase-like »	01c6050g06	hypothetical protein Osl_04250 [Os ind.] « PF01283: Ribosomal protein S26e »	37.1
<i>qSaB5.1</i>	05g05794	Pentatricopeptide repeat domain containing protein.			
	05g05805	Similar to Acidic ribosomal protein (Fragment).	05c4600g01	60S acidic ribosomal protein, putative, expressed	100.0
<i>qSaB11.1</i>	11g05209	hypothetical protein LOC_Os11g31650 [Os jap.]	11c3337g01	Ribosomal protein L31 domain containing protein.	40.5
	11g05333	3-5 exonuclease eri-1, putative, expressed			
<u>Senescence</u>					
<i>qSaB1.1</i>	01g09032	SAG20, putative, expressed	01c6069g09	SAG20, putative, expressed	89.7
	01g09081	leaf senescence related protein, putative, expressed	01c6088g04	leaf senescence related protein, putative, expressed	62.5
<u>Transferases</u>					
<i>qSaB5.1</i>	05g05850	Transferase family protein. « SSF52777: CoA-dependent acyltransferases »	05c4655g03	hypothetical protein Osl_20158 [Os ind.] « SSF52777: CoA-dependent acyltransferases »	44.2
<i>qSaB11.1</i>	11g05128	Omega-hydroxypalmitate O-feruloyl transferase	11c3293g08	Similar to transferase.	33.8
<u>Cell division</u>					
<i>qSaB11.1</i>	11g05234	putative BRUSHY1 [Os jap.]			
	11g05253	coiled-coil domain-containing protein SCD2 isoform X1 [Os jap.]	11c3391g01	hypothetical protein Osl_36288 [Os ind.]	76.8
	11g05295	chromophore lyase CRL, chloroplastic [Os jap.]	11c3437g04	chromophore lyase CRL, chloroplastic [Os jap.]	100.0
<u>Miscellaneous</u>					
<i>qSaB1.1</i>	01g09069	lactate/malate dehydrogenase, putative, expressed	01c6083g02	lactate/malate dehydrogenase, putative, expressed	99.5
	01g09037	Epstein-Barr virus EBNA-1-like protein [Os jap.]			
	01g09038	Epstein-Barr virus EBNA-1-like protein [Os jap.]			
<i>qSaB5.1</i>	05g05913	hypothetical protein BAE44_0002966 [Dichanthelium oligosanthes] « PF09778: Guanylylate cyclase »	05c4698g02	guanylyl cyclase, putative, expressed	81.9
<i>qSaB11.1</i>	11g05245	NLI interacting factor-like phosphatase, putative, expressed	11c3388g01	Similar to NLI interacting factor-like phosphatase family protein, expressed.	81.4

and 2OG-Fe(II) oxygenase domain-containing protein, protein kinases, and a cell wall modification enzyme alpha-expansin, OsEXPA2 (Figure 4.9A–B; Table 4.7; Supplementary Table S4.1).

Also, within the region of *qSaB1.1* of both rice genotypes, several proteins contain known protein domains, such as various zing finger domains, phox domain, ankyrin repeat-containing domain and domains of unknown function (DUFs) (Figure 4.9A–B; Supplementary Table S4.1). Hypothetical proteins and TEs respectively accounted for 19 and 8 % of all IR64 proteins predicted within the region, and 12 and 3 % of all CT8556-37-2-3-1-M proteins within the region (Figure 4.9A–B). The lower number of TEs was expected since CT8556-37-2-3-1-M contigs often did not align to the regions of IR64 that the CT8556-37-2-3-1-M read depth was variable (Figure 4.6–4.8) and often coincided with the prediction of TEs on the IR64 genome.

For QTL on chromosome 5 (*qSaB5.1*), the numbers of proteins in different functional groups were again similar between the two rice genotypes (Figure 4.9C–D). Functional groups of genes in the 2 cM distance from the peak included proteins involved in growth and development, such as a no apical meristem (NAM) protein and plastid division protein *ftsZ1*, as well as, plant hormone-related proteins, such as auxin-independent growth promoter and indole-3-acetic acid (IAA)-amido synthetase, transporters including two solute carrier family 35 members and a major facilitator family transporter, biotic/abiotic stress-related proteins, such as a universal stress protein domain-containing protein, a plant disease resistance polyprotein-like protein and a component of the Skp1-cullin 1-F-box (SCF) E3 ubiquitin ligase complex involved in jasmonate-regulated defence responses. There were also genes encoding proteins involved cell wall biogenesis and modification, such as glycosyl hydrolases family 17, proteins involved in transcription, such as TF TGAL5, MYB TF, ethylene-responsive TF ERF094 and a mediator of RNA polymerase II, proteins involved in DNA processes, such as DNA replication (e.g. a DNA repair protein and the DNA replication

complex GINS protein PSF1) and methylation (e.g. 5-methylcytosine DNA glycosylase/lyase); and a cytochrome P450 (Figure 4.9C–D; Table 4.7; Supplementary Table S4.2).

Besides, a few proteins from the region of *qSaB5.1* from both rice genotypes were annotated with protein domains, including zinc finger domains, pleckstrin homology (PH) domain, tetratricopeptide-like helical domain and DUFs (Figure 4.9C–D; Supplementary Table S4.2). Also, 23 and 10 genes for IR64 (Figure 4.9C), and 25 and 6 genes for CT8556-37-2-3-1-M (Figure 4.9D) were annotated as TEs and hypothetical proteins, respectively (Supplementary Table S4.2).

Within the 2 cM distance from the peak of *qSaB11.1* on chromosome 11, similar numbers of proteins in each functional group were found in both rice genotypes (Figure 4.9E–F). Functional groups included abiotic and biotic stress-related proteins, such as calcineurin B-like protein, *Rice stripe tenuivirus* resistance genes, stripe rust resistance protein YR10, chalcone and stilbene synthases and bisdemethoxycurcumin synthases; proteins involved in growth and development, such as NAM proteins, TAPETUM DETERMINANT 1 (TPD1) protein homolog, NO EXINE FORMATION 1 (NEF1) and chloroplastic COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B (CCB1) protein. There were also plant hormone-related proteins, including DELLA protein RHT-1, BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 (BAK1), transport inhibitor response 1 (TIR1)-like protein, auxin response factor 1 and IAA-amido synthetase, as well as transporters, including sugar transporter and root cap protein 1-like protein, a bHLH TF regulating chill tolerance and stomatal development, proteins involved in the biogenesis and modification of cell wall, including omega-hydroxypalmitate O-feruloyl transferase, lysosomal alpha-mannosidase and xyloglucan fucosyltransferase and a large cluster of sulfotransferases (Figure 4.9E–F; Table 4.7; Supplementary Table S4.3).

Additionally, several proteins within the region of *qSaB11.1* were annotated as proteins containing domains, such as armadillo repeats, HGWP repeat, tetratricopeptide repeat

(TPR) and fibrin-binding domain (FBD). Also, 78 and 75 for IR64 (Figure 4.9E), and 33 and 56 for CT8556-37-2-3-1-M genes (Figure 4.9F) were annotated as TEs and hypothetical proteins, respectively (Supplementary Table S4.3).

Finally, in the region of *qSaB11.1* (where it was likely that a sequence deletion had occurred on the CT8556-37-2-3-1-M genome, as mentioned in Section 4.3.3; Figure 4.8), 13 genes were predicted on the IR64 sequence. Two of these were annotated as retrotransposons, one as a protein kinase-like domain-containing protein, seven as hypothetical proteins and the remaining three had no annotation (Supplementary Table S4.3).

4.4 Discussion

This chapter aimed to identify underlying mechanisms and candidate genes around the peaks of the three major QTL (*SaB1.1*, *qSaB5.1* and *qSaB11.1* on chromosomes 1, 5 and 11, respectively) for the resistance phenotype of *S. asiatica* on CT8556-37-2-3-1-M, compared to IR64. Gene prediction and annotation were carried out on IR64 chromosomal sequences and CT8556-37-2-3-1-M contig sequences. The translated amino acid sequences from both rice genotypes were compared. Although different numbers of genes were predicted within 2 cM distance from the peaks of the QTL, all three QTL contained genes from similar functional groups, such as genes encoding transmembrane transporters, development-related proteins, and hormone-mediated and biotic and abiotic stress signalling pathways.

These functional groups highlighted important processes that were in line with the hypotheses outlined in the Introduction (Section 4.1). The diversity of functions of genes, in all three QTL, and the similarity of the functional types of genes across multiple QTL, such as genes encoding transporters in all three QTL and genes encoding IAA-amido synthetases in both *qSaB5.1* and *qSaB11.1*, were consistent with the results from Chapter 3 in which each of the three QTL additively contributed to the phenotype. The lack of epistatic interactions between QTL suggested that the causal genes from each QTL function independently or work via separate pathways in eliciting the phenotype.

4.4.1 Could limitations in the acquisition of nutrients contribute to the slow growth of parasites on CT8556-37-2-3-1-M?

The first hypothesis was that *S. asiatica* individuals from CT8556-37-2-3-1-M received less water and nutrients from the host, thus they grew more slowly, compared to those from IR64. Two amino acid transporters, two ammonium transporters and a nitrate reductase were annotated within 2 cM distance from the peaks of *qSaB1.1* and *qSaB11.1* (Table 4.7). These proteins are involved in the uptake and translocation of nitrogen in the host and to *Striga* parasites. *Striga* plants completely rely on nitrogen from the host for growth and survival. Rice takes up nitrogen from the soil in the forms of both nitrate and ammonium (Yi *et al.*, 2019). Uptake of nitrate begins with active transport from the soil into root cells via nitrate transporters (Bloom, 2015). Nitrate is then reduced into nitrite by nitrate reductase in the cytosol before the nitrite is transported into plastids where it is further reduced into ammonium and assimilated into amino acids (Noguero & Lacombe, 2016; Tegeder & Rentsch, 2010). Alternatively, nitrate is also transported from roots to be assimilated into amino acids in leaves via xylem vessels (Noguero & Lacombe, 2016; Tegeder & Rentsch, 2010). Ammonium is also actively taken up from the soil, by ammonium transporters, and can be assimilated into amino acids in roots (Tabuchi *et al.*, 2007). Besides, plants can also take up amino acids directly from the soil via amino acid transporters (Näsholm *et al.*, 2009; Paungfoo-Lonhienne *et al.*, 2008).

Ammonium and amino acids are also transported away from roots via xylem vessels. Xylem loading of amino acids involves apoplastically release of amino acids from endodermis cells and cells inside root stele via amino acid transporters (Yao *et al.*, 2020). The movement of ammonium across the plasma membrane also requires ammonium transporters (Schjoerring *et al.*, 2002). In most plants, the majority of amino acids found in xylem sap are asparagine and glutamine (Yao *et al.*, 2020). A study in *S. hermonthica*–sorghum system, Pageau *et al.* (2003) identified the forms of nitrogen that were translocated in xylem sap and received by the parasite to be mainly nitrate and free amino acids, such as asparagine, glutamine and

glutamate, using nitrogen isotope labelling. The high concentration of host-derived nitrate was found in *S. hermonthica* xylem sap. The nitrate was reduced and stored as nitrogen-rich asparagine in the shoot of parasites once they are emerged (Pageau *et al.*, 2003).

Although the nitrate reductase genes in *qSaB11.1* from the two rice genotypes had identical amino acid sequences (Table 4.7), a higher expression level of the gene in CT8556-37-2-3-1-M could reduce the availability of nitrates and increase the concentration of ammonium in the host xylem sap. The growth of *S. hermonthica* was severely hindered when the parasite was supplied mainly with ammonium as the source of nitrogen in aseptic culture without a host (Igbinnosa & Thalouarn, 1996). The reduction in growth was attributed to *S. hermonthica* inability to detoxify accumulated toxic ammonium (Simier *et al.*, 2006; Igbinnosa & Thalouarn, 1996). Under normal circumstances, parasites turn excess ammonium into asparagine and store in their leaves, leaving subterranean parasites more vulnerable to ammonium toxicity.

Furthermore, the difference in amino acid sequences (e.g. between IR64_01g09030 and CT8556_01c6069g07 which encode an amino acid transporter) or differential expression of the amino acid transporters and ammonium transporter genes in *qSaB1.1* could impact the availability of nitrogen for the parasite via changes in the concentration of amino acids in host xylem sap. In rice, Swarbrick *et al.* (2008) analysed global gene expression pattern using microarrays and demonstrated that genes encoding permeases and transporters of amino acids and ammonium were transcriptionally upregulated in the susceptible rice genotype IAC165 after *S. hermonthica* infection. Besides, Huang *et al.* (2012) also showed the upregulation of the transcripts of transporter genes, including amino acid permeases and ammonium transporters, in a compatible interaction between the cowpea genotype B301 and *S. gesnerioides* race SG4z, using microarrays. The transcriptional upregulation of these transporter genes was thought to facilitate parasite nutrient extraction (Huang *et al.*, 2012).

Also, the acquisition of nutrients could be limited by inefficient or fewer connections made between the parasite and the host. Several genes were annotated as proteins involved cell wall biogenesis and modification within the three QTL (Table 4.7). Upon attaching of haustoria to host epidermis, parasitic plants weaken host cell walls by secretion of plant cell wall degrading enzymes, such as cellulases, pectinases and xylanases, while penetrating host root via intrusive cells before the formation of xylem continuity (Mitsumasu *et al.*, 2015). The ability of the host to repair or resist parasite-elicited cell wall modification could lead to a reduction in the number of connections, which in turn lead to restricted flow of water and nutrients. While it was shown in Chapter 2 that *S. asiatica* could form xylem connections to both IR64 and CT8556-37-2-3-1-M rice genotypes, the number and the efficiency of these connections could not be assessed with the microscopy technique used. IR64_01g08991 from *qSaB1.1* was annotated as the α -expansin OsEXPA2, which is involved in cell wall loosening. The parasite could upregulate this gene to disrupt the integrity of the cell wall. A gene from *qSaB5.1* and two genes from *qSaB11.1* were annotated as proteins involved in biosynthetic processes of xylan, suberin and xyloglucan, which are cell wall components. Changes in the expression of these genes can result in alteration in the composition of the cell wall, affecting the integrity of the cell wall and allowing the parasite to form xylem–xylem connections more easily.

4.4.2 Could differences in host hormone status/regulation indirectly affect the growth of the parasite?

The second hypothesis was that the morphological changes or the water status of CT8556-37-2-3-1-M rice after *S. asiatica*-infection, caused by developmental genes or host hormone balance, did not fully support the growth of the parasite, compared to IR64. Several genes were annotated as genes encoding hormone-related proteins, and growth and development-related proteins within the three major *S. asiatica*-resistance QTL (Table 4.7). Within *qSaB1.1* on chromosome 1, IR64_01g09051 and CT8556_01c6073g03 were annotated as a receptor of the plant hormone ABA (Table 4.7). ABA induces the closure of stomata in

response to drought stress, resulting in the reduction of transpiration (Finkelstein, 2013). *Striga* infection causes a rise in the level of ABA in the host, which results in drought stress-like symptoms and stomatal closure in the host (Fujioka *et al.*, 2019b; Frost *et al.*, 1997). Swarbrick *et al.* (2008) used microarrays to show that ABA-responsive genes were upregulated in the susceptible rice genotype IAC165 upon being infected by *S. hermonthica*. *Striga* plants also have an elevated ABA concentration — two orders of magnitude higher than the host (Fujioka *et al.*, 2019b) — and maintain a high rate of transpiration, even under drought conditions (Inoue *et al.*, 2013), by keeping their stomata open as they are ABA insensitive through an abnormal ABA signal transduction (Fujioka *et al.*, 2019b). The discrepancy in the transpiration rates between the host and the parasite is thought to be the driving force behind the translocation of water and nutrients from host to parasite via xylem bridges (Fujioka *et al.*, 2019a; Ackroyd & Graves, 1997; Taylor *et al.*, 1996).

Also, several genes from *qSaB5.1* and *qSaB11.1* were annotated as genes encoding proteins involved in auxin signalling pathway (Table 4.7), including the AUX1-like auxin influx carrier OsAUX3 (Wang *et al.*, 2019), an auxin independent growth promoter, the auxin receptor OsTIR1 (Guo *et al.*, 2020), the auxin response factor OsARF1 (Attia *et al.*, 2009) and three IAA-amido synthetases, which are GH3 family proteins that catalyse the conjugation of auxin to amino acids. Auxin is an important plant hormone controlling all aspects of normal growth and development (Majda & Robert, 2018; Velasquez *et al.*, 2016; Grones & Friml, 2015; Perrot-Rechenmann, 2010). Auxin is involved in cell elongation via polarised growth by promoting loosening up of cell wall and an increase in turgor pressure that prompt cellular expansion (Majda & Robert, 2018; Velasquez *et al.*, 2016; Ren & Gray, 2015; Perrot-Rechenmann, 2010). An alteration in the level of active auxin in the host through the conjugation of IAA (rendering IAA inactive) or changes in auxin transport could affect the morphology and water status of the host, resulting in ineffective translocation of water and nutrients to the parasite. Louden (2017) used quantitative PCR to identify the upregulation of the IAA-amido synthetase gene *OsGH3.2* and the small auxin-up RNA gene *OsSAUR39*

in the stem tissues of the susceptible rice genotype IAC165, after being infected with *S. hermonthica*. *OsGH3.2* and *OsSAUR39* negatively regulate the level of free auxin in the host. Overexpression of either of them led to morphological changes linked to IAA deficiency and *Striga* infection, including stunting and reduced shoot growth (Du *et al.*, 2012; Kant *et al.*, 2009). Possible non-responsive to *Striga*-infection IAA-amido synthetases in CT8556-37-2-3-1-M could sustain shoot growth and limit the pool of water and nutrients available for the parasite in the roots.

Besides, auxin transporters have been shown recently to play a crucial role in the formation of xylem bridge connecting vascular systems between the facultative parasite *P. japonicum* and the host *A. thaliana* (Wakatake *et al.*, 2020). Application of the auxin influx inhibitor 3-chloro-4-hydroxyphenylacetic acid to the parasite haustoria resulted in an aberrant formation of xylem connections to the host (Wakatake *et al.*, 2020). Genes encoding LAX auxin influx transporters were found to be highly expressed at the tip of haustorium and the site of xylem plate towards the centre of the parasite root (Wakatake *et al.*, 2020). These influx transporters facilitate the formation of xylem bridge by maintaining the high concentration of auxin at the two sites (Wakatake *et al.*, 2020). In an older study, *A. thaliana* which had been transformed to express the IAA-lysine synthetase from *Pseudomonas savastanoi* to decrease the level of free auxin resulted in a reduction in the xylem–xylem connections between it and the facultative parasite *Triphysaria* sp. (Goldwasser *et al.*, 2002). Although it was shown in Chapter 2 that there was no significant difference in the frequency of the formation of xylem continuity between *S. asiatica* and either of the two rice genotypes, the quantity and quality of each attachment could not be assessed with the technique used. It was plausible that the difference in parasite growth originated from fewer or inefficient connections between the parasites and CT8556-37-2-3-1-M. A lower concentration of free IAA from more auxin conjugation activity in CT8556-37-2-3-1-M than in IR64 could cause it.

4.4.3 Is there any evidence for the production of toxic compounds in rice genotype CT8556-37-2-3-1-M?

The third hypothesis proposed was that CT8556-37-2-3-1-M could produce a toxic compound or compounds that inhibited the growth of *S. asiatica*. From considering the predicted genes within the 2 cM distance from the peaks of the three QTL, no genes were found to have the annotations that fitted this hypothesis.

However, there was a transporter gene that was annotated with a detoxification domain. IR64_01g09033 and CT8556_01c6069g10 from *qSaB1.1* were annotated as heavy metal transport/detoxification protein domain-containing protein (Os01g0826000; Table 4.7). In another pathosystem, the PI21 protein, which contains a heavy metal-transport/detoxification domain, was identified to confer quantitative resistance in rice against the *Magnaporthe oryzae* fungus, which causes blast disease (Fukuoka *et al.*, 2009). Rice plants carrying a mutated copy of the *Pi21* gene were partially resistant to the fungal infection (Fukuoka *et al.*, 2009). The fungus could penetrate host cells with its hyphae but the rate at which the invasion of neighbouring host cells was significantly slower in resistant hosts, reflecting in slow hyphal growth (Fukuoka *et al.*, 2009). There was a parallel between the slow hyphal growth and the slow *Striga* growth. The molecular mechanism of the *Pi21* gene remains largely unknown, but transcriptional profiling of the *Pi21*-RNA interference (RNAi) line revealed the involvement of jasmonate/ethylene signalling pathway through vigorous pattern-triggered immunity (Zhang *et al.*, 2016). Interestingly, Os01g0826000 was also upregulated in rice plants challenged with the same blast fungus, *M. oryzae* (Agrawal *et al.*, 2002).

4.4.4 Host defence responses against *Striga* infection

In addition to the three hypotheses, relating to the slow growth phenotype of *S. asiatica* on CT8556-37-2-3-1-M was the number of parasites, which was also lower on this rice genotype. The biomass of *S. asiatica* and the number of parasites were strongly correlated, as described

in the phenotypic screen of RILs in Chapter 3 (Section 3.2.2.3). Some of the underlying genes of the QTL mapped using parasite biomass in Chapter 3 might have contributed to post-attachment resistance.

Infection of host roots by parasitic plants such as *Striga* spp. can lead to a host defence response. Several genes encoding proteins involved in defence against pathogens were annotated within the three QTL, especially *qSaB5.1* and *qSaB11.1*. Among these were three versions of *Rice stripe tenuivirus*-resistance genes, Os11g0514000 (IR64_11g05187 & CT8556_11c3320g01), *STV11* (IR64_11g05189) and *STV11-pa1* (IR64_11g05192 & CT8556_11c3327g01). *STV11*, which is a sulfotransferase, elicits *Rice stripe tenuivirus*-resistance by turning SA into sulfonated SA, which prevents the replication of the virus (Qin *et al.*, 2019; Q. Wang *et al.*, 2014). SA signalling pathway is induced during infection of *S. hermonthica* in rice, and differences in the ratio of SA and jasmonic acid (JA) are involved in resistance in Nipponbare to *S. hermonthica* (Mutuku *et al.*, 2015).

Genes in *qSaB5.1* (IR64_05g05858 and CT8556_05c4660g03) were annotated as *OsCOI1b*, encoding a component of the SCF E3 ubiquitin ligase complex (Table 4.7). *OsCOI1b* is a principle component of JA receptor (Yang *et al.*, 2012). The JA-mediated signalling pathway is important in basal defence against herbivores and necrotrophic pathogens in rice (Yang *et al.*, 2013). *OsCOI1b*-RNAi transgenic rice was less sensitive to the application of methyl jasmonate (MeJA) and had lower expression levels of JA-responsive genes (Yang *et al.*, 2012). For parasitic plant interactions, activation of JA-mediated signalling pathway through an application of MeJA increase the level of resistance to *S. hermonthica* in rice (Mutuku *et al.*, 2015). Differences in expression levels of *OsCOI1b* or differences in its sensitivity to JA between IR64 and CT8556-37-2-3-1-M could explain the different levels of susceptibility of the two rice genotypes.

4.4.5 Lack of a complete genome sequence for CT8556-37-2-3-1-M hindered the comparison of the two rice genomes.

One of many ways to identify candidate genes underlying the QTL is to identify polymorphisms both in the coding regions and the regulatory regions (Members of the Complex Trait Consortium, 2003). This can be achieved by comparing the genomes of the two parental genotypes of the mapping population. This was complicated by the unavailability of the complete, assembled genome sequence of CT8556-37-2-3-1-M, which is arguably the more important parental genotype as CT8556-37-2-3-1-M alleles were responsible for the slow growth phenotype of the parasite in most of the mapped QTL. In this study, Illumina sequences from CT8556-37-2-3-1-M were assembled into contigs, which were then mapped onto the IR64 genome. Although this method was unable to identify all genes unique to CT8556-37-2-3-1-M, many genes would likely be present in both rice genomes. Only ~ 40 % of the genes predicted on both IR64 genome and CT8556-37-2-3-1-M contigs were identical or near-identical (> 98 % identity) at the amino acid sequence level, leaving the rest with polymorphisms that may affect their functions as potential candidate genes underlying the differences in phenotype.

It is well known that different gene prediction algorithms produce different gene models from the same sequence as it was shown clearly in this study, underlining the necessity to use a comprehensive gene prediction and annotation pipeline. Transcriptomic data was used to verify the gene predictions in the present study. Although newer approaches that incorporate transcriptomic data into the gene prediction algorithms were available (Keilwagen *et al.*, 2018; Hoff *et al.*, 2016), they were not chosen because only the genome sequence and RNA-Seq data of IR64 were available whilst those of CT8556-37-2-3-1-M were not, which would render comparison of resulting gene predictions of the two rice genotypes uninterpretable. Also, the gene prediction on CT8556-37-2-3-1-M contigs resulted from running prediction algorithms on such short sequences, which is not a standard practice since it is more error-prone than gene prediction over whole chromosomal sequences (Bouchot *et al.*, 2013).

Hence, the differences in the number of genes, the positions and the annotations of predicted genes on the two rice genomes must be interpreted with care. Apparent differences in the amino acid sequences of predicted genes from the two rice genotypes could arise from the prediction errors. A better practice is to first assemble CT8556-37-2-3-1-M sequences using longer reads such as those from PacBio or Nanopore sequencing, and then compare the sequences of the two rice genotypes. When the genome sequence of CT8556-37-2-3-1-M becomes available, a new comparison of the genes with the complete region of all six significant QTL will be undertaken.

4.4.6 Prospects

Quantitative resistance for plant pathogens are complex and are controlled by a vast array of genes with diverse functions, not only limited to the detection of pathogen invasion (Corwin & Kliebenstein, 2017; Pilet-Nayel *et al.*, 2017; Poland *et al.*, 2009). Examples of resistance underlying genes include genes regulating development and morphology of the host, downstream defence signal transduction genes, genes controlling cell wall synthesis and modification and genes involved in defence mechanisms, such as the biosynthesis of defence compounds (Corwin & Kliebenstein, 2017; Poland *et al.*, 2009). Despite numerous discoveries of QTL for quantitative disease resistance, the underlying mechanisms and causal genes were identified for only a handful of these QTL due to the complexity of the traits (Corwin & Kliebenstein, 2017; Pilet-Nayel *et al.*, 2017).

The slow growth phenotype of *S. asiatica* was a highly quantitative trait that is likely to be controlled by a myriad of genes with diverse functions, each with a minor effect, additively contributing to the phenotype. It is without a doubt that more genes within the three QTL other than the more likely ones, such as nutrient transporters that were discussed earlier, are possible candidate genes underlying the *S. asiatica*-resistance. To decrease the number of candidate genes within the QTL, measurement of expression levels of the genes within the QTL would clarify which genes in which pathways are upregulated or downregulated on

either IR64 or CT8556-37-2-3-1-M and are more likely to be involved in the reduction of the growth of parasites. This approach of combine mapping with gene expression data has been successful in the identification of candidate genes controlling traits of interest (Huang *et al.*, 2012; Swarbrick *et al.*, 2008; Wayne & McIntyre, 2002). This approach was not pursued in this study because of time constraints and the unavailability of the assembled CT8556-37-2-3-1-M genome onto which the transcripts will be mapped. Once a smaller set of candidate genes are identified, verification of gene functions through a knockout or knockdown of candidate genes is required to confirm their involvement in the control of the trait. This can be done through techniques such as gene knockout via CRISPR/Cas9 technology or gene knockdown via RNAi of candidate resistance genes in CT8556-37-2-3-1-M background and assessing whether the susceptibility to *S. asiatica* increases or not. Conversely, knockout or knockdown of candidate susceptibility genes in IR64 background can be carried out to assess whether the susceptibility to the parasite decrease or not.

Chapter 5

A genome-wide association study to identify resistance to *Striga hermonthica* in the African rice species, *Oryza glaberrima*

5.1 Introduction

Oryza glaberrima, also known as African rice, is one of the two domesticated rice species, along with Asian rice (*O. sativa*). *Oryza glaberrima* was independently domesticated in the Inner Niger Delta in West Africa ~ 3000 years before present (Cubry *et al.*, 2018; Meyer *et al.*, 2016; M. Wang *et al.*, 2014). Over the past 500 years, the cultivation of *O. glaberrima* has been on a decline because of the introduction of *O. sativa* varieties, which produce higher yields and are less labour intensive to cultivate (Cubry *et al.*, 2018; Linares, 2002). *Oryza glaberrima* germplasm contains valuable genetic materials for resistance to biotic and abiotic stresses, which are locally present in Africa, and was used to cross with *O. sativa* to create stress-resistant interspecific NERICA rice varieties (Jones *et al.*, 1997a, b).

Striga infestation is one of the greatest challenges to the production of cereal crops in sub-Saharan Africa (Scholes & Press, 2008). Many studies have found that *O. glaberrima* genotypes are a good source of resistance to *S. hermonthica*, *S. asiatica* and *S. aspera* Benth. Johnson *et al.* (1997) showed that five *O. glaberrima* genotypes (M27, T2, ACC102196, Makassa and IG10) supported fewer *S. aspera* and *S. hermonthica* plants than *O. sativa* genotypes in fields in Côte d'Ivoire. The highly resistant *O. glaberrima* genotype CG14, which was used as the *O. glaberrima* parental genotype in the development of NERICA rice varieties, was revealed to have both pre- and post-attachment resistance against *S. asiatica* and *S. hermonthica* (Cissoko *et al.*, 2011; Jamil *et al.*, 2011a). Jamil *et al.* (2011a) found that CG14 produced less *Striga* germination stimulants (strigolactones) resulting in lower infection of *S. hermonthica*. Cissoko *et al.* (2011) demonstrated that only a few *S. hermonthica* and *S. asiatica* individuals, of different accessions of parasite seeds, could infect resistant CG14 and had low biomass on the host. Most parasites attached and penetrated the root cortex but failed to breach the endodermis of CG14 (Cissoko *et al.*, 2011). Moreover, Kaewchumnong & Price (2008) found that no *S. hermonthica* plants emerged from pots with CG14 even at 100 DAP. Furthermore, Rodenburg *et al.* (2015) also discovered that CG14 was resistant in fields infested with *S. asiatica* in Kyela, Tanzania and *S. hermonthica* in Mbita, Kenya.

Another field screen by Rodenburg *et al.* (2017) also identified more *O. glaberrima* genotypes, such as MG12, Anaklia and Agee, that were resistant to *S. hermonthica* in infested fields in Mbita, Kenya and Namutumba, Uganda. Cissoko (2012) also showed that MG12 and TOG5681 were highly resistant to *S. hermonthica* from different locations and the resistance in these two genotypes was post-attachment. The discovery of the genetic basis underlying the resistance to *Striga* in these *O. glaberrima* genotypes would contribute to the understanding of complex *Striga*–host interactions and identify targets for *Striga*-resistance breeding in rice varieties.

GWAS is a powerful tool for the identification of loci that are associated with the trait of interest in a natural population, taking advantage of abundant historical recombination events which allow for the discovery of associated loci at a finer scale than in linkage mapping (Nordborg & Weigel, 2008). Recently, GWASes have been successfully used in mapping of the undying loci for resistance to *S. hermonthica* in maize (Adewale *et al.*, 2020) and sorghum (Kavuluko *et al.*, 2020). Adewale *et al.* (2020) identified 24 SNPs that were significantly associated with *S. hermonthica*-tolerance and *S. hermonthica*-resistance traits, such as the grain yield under *S. hermonthica* infestation, the parasite damage score and the number of parasites, using a panel of 132 early-maturing maize genotypes that were screened in artificially infested fields. The study identified candidate genes, such as *ZmCCD1* and *amt5* gene which are involved in mycorrhizal association through a change in the secretion of strigolactones and the transport of ammonium ions, respectively (Adewale *et al.*, 2020). In sorghum, Kavuluko *et al.* (2020) identified candidate genes controlling post-attachment resistance to *S. hermonthica*, using a collection of 206 sorghum genotypes that were phenotyped in rhizotrons under laboratory conditions. The identified candidate genes included genes encoding proteins involved in the modification of the host cell wall, the activation of hormone-mediated innate immune responses and the biosynthesis and the delivery of secondary metabolites, such as isoflavone (Kavuluko *et al.*, 2020). The same technique could be used to dissect resistance to *Striga* in *O. glaberrima*.

Recently, an effort to locate the domestication centre and to study the evolution of *O. glaberrima* genotypes made the genotypic data of a diverse panel consisting of 246 *O. glaberrima* genotypes publicly available (Cubry *et al.*, 2018). After the whole genomes of the 246 *O. glaberrima* genotypes were sequenced, SNPs and indels were identified (Cubry *et al.*, 2018). Also, a subset of the *O. glaberrima* panel, consisting of 163 genotypes, has been used for the identification of genetic loci associated with flowering time, the architecture of panicles and resistance to *Rice yellow mottle virus*, using GWASes (Cubry *et al.*, 2020). This *O. glaberrima* diversity panel was suitable for the detection of *S. hermonthica*-resistance using genome-wide association mapping.

5.1.1 Aim and objectives

This chapter aimed to identify SNPs associated with resistance in *O. glaberrima* genotypes to *S. hermonthica* (Kibos accession), using a GWAS. To achieve the aim: (i) phenotypic scores of *S. hermonthica* resistance traits (obtained prior to this PhD) were analysed for their repeatability; (ii) the conditional mean phenotypic scores that represent the level of resistance for *O. glaberrima* genotypes were obtained by quantifying and taking into account inter-batch variation; (iii) the marker information of the *O. glaberrima* genotypes was retrieved from the **AfricanRice** database hosted by the IRD server and prepared for subsequent analyses by filtering out low-quality markers; (iv) the kinship and the population structure of the *O. glaberrima* genotypes were assessed and the genome-wide association analysis for different resistance traits (the biomass of *S. hermonthica* and the number of parasites per rice plant) was performed.

5.2 Methodology

5.2.1 Plant materials

An accession of *S. hermonthica* seeds was collected from a local population of *S. hermonthica* individuals parasitising maize plants in farmers' fields in Kibos, Kenya in 2009.

Seeds of 163 diverse *O. glaberrima* rice genotypes (Orjuela *et al.*, 2014) from IRD, Montpellier, France were provided by Dr Mathias Lorieux. The details of these rice genotypes can be found in Supplementary Table S5.1. Cubry *et al.* (2018) sequenced a subset of the *O. glaberrima* genotypes from the work by Orjuela *et al.* (2014), which included 154 genotypes in the current study. The rice genotypes were sequenced using Illumina resequencing at 37-fold coverage. The SNPs and indels information of the sequenced *O. glaberrima* genotypes were available in the **AfricanRice** database on a server at IRD. Seeds of two *O. sativa* rice genotypes, IAC165 [*O. sativa* subsp. *japonica* (subgroup *tropical japonica*)] and Koshihikari [*O. sativa* subsp. *japonica* (subgroup *temperate japonica*)], were obtained from NIAS, Ibaraki, Japan.

5.2.2 Preparation of phenotypic data

The experimental design, data preparation and statistical analyses of data are shown in Figure 5.1. The design and execution of the phenotypic screen for *S. hermonthica* resistance took place before the start of this PhD programme. A total of 163 *O. glaberrima* genotypes were divided into three groups, each contained 54–55 genotypes. The *O. glaberrima* genotypes within each group were phenotyped for resistance to *S. hermonthica* in four separate batches. (Twelve batches of plants were screened in total.) Each *O. glaberrima* genotype was replicated four times with one replicate in each batch. Also, eight replicates of *O. sativa* cv. IAC165 and eight replicates of *O. sativa* cv. Koshihikari were also phenotyped in each batch. These two *O. sativa* varieties were intended to be used as susceptible checks to normalise possible differences in phenotypic scores across batches. In total, 864 rice plants were screened. The phenotypic screen was performed according to the protocol described in Section 2.2.2. Twenty-one days after inoculation of the rice roots with germinated *S. hermonthica* seeds, the root systems were either photographed using a Nikon COOLPIX P100 digital camera or scanned using a Canon CanoScan 9000F scanner. *Striga hermonthica* individuals were then harvested from the rice roots as described in Section 2.2.3 and the biomass (dry weight), cumulative length and number of parasites were recorded.

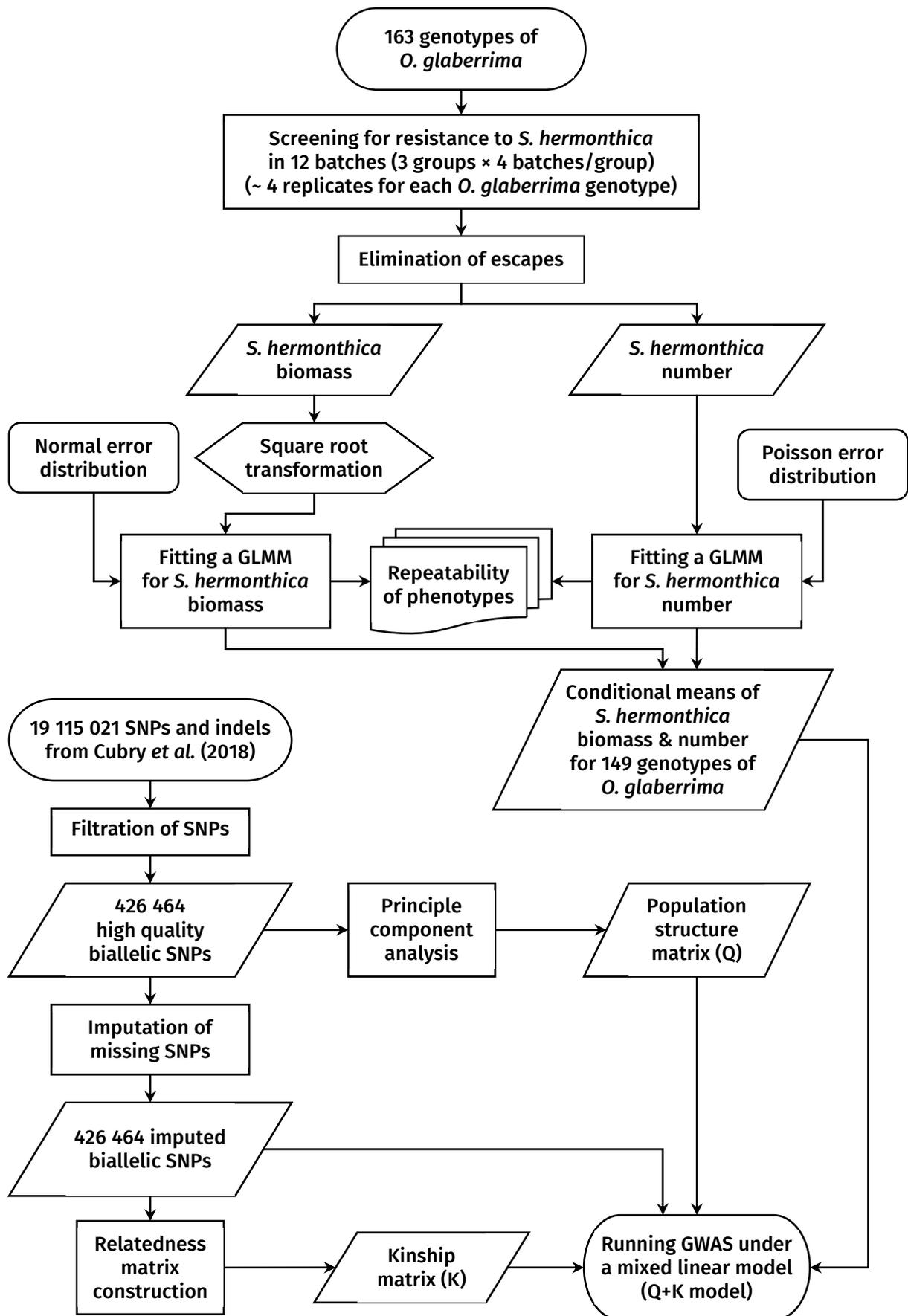


Figure 5.1 Flowchart of experimental design, data preparation, statistical and association analyses. GLMM = generalised linear mixed-effect model.

In some cases, rice plants did not grow properly. The root scans of the four replicate plants of each rice genotypes were inspected and any with aberrant root architecture were removed from the dataset before data analyses as described in Section 3.2.3.2. The omission of replicate plants resulted in the removal of 100 rice plants from 43 *O. glaberrima* genotypes and the complete exclusion of 7 *O. glaberrima* genotypes, leaving 156 *O. glaberrima* genotypes in subsequent analyses.

5.2.2.1 Selection of the traits to represent the resistance phenotype

As detailed in Section 3.2.2.3, biomass (dry weight), cumulative length and number of parasites are commonly used to quantify the level of resistance to *S. hermonthica*. There was a very strong positive correlation ($r = 0.917$, $n = 764$, $P < 0.001$) between the biomass and the cumulative length of the parasites (Figure 5.2A), suggesting that association analyses of both traits would likely result in the same set associated genetic loci. Hence, the cumulative length was not used in following analyses. On the other hand, the positive significant correlation between the biomass and number of parasites was also strong, but weaker, ($r = 0.763$, $n = 764$, $P < 0.001$) (Figure 5.2B), and therefore the number of parasites might yield an association that was not detected by using the parasite biomass. As a result, subsequent association analyses were carried out on both the biomass and the number of parasites.

5.2.2.2 Removal of *Oryza sativa* checks from analyses

Two *O. sativa* varieties, IAC165 and Koshihikari, were included in all twelve batches of the phenotypic screen to normalise possible differences in phenotypic scores across batches. However, these checks were excluded from the dataset for several reasons. Firstly, there was a high level of variation in the susceptibility of replicates of both IAC165 and Koshihikari, within a batch. The intraclass correlation coefficients (ICCs) of both the biomass and number of parasites on either IAC165 or Koshihikari from different batches were less than 0.5 (Table 5.1). ICCs for the parasite biomass were much lower than the ICCs for the numbers of parasites. Low ICCs suggested that the phenotypic scores of the replicates of each

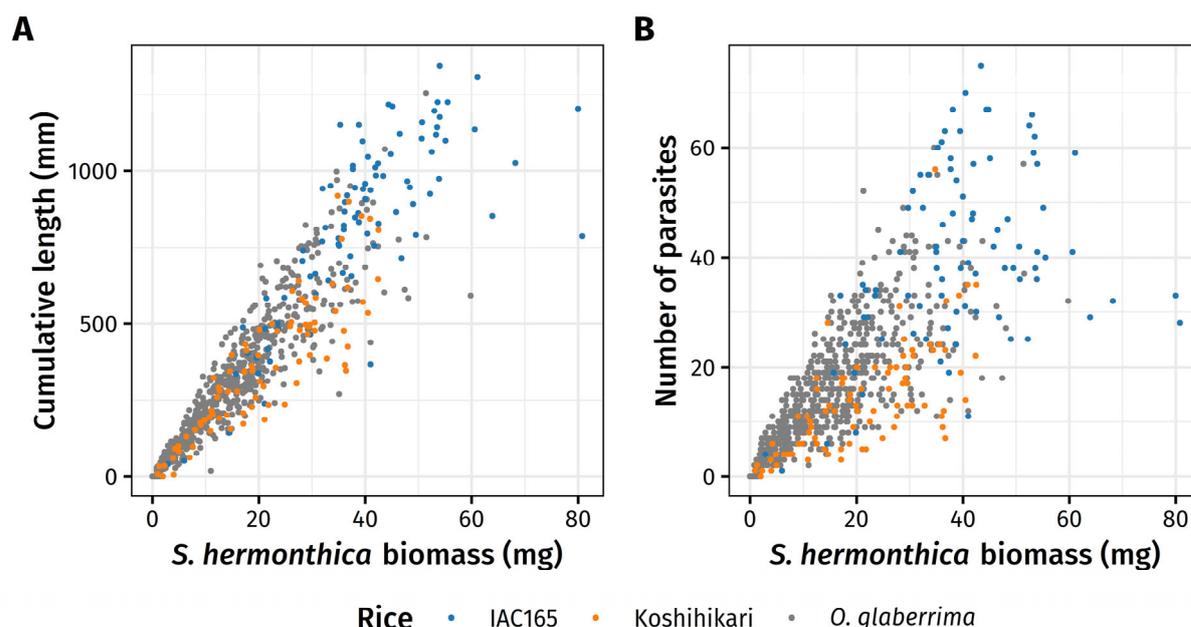


Figure 5.2 Relationships between the biomass, the cumulative length and the number of *Striga hermonthica* individuals per host. (A) A very strongly positive correlation ($r = 0.917$, $n = 764$, $P < 0.001$) between the biomass and the cumulative length of *S. hermonthica* plants harvested from each rice host. (B) A strong, but weaker, positive correlation ($r = 0.763$, $n = 764$, $P < 0.001$) between the biomass and the number of *S. hermonthica* individuals harvested from each rice host. Colours of the symbols correspond to the rice species/varieties: blue, *Oryza sativa* cv. IAC165; orange, *O. sativa* cv. Koshihikari; grey, *O. glaberrima* genotypes.

check in each batch had a poor tendency to be similar to each other. The ICCs were calculated using a GLMM with an identity link function and a normal error distribution on the square root transformed data for the parasite biomass, and a GLMM with a log link function and a Poisson error distribution on the untransformed data for the number of parasites, as described in Section 3.2.3.4.

Table 5.1 Intraclass (intra-batch) correlation coefficient for resistance to *Striga hermonthica* of *Oryza sativa* check varieties, Koshihikari and IAC165.

Trait	Checks	Intra-batch correlation coefficient	Confidence Interval*	P-value [†]
Biomass of parasites	Koshihikari	0.282	[0.039, 0.506]	< 0.001
	IAC165	0.193	[0.000, 0.406]	0.004
Number of parasites	Koshihikari	0.384	[0.076, 0.595]	0.006
	IAC165	0.436	[0.125, 0.630]	0.001

* 95 % confidence intervals were calculated from parametric bootstrap (10 000 iterations).

[†] P-values were calculated from a permutation test (10 000 permutations).

Secondly, IAC165 was far more susceptible to *S. hermonthica* than any of the *O. glaberrima* genotypes (Figure 5.3). The interquartile ranges of both the biomass and number of parasites from IAC165 and those from *O. glaberrima* genotypes had no overlaps, except for Group 3 Batch 3 where IAC165 was, unusually, as susceptible as the *O. glaberrima* genotypes in terms of the parasite biomass (Figure 5.3A) and less susceptible than the *O. glaberrima* genotypes in terms of the number of parasites (Figure 5.3B). The levels of resistance to *S. hermonthica* between IAC165 and *O. glaberrima* genotypes did not scale linearly. Therefore, to use this variety to normalise for inter-batch variation was inappropriate.

Thirdly, Koshihikari could not be used as a check because its levels of susceptibility were highly variable with respect to the levels of susceptibility of the *O. glaberrima* genotypes from different batches within the same group (Figure 5.3). For example, Koshihikari was more susceptible to *S. hermonthica* than the *O. glaberrima* genotypes in Batch 1 and 2 of Group 3 but was more resistance to the same *O. glaberrima* genotypes in Batch 3 and 4 of the same group. Subsequent analyses were performed only on *O. glaberrima* data.

5.2.2.3 Assessment of the phenotypic consistency of the *Oryza glaberrima* genotypes between batches

The pipeline that was developed in Section 3.2.2.4 was used to assess the consistency of the phenotypic scores of both the biomass and the number of parasites. The repeatability of the *O. glaberrima* genotypes was calculated using GLMMs (Equation 3.1). To ensure that the assumptions about the distribution and variance of the models were observed, diagnostic plots were constructed to assess the suitability of different data transformations and selections of the error distribution. For the parasite biomass, GLMMs with a normal error distribution and an identity link function were fitted to either the untransformed data or the square root transformed data (Figure 5.4). The Shapiro–Wilk’s test of normality showed that the distribution of the untransformed biomass of parasites was significantly different from a normal distribution ($W = 0.948$, $n = 589$, $P < 0.001$), whilst the distribution of the square root transformed data did not significantly differ from a normal distribution ($W = 0.997$,

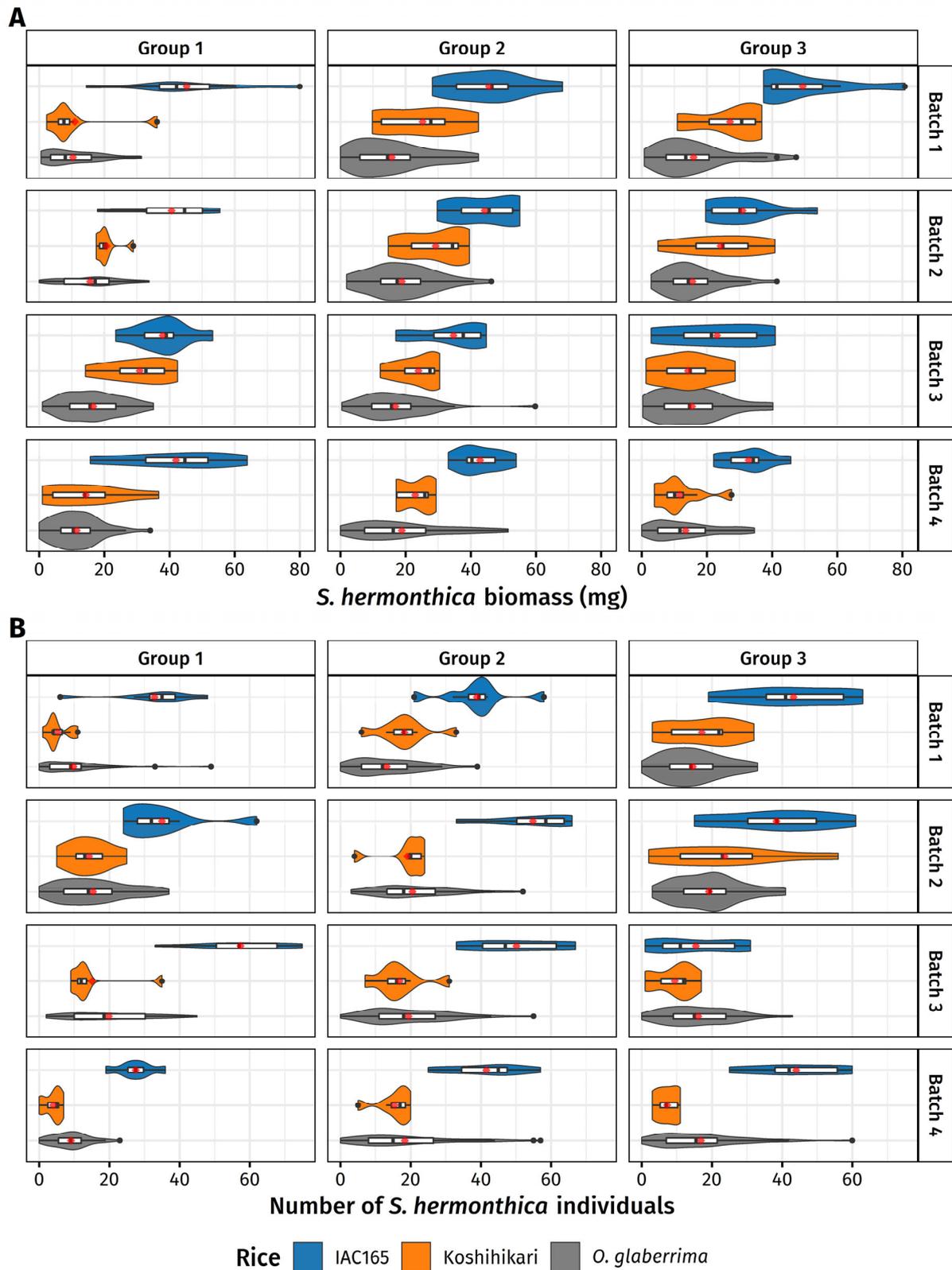


Figure 5.3 Comparison of the susceptibility of the *Oryza glaberrima* genotypes and the two *O. sativa* checks (IAC165 and Koshihikari) to *Striga hermonthica*. Violin and box plots of the biomass (A) and number (B) of *S. hermonthica* individuals per host are shown in groups and batches in which the host plants were screened. Red diamonds mark the measurement means. Coloured areas in the same panel have an equal area. Blue, IAC165 ($n = 6$ to 8); orange, Koshihikari ($n = 7$ to 8); grey, *O. glaberrima* genotypes ($n = 45$ to 59).

$n = 589$, $P = 0.332$). This was also supported by the frequency histograms of residuals (Figure 5.4A–B) and the Q–Q plots (Figure 5.4C–D). Both GLMMs also showed linearity between the responses of the link function (identity function) and the predictors (Figure 5.4E–F). The square root transformation achieved homoscedasticity of data by eliminating a systematic increase of variance as the values of the parasite biomass increased which was found in the untransformed data (Figure 5.4G–H). Therefore, subsequent analyses were carried out using the GLMM that fitted the square root–transformed data for the biomass of *S. hermonthica* per host plant.

For the number of parasites, the GLMM with a Poisson distribution of errors and a log link function was compared to the GLMM with a normal error distribution and an identity link function (Figure 5.5). The parasite number had a distribution that significantly differed from a normal distribution ($W = 0.938$, $n = 589$, $P < 0.001$). The Poisson GLMM matched the distribution of the parasite number better than the normal GLMM, resulting in a normal distribution of residuals (Figure 5.5A–D). Both GLMMs showed linearity between the responses of the link function and the predictors (Figure 5.5E–F). The scale–location plots (Figure 5.5G–H) suggested that the Poisson GLMM fitted the data better than the normal GLMM since there was a systemic increase of variance in the normal GLMM. However, there was a slight overdispersion of data in the Poisson model (dispersion parameter = 1.94), as evidenced by the average standardised residuals, which was over 1.0 (Figure 5.5H). Nevertheless, an overdispersion in the Poisson model was modelled additively (Stoffel *et al.*, 2017; Nakagawa & Schielzeth, 2010) in the calculation of repeatability and the conditional mean number of parasites. Therefore, subsequent analyses were carried out using the Poisson GLMM for the number of *S. hermonthica* plants.

The conditional means of the phenotypic scores of *O. glaberrima* genotypes were extracted from the Poisson GLMM as described in Section 3.2.2.6 and used as phenotypic inputs for subsequent genome-wide association analyses.

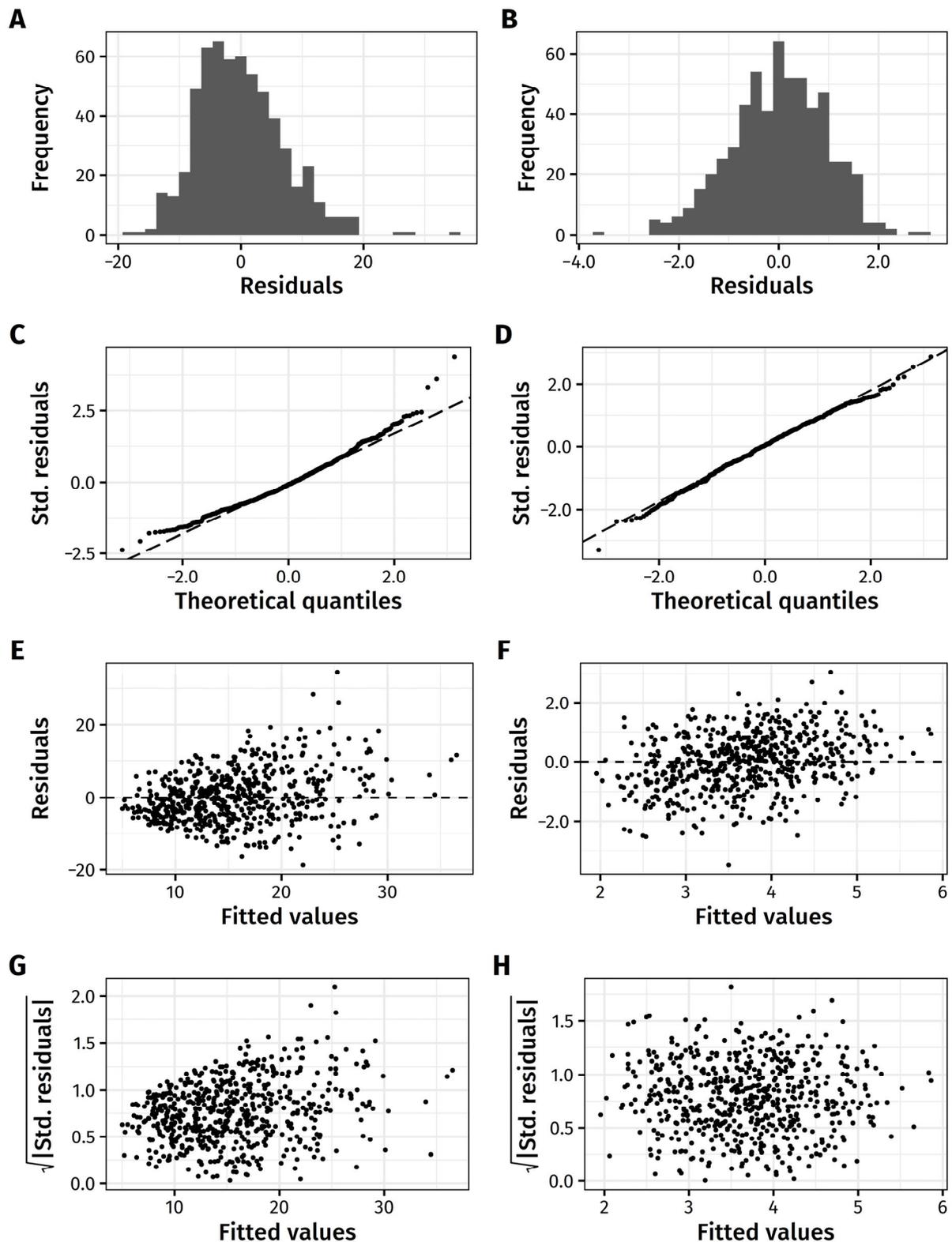


Figure 5.4 Diagnostic plots of the GLMMs for the biomass of *Striga hermonthica* plants per host. Diagnostic plots include the frequency histogram of residuals (A–B), the normal quantile–quantile plot (C–D), the residuals *versus* fitted values plots (E–F) and the scale–location plots (G–H). All GLMMs had the error distribution of a normal distribution. The untransformed data were fitted in a model that produced diagnostic plots A, C, E and G. The square root transformed data were fitted in a model that produced diagnostic plots B, D, F and H. Std., standardised.

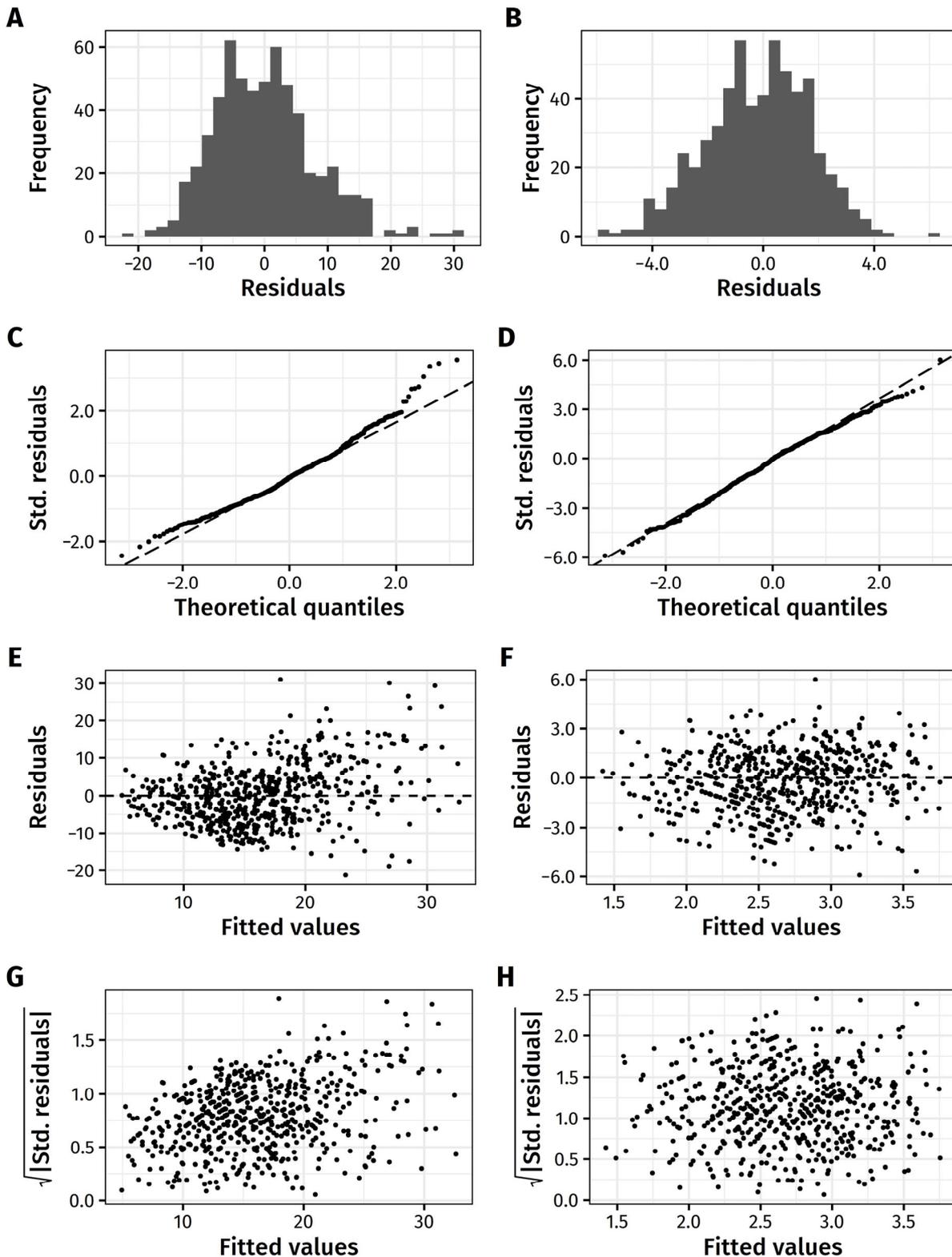


Figure 5.5 Diagnostic plots of the GLMMs for the number of *Striga hermonthica* individuals per host. Diagnostic plots include the frequency histogram of residuals (A–B), the normal quantile–quantile plot (C–D), the residuals *versus* fitted values plots (E–F) and the scale–location plots (G–H). A model that had the error distribution of a normal distribution with the identity link function produced diagnostic plots A, C, E and G. A model that had the error distribution of a Poisson distribution with the log link function produced diagnostic plots B, D, F and H. Std., standardised.

5.2.3 Preparation of genotypic data

A VCF file of the AfricanRice database, containing 19 115 021 SNPs and indels of 163 *O. glaberrima* and 83 *O. barthii* A.Chev. genotypes (Cubry *et al.*, 2018), was downloaded from <https://gigwa.ird.fr/gigwa/>. Out of these 246 genotypes, 149 *O. glaberrima* genotypes that were screened for *S. hermonthica*-resistance were selected using the `view` command of `BCFtools` software version 1.9 (Li, 2011). High-quality biallelic SNPs were filtered by the `filter` command of the `BCFtools` software and the `VariantFiltration` command of `GATK` software version 4.1.4.0 (McKenna *et al.*, 2010) according to the following criteria. A biallelic SNP was considered high quality when the quality flag, `QUAL`, was greater than 200, the read depth (`INFO/DP`) was inclusively between 10 and 20 000, the total mapping quality zero read (`MQ0`) was less than 4, the ratio between `MQ0` and `INFO/DP` was no larger than 0.1, the missing markers was less than 10 %, the frequency of minor alleles (`MAF`) was greater than 5 %, the heterozygosity was less than 10 % and the SNP was not in a cluster of more than three SNPs within any 10 bp windows. The clustering criterion was realised by `GATK`, whilst the rest were done by `BCFtools`. As a result, 426 464 SNPs were left after the filtration.

The population stratification and relatedness within the sampling population can influence association analyses and introduce type I errors if not taken into account (Pritchard & Donnelly, 2001; Pritchard *et al.*, 2000; Devlin & Roeder, 1999). The population stratification structure was inferred from a principal component analysis (PCA) and the admixture coefficients from the sparse non-negative matrix factorization (sNMF) algorithm. These two analyses were performed using package `LEA` version 2.6.0 (Frichot & François, 2015) in R software environment (R Core Team, 2017). Firstly, the `pca` function was used to perform PCA on filtered SNP data. The number of major ancestral populations was inferred by Tracy–Widom test of the eigenvalues from PCA using the `tracy.widom` function. The ‘elbow’ of the curve located on the fourth principle component indicates that there were five major ancestral populations (Figure 5.6A). Secondly, the number of ancestral populations

(K) were tested in 20 runs of sNMF algorithm for each K , from $K = 1$ to $K = 15$, using the `snmf` function. The cross-entropy criterion from each run was used to evaluate the error of ancestry estimate. The best run for each K was the run that had the lowest cross-entropy criterion. The cross-entropy criterion value began to plateau when the number of ancestral populations equalled to five (Figure 5.6B). Figure 5.7 shows the ancestry coefficients obtained from the sNMF algorithms (Figure 5.7A) and the PCA plot of genotypes against the first and the second components (Figure 5.7B).

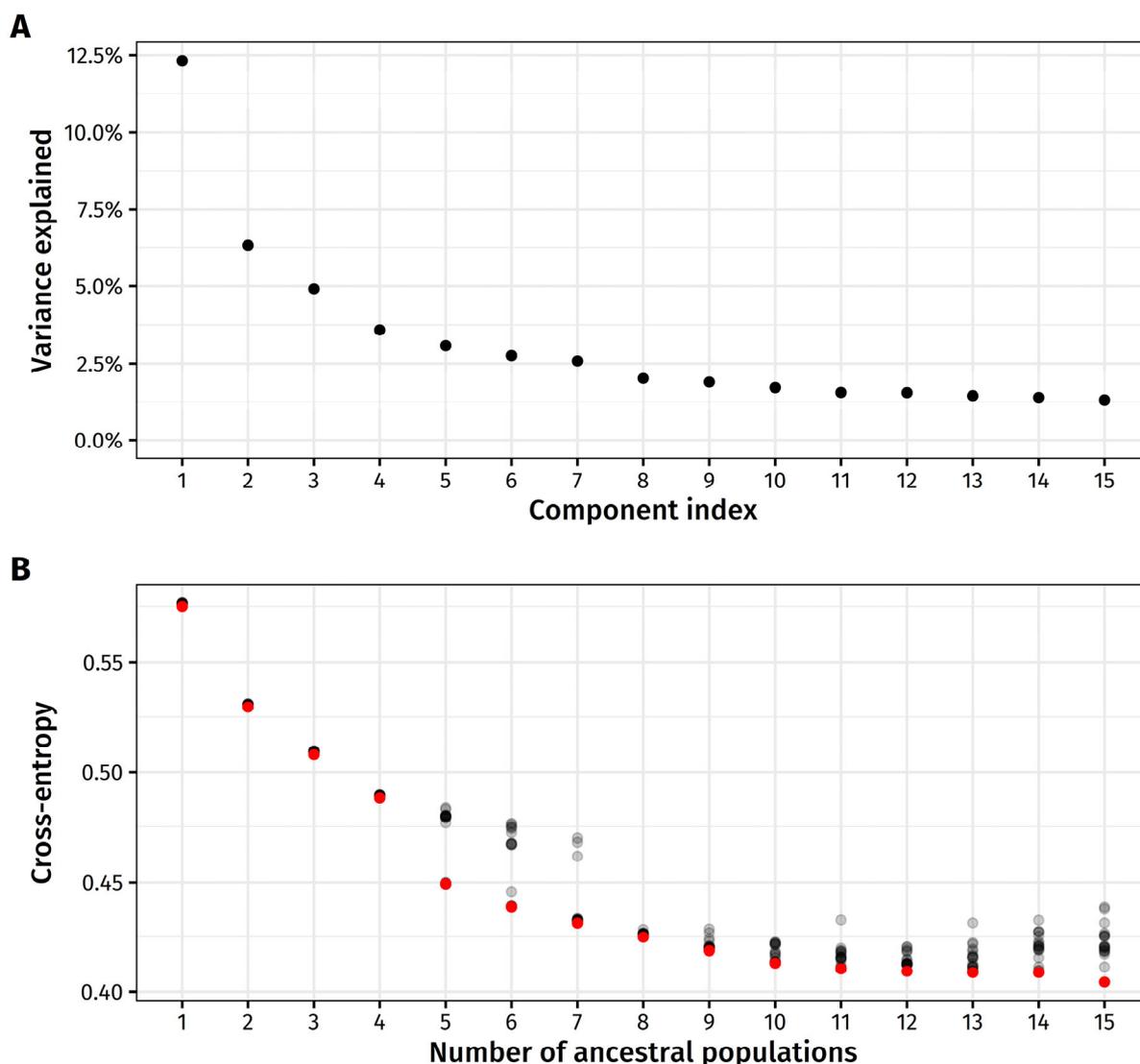
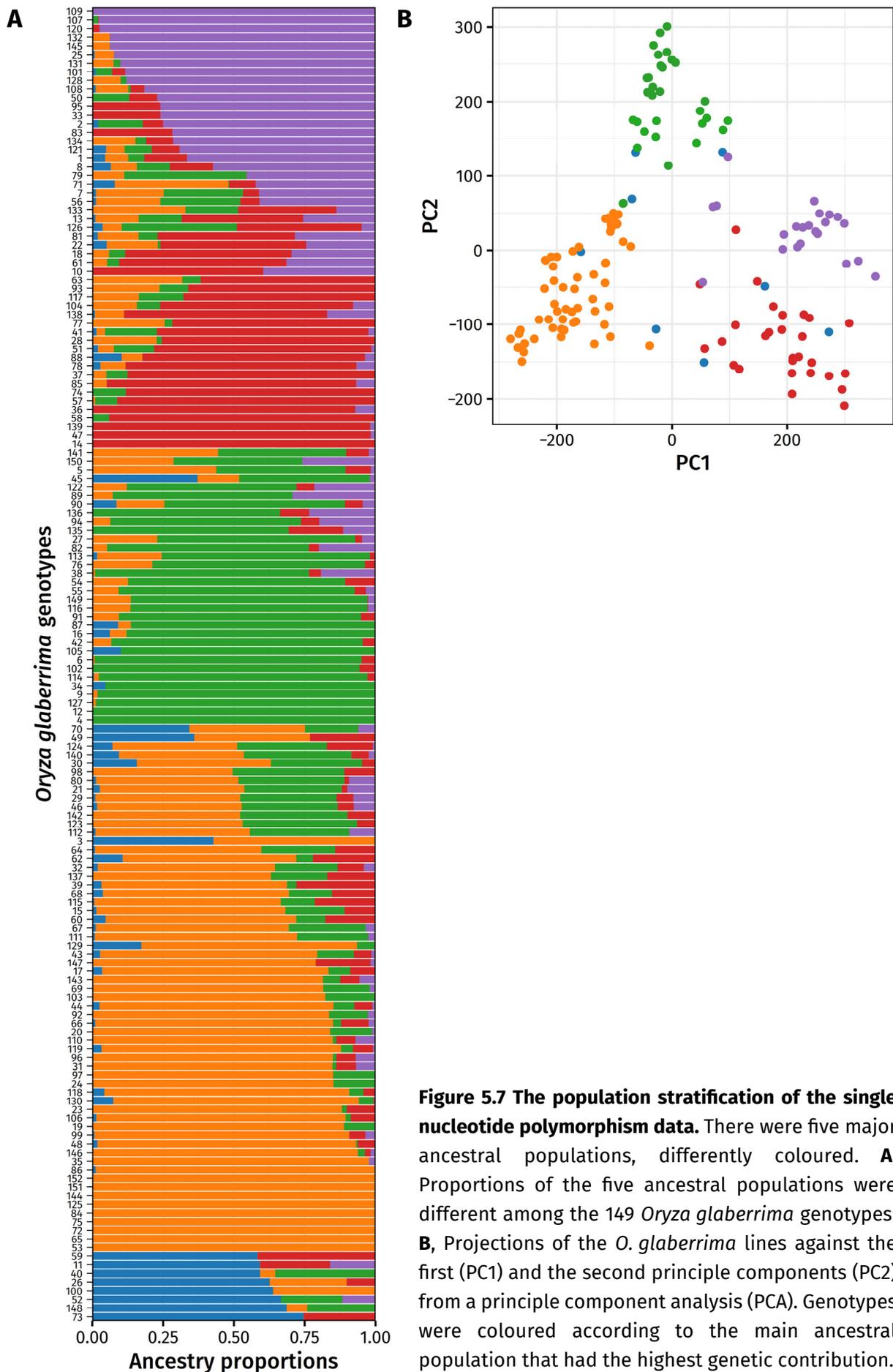


Figure 5.6 Assessment of the number of ancestral populations of the 149 *Oryza glaberrima* genotypes. (A) Scree plot of the percentage variance explained by each component in a principle component analysis of the filtered SNP data. (B) The cross-entropy criterion for each run of the sparse non-negative matrix factorization (sNMF) algorithm at different numbers of ancestral populations. Grey dots are the cross-entropy criterion for each of the 20 runs of sNMF algorithm. Red dots mark the lowest cross-entropy criterion value for each number of ancestral populations.



Missing markers can reduce the detection power of association studies (Marchini & Howie, 2010; Browning, 2008). The missing markers in the filtered SNP data were imputed by the sNMF method using function `impute` in `LEA` package. The number of clusters was set to $K = 5$ and the best run which had the lowest cross-entropy criterion from the 20 sNMF runs in the ancestral population inference was used.

Lastly, the relatedness of the population was represented by a kinship matrix, which was constructed from the imputed SNP data. Software `EMMAX-kin` version 20120205 with `INTEL64` implementation (Kang *et al.*, 2010) was used to generate a Balding–Nichols kinship matrix to the tenth digit precision, with the option ‘`-d 10`’ (Kang *et al.*, 2010).

5.2.4 Association analyses

Two mixed linear models (MLMs) (Yu *et al.*, 2006) were constructed, one for the biomass and the other for the number of *S. hermonthica* individuals per host plant. The ‘Q+K model’ where the imputed SNP data and the population stratification structure (Q-matrix) were fitted as fixed effects whilst the relatedness matrix (K-matrix) was fitted as a random effect (Z. Zhang *et al.*, 2010) was used. The projections of the SNP data onto the first five principal components were used as the Q-matrix. The kinship matrix was used as the K-matrix. Software `PLINK` version 1.90b6.5 (Purcell *et al.*, 2007) was used to convert the imputed SNP data to `tped` format. Association analyses were performed using `EMMAX` software version 20120205 with `INTEL64` implementation (Kang *et al.*, 2010), using command ‘`emmax-intel64 -v -d 10 -t [tped_prefix] -p [phenotype_file] -k [kinship_file] -c [covariate_file] -o [output_prefix]`’. The threshold for a significant association at the significance level $\alpha = 0.05$, after Bonferroni correction for multiple tests, was at $P < 1.17 \times 10^{-7}$, which is equivalent to $-\log_{10} P > 6.93$. Local FDR and Q-values (Storey & Tibshirani, 2003) were calculated from the P -values of the associations using the software package `qvalue` version 2.16.0 (Storey *et al.*, 2019) in R software environment.

5.3 Results

5.3.1 *Oryza glaberrima* genotypes were mostly susceptible to an accession of *Striga hermonthica* and had low repeatabilities.

To assess the resistance of the *O. glaberrima* genotypes to the accession of *S. hermonthica* seeds from Kibos, Kenya, the biomass and the number of parasites were measured at 21 DAI. Most of the *O. glaberrima* genotypes were susceptible to very susceptible (Figure 5.8). Only eight genotypes were resistant to *S. hermonthica*, which had the parasite biomass of less than 10 mg or fewer than ten parasites growing on each host plant across all replicates. Two genotypes (97 and 475) were highly resistant, with the parasite biomass less than 5 mg on each of their replicates.

The repeatabilities (ICC) were calculated to assess the consistency of phenotypic scores and the impact of inter-batch variations on the level of resistance/susceptibility for the rice genotypes. Rice genotypes had low repeatability, < 40 %, for both the biomass and the number of parasites (Table 5.2). The biomass of parasites of the genotypes — explaining 36.3 % of total variance — was more consistent than the number of parasites of the genotypes, explaining only 21.9 % of the total variance. The inter-batch variations explained lower percentages of the total variance than the genetic variations between rice genotypes and affected the number of *S. hermonthica* individuals more than the biomass of parasites, at 11.6 % and 4.1 % of the total variance, respectively (Table 5.2).

Table 5.2 Summary of repeatabilities of the *Striga hermonthica*-resistance.

Trait	Factor	Repeatability	95 % confidence interval*	P-value†
<i>S. hermonthica</i> biomass	Rice genotypes	0.363	[0.271, 0.447]	< 0.001
	Batches	0.041	[0.004, 0.095]	< 0.001
<i>S. hermonthica</i> number	Rice genotypes	0.219	[0.138, 0.310]	0.003
	Batches	0.116	[0.027, 0.208]	< 0.001

* 95 % confidence intervals were calculated from parametric bootstrap (10 000 iterations).

† P-values were calculated from a permutation test (10 000 permutations).

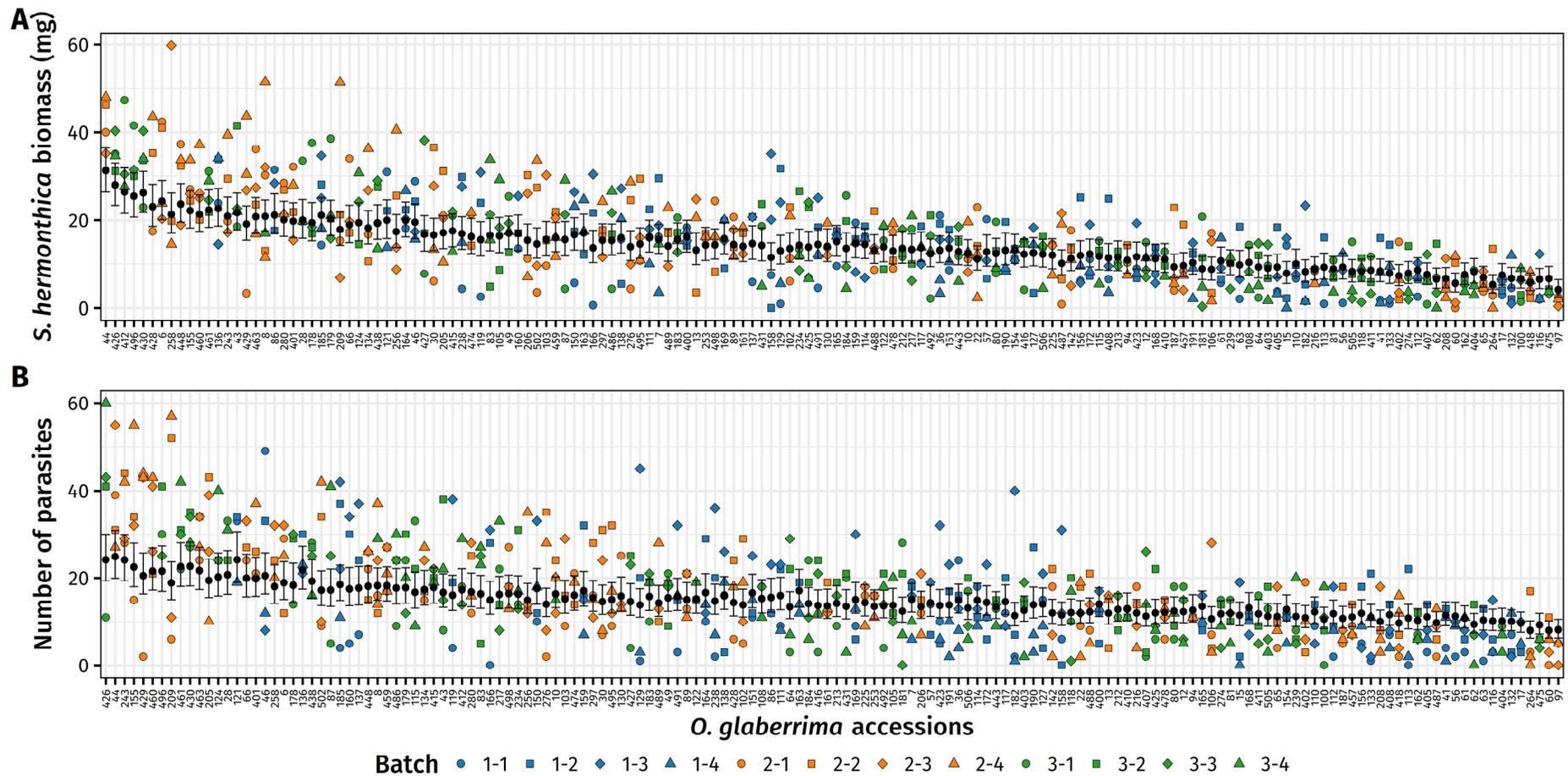


Figure 5.8 Resistance to *Striga hermonthica* of a panel of 156 *Oryza glaberrima* genotypes. The biomass (dry weight) (A) and the number of *S. hermonthica* plants (B) harvested from each rice host plant at 21 d after inoculation of pre-germinated *S. hermonthica* seeds collected from Kibos, Kenya. Each black dot represents the back-transformed conditional mean of the measurements for each genotype from (A) a GLMM that had a normal error distribution and was fitted with the square root transformed *S. hermonthica* biomass and (B) a GLMM that had the error distribution of a Poisson distribution with the log link function and was fitted with the number of parasites (error bar = mean \pm SD). In each trait, the rice genotypes were ranked from the most susceptible to the most resistant. Differently coloured and shaped dots represent measurements from different batches (Group-Batch).

5.3.2 There was no significant association between the molecular markers and the biomass or the number of parasites.

A total of 426 464 SNPs from 149 genotypes of *O. glaberrima* were used in this genome-wide association study. Two MLMs, for which the population stratification and the relatedness among genotypes were controlled, were employed in the association analyses for the biomass and the number of parasites. There was no significant association neither between the biomass of *S. hermonthica* and the SNP markers (Figure 5.9A–B) nor between the number of *S. hermonthica* individuals and the SNP markers (Figure 5.9C–D). The highest values of the $-\log_{10} P$ for the parasite biomass and parasite number were 4.29 and 4.31, respectively (Figure 5.9A & C). These values were lower than the significance threshold with Bonferroni correction for multiple tests at 6.93. The local FDRs at the loci that had the lowest P -value for the biomass and the number of parasites were 100 %, with the Q -values of 0.705 and 0.835, respectively. The significance levels of the association from both traits (Figure 5.9B & D) were lower than the expected values under the null hypothesis of no association, in which P -values followed a uniform distribution. There was no sign of an early divergence of the observed $-\log_{10} P$ values from the expected values towards the y-axis in both Q–Q plots (Figure 5.9B & D), indicating that the population structure had been effectively controlled by the software through the MLMs.

5.4 Discussion

This chapter aimed to establish the genetic basis of resistance to *S. hermonthica* in a diverse panel of *O. glaberrima* genotypes, using the genome-wide association mapping approach in conjunction with the pipeline to calculate repeatability, developed in Chapter 3. The level of resistance to *S. hermonthica* was measured by the biomass (dry weight) and the number of parasites per host plant at 21 DAI. Unfortunately, no significant association was found between the genetic loci and either the biomass or the number of parasites at the Bonferroni corrected significance level ($\alpha = 0.05$). Even though Bonferroni correction is considered

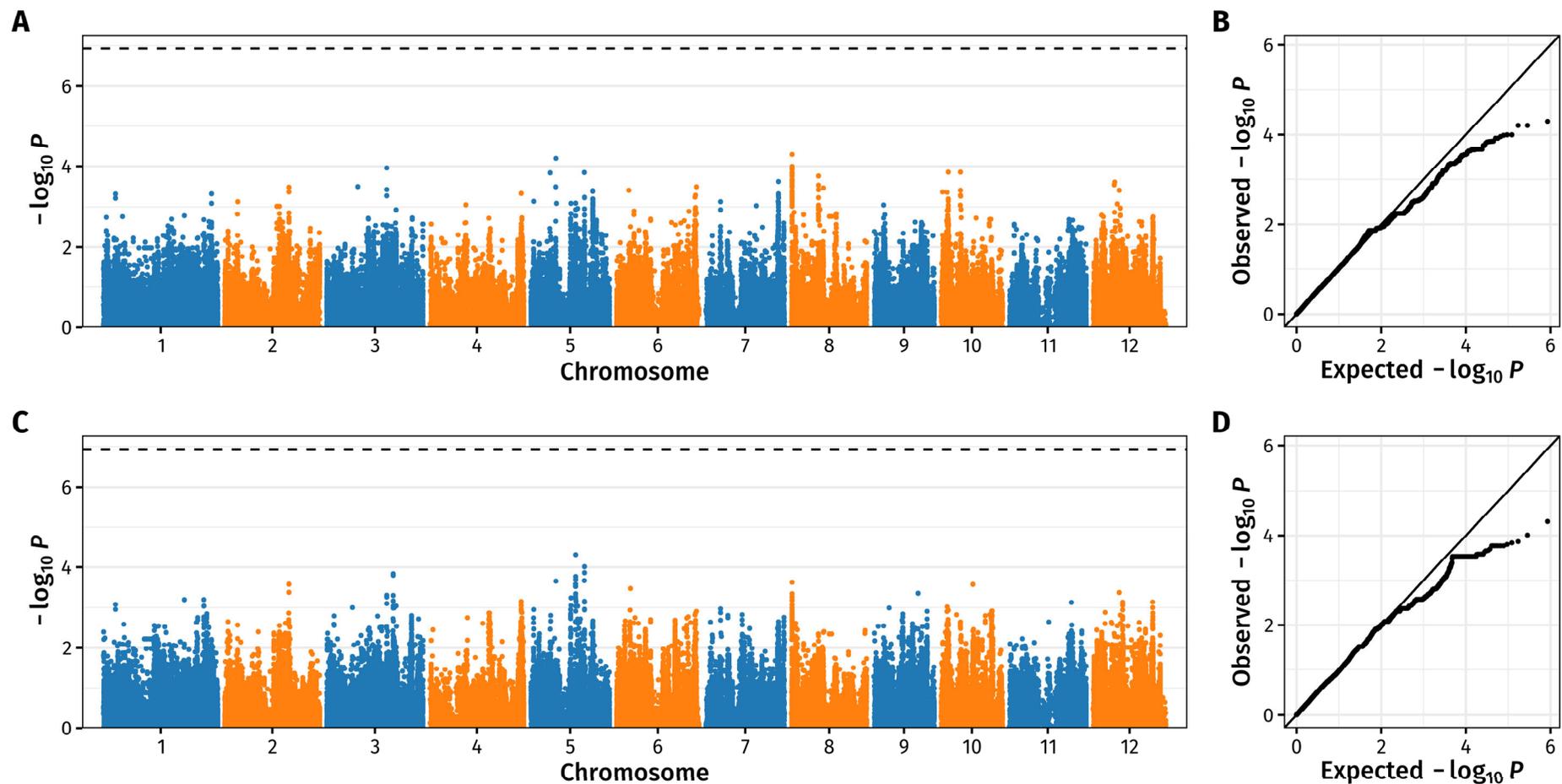


Figure 5.9 Genome-wide association studies of resistance to *Striga hermonthica* in *Oryza glaberrima*. The Manhattan plot (A) and the quantile–quantile (Q–Q) plot (B) of the association results for the biomass of *S. hermonthica*. The Manhattan plot (C) plot and the Q–Q plot (D) of the association results for the number of *S. hermonthica* individuals per host. A mixed linear models (MLM) where the population stratification structure and the relatedness of rice lines were incorporated in the ‘Q+K model’ was fitted for each trait. The dashes lines in Manhattan plots (A & C) were the genome-wide significance level at $\alpha = 0.05$ after Bonferroni correction for multiple tests ($-\log_{10} P = 6.93$). In the Q–Q plots (B & D), the observed values of $-\log_{10} P$ were compared to the expected values under the uniform distribution of P -values.

conservative and the selection of a less conservative threshold such as the use of FDR has become a common practice in rice GWASes [e.g. de Abreu Neto *et al.* (2017), Matthus *et al.* (2015) and Wissuwa *et al.* (2015)], this method was not relevant to this work because the local FDR of 100 % was observed at the SNP with the lowest *P*-value for both resistance traits.

Among the *O. glaberrima* genotypes, both the biomass and the number of *S. hermonthica* individuals had low repeatability values. These low repeatability values affected the association analyses that followed since the repeatability of a trait sets an upper bound of the broad-sense heritability of that trait (Falconer & Mackay, 1996; Lynch & Walsh, 1996), which is always larger than the additive contribution of genetic variation (a narrow-sense heritability) that is detected in association analysis. A significant association for a trait with low repeatability is unlikely to be identified when the population size is small. By definition, low repeatability is a result of (i) lack of variation between individuals (genotypes) or (ii) large variation of phenotypic scores within genotypes, also called the residual variance (Stoffel *et al.*, 2017; Lynch & Walsh, 1996). As observed in the phenotypic screen (Figure 5.8), there was a lack of variation in the levels of *S. hermonthica*-resistance of the *O. glaberrima* panel. The majority of the *O. glaberrima* genotypes were susceptible to *S. hermonthica* to the same level as susceptible *O. sativa* cultivars, such as Azucena and Koshihikari, which have been used for mapping of loci underlying *S. hermonthica* resistance and supported more than 10–15 mg biomass of the same accession of *S. hermonthica* at 21 DAI (Beardon, 2018).

Of the 156 *O. glaberrima* genotypes screened, only four, TOG5429, TOG6356, TOG7195 and OR09849 (genotype numbers 17, 97, 404 and 475), exhibited clear and strong resistance to the *S. hermonthica* accession, consistently having fewer than ten parasites per host and less than 10 mg per host of parasites biomass at 21 DAI (Figure 5.8). The levels of resistance of these four rice genotypes were comparable to other rice genotypes, which their post-attachment resistance to the same accession of *S. hermonthica* from Kibos, Kenya had been found under rhizotron systems, including *O. sativa* varieties Nipponbare and IR64 (Beardon, 2018), *O. glaberrima* genotypes MG12 (Cissoko, 2012) and CG14 (Cissoko *et al.*,

2011) and interspecific varieties (between *O. sativa* and *O. glaberrima*) NERICA 1 and 10 (Cissoko *et al.*, 2011). The results contradicted a report from field screening that the plots that were cultivated with *O. glaberrima* genotypes, which were more vigorous and more competitive with weeds than *O. sativa*, had far fewer *Striga* parasites than in plots planted with *O. sativa* genotypes (Johnson *et al.*, 1997). However, the apparent lack of resistant genotypes could be a result of sampling bias as these cultivated *O. glaberrima* genotypes were primarily selected from West Africa to locate the geographic region that *O. glaberrima* was domesticated in the original study by Cubry *et al.* (2018) and did not include all the variations present in *O. glaberrima* germplasm, thus the population was not balanced.

In addition to the host genetics, the compatibility or incompatibility of a *Striga*–host interaction also depends on the genetic makeup of the parasites. It is well documented that host resistance to *Striga* spp. is specific to the accessions/ecotypes of parasites. Rodenburg *et al.* (2017, 2015) and Cissoko *et al.* (2011) showed that a rice genotype—*O. sativa* and *O. glaberrima* alike—responded differently to *S. hermonthica* and *S. asiatica* parasites from different geographic locations, in terms of the biomass and the number of parasites, both in the field and under controlled laboratory conditions. The Kibos accession of *S. hermonthica* used in this study was evidently virulent to most genotypes in the *O. glaberrima* panel. These *O. glaberrima* genotypes can potentially respond differently to a different accession of *S. hermonthica* or other *Striga* species.

It is possible that the small sample size of resistant individuals from the *O. glaberrima* panel could have prevented association algorithms from identifying any significant loci from a genome-wide analysis as the signal would have been lost to the noise. Moreover, alleles underlying the resistance and were unique to the four resistant genotypes would have been filtered out in the preparation of SNP data (MAF < 5 % in this study). This is an inherent limitation to association analyses as alleles with very low occurrences in the population are not statistically informative and have to be removed (Zhou & Huang, 2019; Nordborg & Weigel, 2008). To overcome the problem of minor alleles, the number of genotypes, especially

those possessing resistance, must be increased in the study in hope that the number of rare alleles exceeds the cut-off threshold. An additional benefit of the larger population size is an increase in the detection power of loci with smaller effects. Nevertheless, many GWASes in rice have been successfully executed on a population with a comparable size to the *O. glaberrima* panel used in this study. For example, Lekklar *et al.* (2019) identified loci associated with salt tolerance at the flowering stage using 104 *O. sativa* genotypes. Over 70 % of the loci had been reported in QTL for salt tolerance in other studies (Lekklar *et al.*, 2019). Another example was the use of genome-wide association on 169 *O. sativa* varieties to isolate the rice *SPINDLY* gene (*OsSPY*) that was involved in the regulation of rice plant architecture through gibberellin-mediated signalling pathway (Yano *et al.*, 2019). However, these studies tend to identify SNPs with large effects from highly repeatable traits. Alternatively to GWASes, a mapping population, such as F₂ and RIL populations, could be constructed by crossing one of the more resistant genotypes, such as TOG6356 or OR09849, with the more susceptible genotypes, such as TOG5672 or IRGC103593. Creating a mapping population, though being more time consuming, guarantees that the problem of rare variants is eliminated as half of the mapping population would process those rare alleles.

Another cause of low repeatability is a high variance of the phenotypic data from within an individual genotype. Low repeatability affects the accuracy of the conditional means from the GLMMs because of the large residual variance. Despite performing the phenotypic screen with four repeated measurements per genotype under a controlled environment and the exclusion of replicates with aberrant root architecture, the residual variance remained high in the dataset. The experimental design incorporated *O. sativa* check cultivars to calibrate the level of resistance between the batches. However, the check cultivars could not be included in the data analyses, as discussed in the Methodology (Section 5.2.2.2), since the levels of resistance/susceptibility were not comparable to those of the *O. glaberrima* genotypes. This led to high uncertainty of conditional means from the GLMMs for repeatability because each genotype score could only be compared to other genotypes within the same group (a third of

the population). As a result, the conditional means of the most resistant genotypes were less resistant than the mean and the conditional means of the most susceptible genotypes were less susceptible than the mean. A change in conditional means directly impacts the GWAS results. A better experimental design would involve placing the replicates of each genotype randomly across the twelve batches, like the phenotypic screen in Chapter 3 (Section 3.2.2.1). This would allow for direct comparisons between a genotype and more than a third of the population, decreasing the importance of the checks and obtaining a higher certainty of the conditional means of phenotypic scores.

Nonetheless, GWASes are still a powerful tool for the identification of *Striga*-resistance. The diversity panel must consist of good representations of both resistant and susceptible genotypes, and the heritability of the phenotype used must be high enough. Successful examples of association studies for resistance and tolerance to *S. hermonthica* have recently been carried out in maize and sorghum. In maize, Adewale *et al.* (2020) performed a GWAS and identified loci associated with the number of emerged *S. hermonthica* plants and the parasite damage scores in artificially infested fields, using a panel of 132 maize inbred lines and 47 440 DArT markers. The broad-sense heritability of the traits was high, at 47–71 %. Candidate genes, such as the *ZmCCD1* gene which encodes a maize carotenoid cleavage dioxygenase and is linked to the production of *Striga* germination stimulant strigolactones, were identified (Adewale *et al.*, 2020). In sorghum, Kavuluko *et al.* (2020) identified candidate genes for post-attachment resistance, such as xylanase inhibitors, which are involved in cell wall strengthening by inhibiting xylanase enzyme secreted by the parasites to weaken host cell wall during the penetration of host endodermis, and resistance response proteins, which are involved in phytohormone signalling. This was done using a GWAS on a panel of 206 diverse sorghum genotypes from various geographical locations and 247 975 SNPs. The panel was balanced and harboured several resistant (45 %) and susceptible genotypes (34 %) (Kavuluko *et al.*, 2020). The population size and the number of SNPs were not the limiting factors for the discovery of resistance loci in the current study since the

population sizes of these two successful examples were comparable to the *O. glaberrima* panel (149 genotypes) and the current study had a higher density of high-quality genetic markers (426 464 SNPs over a smaller rice genome) than the two examples.

In future, a subset of the *O. glaberrima* genotypes could be screened against different accessions of *Striga* spp. to ensure there is a balanced distribution of resistance/susceptibility among the genotypes within the panel. Moreover, the addition of more *O. glaberrima* genotypes to capture a wider genetic background of *O. glaberrima* could also potentially introduce more resistant genotypes to the panel.

Chapter 6

General discussion

6.1 Introduction

An infestation of *Striga* spp., such as *S. hermonthica* and *S. asiatica*, is a major biotic threat to crop production in sub-Saharan Africa, especially to subsistence, resource-limited small-holder farmers (Atera & Itoh, 2011; Scholes & Press, 2008). These parasitic weeds are found in at least 40 % of cereal farmland in sub-Saharan Africa, resulting in 20–100 % yield losses, costing over US\$1 billion per year and affecting more than 100 million people (Spallek *et al.*, 2013; Scholes & Press, 2008). Rice is among cereal crops affected by *Striga* weeds (Atera & Itoh, 2011). As rice gains popularity as a staple food in Africa in recent years (Seck *et al.*, 2012), its production has expanded into rain-fed upland arable lands previously used for the production of other cereal crops, such as sorghum and maize, and is infested with *Striga* spp. (Rodenburg *et al.*, 2016). This results in an annual loss of 488 000 tonnes of rice, costing more than US\$100 million (Rodenburg *et al.*, 2016; N'cho *et al.*, 2014). Effective *Striga* control measures are required to be affordable, widely available and compatible with current farmer practices to guarantee wide adoption by small-holder farmers (Mrema *et al.*, 2017; Teka, 2014; Hearne, 2009). The use of *Striga*-resistant rice varieties satisfies these requirements and can be used as in conjunction with other control methods that aim at improving soil fertility and depleting the size of *Striga* seed bank in the field to further reduce the severity of *Striga* infestation (Rodenburg *et al.*, 2010; Yoder & Scholes, 2010).

Despite the identification of sources of *Striga*-resistance in rice germplasm, little is known about the genetic basis and the mechanisms the resistance found in these resistant varieties. This prevents the development of rice varieties that incorporate multiple resistance genes/loci. Stacking different resistance genes that use multiple resistance mechanisms into a single variety is important to ensure durable resistance as the *Striga* seed bank contains vast genetic diversity in countless seeds that are long-lasting (Scholes & Press, 2008). Most work on the genetic basis of *Striga*-resistance in rice has focused on *S. hermonthica*, resulting in several discoveries of different QTL for *S. hermonthica*-resistance in rice. Only candidate genes for post-attachment resistance identified in rice against *S. hermonthica* are the RLP genes

identified in resistance rice cultivars, Nipponbare and IR64 (Beardon, 2018). No resistance QTL or genes for post-attachment resistance to *S. asiatica* has been identified in rice.

The main aim of this thesis is to identify and characterise resistance phenotypes in rice against *S. asiatica* and to identify genetic loci and candidate genes underlying the resistance phenotype, using quantitative genetic approaches. Also, the secondary aim is to perform a GWAS on an existing dataset of the levels of resistance/susceptibility in an *O. glaberrima* diversity panel to *S. hermonthica*, to identify loci underlying resistance. Figure 6.1 summarises the work carried out in this PhD programme, the main findings, conclusions and future directions from the work.

6.2 Slow growth resistance to *Striga asiatica*

To utilise available RIL populations in mapping the genetic basis of resistance to *S. asiatica* in rice, four parental genotypes from three RIL populations were phenotyped for their resistance/susceptibility to an accession of *S. asiatica* seeds from Ethiopia (Chapter 2). This led to the identification of the slow growth resistance phenotype of *S. asiatica* in the rice genotype CT8556-37-2-3-1-M, compared with the rice genotype IR64 (Figure 6.1A). The slow growth *S. asiatica*-resistance phenotype was characterised in detail in Chapter 2, which revealed its two components: (i) parasites grew significantly more slowly on the resistant CT8556-37-2-3-1-M rice genotype compared to the susceptible IR64 rice genotype; and (ii) fewer parasites successfully infected CT8556-37-2-3-1-M, compared to IR64 (Figure 6.1B). Upon further microscopic study of the infection of parasites on the two rice genotypes, no significant difference was found in the timing of each stage of parasite infection through the host root. The same proportion of *S. asiatica* individuals formed xylem–xylem connections with both CT8556-37-2-3-1-M and IR64. The shorter length and lower biomass of parasites growing on CT8556-37-2-3-1-M, compared to IR64, was a result of the slower growth of parasites after the formation of host–parasite xylem continuity. This was particularly interesting because this resistance phenotype was different from previously identified

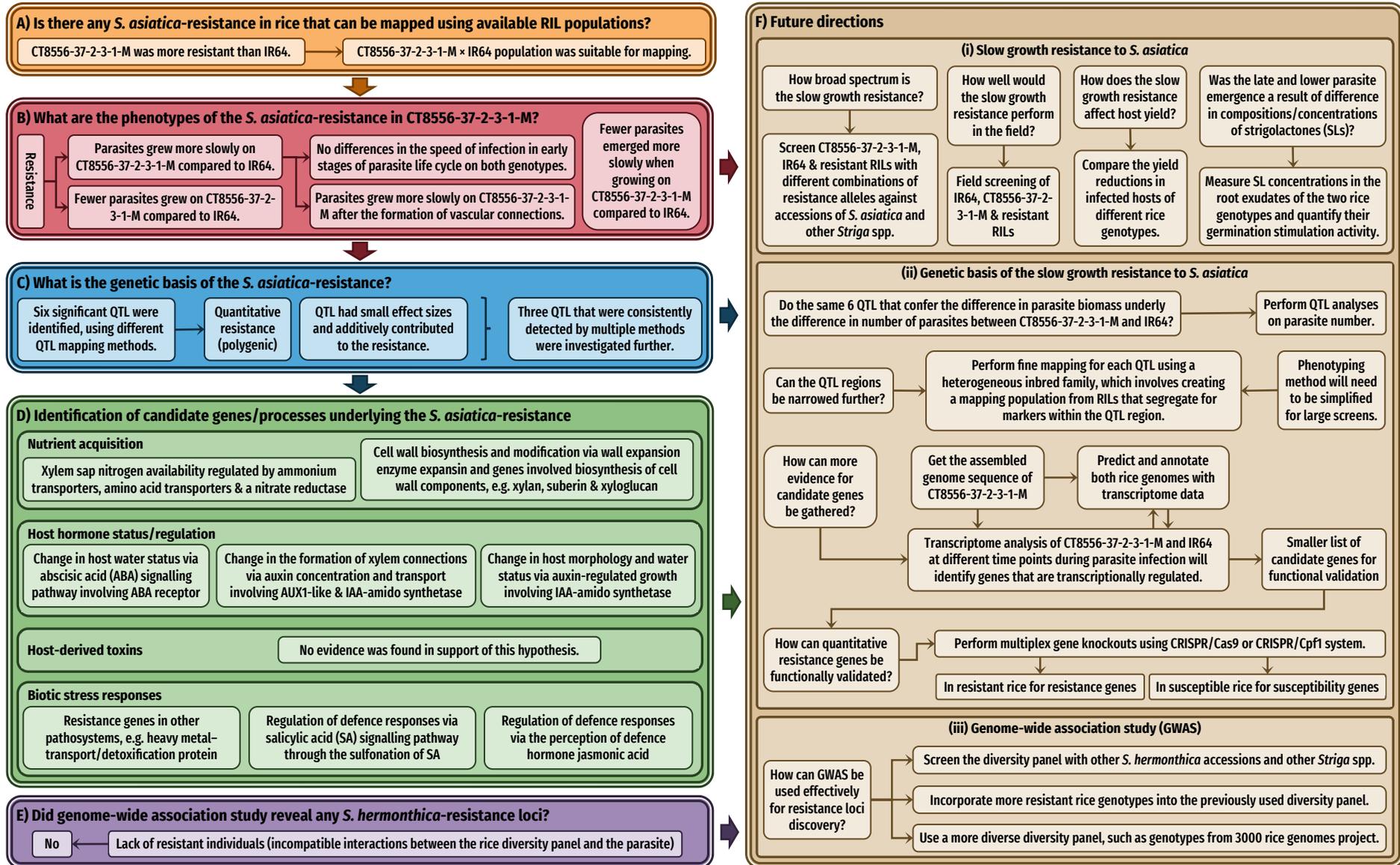


Figure 6.1 Overview of work carried out in this PhD programme, findings and future directions from the work in this thesis.

resistance to both *S. asiatica* and *S. hermonthica* in rice, where the majority of parasite endophytes failed to breach the endodermal layer of host root, thus they were unable to form vascular connections with the host, e.g. Beardon (2018), Cissoko *et al.* (2011), and Gurney *et al.* (2006). More importantly, the identification of a new form of *Striga*-resistance that relies on different mechanisms will provide additional materials to be used in pyramiding multiple resistance genes to produce durable rice varieties (Dormatey *et al.*, 2020; Mundt, 2014; Scholes & Press, 2008).

There are still several unanswered questions regarding the nature of the slow growth resistance to *S. asiatica* (Figure 6.1Fi). It is important to know whether this *Striga*-resistance in CT8556-37-2-3-1-M is effective against a wide range of *S. asiatica* accessions from other geographical locations or to other *Striga* species, such as *S. hermonthica*. Broad-spectrum *Striga*-resistance is desirable for the protection of crop plants from enormous genetic diversity in the parasite seed bank (Mohamed *et al.*, 2007), and also means that resistant varieties would work in different geographical regions. Knowledge on the specificity of the resistance is needed for the breeding of resistant crop varieties appropriate for different regions (Scholes & Press, 2008). Further screens of CT8556-37-2-3-1-M against available *Striga* accessions would help answer this question.

The work in this thesis showed that the slow growth *S. asiatica*-resistance was also associated with late emergence and fewer emerged parasites in a pot-based experiment (Figure 6.1B). This showed that CT8556-37-2-3-1-M imposed a fitness cost onto the parasites, in comparison to IR64. This has a direct implication on the applicability of the resistance in the field where a reduction in the returning of parasite seeds into the soil is desirable (Rodenburg *et al.*, 2010). A question regarding the involvement of the germination stimulant activity of the host root exudate remains unanswered. Certainly, the root exudate of CT8556-37-2-3-1-M could germinate *S. asiatica* seeds and induce the differentiation of haustoria and subsequent infection because few parasites were found on the host roots (Figure 2.7H & J). However, it is possible that the lower number of emerged parasites in CT8556-37-2-3-1-M

pots, compared to IR64, was also controlled by a lower production or a different composition of strigolactones, similar to rice varieties with pre-attachment resistance (Cardoso *et al.*, 2014; Jamil *et al.*, 2011a). In a study of two maize inbred lines that compared the phenotype of the more resistant maize genotype TZSTRI108 to that of the more susceptible maize genotype 5057, the slow growth resistance to *S. hermonthica*—similar to the one identified in this thesis—was found in the TZSTRI108 genotype and was associated with lower germination stimulation ability of the host root exudate (Unachukwu *et al.*, 2020). The quantification of the germination stimulation activity of the root exudates from the two rice genotypes would answer the question regarding the involvement of pre-attachment mechanisms and determine whether further analysis of strigolactone composition is needed.

Also, it is not known how stable the slow growth resistance phenotype is because the phenotypic screens in this PhD were carried out under controlled laboratory conditions. Several studies have found that the level of resistance/susceptibility of the host is affected by the infestation level and environmental conditions, on top of the genetic variations of both the host and the parasite (Rodenburg *et al.*, 2017, 2015; Haussmann *et al.*, 2001b, 2000). Haussmann *et al.* (2004) reported highly significant QTL × environmental interactions from the mapping of QTL for resistance to *S. hermonthica* in sorghum at multiple locations, using the emergence time and the number of emerged parasites to assess the level of host resistance. In contrast, the study by Unachukwu *et al.* (2020), which identified the slow growth *S. hermonthica*-resistance in the maize inbred genotype TZSTRI108, found that the resistant TZSTRI108 genotype yielded reliably higher and was consistently less affected by *Striga* infection, compared to the more susceptible maize genotype 5057, when screened at two locations in Nigeria over two consecutive seasons. The slow growth *S. asiatica*-resistance in rice identified in this thesis could have similar stability and could be similarly unaffected by environmental conditions. A field trial of CT8556-37-2-3-1-M and IR64 plants at multiple locations for a few seasons would elucidate on the stability of the resistance phenotypes.

Another aspect of the parasite–host interaction that is relevant to *Striga* control in agriculture is host tolerance/sensitivity (Rodenburg & Bastiaans, 2011). Although tolerance and resistance are not often found in the same host genotype (Rodenburg *et al.*, 2017; Rodenburg & Bastiaans, 2011), resistant CT8556-37-2-3-1-M plants that were grown in pots had taller stature than susceptible IR64 and IAC165 counterparts (Figure 2.7). It would be interesting to see whether CT8556-37-2-3-1-M possesses any tolerance by actually comparing the reduction in the yield or the photosynthetic rate of CT8556-37-2-3-1-M plants under *Striga*-infection to those of the other two rice genotypes.

6.3 Genetic basis of the slow growth resistance to *Striga asiatica*

Utilising the available RIL population derived from a cross between CT8556-37-2-3-1-M and IR64, a mapping approach was taken to explore the genetic basis of the slow growth resistance to *S. asiatica*, using the biomass of *S. asiatica* per host plant as the measurement of the level of resistance (Figure 6.1C). The slow growth *S. asiatica*-resistance was revealed to be a genetically quantitative trait, controlled by several QTL; six significant QTL were identified using several QTL mapping methods (Table 3.6). Five of these QTL had the resistance alleles derived from CT8556-37-2-3-1-M background, whilst the IR64 allele contributed to the resistance at *qSaB11.2* on chromosome 11. The effect sizes of these QTL were small, each explaining 7.5–17.5 % of the phenotypic variance. Also, these QTL were largely additive; QTL mapping approaches identified no significant epistatic interactions between them.

These QTL did not share the physical location with the major *S. hermonthica*-resistance QTL *qShR12.1* on chromosome 12, previously identified multiple times in diverse rice genotypes (*O. sativa* subsp. *japonica* cv. Nipponbare and *O. sativa* subsp. *indica* cv. IR64) (Beardon, 2018; Gurney *et al.*, 2006) (Table 3.1). The candidate resistance genes within *qShR12.1* included genes encoding RLPs, which were annotated as orthologs of the wounding-inducing tomato gene *Ve1* that confers *Verticillium* wilt resistance in tomato (Nazar *et al.*,

2019; Beardon, 2018). Also, no RLP gene was found within the regions of the three QTL that were analysed in Chapter 4. Taken together, the three QTL for slow growth resistance have the potential to add another layer of protection to crop plants on top of the RLP genes within *qShR12.1*.

Another component of resistance phenotype was the lower number of parasites growing on CT8556-37-2-3-1-M plants. Due to limitation of time, the underlying genetics of the difference in the number of parasites between CT8556-37-2-3-1-M and IR64 was not explored. It would be interesting to see whether the same six QTL also contributed to fewer parasites growing on CT8556-37-2-3-1-M hosts (Figure 6.1Fii). However, there was a strong positive correlation between the number of parasites and the biomass of parasite per host plant (Figure 3.3). Some of the six QTL for the slow growth resistance phenotype are likely to underlie the lower number of parasites in CT8556-37-2-3-1-M as well. Counting the number of parasites on each RIL and conducting an independent QTL analysis will be carried out in the future to see whether there are new QTL that underlie the lower number of parasites on CT8556-37-2-3-1-M, compared to IR64.

The genomic sequences of three QTL (*qSaB1.1*, *qSaB5.1* and *qSaB11.1*) with the largest effect sizes that were consistently mapped using different QTL mapping methods were further analysed to see whether candidate genes or processes can be identified as potential mechanisms of the resistance (Figure 6.1D). The identification of physical locations of the three QTL on IR64 genome revealed that the QTL regions were quite large (1.4–2.2 Mbp). Though not pursued in this PhD, these QTL regions could be narrowed further by fine mapping of each QTL region. This can be done using heterogeneous inbred families (Tuinstra *et al.*, 1997). In this method, a RIL (in F₇ generation in this case) that still retains heterozygosity over the QTL region of interest is selfed to create an F_{7:9} segregating population of ~ 150–200 individuals that can be used for fine mapping. For each of the six QTL, one to two RILs segregated over markers within the QTL region and were suitable for this approach. A large number of genotypes in the newly created heterogeneous inbred

families will need to be typed and screened for their resistance levels. Hence, the phenotyping method might need to be simplified to facilitate large phenotyping effort. However, many attempts to fine-map quantitative resistance traits have met with a difficulty due to a lack of recombination to separate a large QTL region apart (Corwin & Kliebenstein, 2017). Alternatively, more studies have switched to use more advance mapping populations that derived from multiple parental genotypes and have larger population sizes, such as NAM and multi-parent advanced generation inter-cross (MAGIC) populations (Scott *et al.*, 2020), to untangle resistance traits (Corwin & Kliebenstein, 2017).

Instead, a genomic approach was taken in Chapter 4, utilising the available genomic sequences of the two rice genotypes. However, the assembled genome sequence was available for only one of the rice genotypes, IR64. This complicated the gene prediction analysis as the short sequence reads of CT8556-37-2-3-1-M could only be assembled into short contigs. Therefore, a decision was made to focus the analysis to the peak regions by prioritising the analysis boundaries to 2 cM from the peak of each of the three QTL with the largest effect sizes. This was because it had been reported that most causal genes of major QTL were found within this genetic distance from the mapped QTL peak (Price, 2006).

Interestingly, similar genes or genes that belong to the same functional group were predicted within the regions of the three QTL (Figure 4.9). This finding was consistent with the additive nature of the QTL. Three groups of processes were identified as the candidates for the mechanisms of resistance (Figure 6.1D). The first was the acquisition of nutrients. Transporters of nitrogenous nutrients, such as ammonium and amino acid, could alter the availability of these essential nutrients for parasite growth in host xylem sap. This is because xylem loading of these nutrients requires the activity of corresponding transporters (Yao *et al.*, 2020; Schjoerring *et al.*, 2002), and subterranean *Striga* completely rely on host supplied nitrogen in the form of free amino acids — especially, asparagine, glutamine and glutamate — and nitrate (Pageau *et al.*, 2003). Also, within the same group, several cell wall biosynthesis and modification proteins were annotated within the QTL regions. Host maintenance of cell

wall integrity could indirectly affect parasite nutrient acquisition by preventing the formation of effective vascular connections since it is known that root parasites, such as *Striga*, secrete cell wall degrading enzymes to weaken host cell wall during parasite intrusion and formation of vascular connections (Mitumasu *et al.*, 2015).

The second group was the host responses to hormones involved in growth and development, such as auxin and ABA. The transport and the local pool of these hormones had a potential to alter the morphology and physiology of the host that could result in a change in host water status that is less suitable to the transfer of water and nutrients to the parasite. For example, insensitive ABA receptor or lower expression level of an ABA receptor gene could make the host less responsive to elevated ABA concentration associated with *Striga* infection (Fujioka *et al.*, 2019b; Frost *et al.*, 1997). This could result in the maintenance of higher transpiration rate in CT8556-37-2-3-1-M, compared to IR64, and inefficient transfer of water and nutrients to the parasite because the high transpiration of the parasite—coupled with restricted host transpiration—was thought to be the main driving force of the flow of nutrient towards the parasite (Fujioka *et al.*, 2019a; Ackroyd & Graves, 1997; Taylor *et al.*, 1996). Also, within this group were auxin-related genes, such as an auxin transporter and IAA-amino conjugator genes that reduce the availability of active auxin. Any changes in the transport or local concentration of auxin could lead to an alteration of host morphology and water status. The upregulation of auxin-amino conjugator genes has been shown to link to *Striga*-infection and result in a reduction in shoot growth (Louden, 2017; Du *et al.*, 2012).

Lastly, the third group was the genes that are involved in biotic stress responses in other pathosystem or genes that are known to be associated with *Striga*-resistance, such as genes in SA- and JA-mediated signalling pathways. These included genes encoding a sulfotransferase that convert SA into sulfonated SA and a component of JA receptor, COI1 protein. A loss of SA accumulation (which leads to a strong induction of JA defence pathway) and the activation of JA-mediated signalling pathway have been shown to increase the level of resistance to *S. hermonthica* in rice (Mutuku *et al.*, 2015). These genes related to *Striga*-

resistance within this group were likely to also underlie the lower number of parasites on CT8556-37-2-3-1-M as the parasite biomass and the parasite number strongly correlated.

Candidate genes for qualitative resistance often include genes such as NB-LRR genes in the case of *R* genes (Jones *et al.*, 2016) and RLK or RLP genes in the case of plant surveillance systems which recognise molecular patterns produced by plant pathogens (Tang *et al.*, 2017). In the case of the major QTL for *S. hermonthica*-resistance on chromosome 12 of the rice genotypes Nipponbare and IR64, RLP genes, which were orthologs of the *Verticillium* wilt resistance *Ve1* gene in tomato, were identified as the candidate genes (Beardon, 2018). Unlike candidate genes for qualitative resistance, the identification of candidate genes for quantitative resistance, which is the case of the slow grow *S. asiatica*-resistance, is more complicated as there are many ways the resistance could arise (Pilet-Nayel *et al.*, 2017; Poland *et al.*, 2009). More evidence is needed to produce a list of candidate genes underlying the slow growth resistance to *S. asiatica*. A transcriptome analysis could shed some light onto the question of which genes within the QTL are likely involved in the resistance at different stages of infection (Figure 6.1Fii). Swarbrick *et al.* (2008) identified several genes that were upregulated or downregulated at the transcriptional level in the resistant rice genotype Nipponbare and the susceptible rice genotypes IAC165 when infected with *S. hermonthica*. Among these was an ortholog of *Ve1*, which was upregulated in Nipponbare in the incompatible interaction with the parasite (Swarbrick *et al.*, 2008) and was later identified as one of the strong candidate genes for resistance to *S. hermonthica* in Nipponbare in a study by Beardon (2018).

For the slow growth resistance, changes in the expression of genes at the transcriptional level within the whole regions of all six QTL during the infection of *S. asiatica* should be investigated using, for example, RNA-Seq analysis. The time points of measurements should cover the time before the formation of host–parasite xylem connections at 2 DAI, during the formation of vascular connections at 3–5 DAI and after the establishment of xylem continuity when the difference in the growth of parasites begins to appear at 8–12 DAI. This

is because the underlying change in the expression of causal genes may precede observable changes in the phenotype.

In order to carry out effective transcriptome analyses, assembled genome and correct gene prediction and annotation are needed. The transcriptome data can be utilised in the gene prediction analyses of both rice genotypes, after the assembled CT8556-37-2-3-1-M sequence has become available, for better prediction of gene models (Keilwagen *et al.*, 2018; Hoff *et al.*, 2016) (Figure 6.1Fii). Available complete genome sequences would also enable more detailed comparison of genes from IR64 and CT8556-37-2-3-1-M. Genes that contain non-synonymous polymorphisms in their coding regions have the potential to be the causal genes. Also, upstream promoter regions should not be neglected as some quantitative resistance genes contain mutations in its promoter regions that prevent the binding of pathogen effectors. For example, quantitative resistance to the bacterium *Xanthomonas oryzae* pv. *oryzae* that cause leaf blight in rice is a result of a mutation of the sugar transporter gene *OsSWEET11* in its promoter region that prevents the binding of bacterial transcription activator-like (TAL) effector to the gene promoter, resulting in a low expression of the gene and limited sugar supply to the pathogen (Niks *et al.*, 2015; Chen *et al.*, 2010).

Once more evidence is gathered for the candidate genes of slow growth resistance, gene validation can be considered (Figure 6.1Fii). The polygenic nature of the slow growth *S. asiatica*-resistance may complicate the validation of candidate genes (Corwin & Kliebenstein, 2017). Cloning of a single candidate resistance gene into susceptible genotypes or a knockout of a single candidate resistance gene in resistant genotype may not result in a measurable change in the level of resistance as the effect size of each locus is small. Instead, functional validation of candidate genes may require simultaneous knockout of multiple genes. This multiplex gene knockout can be achieved by CRISPR/Cas9 or CRISPR/Cpf1 system where multiple guide RNA cassettes are used for targeting multiple genes at the same time (Mishra *et al.*, 2018; Wang *et al.*, 2017; Endo *et al.*, 2016).

6.4 Genome-wide association study for the detection of loci for resistance to *Striga hermonthica* in rice

GWASes have been successfully used for the identification of qualitative and quantitative resistance loci in many plant pathosystems (Bartoli & Roux, 2017; Corwin & Kliebenstein, 2017; Pilet-Nayel *et al.*, 2017), including *S. hermonthica*–maize (Adewale *et al.*, 2020) and *S. hermonthica*–sorghum (Kavuluko *et al.*, 2020). This approach was explored for *S. hermonthica*-resistance, using a panel of 163 diverse *O. glaberrima* genotypes in Chapter 5 (Figure 6.1E). Unfortunately, no significant loci for *S. hermonthica*-resistance were identified in the association analysis. This was mainly because almost all of the *O. glaberrima* genotypes were susceptible to the accession of *S. hermonthica* seeds, collected from Kibos (Kenya), that was used for the phenotypic screen.

Nonetheless, GWASes remain a useful tool to dissect the genetic basis of resistance, especially for quantitative resistance. After all, it has a potential to pinpoint resistance loci because it takes advantage of fine historical recombination breakpoints (Descalsota *et al.*, 2018; Corwin & Kliebenstein, 2017; Nordborg & Weigel, 2008). Since *Striga*-resistance is a function of both host genetics and parasite genetics (Rodenburg *et al.*, 2017), the same diversity panel can be used again with new accessions of *S. hermonthica* seeds or a different *Striga* species (Figure 6.1Fiii). To prevent a similar situation from happening again, a small subset of the diversity panel should be screened to ensure a wide range of resistance/susceptibility levels to maximise the chance of identifying resistance loci.

In addition, more diversity should be included in the diversity panel (Figure 6.1Fiii). For example, some *O. glaberrima* genotypes that are known to be highly resistant to the Kibos accession of *S. hermonthica*, such as CG14 (Cissoko *et al.*, 2011), were not included in the diversity panel used in Chapter 5. Alternatively, a more diverse diversity panel should be used instead to broaden the genetic basis and increase the probability of identifying new sources of resistance to *Striga*. Examples of this include genotypes from the 3000 Rice Genome

Project (Li *et al.*, 2014) and a diversity panel that consists of different rice species and their wild relative, such as *O. barthii*, which is the progenitor of *O. glaberrima* (Choi *et al.*, 2019; Cubry *et al.*, 2018).

References

- Ackroyd, R. D. & Graves, J. D. (1997) The regulation of the water potential gradient in the host and parasite relationship between *Sorghum bicolor* and *Striga hermonthica*. *Annals of Botany*, 80(5), pp. 649–656. doi: 10.1006/anbo.1997.0506
- Adewale, S. A., Badu-Apraku, B., Akinwale, R. O., Paterne, A. A., Gedil, M. & Garcia-Oliveira, A. L. (2020) Genome-wide association study of *Striga* resistance in early maturing white tropical maize inbred lines. *BMC Plant Biology*, 20(1), pp. 1–16. doi: 10.1186/s12870-020-02360-0
- Agrawal, G. K., Rakwal, R., Jwa, N. & Agrawal, V. P. (2002) Characterization of a novel rice gene *OsATX* and modulation of its expression by components of the stress signalling pathways. *Physiologia Plantarum*, 116(1), pp. 87–95. doi: 10.1034/j.1399-3054.2002.1160111.x
- Al-Babili, S. & Bouwmeester, H. J. (2015) Strigolactones, a novel carotenoid-derived plant hormone. *Annual Review of Plant Biology*, 66, pp. 161–186. doi: 10.1146/annurev-arplant-043014-114759
- Amusan, I. O. (2011) *Mechanisms and quantitative trait loci for Striga hermonthica resistance in maize (Zea mays L.) inbred line*. PhD thesis, Purdue University, United States.
- Amusan, I. O., Rich, P. J., Menkir, A., Housley, T. & Ejeta, G. (2008) Resistance to *Striga hermonthica* in a maize inbred line derived from *Zea diploperennis*. *New Phytologist*, 178(1), pp. 157–166. doi: 10.1111/j.1469-8137.2007.02355.x
- Andreeva, A., Howorth, D., Chothia, C., Kulesha, E. & Murzin, A. G. (2014) SCOP2 prototype: A new approach to protein structure mining. *Nucleic Acids Research*, 42(D1), pp. 310–314. doi: 10.1093/nar/gkt1242
- Arends, D., Prins, P., Jansen, R. C. & Broman, K. W. (2010) R/qtl: High-throughput multiple QTL mapping. *Bioinformatics*, 26(23), pp. 2990–2992. doi: 10.1093/bioinformatics/btq565
- Arnaud, M. C., Véronési, C. & Thalouarn, P. (1999) Physiology and histology of resistance to *Striga hermonthica* in *Sorghum bicolor* var. Framida. *Australian Journal of Plant Physiology*, 26(1), pp. 63–70. doi: 10.1071/PP98070
- Arouna, A., Lokossou, J. C., Wopereis, M. C. S., Bruce-Oliver, S. & Roy-Macauley, H. (2017) Contribution of improved rice varieties to poverty reduction and food security in sub-Saharan Africa. *Global Food Security*, 14(January), pp. 54–60. doi: 10.1016/j.gfs.2017.03.001
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000) Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), pp. 25–29. doi: 10.1038/75556
- Atera, E. & Itoh, K. (2011) Evaluation of ecologies and severity of *Striga* weed on rice in sub-Saharan Africa. *Agriculture and Biology Journal of North America*, 2(5), pp. 752–760. doi: 10.5251/abjna.2011.2.5.752.760
- Atera, E. A., Itoh, K., Azuma, T. & Ishii, T. (2012a) Response of NERICA rice to *Striga hermonthica* infections in western Kenya. *International Journal of Agriculture and Biology*, 14(2), pp. 271–275.

- Atera, E. A., Itoh, K., Azuma, T. & Ishii, T. (2012b) Farmers' perspectives on the biotic constraint of *Striga hermonthica* and its control in western Kenya. *Weed Biology and Management*, 12(1), pp. 53–62. doi: 10.1111/j.1445-6664.2012.00435.x
- Attia, K. A., Abdelkhalik, A. F., Ammar, M. H., Wei, C., Yang, J., Lightfoot, D. A., *et al.* (2009) Antisense phenotypes reveal a functional expression of OsARF1, an auxin response factor, in transgenic rice. *Current Issues in Molecular Biology*, 11(SUPPL. 1), pp. 29–34. doi: 10.21775/9781912530069.04
- Awad, A. A., Sato, D., Kusumoto, D., Kamioka, H., Takeuchi, Y. & Yoneyama, K. (2006) Characterization of strigolactones, germination stimulants for the root parasitic plants *Striga* and *Orobanche*, produced by maize, millet and sorghum. *Plant Growth Regulation*, 48(3), pp. 221–227. doi: 10.1007/s10725-006-0009-3
- Balasubramanian, V., Sie, M., Hijmans, R. J. & Otsuka, K. (2007) Increasing rice production in sub-Saharan Africa: Challenges and opportunities, In: Sparks, D. L. (Ed.), *Advances in Agronomy*. Elsevier Masson SAS. pp.55–133.
- Bartoli, C. & Roux, F. (2017) Genome-wide association studies in plant pathosystems: Toward an ecological genomics approach. *Frontiers in Plant Science*, 8(763). doi: 10.3389/fpls.2017.00763
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). doi: 10.18637/jss.v067.i01
- Beardon, E. G. (2018) *Discovery and analysis of novel resistance genes in rice to the parasitic weed Striga hermonthica*. PhD thesis, the University of Sheffield, United Kingdom.
- Beavis, W. D. (1994) The power and deceit of QTL experiments: Lessons from comparative QTL studies, In: *49th Annual Corn & Sorghum Research Conference*. pp.250–266.
- Beavis, W. D. (1998) QTL analyses: Power, precision, and accuracy, In: Paterson, A. H. (Ed.), *Molecular Dissection of Complex Traits*. pp.145–162.
- Bennett, J. R. & Mathews, S. (2006) Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *American Journal of Botany*, 93(7), pp. 1039–1051. doi: 10.3732/ajb.93.7.1039
- Bernardo, R. (2004) What proportion of declared QTL in plants are false? *Theoretical and Applied Genetics*, 109(2), pp. 419–424. doi: 10.1007/s00122-004-1639-3
- Berner, D. K., Kling, J. G. & Singh, B. B. (1995) *Striga* research and control: A perspective from Africa. *Plant Disease*, 79(7), pp. 652. doi: 10.1094/PD-79-0652
- Bloom, A. J. (2015) The increasing importance of distinguishing among plant nitrogen sources. *Current Opinion in Plant Biology*, 25(2), pp. 10–16. doi: 10.1016/j.pbi.2015.03.002
- Botanga, C. J. & Timko, M. P. (2005) Genetic structure and analysis of host and nonhost interactions of *Striga gesnerioides* (witchweed) from Central Florida. *Phytopathology*, 95(10), pp. 1166–1173. doi: 10.1094/PHYTO-95-1166
- Bouchot, J. L., Trimble, W. L., Ditzler, G., Lan, Y., Essinger, S. & Rosen, G. (2013) Advances in machine learning for processing and comparison of metagenomic data, In: Kriete, A. and Eils, R. (Eds.), *Computational Systems Biology: From Molecular Mechanisms to Disease*. Academic Press. pp.295–329.
- Bouwmeester, H. J., Fonne-Pfister, R., Screpanti, C. & De Mesmaeker, A. (2019) Strigolactones: Plant hormones with promising features. *Angewandte Chemie - International Edition*, 58(37), pp. 12 778–12 786. doi: 10.1002/anie.201901626
- Bouwmeester, H. J., Roux, C., Lopez-Raez, J. A. & Bécard, G. (2007) Rhizosphere communication of plants, parasitic plants and AM fungi. *Trends in Plant Science*, 12(5), pp. 224–230. doi: 10.1016/j.tplants.2007.03.009

REFERENCES

- Bouwmeester, H. J., Matusova, R., Zhongkui, S. & Beale, M. H. (2003) Secondary metabolite signalling in host–parasitic plant interactions. *Current Opinion in Plant Biology*, 6(4), pp. 358–364. doi: 10.1016/S1369-5266(03)00065-7
- Bowden, R., Davies, R. W., Heger, A., Pagnamenta, A. T., Cesare, M. De, Oikkonen, L. E., *et al.* (2019) Sequencing of human genomes with nanopore technology. *Nature Communications*, 10(1), pp. 1869. doi: 10.1038/s41467-019-09637-5
- Broad Institute. (2019) *Picard Toolkit*. Broad Institute, GitHub Repository, [available at <http://broadinstitute.github.io/picard/>].
- Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7), pp. 889–890. doi: 10.1093/bioinformatics/btg112
- Broman, K. W. (2001) Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal*, 30(7), pp. 44–52. doi: 10.1038/5000133
- Bromham, L., Cowman, P. F. & Lanfear, R. (2013) Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evolutionary Biology*, 13(1). doi: 10.1186/1471-2148-13-126
- Browning, S. R. (2008) Missing data imputation and haplotype phase inference for genome-wide association studies. *Human Genetics*, 124(5), pp. 439–450. doi: 10.1007/s00439-008-0568-7
- Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., Lewis, S. E., *et al.* (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47, pp. D330–D338. doi: 10.1093/nar/gky1055
- Cardoso, C., Zhang, Y., Jamil, M., Hepworth, J., Charnikhova, T., Dimkpa, S. O. N., *et al.* (2014) Natural variation of rice strigolactone biosynthesis is associated with the deletion of two *MAX1* orthologs. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), pp. 2379–2384. doi: 10.1073/pnas.1317360111
- Cardoso, C., Ruyter-Spira, C. & Bouwmeester, H. J. (2011) Strigolactones and root infestation by plant-parasitic *Striga*, *Orobanch*e and *Phelipanche* spp. *Plant Science*, 180(3), pp. 414–420. doi: 10.1016/j.plantsci.2010.11.007
- Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., *et al.* (2010) Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, 468(7323), pp. 527–532. doi: 10.1038/nature09606
- Chikhi, R. & Medvedev, P. (2014) Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, 30(1), pp. 31–37. doi: 10.1093/bioinformatics/btt310
- Choi, J. Y., Zaidem, M., Gutaker, R., Dorph, K., Singh, R. K. & Purugganan, M. D. (2019) The complex geography of domestication of the African rice *Oryza glaberrima*. *PLoS Genetics*, 15(3), pp. e1007414. doi: 10.1371/journal.pgen.1007414
- Churchill, G. A. & Doerge, R. W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, 138(3), pp. 963–971. doi: 10.1007/s11703-007-0022-y
- Cissoko, M. (2012) *Understanding resistance in inter-specific rice cultivars to the parasitic witchweed Striga*. PhD thesis, the University of Sheffield, United Kingdom.
- Cissoko, M., Boissard, A., Rodenburg, J., Press, M. C. & Scholes, J. D. (2011) New Rice for Africa (NERICA) cultivars exhibit different levels of post-attachment resistance against the parasitic weeds *Striga hermonthica* and *Striga asiatica*. *New Phytologist*, 192(4), pp. 952–963. doi: 10.1111/j.1469-8137.2011.03846.x

- Clarke, C. R., Timko, M. P., Yoder, J. I., Axtell, M. J. & Westwood, J. H. (2019) Molecular dialog between parasitic plants and their hosts. *Annual Review of Phytopathology*, 57, pp. 279–299. doi: 10.1146/annurev-phyto-082718-100043
- Cochrane, V. & Press, M. C. (1997) Geographical distribution and aspects of the ecology of the hemiparasitic angiosperm *Striga asiatica* (L.) Kuntze: A herbarium study. *Journal of Tropical Ecology*, 13(3), pp. 371–380. doi: 10.1017/S0266467400010579
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. & Pang, E. C. K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1–2), pp. 169–196. doi: 10.1007/s10681-005-1681-5
- Corwin, J. A. & Kliebenstein, D. J. (2017) Quantitative resistance: More than just perception of a pathogen. *Plant Cell*, 29(4), pp. 655–665. doi: 10.1105/tpc.16.00915
- Cubry, P., Pidon, H., Ta, K. N., Tranchant-Dubreuil, C., Thuillet, A.-C., Holzinger, M., *et al.* (2020) Genome wide association study pinpoints key agronomic QTLs in African rice *Oryza glaberrima*. *Rice*, 13(66). doi: 10.1186/s12284-020-00424-1
- Cubry, P., Tranchant-Dubreuil, C., Thuillet, A. C., Monat, C., Ndjiondjop, M. N., Labadie, K., *et al.* (2018) The rise and fall of African rice cultivation revealed by analysis of 246 new genomes. *Current Biology*, 28(14), pp. 2274–2282. doi: 10.1016/j.cub.2018.05.066
- Cui, S., Wada, S., Tobimatsu, Y., Takeda, Y., Saucet, S. B., Takano, T., *et al.* (2018) Host lignin composition affects haustorium induction in the parasitic plants *Phtheirospermum japonicum* and *Striga hermonthica*. *New Phytologist*, 218(2), pp. 710–723. doi: 10.1111/nph.15033
- Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14(7), pp. 1394–1403. doi: 10.1101/gr.2289704
- de Abreu Neto, J. B., Hurtado-Perez, M. C., Wimmer, M. A. & Frei, M. (2017) Genetic factors underlying boron toxicity tolerance in rice: Genome-wide association study and transcriptomic analysis. *Journal of Experimental Botany*, 68(3), pp. 687–700. doi: 10.1093/jxb/erw423
- De Groote, H., Wangare, L., Kanampiu, F., Odeno, M., Diallo, A., Karaya, H., *et al.* (2008) The potential of a herbicide resistant maize technology for *Striga* control in Africa. *Agricultural Systems*, 97(1–2), pp. 83–94. doi: 10.1016/j.agsy.2007.12.003
- Demont, M. (2013) Reversing urban bias in African rice markets: A review of 19 national rice development strategies. *Global Food Security*, 2(3), pp. 172–181. doi: 10.1016/j.gfs.2013.07.001
- Descalsota, G. I. L., Swamy, B. P. M., Zaw, H., Inabangan-Asilo, M. A., Amparado, A., Mauleon, R., *et al.* (2018) Genome-wide association mapping in a rice magic plus population detects qtls and genes useful for biofortification. *Frontiers in Plant Science*, 9(1347), pp. 1–20. doi: 10.3389/fpls.2018.01347
- Devlin, B. & Roeder, K. (1999) Genomic control for association studies. *Biometrics*, 55(4), pp. 997–1004. doi: 10.1111/j.0006-341X.1999.00997.x
- Diagne, A., Midingoyi, S.-K. G. & Kinkinginhoun-Medagbe, F. M. (2013) Impact of NERICA adoption on rice yield: Evidence from West Africa, In: Otsuka, K. and Larson, D. F. (Eds.), *An African Green Revolution*. Dordrecht: Springer Netherlands. pp.143–163.
- Dormatey, R., Sun, C., Ali, K., Coulter, J. A., Bi, Z. & Bai, J. (2020) Gene pyramiding for sustainable crop improvement against biotic and abiotic stresses. *Agronomy*, 10(9), pp. 1–21. doi: 10.3390/agronomy10091255
- Dorr, I. (1997) How *Striga* parasitizes its host: A TEM and SEM study. *Annals of Botany*, 79(5), pp. 463–472. doi: 10.1006/anbo.1996.0385

REFERENCES

- Du, H., Wu, N., Fu, J., Wang, S., Li, X., Xiao, J., *et al.* (2012) A GH3 family member, OsGH3-2, modulates auxin and abscisic acid levels and differentially affects drought and cold tolerance in rice. *Journal of Experimental Botany*, 63(18), pp. 6467–6480. doi: 10.1093/jxb/ers300
- Duriez, P., Vautrin, S., Auriac, M. C., Bazerque, J., Boniface, M. C., Callot, C., *et al.* (2019) A receptor-like kinase enhances sunflower resistance to *Orobanche cumana*. *Nature Plants*, 5(12), pp. 1211–1215. doi: 10.1038/s41477-019-0556-z
- Ejeta, G. (2007) The *Striga* scourge in Africa, In: Ejeta, G. and Gressel, J. (Eds.), *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt*. World Scientific Publishing. pp.3–16.
- Ejeta, G. & Butler, L. G. (1993) Host plant resistance to *Striga*, In: Buxton, D. R., Shibles, R., Forsberg, R. A., Blad, B. L., Asay, K. H., Paulsen, G. M., *et al.* (Eds.), *International Crop Science I*. Madison, WI, USA: Crop Science Society of America. pp.561–569.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), pp. D427–D432. doi: 10.1093/nar/gky995
- Ellis, J. G., Lagudah, E. S., Spielmeier, W. & Dodds, P. N. (2014) The past, Present and future of breeding rust resistant wheat. *Frontiers in Plant Science*, 5(NOV), pp. 1–13. doi: 10.3389/fpls.2014.00641
- Endo, M., Mikami, M. & Toki, S. (2016) Biallelic gene targeting in rice. *Plant Physiology*, 170(2), pp. 667–677. doi: 10.1104/pp.15.01663
- Falconer, D. S. & Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics*. Harlow, Essex, UK: Addison Wesley Longman Limited.
- Feenstra, B., Skovgaard, I. M. & Broman, K. W. (2006) Mapping quantitative trait loci by an extension of the Haley–Knott regression method using estimating equations. *Genetics*, 173(4), pp. 2269–2282. doi: 10.1534/genetics.106.058537
- Fernández-Aparicio, M., Reboud, X. & Gibot-Leclerc, S. (2016) Broomrape weeds. Underground mechanisms of parasitism and associated strategies for their control: A review. *Frontiers in Plant Science*, 7(135). doi: 10.3389/fpls.2016.00135
- Fernie, A. R. & Keurentjes, J. J. B. (2018) Genetics, genomics and metabolomics, In: Fernie, A. R. and Keurentjes, J. J. B. (Eds.), *Annual Plant Reviews*. Chichester, UK: John Wiley & Sons, Ltd. pp.219–259.
- Finkelstein, R. (2013) Abscisic Acid Synthesis and Response. *The Arabidopsis Book*, 2013(11). doi: 10.1199/tab.0166
- Foissac, S., Gouzy, J., Rombauts, S., Mathe, C., Amsalem, J., Sterck, L., *et al.* (2008) Genome annotation in plants and fungi: EuGene as a model platform. *Current Bioinformatics*, 3(2), pp. 87–97. doi: 10.2174/157489308784340702
- Fragoso, C. A., Moreno, M., Wang, Z., Heffelfinger, C., Arbelaez, L. J., Aguirre, J. A., *et al.* (2017) Genetic architecture of a rice nested association mapping population. *G3: Genes, Genomes, Genetics*, 7(6), pp. 1913–1926. doi: 10.1534/g3.117.041608
- Fragoso, C. A., Heffelfinger, C., Zhao, H. & Dellaporta, S. L. (2016) Imputing genotypes in biallelic populations from low-coverage sequence data. *Genetics*, 202(2), pp. 487–495. doi: 10.1534/genetics.115.182071
- Frichot, E. & François, O. (2015) LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), pp. 925–929. doi: 10.1111/2041-210X.12382
- Frost, D. L., Gurney, A. L., Press, M. C. & Scholes, J. D. (1997) *Striga hermonthica* reduces photosynthesis in sorghum: the importance of stomatal limitations and a potential role for ABA? *Plant, Cell and Environment*, 20(4), pp. 483–492. doi: 10.1046/j.1365-3040.1997.d01-87.x

- Fujioka, H., Samejima, H., Mizutani, M., Okamoto, M. & Sugimoto, Y. (2019a) How does *Striga hermonthica* bewitch its hosts? *Plant Signaling and Behavior*, *14*(7), pp. 1–2. doi: 10.1080/15592324.2019.1605810
- Fujioka, H., Samejima, H., Suzuki, H., Mizutani, M., Okamoto, M. & Sugimoto, Y. (2019b) Aberrant protein phosphatase 2C leads to abscisic acid insensitivity and high transpiration in parasitic *Striga*. *Nature Plants*, *5*(3), pp. 258–262. doi: 10.1038/s41477-019-0362-7
- Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., *et al.* (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science*, *325*(5943), pp. 998–1001. doi: 10.1126/science.1175550
- Gasura, E., Setimela, P., Mabasa, S., Rwafa, R., Kageler, S. & Nyakurwa, C. (2019) Response of IITA maize inbred lines bred for *Striga hermonthica* resistance to *Striga asiatica* and associated resistance mechanisms in southern Africa. *Euphytica*, *215*(10), pp. 1–15. doi: 10.1007/s10681-019-2467-5
- Gethi, J. G. & Smith, M. E. (2004) Genetic responses of single crosses of maize to *Striga hermonthica* (Del.) Benth. and *Strigan asiatica* (L.) Kuntze. *Crop Science*, *44*(6), pp. 2068–2077. doi: 10.2135/cropsci2004.2068
- Gobena, D., Shimels, M., Rich, P. J., Ruyter-Spira, C., Bouwmeester, H., Kanuganti, S., *et al.* (2017) Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes *Striga* resistance. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(17), pp. 4471–4476. doi: 10.1073/pnas.1618965114
- Goldwasser, Y., Westwood, J. H. & Yoder, J. I. (2002) The use of *Arabidopsis* to study interactions between parasitic angiosperms and their plant hosts. *The Arabidopsis Book*, *2002*(1). doi: 10.1199/tab.0035
- Goyet, V., Wada, S., Cui, S., Wakatake, T., Shirasu, K., Montiel, G., *et al.* (2019) Haustorium inducing factors for parasitic Orobanchaceae. *Frontiers in Plant Science*, *10*(1056), pp. 1–8. doi: 10.3389/fpls.2019.01056
- Graves, J. D., Press, M. C. & Stewart, G. R. (1989) A carbon balance model of the sorghum–*Striga hermonthica* host–parasite association. *Plant, Cell and Environment*, *12*(1), pp. 101–107. doi: 10.1111/j.1365-3040.1989.tb01921.x
- Grenier, C., Ibrahim, Y., Haussmann, B. I. G., Kiambi, D. & Ejeta, G. (2007) Marker-assisted selection for *Striga* resistance in sorghum, In: Ejeta, G. and Gressel, J. (Eds.), *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt*. World Scientific Publishing. pp.159–171.
- Grones, P. & Friml, J. (2015) Auxin transporters and binding proteins at a glance. *Journal of Cell Science*, *128*(1), pp. 1–7. doi: 10.1242/jcs.159418
- Guo, F., Huang, Y., Qi, P., Lian, G., Hu, X., Han, N., *et al.* (2020) Functional analysis of auxin receptor *OsTIR1/OsAFB* family members in rice grain yield, tillering, plant height, root system, germination, and auxinic herbicide resistance. *New Phytologist*. doi: 10.1111/nph.17061
- Gurney, A. L., Slate, J., Press, M. C. & Scholes, J. D. (2006) A novel form of resistance in rice to the angiosperm parasite *Striga hermonthica*. *New Phytologist*, *169*(1), pp. 199–208. doi: 10.1111/j.1469-8137.2005.01560.x
- Gurney, A. L., Grimanelli, D., Kanampiu, F., Hoisington, D., Scholes, J. D. & Press, M. C. (2003) Novel sources of resistance to *Striga hermonthica* in *Tripsacum dactyloides*, a wild relative of maize. *New Phytologist*, *160*(3), pp. 557–568. doi: 10.1046/j.1469-8137.2003.00904.x
- Gurney, A. L., Taylor, A., Mbwaga, A., Scholes, J. D. & Press, M. C. (2002) Do maize cultivars demonstrate tolerance to the parasitic weed *Striga asiatica*? *Weed Research*, *42*(4), pp. 299–306. doi: 10.1046/j.1365-3180.2002.00287.x

REFERENCES

- Gurney, A. L., Press, M. C. & Scholes, J. D. (1999) Infection time and density influence the response of sorghum to the parasitic angiosperm *Striga hermonthica*. *New Phytologist*, 143(3), pp. 573–580. doi: 10.1046/j.1469-8137.1999.00467.x
- Gworgwor, N. A. (2003) Resistance of sorghum varieties to *Striga hermonthica*. *Tropical Science*, 43, pp. 48–52.
- Haley, C. S. & Knott, S. A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4), pp. 315–324. doi: 10.1038/hdy.1992.131
- Harahap, Z., Ampong-Nyarko, K. & Olela, J. C. (1993) *Striga hermonthica* resistance in upland rice. *Crop Protection*, 12(3), pp. 229–231. doi: 10.1016/0261-2194(93)90114-X
- Hausmann, B. I. G., Hess, D. E., Omany, G. O., Reddy, B. V. S., Welz, H. G. & Geiger, H. H. (2001a) Major and minor genes for stimulation of *Striga hermonthica* seed germination in sorghum, and interaction with different *Striga* populations. *Crop Science*, 41(5), pp. 1507–1512. doi: 10.2135/cropsci2001.4151507x
- Hausmann, B. I. G., Hess, D. E., Reddy, B. V. S., Mukuru, S. Z., Kayentao, M., Welz, H. G., *et al.* (2001b) Pattern analysis of genotype × environment interaction for *Striga* resistance and grain yield in African sorghum trials. *Euphytica*, 122(2), pp. 297–308. doi: 10.1023/A:1012909719137
- Hausmann, B. I. G., Hess, D. E., Reddy, B. V. S., Mukuru, S. Z., Kayentao, M., Welz, H. G., *et al.* (2001c) Quantitative-genetic parameters of sorghum growth under *Striga* infestation in Mali and Kenya. *Plant Breeding*, 120(1), pp. 49–56. doi: 10.1046/j.1439-0523.2001.00546.x
- Hausmann, B. I. G., Hess, D. E., Omany, G. O., Folkertsma, R. T., Reddy, B. V. S., Kayentao, M., *et al.* (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *Theoretical and Applied Genetics*, 109(5), pp. 1005–1016. doi: 10.1007/s00122-004-1706-9
- Hausmann, B. I. G., Hess, D. E., Welz, H. G. & Geiger, H. H. (2000) Improved methodologies for breeding *Striga*-resistant sorghums. *Field Crops Research*, 66(3), pp. 195–211. doi: 10.1016/S0378-4290(00)00076-9
- Hautier, Y., Hector, A., Vojtech, E., Purves, D. & Turnbull, L. A. (2010) Modelling the growth of parasitic plants. *Journal of Ecology*, 98(4), pp. 857–866. doi: 10.1111/j.1365-2745.2010.01657.x
- Hearne, S. J. (2009) Control-the *Striga* conundrum. *Pest Management Science*, 65(5), pp. 603–614. doi: 10.1002/ps.1735
- Heffelfinger, C., Fragoso, C. A. & Lorieux, M. (2017) Constructing linkage maps in the genomics era with MapDisto 2.0. *Bioinformatics*, 33(14), pp. 2224–2225. doi: 10.1093/bioinformatics/btx177
- Hegenauer, V., Fürst, U., Kaiser, B., Smoker, M., Zipfel, C., Felix, G., *et al.* (2016) Detection of the plant parasite *Cuscuta reflexa* by a tomato cell surface receptor. *Science*, 353(6298), pp. 478–481. doi: 10.1126/science.aaf3919
- Hewitt, E. J. (1966) *Sand and Water Culture Methods Used in the Study of Plant Nutrition*. Common Agricultural Bureau.
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. (2016) BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32(5), pp. 767–769. doi: 10.1093/bioinformatics/btv661
- Hood, M. E., Condon, J. M., Timko, M. P. & Riopel, J. L. (1998) Primary haustorial development of *Striga asiatica* on host and nonhost species. *Phytopathology*, 88(1), pp. 70–75. doi: 10.1094/phyto.1998.88.1.70

- Huang, K., Mellor, K. E., Paul, S. N., Lawson, M. J., Mackey, A. J. & Timko, M. P. (2012) Global changes in gene expression during compatible and incompatible interactions of cowpea (*Vigna unguiculata* L.) with the root parasitic angiosperm *Striga gesnerioides*. *BMC Genomics*, *13*(1), pp. 402. doi: 10.1186/1471-2164-13-402
- Hyndman, R. J. & Fan, Y. (1996) Sample quantiles in statistical packages. *American Statistician*, *50*(4), pp. 361–365. doi: 10.1080/00031305.1996.10473566
- Ichihashi, Y., Mutuku, J. M., Yoshida, S. & Shirasu, K. (2015) Transcriptomics exposes the uniqueness of parasitic plants. *Briefings in Functional Genomics*, *14*(4), pp. 275–282. doi: 10.1093/bfpg/elv001
- Igbinnosa, I. & Thalouarn, P. A. (1996) Nitrogen assimilation enzyme activities in witchweed (*Striga*) in hosts presence and absence. *Weed Science*, *44*(2), pp. 224–232. doi: 10.1017/s0043174500093826
- Inoue, T., Yamauchi, Y., Eltayeb, A. H., Samejima, H., Babiker, A. G. T. & Sugimoto, Y. (2013) Gas exchange of root hemi-parasite *Striga hermonthica* and its host *Sorghum bicolor* under short-term soil water stress. *Biologia Plantarum*, *57*(4), pp. 773–777. doi: 10.1007/s10535-013-0348-7
- International Rice Research Institute. (2013) *World Rice Statistics* [online]. Available from: <http://ricestat.irri.org:8080/wrsv3/entrypoint.htm> [Accessed 15 December 2020].
- Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., *et al.* (2017) ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. *Genome Research*, *27*, pp. 768–777. doi: 10.1101/gr.214346.116.Freely
- Jamil, M., Rodenburg, J., Charnikhova, T. & Bouwmeester, H. J. (2011a) Pre-attachment *Striga hermonthica* resistance of New Rice for Africa (NERICA) cultivars based on low strigolactone production. *The New phytologist*, *192*(4), pp. 964–75. doi: 10.1111/j.1469-8137.2011.03850.x
- Jamil, M., Charnikhova, T., Cardoso, C., Jamil, T., Ueno, K., Verstappen, F., *et al.* (2011b) Quantification of the relationship between strigolactones and *Striga hermonthica* infection in rice under varying levels of nitrogen and phosphorus. *Weed Research*, *51*(4), pp. 373–385. doi: 10.1111/j.1365-3180.2011.00847.x
- Jansen, R. C. (2007) Quantitative trait loci in inbred lines, In: Balding, D. J., Bishop, M., and Cannings, C. (Eds.), *Handbook of Statistical Genetics*. Chippingham, UK: John Wiley & Sons, Ltd. pp.589–622.
- Jansen, R. C. (1994) Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, *138*(3), pp. 871–881.
- Jansen, R. C. & Stam, P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, *136*(4), pp. 1447–1455.
- Jansen, R. C. (1993) Interval mapping of multiple quantitative trait loci. *Genetics*, *135*(1), pp. 205–211.
- Jia, K. P., Baz, L. & Al-Babili, S. (2018) From carotenoids to strigolactones. *Journal of Experimental Botany*, *69*(9), pp. 2189–2204. doi: 10.1093/jxb/erx476
- Joehanes, R. & Nelson, J. C. (2008) QGene 4.0, an extensible Java QTL-analysis platform. *Bioinformatics*, *24*(23), pp. 2788–2789. doi: 10.1093/bioinformatics/btn523
- Joel, D. M., Hershenhorn, J., Eizenberg, H., Aly, R., Ejeta, G., Rich, P. J., *et al.* (2007) Biology and management of weedy root parasites, In: Janick, J. (Ed.), *Horticultural Reviews*. Hoboken, NJ, USA: John Wiley & Sons, Inc. pp.267–349.
- Johnson, D. E., Riches, C. R., Diallo, R. & Jones, M. J. (1997) *Striga* on rice in West Africa; Crop host range and the potential of host resistance. *Crop Protection*, *16*(2), pp. 153–157. doi: 10.1016/S0261-2194(96)00079-8
- Jones, J. D. G., Vance, R. E. & Dangl, J. L. (2016) Intracellular innate immune surveillance devices in plants and animals. *Science*, *354*(6316), pp. aaf6395–aaf6395. doi: 10.1126/science.aaf6395

REFERENCES

- Jones, M. P., Dingkuhn, M., Aluko, G. & Semon, M. (1997a) Interspecific *O. sativa* L. × *O. glaberrima* Steud: Progenies in upland rice improvement. *Euphytica*, 92, pp. 237–246. doi: 10.1023/A:1002969932224
- Jones, M. P., Mande, S. & Aliko, K. (1997b) Diversity and potential of *Oryza glaberrima* Steud in upland rice breeding. *Breeding Science*, 47, pp. 395–398.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., *et al.* (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), pp. 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jordá, L., Sopena-Torres, S., Escudero, V., Nuñez-Corcuera, B., Delgado-Cerezo, M., Torii, K. U., *et al.* (2016) ERECTA and BAK1 receptor like kinases interact to regulate immune responses in *Arabidopsis*. *Frontiers in Plant Science*, 7(897), pp. 1–15. doi: 10.3389/fpls.2016.00897
- Kaewchumnong, K. & Price, A. H. (2008) A study on the susceptibility of rice cultivars to *Striga hermonthica* and mapping of *Striga* tolerance quantitative trait loci in rice. *New Phytologist*, 180(1), pp. 206–216. doi: 10.1111/j.1469-8137.2008.02568.x
- Kanampiu, F. K., Kabambe, V., Massawe, C., Jasi, L., Friesen, D., Ransom, J. K., *et al.* (2003) Multi-site, multi-season field tests demonstrate that herbicide seed-coating herbicide-resistance maize controls *Striga* spp. and increases yields in several African countries. *Crop Protection*, 22(5), pp. 697–706. doi: 10.1016/s0261-2194(03)00007-3
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., *et al.* (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), pp. 348–354. doi: 10.1038/ng.548
- Kant, S., Bi, Y. M., Zhu, T. & Rothstein, S. J. (2009) *SAUR39*, a small auxin-up RNA gene, acts as a negative regulator of auxin synthesis and transport in rice. *Plant Physiology*, 151(2), pp. 691–701. doi: 10.1104/pp.109.143875
- Kao, C. H., Zeng, Z. B. & Teasdale, R. D. (1999) Multiple interval mapping for quantitative trait loci. *Genetics*, 152(3), pp. 1203–1216. doi: 10.1007/s11434-010-0032-7
- Kavuluko, J., Kibe, M., Sugut, I., Kibet, W., Masanga, J., Mutinda, S., *et al.* (2020) *GWAS provides biological insights into mechanisms of the parasitic plant (Striga) resistance in sorghum* [online]. Available from: <https://doi.org/10.21203/rs.3.rs-54024/v1>
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., *et al.* (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6(1), pp. 4. doi: 10.1186/1939-8433-6-4
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. (2018) Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, 19(1), pp. 189. doi: 10.1186/s12859-018-2203-5
- Khan, Z. R., Hassanali, A., Overholt, W., Khamis, T. M., Hooper, A. M., Pickett, J. A., *et al.* (2002) Control of witchweed *Striga hermonthica* by intercropping with *Desmodium* spp., and the mechanism defined as allelopathic. *Journal of Chemical Ecology*, 28(9), pp. 1871–1885. doi: 10.1023/A:1020525521180
- Kifuko-Koech, M., Pypers, P., Okalebo, J. R., Othieno, C. O., Khan, Z. R., Pickett, J. A., *et al.* (2012) The impact of *Desmodium* spp. and cutting regimes on the agronomic and economic performance of *Desmodium*–maize intercropping system in western Kenya. *Field Crops Research*, 137, pp. 97–107. doi: 10.1016/j.fcr.2012.08.007
- Kosambi, D. D. (1944) The estimation of map distances from recombination values. *Annals of Eugenics*, 12(1), pp. 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

- Kountche, B. A., Al-Babili, S. & Haussmann, B. I. G. (2016) Striga: A persistent problem on millets, In: Das, I. K. and Padmaja, P. G. (Eds.), *Biotic Stress Resistance in Millets*. Academic Press. pp.173–203.
- Krattinger, S. G., Kang, J., Bräunlich, S., Boni, R., Chauhan, H., Selter, L. L., *et al.* (2019) Abscisic acid is a substrate of the ABC transporter encoded by the durable wheat disease resistance gene *Lr34*. *New Phytologist*, 223(2), pp. 853–866. doi: 10.1111/nph.15815
- Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., *et al.* (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science*, 323(5919), pp. 1360–1363. doi: 10.1126/science.1166453
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., *et al.* (2004) Versatile and open software for comparing large genomes. *Genome biology*, 5(2). doi: 10.1186/gb-2004-5-2-r12
- Lander, E. S. & Botstein, S. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121(1), pp. 185.
- Lane, J. A. (1996) Characterization of virulence and geographic distribution of *Striga gesnerioides* on cowpea in West Africa. *Plant Disease*, 80(3), pp. 299. doi: 10.1094/PD-80-0299
- Lane, J. A., Bailey, J. A., Butler, R. C. & Terry, P. J. (1993) Resistance of cowpea [*Vigna unguiculata* (L.) Walp.] to *Striga gesnerioides* (Willd.) Vatke, a parasitic angiosperm. *New Phytologist*, 125(2), pp. 405–412. doi: 10.1111/j.1469-8137.1993.tb03893.x
- Lane, J. A., Moore, T. H. M., Child, D. V., Cardwell, K. F., Singh, B. B. & Bailey, J. A. (1994) Virulence characteristics of a new race of the parasitic angiosperm, *Striga gesnerioides*, from southern Benin on cowpea (*Vigna unguiculata*). *Euphytica*, 72(3), pp. 183–188. doi: 10.1007/BF00034156
- Lekklar, C., Pongpanich, M., Suriya-Arunroj, D., Chinpongpanich, A., Tsai, H., Comai, L., *et al.* (2019) Genome-wide association study for salinity tolerance at the flowering stage in a panel of rice accessions from Thailand. *BMC Genomics*, 20(1), pp. 1–18. doi: 10.1186/s12864-018-5317-2
- Lemarié, S., Robert-Seilantantz, A., Lariagon, C., Lemoine, J., Marnet, N., Levrel, A., *et al.* (2015) Camalexin contributes to the partial resistance of *Arabidopsis thaliana* to the biotrophic soilborne protist *Plasmodiophora brassicae*. *Frontiers in Plant Science*, 6(539), pp. 1–11. doi: 10.3389/fpls.2015.00539
- Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), pp. 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, H., Ribaut, J. M., Li, Z. & Wang, J. (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theoretical and Applied Genetics*, 116(2), pp. 243–260. doi: 10.1007/s00122-007-0663-5
- Li, H., Ye, G. & Wang, J. (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics*, 175(1), pp. 361–374. doi: 10.1534/genetics.106.066811
- Li, J. Y., Wang, J. & Zeigler, R. S. (2014) The 3,000 rice genomes project: New opportunities and challenges for future rice research. *GigaScience*, 3(1), pp. 1–3. doi: 10.1186/2047-217X-3-8
- Li, J. & Timko, M. P. (2009) Gene-for-gene resistance in *Striga*–cowpea associations. *Science*, 325(5944), pp. 1094–1094. doi: 10.1126/science.1174754
- Li, J., Lis, K. E. & Timko, M. P. (2009) Molecular genetics of race-specific resistance of cowpea to *Striga gesnerioides* (Willd.). *Pest Management Science*, 65(5), pp. 520–527. doi: 10.1002/ps.1722

REFERENCES

- Linares, O. F. (2002) African rice (*Oryza glaberrima*): History and future potential. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(25), pp. 16360–16365. doi: 10.1073/pnas.252604599
- Lischer, H. E. L. & Shimizu, K. K. (2017) Reference-guided *de novo* assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, *18*(1), pp. 1–12. doi: 10.1186/s12859-017-1911-6
- Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O. & Borodovsky, M. (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), pp. 6494–6506. doi: 10.1093/nar/gki937
- Lorieux, M. (2018) Is this a fake QTL? A short note on analysis of multiple QTL peaks on the same chromosome. *bioRxiv*, pp. 270736. doi: 10.1101/270736
- Lorieux, M. (2012) MapDisto: Fast and efficient computation of genetic linkage maps. *Molecular Breeding*, *30*(2), pp. 1231–1235. doi: 10.1007/s11032-012-9706-y
- Louden, R. (2017) *The role of plant growth regulators in the interaction between the parasitic weed Striga hermonthica and rice*. PhD thesis, the University of Sheffield, United Kingdom.
- Lozano-Baena, M. D., Prats, E., Moreno, M. T., Rubiales, D. & Pérez-de-Luque, A. (2007) *Medicago truncatula* as a model for nonhost resistance in legume–parasitic plant interactions. *Plant Physiology*, *145*(2), pp. 437–449. doi: 10.1104/pp.107.097089
- Lumba, S., Hollbrook-Smith, D. & McCourt, P. (2017a) The perception of strigolactones in vascular plants. *Nature Chemical Biology*, *13*(6), pp. 599–606. doi: 10.1038/nchembio.2340
- Lumba, S., Subha, A. & McCourt, P. (2017b) Found in translation: Applying lessons from model systems to strigolactone signaling in parasitic plants. *Trends in Biochemical Sciences*, *42*(7), pp. 556–565. doi: 10.1016/j.tibs.2017.04.006
- Lynch, M. & Walsh, B. (1996) *Genetics and Analysis of Quantitative Traits*. Sunderland, MA, USA: Sinauer Associates.
- Maiti, R. K., Ramaiah, K. V., Bisen, S. S. & Chidley, V. L. (1984) A comparative study of the haustorial development of *Striga asiatica* (L.) Kuntze on sorghum cultivars. *Annals of Botany*, *54*(4), pp. 447–457. doi: 10.1093/oxfordjournals.aob.a086816
- Majda, M. & Robert, S. (2018) The role of auxin in cell wall expansion. *International Journal of Molecular Sciences*, *19*(4). doi: 10.3390/ijms19040951
- Manichaikul, A., Moon, J. Y., Sen, S., Yandell, B. S. & Broman, K. W. (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics*, *181*(3), pp. 1077–1086. doi: 10.1534/genetics.108.094565
- Manyong, V. M., Nindi, S. J., Alene, A. D., Odhiambo, G. D., Omany, G., D, M. H., *et al.* (2008) *Farmer Perceptions of Imazapyr-Resistant (IR) Maize Technology on the Control of Striga in Western Kenya*. Nairobi, Kenya: African Agricultural Technology Foundation.
- Marchini, J. & Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), pp. 499–511. doi: 10.1038/nrg2796
- Masumoto, N., Suzuki, Y., Cui, S., Wakazaki, M., Sato, M., Kumaishi, K., *et al.* (2020) Three-dimensional reconstructions of the internal structures of haustoria in parasitic Orobanchaceae. *bioRxiv*, pp. 2020.05.07.083055. doi: 10.1101/2020.05.07.083055
- Matthies, D. (2017) Interactions between a root hemiparasite and 27 different hosts: Growth, biomass allocation and plant architecture. *Perspectives in Plant Ecology, Evolution and Systematics*, *24*, pp. 118–137. doi: 10.1016/j.ppees.2016.12.006

- Matthus, E., Wu, L. B., Ueda, Y., Höller, S., Becker, M. & Frei, M. (2015) Loci, genes, and mechanisms associated with tolerance to ferrous iron toxicity in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 128(10), pp. 2085–2098. doi: 10.1007/s00122-015-2569-y
- Mauricio, R. (2001) Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nature Reviews Genetics*, 2(5), pp. 370–381. doi: 10.1038/35072085
- Mbuvi, D. A., Masiga, C. W., Kuria, E., Masanga, J., Wamalwa, M., Mohamed, A., *et al.* (2017) Novel sources of witchweed (*Striga*) resistance from wild sorghum accessions. *Frontiers in Plant Science*, 8(FEBRUARY). doi: 10.3389/fpls.2017.00116
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), pp. 1297–1303. doi: 10.1101/gr.107524.110
- Melotto, M., Zhang, L., Oblessuc, P. R. & He, S. Y. (2017) Stomatal defense a decade later. *Plant Physiology*, 174(2), pp. 561–571. doi: 10.1104/pp.16.01853
- Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S. Y. (2006) Plant stomata function in innate immunity against bacterial invasion. *Cell*, 126(5), pp. 969–980. doi: 10.1016/j.cell.2006.06.054
- Members of the Complex Trait Consortium. (2003) The nature and identification of quantitative trait loci: A community's view. *Nature Reviews Genetics*, 4(11), pp. 911–916. doi: 10.1038/nrg1206
- Meng, L., Li, H., Zhang, L. & Wang, J. (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop Journal*, 3(3), pp. 269–283. doi: 10.1016/j.cj.2015.01.001
- Menkir, A. (2006) Assessment of reactions of diverse maize inbred lines to *Striga hermonthica* (Del.) Benth. *Plant Breeding*, 125(2), pp. 131–139. doi: 10.1111/j.1439-0523.2006.01175.x
- Menkir, A. & Meseka, S. (2019) Genetic improvement in resistance to *Striga* in tropical maize hybrids. *Crop Science*, 59(6), pp. 2484–2497. doi: 10.2135/cropsci2018.12.0749
- Meyer, R. S., Choi, J. Y., Sanches, M., Plessis, A., Flowers, J. M., Amas, J., *et al.* (2016) Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nature Genetics*, 48(9), pp. 1083–1088. doi: 10.1038/ng.3633
- Midega, C. A. O., Pickett, J., Hooper, A., Pittchar, J. & Khan, Z. R. (2016) Maize landraces are less affected by *Striga hermonthica* relative to hybrids in Western Kenya. *Weed Technology*, 30(1), pp. 21–28. doi: 10.1614/wt-d-15-00055.1
- Mishra, R., Joshi, R. K. & Zhao, K. (2018) Genome editing in rice: Recent advances, challenges, and future implications. *Frontiers in Plant Science*, 9(1361). doi: 10.3389/fpls.2018.01361
- Mitsumasu, K., Seto, Y. & Yoshida, S. (2015) Apoplastic interactions between plants and plant root intruders. *Frontiers in Plant Science*, 6(August), pp. 1–17. doi: 10.3389/fpls.2015.00617
- Mohamed, A., Ellicott, A., Housley, T. L. & Ejeta, G. (2003) Hypersensitive response to *Striga* infection in *Sorghum*. *Crop Science*, 43(4), pp. 1320–1324. doi: 10.2135/cropsci2003.1320
- Mohamed, A. H., Ejeta, G., Butler, L. G. & Housley, T. L. (1998) Moisture content and dormancy in *Striga asiatica* seeds. *Weed Research*, 38(4), pp. 257–265. doi: 10.1046/j.1365-3180.1998.00102.x
- Mohamed, A. H., Housley, T. L. & Ejeta, G. (2010) Inheritance of hyper sensitive response to *Striga* parasitism in sorghum [*Sorghum bicolor* (L.) Moench]. *African Journal of Agricultural Research*, 5(19), pp. 2720–2729.
- Mohamed, K. I., Bolin, J. F., Musselman, L. J. & Peterson, A. T. (2007) Genetic diversity of *Striga* and implications for control and modeling future distributions, In: Ejeta, G. and Gressel, J. (Eds.),

REFERENCES

- Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt*. World Scientific Publishing. pp.71–84.
- Mohamed, K. I., Papes, M., Williams, R., Benz, B. W. & Peterson, A. T. (2006) Global invasive potential of 10 parasitic witchweeds and related orobanchaceae. *AMBIO: A Journal of the Human Environment*, 35(6), pp. 281–288. doi: 10.1579/05-R-051R.1
- Mohamed, K. I., Musselman, L. J. & Riches, C. R. (2001) Genus *Striga* (Scrophulariaceae) in Africa. *Annals of Missouri Botanical Garden*, 88(1), pp. 60–103.
- Mohamed, N., Charnikhova, T., Fradin, E. F., Rienstra, J., Babiker, A. G. T. & Bouwmeester, H. J. (2018) Genetic variation in *Sorghum bicolor* strigolactones and their role in resistance against *Striga hermonthica*. *Journal of Experimental Botany*, 69(9), pp. 2415–2430. doi: 10.1093/jxb/ery041
- Mohamed, N., Charnikhova, T., Bakker, E. J., van Ast, A., Babiker, A. G. & Bouwmeester, H. J. (2016) Evaluation of field resistance to *Striga hermonthica* (Del.) Benth. in *Sorghum bicolor* (L.) Moench. The relationship with strigolactones. *Pest management science*, 72(11), pp. 2082–2090. doi: 10.1002/ps.4426
- Mrema, E., Shimelis, H., Laing, M. & Bucheyeki, T. (2017) Farmer's perceptions of sorghum production constraints and *Striga* control practices in semi-arid areas of Tanzania. *International Journal of Pest Management*, 63(2), pp. 146–156. doi: 10.1080/09670874.2016.1238115
- Mundt, C. C. (2014) Durable resistance: A key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution*, 27, pp. 446–455. doi: 10.1016/j.meegid.2014.01.011
- Muthayya, S., Sugimoto, J. D., Montgomery, S. & Maberly, G. F. (2014) An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences*, 1324(1), pp. 7–14. doi: 10.1111/nyas.12540
- Mutinda, S. M., Masanga, J., Mutuku, J. M., Runo, S. & Alakonya, A. (2018) KSTP 94, an open-pollinated maize variety has postattachment resistance to purple witchweed (*Striga hermonthica*). *Weed Science*, 66(4), pp. 525–529. doi: 10.1017/wsc.2018.24
- Mutuku, J. M., Cui, S., Yoshida, S. & Shirasu, K. (2020) Orobanchaceae parasite–host interactions. *New Phytologist*. doi: 10.1111/nph.17083
- Mutuku, J. M., Yoshida, S., Shimizu, T., Ichihashi, Y., Wakatake, T., Takahashi, A., *et al.* (2015) The *WRKY45*-dependent signaling pathway is required for resistance against *Striga hermonthica* parasitism. *Plant Physiology*, 168(3), pp. 1152–1163. doi: 10.1104/pp.114.256404
- N'cho, S. A., Mourits, M., Rodenburg, J. & Oude Lansink, A. (2019) Inefficiency of manual weeding in rainfed rice systems affected by parasitic weeds. *Agricultural Economics*, 50(2), pp. 151–163. doi: 10.1111/agec.12473
- N'cho, S. A., Mourits, M., Rodenburg, J., Demont, M. & Oude Lansink, A. (2014) Determinants of parasitic weed infestation in rainfed lowland rice in Benin. *Agricultural Systems*, 130, pp. 105–115. doi: 10.1016/j.agsy.2014.07.003
- Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews*, 85(4), pp. 935–956. doi: 10.1111/j.1469-185X.2010.00141.x
- Näsholm, T., Kielland, K. & Ganeteg, U. (2009) Uptake of organic nitrogen by plants. *New Phytologist*, 182(1), pp. 31–48. doi: 10.1111/j.1469-8137.2008.02751.x
- Nazar, R. N., Castroverde, C. D. M., Xu, X., Kurosky, A. & Robb, J. (2019) Wounding induces tomato *Ve1* R-gene expression. *Planta*, 249(6), pp. 1779–1797. doi: 10.1007/s00425-019-03121-6
- Nickrent, D. L. & Musselman, L. J. (1979) Autogamy in the American strain of witchweed, *Striga asiatica* (Scrophulariaceae). *Brittonia*, 31(2), pp. 253. doi: 10.2307/2806182

- Nikolov, L. A., Tomlinson, P. B., Manickam, S., Endress, P. K., Kramer, E. M. & Davis, C. C. (2014) Holoparasitic Rafflesiaceae possess the most reduced endophytes and yet give rise to the world's largest flowers. *Annals of Botany*, *114*(2), pp. 233–242. doi: 10.1093/aob/mcu114
- Niks, R. E., Qi, X. & Marcel, T. C. (2015) Quantitative resistance to biotrophic filamentous plant pathogens: Concepts, misconceptions, and mechanisms. *Annual Review of Phytopathology*, *53*(1), pp. 445–470. doi: 10.1146/annurev-phyto-080614-115928
- Noguero, M. & Lacombe, B. (2016) Transporters involved in root nitrate uptake and sensing by *Arabidopsis*. *Frontiers in Plant Science*, *7*(1391), pp. 1–7. doi: 10.3389/fpls.2016.01391
- Nordborg, M. & Weigel, D. (2008) Next-generation genetics in plants. *Nature*, *456*(7223), pp. 720–723. doi: 10.1038/nature07629
- Novocraft. (2019) *Novoalign*. Novocraft Technologies Sdn Bhd, [available at] <http://www.novocraft.com/products/novoalign/>.
- Nwanze, K. F., Mohapatra, S., Kormawa, P., Keya, S. & Bruce-Oliver, S. (2006) Rice development in sub-Saharan Africa. *Journal of the Science of Food and Agriculture*, *86*(5), pp. 675–677. doi: 10.1002/jsfa.2415
- Omoarelojie, L. O., Kulkarni, M. G., Finnie, J. F. & Van Staden, J. (2019) Strigolactones and their crosstalk with other phytohormones. *Annals of Botany*, *124*(5), pp. 749–767. doi: 10.1093/aob/mcz100
- Orjuela, J., Sabot, F., Chéron, S., Vigouroux, Y., Adam, H., Chrestin, H., *et al.* (2014) An extensive analysis of the African rice genetic diversity through a global genotyping. *Theoretical and Applied Genetics*, *127*(10), pp. 2211–2223. doi: 10.1007/s00122-014-2374-z
- Pageau, K., Simier, P., Le Bizec, B., Robins, R. J. & Fer, A. (2003) Characterization of nitrogen relationships between *Sorghum bicolor* and the root-hemiparasitic angiosperm *Striga hermonthica* (Del.) Benth. using $K^{15}NO_3$ as isotopic tracer. *Journal of Experimental Botany*, *54*(383), pp. 789–799. doi: 10.1093/jxb/erg081
- Parker, C. & Riches, C. R. (1993) *Parasitic Weeds of the World: Biology and Control*. Wallingford: CAB International.
- Parker, C. (2013) The parasitic weeds of the Orobanchaceae, In: Joel, D. M., Gressel, J., and Musselman, L. J. (Eds.), *Parasitic Orobanchaceae*. Berlin, Heidelberg: Springer Berlin Heidelberg. pp.313–344.
- Parker, C. (2012) Parasitic weeds: A world challenge. *Weed Science*, *60*(2), pp. 269–276. doi: 10.1614/WS-D-11-00068.1
- Parker, C. (2009) Observations on the current status of *Orobanche* and *Striga* problems worldwide. *Pest Management Science*, *65*(5), pp. 453–459. doi: 10.1002/ps.1713
- Paungfoo-Lonhienne, C., Lonhienne, T. G. A., Rentsch, D., Robinson, N., Christie, M., Webb, R. I., *et al.* (2008) Plants can use protein as a nitrogen source without assistance from other organisms. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), pp. 4524–4529. doi: 10.1073/pnas.0712078105
- Pérez-de-Luque, A., Fondevilla, S., Pérez-Vich, B., Aly, R., Thoiron, S., Simier, P., *et al.* (2009) Understanding *Orobanche* and *Phelipanche*–host plant interactions and developing resistance. *Weed Research*, *49*(SUPPL. 1), pp. 8–22. doi: 10.1111/j.1365-3180.2009.00738.x
- Perrot-Rechenmann, C. (2010) Cellular responses to auxin: Division versus expansion. *Cold Spring Harbor perspectives in biology*, *2*(5), pp. 1–16. doi: 10.1101/cshperspect.a001446
- Pierce, S., Mbwaga, A. M., Press, M. C. & Scholes, J. D. (2003) Xenognosin production and tolerance to *Striga asiatica* infection of high-yielding maize cultivars. *Weed Research*, *43*(2), pp. 139–145. doi: 10.1046/j.1365-3180.2003.00325.x

REFERENCES

- Pilet-Nayel, M. L., Moury, B., Caffier, V., Montarry, J., Kerlan, M. C., Fournet, S., *et al.* (2017) Quantitative resistance to plant pathogens in pyramiding strategies for durable crop protection. *Frontiers in Plant Science*, 8(October), pp. 1–9. doi: 10.3389/fpls.2017.01838
- Poland, J. A., Balint-Kurti, P. J., Wissler, R. J., Pratt, R. C. & Nelson, R. J. (2009) Shades of gray: The world of quantitative disease resistance. *Trends in Plant Science*, 14(1), pp. 21–29. doi: 10.1016/j.tplants.2008.10.006
- Poulin, R. (2011) The many roads to parasitism. A tale of convergence, In: Rollinson, D. and Hay, S. I. (Eds.), *Advances in Parasitology*. Academic Press. pp.1–40.
- Press, A. M. C., Smith, S., Stewart, G. R., Ecology, S. F., Directions, N., Press, M. C., *et al.* (1991) Carbon acquisition and assimilation in parasitic plants. *Functional Ecology*, 5(2), pp. 278–283.
- Price, A. H. (2006) Believe it or not, QTLs are accurate! *Trends in Plant Science*, 11(5), pp. 213–216. doi: 10.1016/j.tplants.2006.03.006
- Pritchard, J. K. & Donnelly, P. (2001) Case-control studies of association in structured or admixed populations. *Theoretical Population Biology*, 60(3), pp. 227–237. doi: 10.1006/tpbi.2001.1543
- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp. 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795
- Qin, J., Wang, C., Wang, L., Zhao, S. & Wu, J. (2019) Defense and counter-defense in rice–virus interactions. *Phytopathology Research*, 1(1), pp. 1–6. doi: 10.1186/s42483-019-0041-7
- R Core Team. (2017) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, [available at] <https://www.r-project.org/>.
- Rasband, W. S. (2016) *ImageJ*. U. S. National Institutes of Health, [available at] <https://imagej.nih.gov/ij/>.
- Reflinur, Kim, B., Jang, S. M., Chu, S. H., Bordiya, Y., Akter, M. B., *et al.* (2014) Analysis of segregation distortion and its relationship to hybrid barriers in rice. *Rice*, 7(1). doi: 10.1186/s12284-014-0003-8
- Ren, H. & Gray, W. M. (2015) SAUR proteins as effectors of hormonal and environmental signals in plant growth. *Molecular Plant*, 8(8), pp. 1153–1164. doi: 10.1016/j.molp.2015.05.003
- Rhoads, A. & Au, K. F. (2015) PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, 13(5), pp. 278–289. doi: 10.1016/j.gpb.2015.08.002
- Rice, W. R. (1989) Analyzing tables of statistical tests. *Evolution*, 43(1), pp. 223–225. doi: 10.1111/j.1558-5646.1989.tb04220.x
- Rich, P. J., Grenier, C. & Ejeta, G. (2004) *Striga* resistance in the wild relatives of sorghum. *Crop Science*, 44(6), pp. 2221–2229. doi: 10.2135/cropsci2004.2221
- Rodenburg, J., Cissoko, M., Kayongo, N., Dieng, I., Bisikwa, J., Irakiza, R., *et al.* (2017) Genetic variation and host-parasite specificity of *Striga* resistance and tolerance in rice: The need for predictive breeding. *New Phytologist*, 214(3), pp. 1267–1280. doi: 10.1111/nph.14451
- Rodenburg, J., Demont, M., Zwart, S. J. & Bastiaans, L. (2016) Parasitic weed incidence and related economic losses in rice in Africa. *Agriculture, Ecosystems and Environment*, 235, pp. 306–317. doi: 10.1016/j.agee.2016.10.020
- Rodenburg, J., Cissoko, M., Kayeke, J., Dieng, I., Khan, Z. R., Midega, C. A. O., *et al.* (2015) Do NERICA rice cultivars express resistance to *Striga hermonthica* (Del.) Benth. and *Striga asiatica* (L.) Kuntze under field conditions? *Field Crops Research*, 170, pp. 83–94. doi: 10.1016/j.fcr.2014.10.010

- Rodenburg, J. & Bastiaans, L. (2011) Host-plant defence against *Striga* spp.: Reconsidering the role of tolerance. *Weed Research*, 51(5), pp. 438–441. doi: 10.1111/j.1365-3180.2011.00871.x
- Rodenburg, J., Riches, C. R. & Kayeke, J. M. (2010) Addressing current and future problems of parasitic weeds in rice. *Crop Protection*, 29(3), pp. 210–221. doi: 10.1016/j.cropro.2009.10.015
- Rodenburg, J., Bastiaans, L., Schapendonk, A. H. C. M., Van Der Putten, P. E. L., Van Ast, A., Dingemans, N. J., *et al.* (2008) CO₂-assimilation and chlorophyll fluorescence as indirect selection criteria for host tolerance against *Striga*. *Euphytica*, 160(1), pp. 75–87. doi: 10.1007/s10681-007-9555-7
- Rodenburg, J., Bastiaans, L., Weltzien, E. & Hess, D. E. (2005) How can field selection for *Striga* resistance and tolerance in sorghum be improved? *Field Crops Research*, 93(1), pp. 34–50. doi: 10.1016/j.fcr.2004.09.004
- Rowntree, J. K., Fisher Barham, D., Stewart, A. J. A. & Hartley, S. E. (2014) The effect of multiple host species on a keystone parasitic plant and its aphid herbivores. *Functional Ecology*, 28(4), pp. 829–836. doi: 10.1111/1365-2435.12281
- Safa, S. B., Jones, B. M. G. & Musselman, L. J. (1984) Mechanisms favouring outbreeding in *Striga hermonthica* [Scrophulariaceae]. *New Phytologist*, 96(2), pp. 299–305. doi: 10.1111/j.1469-8137.1984.tb03567.x
- Samejima, H., Babiker, A. G., Mustafa, A. & Sugimoto, Y. (2016) Identification of *Striga hermonthica*-resistant upland rice varieties in Sudan and their resistance phenotypes. *Frontiers in Plant Science*, 7(634), pp. 1–12. doi: 10.3389/fpls.2016.00634
- Satish, K., Gutema, Z., Grenier, C., Rich, P. J. & Ejeta, G. (2012) Molecular tagging and validation of microsatellite markers linked to the low germination stimulant gene (*lgs*) for *Striga* resistance in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoretical and Applied Genetics*, 124(6), pp. 989–1003. doi: 10.1007/s00122-011-1763-9
- Sattler, F. T., Sanogo, M. D., Kassari, I. A., Angarawai, I. I., Gwadi, K. W., Dodo, H., *et al.* (2018) Characterization of West and Central African accessions from a pearl millet reference collection for agro-morphological traits and *Striga* resistance. *Plant Genetic Resources: Characterisation and Utilisation*, 16(3), pp. 260–272. doi: 10.1017/S1479262117000272
- Schjoerring, J. K., Husted, S., Mäck, G. & Mattsson, M. (2002) The regulation of ammonium translocation in plants. *Journal of Experimental Botany*, 53(370), pp. 883–890. doi: 10.1093/jexbot/53.370.883
- Scholes, J. D. & Press, M. C. (2008) *Striga* infestation of cereal crops — an unsolved problem in resource limited agriculture. *Current Opinion in Plant Biology*, 11(2), pp. 180–186. doi: 10.1016/j.pbi.2008.02.004
- Schön, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S. & Melchinger, A. E. (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics*, 167(1), pp. 485–498. doi: 10.1534/genetics.167.1.485
- Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., *et al.* (2020) Multi-parent populations in crops: A toolbox integrating genomics and genetic mapping with breeding. *Heredity*, 125(6), pp. 396–416. doi: 10.1038/s41437-020-0336-6
- Seck, P. A., Tollens, E., Wopereis, M. C. S., Diagne, A. & Bamba, I. (2010) Rising trends and variability of rice prices: Threats and opportunities for sub-Saharan Africa. *Food Policy*, 35(5), pp. 403–411. doi: 10.1016/j.foodpol.2010.05.003
- Seck, P. A., Diagne, A., Mohanty, S. & Wopereis, M. C. S. (2012) Crops that feed the world 7: Rice. *Food Security*, 4(1), pp. 7–24. doi: 10.1007/s12571-012-0168-1

REFERENCES

- Seel, W. E. & Press, M. C. (1993) Influence of the host on three sub-Arctic annual facultative root hemiparasites. I. Growth, mineral accumulation and above-ground dry-matter partitioning. *New Phytologist*, *125*, pp. 131–138. doi: 10.1111/j.1469-8137.1993.tb03871.x
- Sen, S. & Churchill, G. A. (2001) A statistical framework for quantitative trait mapping. *Genetics*, *159*(1), pp. 371–387.
- Shahid, S., Kim, G., Johnson, N. R., Wafula, E., Wang, F., Coruh, C., *et al.* (2018) MicroRNAs from the parasitic plant *Cuscuta campestris* target host messenger RNAs. *Nature*, *553*(7686), pp. 82–85. doi: 10.1038/nature25027
- Showemimo, F. A. (2003) Selection criteria for combining high yield and *Striga* resistance in sorghum. *Tropicultura*, *21*(3), pp. 157–159.
- Simier, P., Constant, S., Degrande, D., Moreau, C., Robins, R. J., Fer, A., *et al.* (2006) Impact of nitrate supply in C and N assimilation in the parasitic plant *Striga hermonthica* (Del.) Benth (Scrophulariaceae) and its host *Sorghum bicolor* L. *Plant, Cell and Environment*, *29*(4), pp. 673–681. doi: 10.1111/j.1365-3040.2005.01449.x
- Softberry Inc. (2016) *FGenesH*, [available at] <http://www.softberry.com/berry.phtml?topic=fgenes&group=programs&subgroup=gfind>.
- Soller, M., Brody, T. & Genizi, A. (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics*, *47*(1), pp. 35–39. doi: 10.1007/BF00277402
- Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome biology*, *7*(Suppl 1), pp. S10.1–12. doi: 10.1186/gb-2006-7-s1-s10
- Spallek, T., Mutuku, M. & Shirasu, K. (2013) The genus *Striga*: A witch profile. *Molecular Plant Pathology*, *14*(9), pp. 861–869. doi: 10.1111/mpp.12058
- Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), pp. 637–644. doi: 10.1093/bioinformatics/btn013
- Stevens, P. F. (2017) *Angiosperm Phylogeny Website* [online]. Available from: <http://www.mobot.org/mobot/research/apweb/orders/lamialesweb.htm#Orobanchaceae> [Accessed 5 January 2021].
- Stoffel, M. A., Nakagawa, S. & Schielzeth, H. (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *8*(11), pp. 1639–1644. doi: 10.1111/2041-210X.12797
- Storey, J. D., Bass, A. J., Dabney, A. & Robinson, D. (2019) *qvalue: Q-value estimation for false discovery rate control*. R package version 2.16.0, [available at] <http://github.com/jdstorey/qvalue>.
- Storey, J. D. & Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), pp. 9440–9445. doi: 10.1073/pnas.1530509100
- Sun, Z., Hans, J., Walter, M. H., Matusova, R., Beekwilder, J., Verstappen, F. W. A., *et al.* (2008) Cloning and characterisation of a maize carotenoid cleavage dioxygenase (*ZmCCD1*) and its involvement in the biosynthesis of apocarotenoids with various roles in mutualistic and parasitic interactions. *Planta*, *228*(5), pp. 789–801. doi: 10.1007/s00425-008-0781-6
- Swarbrick, P. J., Huang, K., Liu, G., Slate, J., Press, M. C. & Scholes, J. D. (2008) Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant *Striga hermonthica*. *New Phytologist*, *179*(2), pp. 515–529. doi: 10.1111/j.1469-8137.2008.02484.x

- Swarbrick, P. J., Scholes, J. D., Press, M. C. & 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(5), pp. 528–532. doi: 10.1002/ps.1719
- Tabuchi, M., Abiko, T. & Yamaya, T. (2007) Assimilation of ammonium ions and reutilization of nitrogen in rice (*Oryza sativa* L.). *Journal of Experimental Botany*, 58(9), pp. 2319–2327. doi: 10.1093/jxb/erm016
- Tang, D., Wang, G. & Zhou, J. M. (2017) Receptor kinases in plant-pathogen interactions: More than pattern recognition. *Plant Cell*, 29(4), pp. 618–637. doi: 10.1105/tpc.16.00891
- Tateda, C., Obara, K., Abe, Y., Sekine, R., Nekoduka, S., Hikage, T., *et al.* (2019) The host stomatal density determines resistance to *Septoria gentianae* in Japanese gentian. *Molecular Plant-Microbe Interactions*, 32(4), pp. 428–436. doi: 10.1094/MPMI-05-18-0114-R
- Taylor, A., Martin, J. & Seel, W. E. (1996) Physiology of the parasitic association between maize and witchweed (*Striga hermonthica*): Is ABA involved? *Journal of Experimental Botany*, 47(301), pp. 1057–1065. doi: 10.1093/jxb/47.8.1057
- Tegeder, M. & Rentsch, D. (2010) Uptake and partitioning of amino acids and peptides. *Molecular Plant*, 3(6), pp. 997–1011. doi: 10.1093/mp/ssf047
- Teka, H. B. (2014) Advance research on *Striga* control: A review. *African Journal of Plant Science*, 8(11), pp. 492–506. doi: 10.5897/AJPS2014.1186
- The Plant List. (2013) *Version 1.1* [online]. Available from: <http://www.theplantlist.org/> [Accessed 30 November 2020].
- Timko, M. P. & Scholes, J. D. (2013) Host reaction to attack by root parasitic plants, In: Joel, D. M., Gressel, J., and Musselman, L. J. (Eds.), *Parasitic Orobanchaceae*. Berlin, Heidelberg: Springer. pp.115–141.
- Timko, M. P., Gowda, B. S., Ouedraogo, J. & Ousmane, B. (2007) Molecular markers for analysis of resistance to *Striga gesnerioides* in cowpea, In: *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt*. World Scientific Publishing. pp.115–128.
- Tippe, D. E., Rodenburg, J., Schut, M., van Ast, A., Kayeke, J. & Bastiaans, L. (2017) Farmers' knowledge, use and preferences of parasitic weed management strategies in rain-fed rice production systems. *Crop Protection*, 99, pp. 93–107. doi: <https://doi.org/10.1016/j.cropro.2017.05.007>
- Toh, S., Holbrook-Smith, D., Stogios, P. J., Onopriyenko, O., Lumba, S., Tsuchiya, Y., *et al.* (2015) Structure-function analysis identifies highly sensitive strigolactone receptors in *Striga*. *Science*, 350(6257), pp. 203–207. doi: 10.1126/science.aac9476
- Treangen, T. J., Sommer, D. D., Angly, F. E., Koren, S. & Pop, M. (2011) Next generation sequence assembly with AMOS. *Current Protocols in Bioinformatics*, 33, pp. 11.8.1–11.8.18. doi: 10.1002/0471250953.bi1108s33
- Tsymbalyuk, Z. M. & Mosyakin, S. L. (2018) Palynomorphological peculiarities of representatives of tribes *Lindenbergiae* and *Cymbaricae* and pollen evolution in early-branching lineages of *Orobanchaceae*. *Ukrainian Botanical Journal*, 75(2), pp. 123–136. doi: 10.15407/ukrbotj75.02.123
- Tuinstra, M. R., Ejeta, G. & Goldsbrough, P. B. (1997) Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. *Theoretical and Applied Genetics*, 95(5–6), pp. 1005–1011. doi: 10.1007/s001220050654
- Unachukwu, N. N., Menkir, A., Stanley, A., Farombi, E. O. & Gedil, M. (2020) Contrasting response mechanisms of maize lines to *Striga hermonthica*. *Agriculture*, 10(485), pp. 1–18. doi: 10.3390/agriculture10100485

REFERENCES

- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., *et al.* (2013) From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, (43), pp. 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43
- Van Oort, P. A. J., Saito, K., Tanaka, A., Amovin-Assagba, E., Van Bussel, L. G. J., Van Wart, J., *et al.* (2015) Assessment of rice self-sufficiency in 2025 in eight African countries. *Global Food Security*, 5, pp. 39–49. doi: 10.1016/j.gfs.2015.01.002
- Velasquez, S. M., Barbez, E., Kleine-Vehn, J. & Estevez, J. M. (2016) Auxin and cellular elongation. *Plant Physiology*, 170(3), pp. 1206–1215. doi: 10.1104/pp.15.01863
- Vinod, K. K. (2011) Kosambi and the genetic mapping function. *Resonance*, 16(6), pp. 540–550. doi: 10.1007/s12045-011-0060-x
- Vogler, R. K., Ejeta, G. & Butler, L. G. (1996) Inheritance of low production of *Striga* germination stimulant in sorghum. *Crop Science*, 36(5), pp. 1185–1191. doi: 10.2135/cropsci1996.0011183X003600050020x
- Wakatake, T., Ogawa, S., Yoshida, S. & Shirasu, K. (2020) An auxin transport network underlies xylem bridge formation between the hemi-parasitic plant *Phtheirospermum japonicum* and host *Arabidopsis*. *Development*, 147(14). doi: 10.1242/dev.187781
- Wang, M., Qiao, J. Y., Yu, C. L., Chen, H., Sun, C. D., Huang, L. Z., *et al.* (2019) The auxin influx carrier, OsAUX3, regulates rice root development and responses to aluminium stress. *Plant Cell and Environment*, 42(4), pp. 1125–1138. doi: 10.1111/pce.13478
- Wang, M., Mao, Y., Lu, Y., Tao, X. & Zhu, J. kang. (2017) Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Molecular Plant*, 10(7), pp. 1011–1013. doi: 10.1016/j.molp.2017.03.001
- Wang, M., Yu, Y., Haberer, G., Marri, P. R., Fan, C., Goicoechea, J. L., *et al.* (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nature Genetics*, 46(9), pp. 982–988. doi: 10.1038/ng.3044
- Wang, Q., Liu, Yuqiang, He, J., Zheng, X., Hu, J., Liu, Yanling, *et al.* (2014) *STV11* encodes a sulphotransferase and confers durable resistance to rice stripe virus. *Nature Communications*, 5(1), pp. 4768. doi: 10.1038/ncomms5768
- Waruru, M. (2013) *Deadly Striga weed spreading across Eastern Africa* [online]. Available from: <https://www.scidev.net/sub-saharan-africa/news/deadly-striga-weed-spreading-across-eastern-africa/> [Accessed 12 January 2021].
- Waweru, D. N., Kuria, E. K., Bradley, J. M., Scholes, J. D. & Runo, S. (2019) Tissue culture protocols for the obligate parasitic plant *Striga hermonthica* and implications for host-parasite co-cultivation. *Plant Cell, Tissue and Organ Culture*, 138(2), pp. 247–256. doi: 10.1007/s11240-019-01621-7
- Wayne, M. L. & McIntyre, L. M. (2002) Combining mapping and arraying: An approach to candidate gene identification. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), pp. 14903–14906. doi: 10.1073/pnas.222549199
- Welsh, A. B. & Mohamed, K. I. (2011) Genetic diversity of *Striga hermonthica* populations in Ethiopia: evaluating the role of geography and host specificity in shaping population structure. *International Journal of Plant Sciences*, 172(6), pp. 773–782. doi: 10.1086/660104
- Westwood, J. H., Yoder, J. I., Timko, M. P. & dePamphilis, C. W. (2010) The evolution of parasitism in plants. *Trends in Plant Science*, 15(4), pp. 227–235. doi: 10.1016/j.tplants.2010.01.004
- Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., *et al.* (2009) SUPERFAMILY — Sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Research*, 37(Database issue), pp. 380–386. doi: 10.1093/nar/gkn762

- Wissuwa, M., Kondo, K., Fukuda, T., Mori, A., Rose, M. T., Pariasca-Tanaka, J., *et al.* (2015) Unmasking novel loci for internal phosphorus utilization efficiency in rice germplasm through genome-wide association analysis. *PLoS ONE*, *10*(4). doi: 10.1371/journal.pone.0124215
- Wu, B., Han, Z. & Xing, Y. (2013) Genome mapping, markers and QTLs, In: Zhang, Q. and Wing, R. A. (Eds.), *Genetics and Genomics of Rice*. New York, NY, USA: Springer. pp.35–54.
- Xu, S. (2008) Quantitative trait locus mapping can benefit from segregation distortion. *Genetics*, *180*(4), pp. 2201–2208. doi: 10.1534/genetics.108.090688
- Yang, D.-L., Yao, J., Mei, C.-S., Tong, X.-H., Zeng, L.-J., Li, Q., *et al.* (2012) Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. *Proceedings of the National Academy of Sciences*, *109*(19), pp. E1192–E1200. doi: 10.1073/pnas.1201616109
- Yang, D.-L., Yang, Y. & He, Z. (2013) Roles of plant hormones and their interplay in rice immunity. *Molecular Plant*, *6*(3), pp. 675–685. doi: 10.1093/mp/sst056
- Yang, Z., Zhang, Y., Wafula, E. K., Honaas, L. A., Ralph, P. E., Jones, S., *et al.* (2016) Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(45), pp. E7010–E7019. doi: 10.1073/pnas.1608765113
- Yang, Z., Wafula, E. K., Honaas, L. A., Zhang, H., Das, M., Fernandez-Aparicio, M., *et al.* (2015) Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. *Molecular Biology and Evolution*, *32*(3), pp. 767–790. doi: 10.1093/molbev/msu343
- Yano, K., Morinaka, Y., Wang, F., Huang, P., Takehara, S., Hirai, T., *et al.* (2019) GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(42), pp. 21 262–21 267. doi: 10.1073/pnas.1904964116
- Yao, X., Nie, J., Bai, R. & Sui, X. (2020) Amino acid transporters in plants: Identification and function. *Plants*, *9*(8), pp. 972. doi: 10.3390/plants9080972
- Yi, J., Gao, J., Zhang, W., Zhao, C., Wang, Y. & Zhen, X. (2019) Differential uptake and utilization of two forms of nitrogen in *japonica* rice cultivars from north-eastern China. *Frontiers in Plant Science*, *10*(1061), pp. 1–11. doi: 10.3389/fpls.2019.01061
- Yoder, J. I. & Scholes, J. D. (2010) Host plant resistance to parasitic weeds; recent progress and bottlenecks. *Current Opinion in Plant Biology*, *13*(4), pp. 478–484. doi: 10.1016/j.pbi.2010.04.011
- Yogendra, K. N., Kushalappa, A. C., Sarmiento, F., Rodriguez, E. & Mosquera, T. (2015) Metabolomics deciphers quantitative resistance mechanisms in diploid potato clones against late blight. *Functional Plant Biology*, *42*(3), pp. 284–298. doi: 10.1071/FP14177
- Yoneyama, Koichi, Awad, A. A., Xie, X., Yoneyama, Kaori & Takeuchi, Y. (2010) Strigolactones as germination stimulants for root parasitic plants. *Plant and Cell Physiology*, *51*(7), pp. 1095–1103. doi: 10.1093/pcp/pcq055
- Yoneyama, Koichi, Xie, X., Yoneyama, Kaori & Takeuchi, Y. (2009) Strigolactones: Structures and biological activities. *Pest Management Science*, *65*(5), pp. 467–470. doi: 10.1002/ps.1726
- Yoshida, S., Cui, S., Ichihashi, Y. & Shirasu, K. (2016) The haustorium, a specialized invasive organ in parasitic plants. *Annual Review of Plant Biology*, *67*, pp. 643–667. doi: 10.1146/annurev-arplant-043015-111702
- Yoshida, S. & Shirasu, K. (2009) Multiple layers of incompatibility to the parasitic witchweed, *Striga hermonthica*. *New Phytologist*, *183*(1), pp. 180–189. doi: 10.1111/j.1469-8137.2009.02840.x

REFERENCES

- Youens-Clark, K., Buckler, E., Casstevens, T., Chen, C., DeClerck, G., Derwent, P., *et al.* (2011) Gramene database in 2010: Updates and extensions. *Nucleic Acids Research*, 39(Database issue), pp. 1085–1094. doi: 10.1093/nar/gkq1148
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2), pp. 203–208. doi: 10.1038/ng1702
- Zeng, Z. B. (1994) Precision mapping of quantitative trait loci. *Genetics*, 136(4), pp. 1457–68.
- Zeng, Z. B. (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America*, 90(23), pp. 10 972–10 976. doi: 10.1073/pnas.90.23.10972
- Zhang, L., Wang, S., Li, H., Deng, Q., Zheng, A., Li, S., *et al.* (2010) Effects of missing marker and segregation distortion on QTL mapping in F₂ populations. *Theoretical and Applied Genetics*, 121(6), pp. 1071–1082. doi: 10.1007/s00122-010-1372-z
- Zhang, Y., Zhao, J., Li, Y., Yuan, Z., He, H., Yang, H., *et al.* (2016) Transcriptome analysis highlights defense and signaling pathways mediated by rice *pi21* gene with partial resistance to *Magnaporthe oryzae*. *Frontiers in Plant Science*, 7(1834), pp. 1–17. doi: 10.3389/fpls.2016.01834
- Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7(1–2), pp. 203–214. doi: 10.1089/10665270050081478
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., *et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), pp. 355–360. doi: 10.1038/ng.546
- Zhou, X. & Huang, X. (2019) Genome-wide association studies in rice: How to solve the low power problems? *Molecular Plant*, 12(1), pp. 10–12. doi: 10.1016/j.molp.2018.11.010
- Zhou, Y., Chebotarov, D., Kudrna, D., Llaca, V., Lee, S., Rajasekar, S., *et al.* (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. *Scientific Data*, 7(1), pp. 1–11. doi: 10.1038/s41597-020-0438-2
- Zwanenburg, B., Pospíšil, T. & Čavar Zeljković, S. (2016) Strigolactones: New plant hormones in action. *Planta*, 243(6), pp. 1311–1326. doi: 10.1007/s00425-015-2455-5

Appendix

Supplementary Table S3.1 Number of rice replicates of RILs and their parental genotypes in Screen 1 of the two-step phenotypic scoring strategy.

Genotype names (CIAT code)	Abbreviated code	Batches					
		Planned			In experiment		
		1-1	1-2	1-3	1-1	1-2	1-3
IR64	IR	4	4	4	4	4	4
CT8556-37-2-3-1-M	CT	4	4	4	4	4	4
CT22591-3-1/F7	1	1	2	1	1	2	1
CT22591-3-5/F7	5	2	1	1	1	1	1
CT22591-3-6/F7	6	1	1	2	1	2	1
CT22591-3-7/F7	7	1	1	2	1	1	2
CT22591-3-10/F7	10	1	2	1	1	2	1
CT22591-3-12/F7	12	1	1	2	1	1	2
CT22591-3-21/F7	21	1	1	2	1	2	1
CT22591-3-26/F7	26	2	1	1	2	1	1
CT22591-3-36/F7	36	2	1	1	2	—	2
CT22591-3-38/F7	38	1	2	1	1	2	1
CT22591-3-39/F7	39	1	2	1	1	1	1
CT22591-3-52/F7	52	1	2	1	1	2	1
CT22591-3-61/F7	61	1	2	1	1	2	1
CT22591-3-66/F7	66	1	2	1	1	2	—
CT22591-3-69/F7	69	1	1	2	1	2	1
CT22591-3-74/F7	74	1	2	1	1	2	1
CT22591-3-77/F7	77	1	2	1	1	1	1
CT22591-3-83/F7	83	1	2	1	1	1	1
CT22591-3-92/F7	92	2	1	1	2	1	1
CT22591-3-98/F7	98	1	1	2	1	—	3
CT22591-3-103/F7	103	1	1	2	2	1	1
CT22591-3-106/F7	106	1	1	2	1	1	2
CT22591-3-107/F7	107	1	2	1	1	2	1
CT22591-3-108/F7	108	2	1	1	2	1	1
CT22591-3-117/F7	117	2	1	1	2	1	1
CT22591-3-125/F7	125	2	1	1	2	1	1
CT22591-3-126/F7	126	2	1	1	2	1	1
CT22591-3-135/F7	135	1	1	2	1	1	2
CT22591-3-136/F7	136	2	1	1	2	1	1
CT22591-3-138/F7	138	2	1	1	2	1	1
CT22591-3-142/F7	142	2	1	1	2	1	1
CT22591-3-143/F7	143	1	2	1	—	2	2
CT22591-3-144/F7	144	1	2	1	1	2	1
CT22591-3-150/F7	150	2	1	1	2	1	1
CT22591-3-156/F7	156	1	1	2	1	2	1
CT22591-3-166/F7	166	2	1	1	2	1	1
CT22591-3-167/F7	167	1	2	1	1	2	1
CT22591-3-173/F7	173	1	1	2	1	1	2
CT22591-3-176/F7	176	1	1	2	1	1	2
CT22591-3-182/F7	182	1	1	2	1	1	2

Supplementary Table S3.1 (continued)

Genotype names (CIAT code)	Abbreviated code	Batches					
		Planned			In experiment		
		1-1	1-2	1-3	1-1	1-2	1-3
CT22591-3-188/F7	188	1	1	2	1	1	2
CT22591-3-190/F7	190	2	1	1	2	1	1
CT22591-3-193/F7	193	2	1	1	2	1	1
CT22591-3-199/F7	199	1	1	2	1	1	2
CT22591-3-204/F7	204	2	1	1	2	1	1
CT22591-3-211/F7	211	1	2	1	1	2	1
CT22591-3-220/F7	220	1	1	2	2	1	1
CT22591-3-223/F7	223	1	1	2	1	1	2
CT22591-3-231/F7	231	1	2	1	1	2	1
CT22591-3-241/F7	241	1	2	1	1	2	1
CT22591-3-242/F7	242	2	1	1	2	1	1
CT22591-3-247/F7	247	1	1	2	2	1	1
CT22591-3-250/F7	250	1	1	2	2	1	1
CT22591-3-252/F7	252	1	2	1	1	2	1
CT22591-3-253/F7	253	1	2	1	1	2	1
CT22591-3-278/F7	278	2	1	1	2	1	1
CT22591-3-281/F7	281	2	1	1	2	1	1
Total number of rice plants		84	84	84	86	83	78

Supplementary Table S3.2 Number of rice replicates of RILs and their parental genotypes in Screen 2 of the two-step phenotypic scoring strategy.

Genotype names (CIAT code)	Abbreviated code	Screen	Batches								
			2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9
IR64	IR	1 & 2	4	4	4	4	4	4	4	4	4
CT8556-37-2-3-1-M	CT	1 & 2	4	4	4	4	4	4	4	4	4
CT22591-3-1/F7	1	1 & 2	—	—	1	—	1	1	—	—	1
CT22591-3-6/F7	6	1 & 2	1	—	—	—	—	1	1	1	—
CT22591-3-26/F7	26	1 & 2	1	—	—	1	—	—	—	1	1
CT22591-3-36/F7	36	1 & 2	1	—	1	—	1	—	—	—	1
CT22591-3-39/F7	39	1 & 2	1	—	—	1	—	1	—	1	—
CT22591-3-69/F7	69	1 & 2	—	1	1	—	—	—	1	—	1
CT22591-3-74/F7	74	1 & 2	1	1	—	1	—	—	—	—	1
CT22591-3-92/F7	92	1 & 2	—	—	—	1	—	1	—	1	1
CT22591-3-107/F7	107	1 & 2	—	—	—	1	—	1	1	—	1
CT22591-3-173/F7	173	1 & 2	1	—	1	—	1	—	1	—	—
CT22591-3-241/F7	241	1 & 2	—	1	—	—	—	1	—	1	1
CT22591-3-250/F7	250	1 & 2	—	1	1	—	—	—	1	1	—
CT22591-3-2/F7	2	Only 2	1	1	—	—	—	1	—	—	1
CT22591-3-9/F7	9	Only 2	—	1	1	1	—	—	1	—	—
CT22591-3-11/F7	11	Only 2	—	—	—	—	1	1	—	2	1
CT22591-3-13/F7	13	Only 2	—	1	—	1	—	1	—	1	1
CT22591-3-17/F7	17	Only 2	1	1	1	1	—	—	—	—	—
CT22591-3-20/F7	20	Only 2	1	—	1	—	—	—	1	1	—
CT22591-3-21/F7	21	Only 2	1	—	—	—	1	1	—	—	1
CT22591-3-22/F7	22	Only 2	—	1	1	1	—	1	—	—	—
CT22591-3-24/F7	24	Only 2	1	—	—	—	—	1	—	1	1
CT22591-3-25/F7	25	Only 2	—	—	—	1	1	—	1	—	1
CT22591-3-27/F7	27	Only 2	1	—	—	—	1	1	—	1	—
CT22591-3-28/F7	28	Only 2	1	—	—	1	1	—	1	—	—
CT22591-3-29/F7	29	Only 2	—	—	—	—	1	1	1	1	—
CT22591-3-31/F7	31	Only 2	—	1	—	—	1	—	1	1	—
CT22591-3-32/F7	32	Only 2	—	1	1	—	—	—	—	1	—
CT22591-3-33/F7	33	Only 2	1	—	1	—	—	1	1	—	—
CT22591-3-34/F7	34	Only 2	—	—	—	1	1	1	—	—	1
CT22591-3-35/F7	35	Only 2	1	—	—	—	1	—	—	1	1
CT22591-3-37/F7	37	Only 2	—	—	1	—	—	1	1	—	1
CT22591-3-40/F7	40	Only 2	1	—	—	1	—	1	—	—	1
CT22591-3-41/F7	41	Only 2	1	—	—	—	—	—	1	—	3
CT22591-3-42/F7	42	Only 2	—	1	—	—	—	1	—	2	—
CT22591-3-43/F7	43	Only 2	—	1	—	—	1	1	—	1	—
CT22591-3-44/F7	44	Only 2	1	—	—	—	1	—	1	—	1
CT22591-3-45/F7	45	Only 2	1	—	1	—	1	1	—	—	—
CT22591-3-46/F7	46	Only 2	—	—	1	1	1	—	—	—	1
CT22591-3-47/F7	47	Only 2	—	—	—	—	1	1	1	1	—
CT22591-3-48/F7	48	Only 2	1	1	1	—	—	—	—	—	—
CT22591-3-50/F7	50	Only 2	1	1	—	1	—	1	—	—	—
CT22591-3-52/F7	52	Only 2	1	—	—	1	1	1	—	—	—
CT22591-3-53/F7	53	Only 2	—	—	—	1	1	—	—	1	1
CT22591-3-54/F7	54	Only 2	—	1	—	1	—	1	—	—	1
CT22591-3-60/F7	60	Only 2	—	1	—	—	1	1	1	—	—
CT22591-3-66/F7	66	Only 2	1	—	1	—	—	—	1	1	1
CT22591-3-68/F7	68	Only 2	—	—	—	—	—	1	1	1	1
CT22591-3-70/F7	70	Only 2	—	1	—	1	—	—	—	1	1
CT22591-3-71/F7	71	Only 2	—	1	—	1	—	1	1	—	—
CT22591-3-75/F7	75	Only 2	—	—	—	—	1	1	—	1	1
CT22591-3-77/F7	77	Only 2	—	—	1	1	—	1	—	—	1
CT22591-3-78/F7	78	Only 2	—	—	—	1	1	—	1	1	—
CT22591-3-81/F7	81	Only 2	—	—	1	1	—	1	1	—	—
CT22591-3-82/F7	82	Only 2	—	—	1	—	—	—	1	2	—

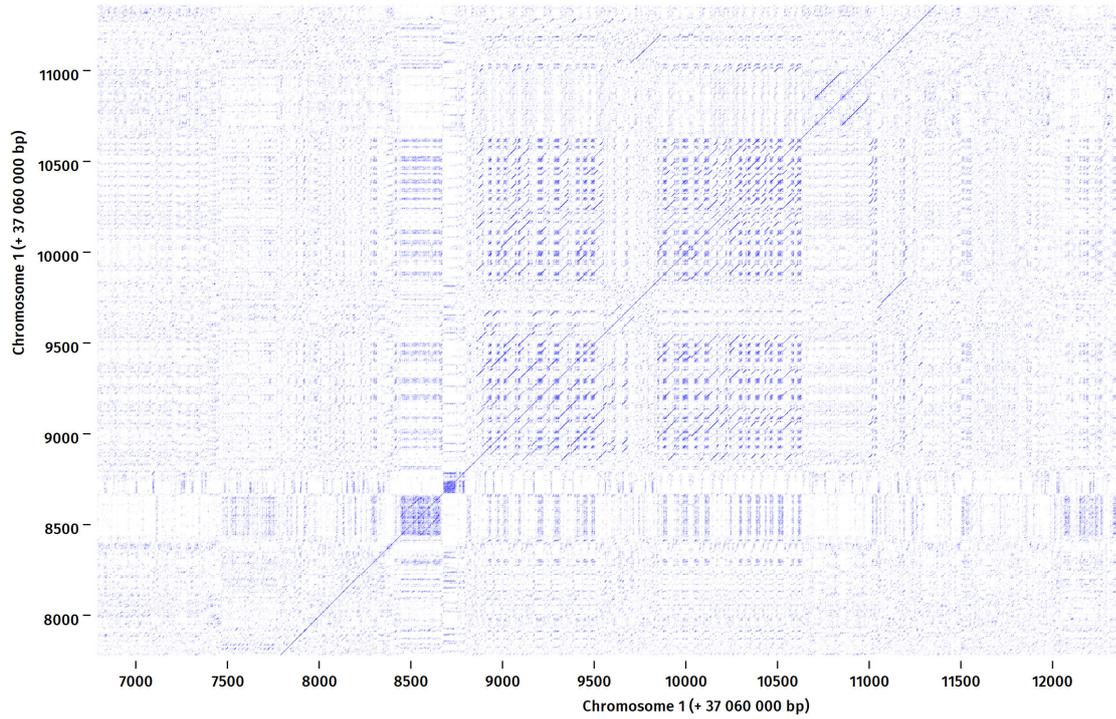
Supplementary Table S3.2 (continued)

Genotype names (CIAT code)	Abbreviated code	Screen	Batches								
			2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9
CT22591-3-84/F7	84	Only 2	—	—	1	1	—	1	—	—	1
CT22591-3-90/F7	90	Only 2	—	—	1	1	—	—	1	—	1
CT22591-3-100/F7	100	Only 2	—	1	—	1	—	—	—	1	1
CT22591-3-103/F7	103	Only 2	—	1	1	1	1	—	—	—	—
CT22591-3-104/F7	104	Only 2	—	1	1	1	—	—	—	—	1
CT22591-3-109/F7	109	Only 2	1	1	—	—	—	1	—	1	—
CT22591-3-111/F7	111	Only 2	—	—	1	—	1	—	1	—	1
CT22591-3-113/F7	113	Only 2	—	—	1	—	—	1	1	1	1
CT22591-3-114/F7	114	Only 2	1	—	—	1	—	—	1	1	—
CT22591-3-115/F7	115	Only 2	1	1	—	—	—	1	—	—	1
CT22591-3-116/F7	116	Only 2	—	1	—	—	1	—	1	—	1
CT22591-3-118/F7	118	Only 2	—	1	—	—	—	—	1	1	1
CT22591-3-120/F7	120	Only 2	1	—	—	1	1	—	—	1	—
CT22591-3-121/F7	121	Only 2	1	—	1	—	—	—	1	—	1
CT22591-3-123/F7	123	Only 2	—	—	1	1	1	—	—	1	—
CT22591-3-127/F7	127	Only 2	—	—	—	1	1	—	1	1	2
CT22591-3-129/F7	129	Only 2	—	1	—	1	1	—	—	—	1
CT22591-3-132/F7	132	Only 2	1	1	—	—	—	—	1	1	—
CT22591-3-133/F7	133	Only 2	1	1	—	1	—	—	1	—	—
CT22591-3-134/F7	134	Only 2	1	—	1	1	—	1	—	—	—
CT22591-3-139/F7	139	Only 2	—	—	1	1	—	1	—	1	—
CT22591-3-141/F7	141	Only 2	1	1	—	1	1	—	—	—	—
CT22591-3-148/F7	148	Only 2	1	1	—	—	—	—	1	—	1
CT22591-3-149/F7	149	Only 2	1	—	—	—	1	—	1	1	—
CT22591-3-153/F7	153	Only 2	—	1	1	—	—	1	—	—	1
CT22591-3-155/F7	155	Only 2	1	1	1	1	—	—	—	—	—
CT22591-3-157/F7	157	Only 2	—	1	—	—	1	—	—	1	1
CT22591-3-160/F7	160	Only 2	—	—	—	1	1	1	1	—	—
CT22591-3-162/F7	162	Only 2	1	—	—	1	1	—	—	—	1
CT22591-3-163/F7	163	Only 2	1	—	1	—	1	—	—	1	—
CT22591-3-164/F7	164	Only 2	—	1	1	—	1	1	—	—	—
CT22591-3-165/F7	165	Only 2	—	—	1	1	—	—	—	1	1
CT22591-3-171/F7	171	Only 2	1	—	—	1	1	—	1	—	—
CT22591-3-172/F7	172	Only 2	—	1	1	—	1	—	1	—	—
CT22591-3-174/F7	174	Only 2	1	1	—	—	1	1	—	—	—
CT22591-3-175/F7	175	Only 2	—	1	—	—	1	1	—	1	—
CT22591-3-177/F7	177	Only 2	1	—	—	—	1	1	1	—	—
CT22591-3-178/F7	178	Only 2	—	1	1	—	1	—	—	—	1
CT22591-3-181/F7	181	Only 2	1	—	—	—	—	—	1	1	1
CT22591-3-183/F7	183	Only 2	1	1	—	—	—	—	—	1	1
CT22591-3-184/F7	184	Only 2	—	1	—	—	—	1	1	—	1
CT22591-3-188/F7	188	Only 2	—	—	—	—	1	—	1	1	1
CT22591-3-189/F7	189	Only 2	—	1	—	1	—	1	—	1	—
CT22591-3-191/F7	191	Only 2	—	—	1	1	1	1	—	—	—
CT22591-3-194/F7	194	Only 2	1	—	1	1	—	—	—	1	—
CT22591-3-195/F7	195	Only 2	—	1	—	—	1	1	1	—	—
CT22591-3-196/F7	196	Only 2	1	—	1	—	1	—	—	1	—
CT22591-3-198/F7	198	Only 2	—	1	1	—	1	—	—	1	—
CT22591-3-201/F7	201	Only 2	1	—	1	—	—	—	—	1	1
CT22591-3-202/F7	202	Only 2	1	—	1	—	—	1	—	—	1
CT22591-3-203/F7	203	Only 2	—	1	1	—	—	—	1	—	1
CT22591-3-205/F7	205	Only 2	—	—	1	—	1	1	1	—	—
CT22591-3-206/F7	206	Only 2	1	1	1	—	—	—	—	—	1
CT22591-3-207/F7	207	Only 2	1	—	—	1	—	1	—	1	—
CT22591-3-208/F7	208	Only 2	1	1	—	—	1	—	1	—	—
CT22591-3-209/F7	209	Only 2	—	1	—	1	1	—	1	—	—
CT22591-3-210/F7	210	Only 2	—	1	—	1	—	—	1	—	1

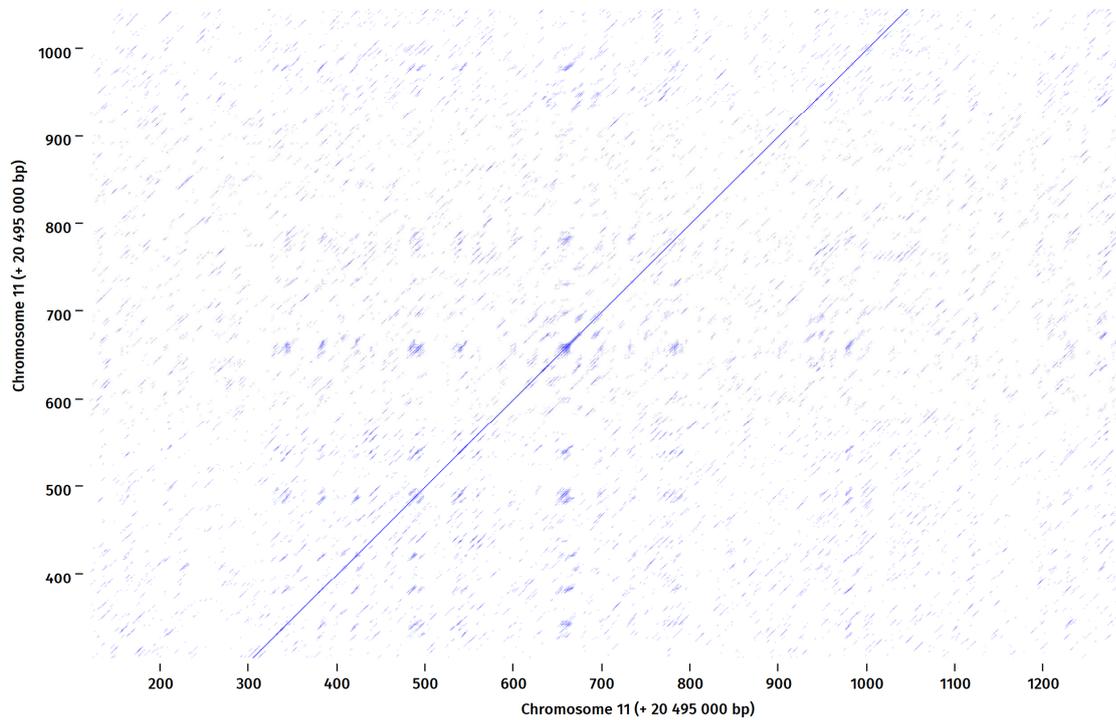
Supplementary Table S3.2 (continued)

Genotype names (CIAT code)	Abbreviated code	Screen	Batches								
			2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9
CT22591-3-212/F7	212	Only 2	—	—	1	—	1	—	1	1	—
CT22591-3-213/F7	213	Only 2	—	—	—	1	1	—	—	1	1
CT22591-3-218/F7	218	Only 2	—	—	—	—	1	1	—	1	1
CT22591-3-221/F7	221	Only 2	1	1	1	—	—	—	—	1	—
CT22591-3-222/F7	222	Only 2	—	—	1	—	—	1	—	1	1
CT22591-3-225/F7	225	Only 2	—	1	—	—	—	1	1	—	1
CT22591-3-226/F7	226	Only 2	1	—	—	—	1	1	—	—	1
CT22591-3-228/F7	228	Only 2	—	1	1	1	—	—	1	—	—
CT22591-3-229/F7	229	Only 2	—	—	—	1	—	—	1	1	1
CT22591-3-230/F7	230	Only 2	—	—	—	1	1	1	—	1	—
CT22591-3-234/F7	234	Only 2	1	—	1	1	—	—	—	—	1
CT22591-3-235/F7	235	Only 2	—	1	1	—	—	1	—	—	—
CT22591-3-236/F7	236	Only 2	—	1	1	—	—	1	—	1	—
CT22591-3-240/F7	240	Only 2	—	1	1	—	—	—	—	1	1
CT22591-3-242/F7	242	Only 2	1	1	1	—	—	—	1	—	—
CT22591-3-244/F7	244	Only 2	1	—	1	—	—	—	1	—	1
CT22591-3-246/F7	246	Only 2	—	—	1	1	1	—	1	—	—
CT22591-3-247/F7	247	Only 2	1	1	—	—	1	—	—	1	—
CT22591-3-251/F7	251	Only 2	1	—	—	1	—	1	1	—	—
CT22591-3-254/F7	254	Only 2	1	1	1	—	1	—	—	—	—
CT22591-3-256/F7	256	Only 2	1	1	—	—	—	1	1	—	—
CT22591-3-259/F7	259	Only 2	—	1	—	—	1	1	—	—	1
CT22591-3-260/F7	260	Only 2	1	—	1	—	—	—	1	1	—
CT22591-3-261/F7	261	Only 2	—	—	—	1	—	1	1	1	—
CT22591-3-262/F7	262	Only 2	—	—	—	1	1	1	—	1	—
CT22591-3-263/F7	263	Only 2	—	1	—	1	—	—	1	1	—
CT22591-3-264/F7	264	Only 2	—	1	—	1	1	1	—	—	—
CT22591-3-266/F7	266	Only 2	—	—	1	—	1	1	—	1	—
CT22591-3-267/F7	267	Only 2	1	—	—	1	1	1	—	—	—
CT22591-3-268/F7	268	Only 2	—	—	1	1	—	—	1	1	—
CT22591-3-269/F7	269	Only 2	1	—	1	1	1	—	—	—	—
CT22591-3-271/F7	271	Only 2	1	1	—	1	—	—	—	1	—
CT22591-3-273/F7	273	Only 2	1	—	1	1	—	—	1	—	—
Total number of rice plants			63	63	63	63	62	64	64	65	62

A



B



Supplementary Figure S4.1 Dot plots of the IR64 sequences at the positions where the depth of coverage of CT8556-37-2-3-1-M reads was very high within 2 cM from the peaks of *Striga asiatica*-resistance quantitative trait loci, *qSaB1.1* on chromosome 1 (A) and *qSaB11.1* on chromosome 11 (B).

Supplementary Table S4.1 Annotations of predicted genes on the sequences of the IR64 genome and CT8556-37-2-3-1-M contigs (> 2442 bp) within 2 cM distance from the peak of the *Striga asiatica*-resistance QTL, *qSaB1.1*. Genes at the peak of the QTL are highlighted. Pos., Start and stop positions of coding region (CDS); S., Strand orientation; #E, Number of exons; #AA, Number of amino acids in the translated sequence; DB., Database; Cov., Percentage query coverage of BLAST hit; Iden., Percentage identity of BLAST hit; SSF, Superfamily; PF, Pfam domains; GO, Gene Ontology; C., Is the gene complete?; Sim., Percentage similarity between amino acid sequences of IR64 and CT8556-37-2-3-1-M genes; RAP-DB, Rice Annotation Project Database; MSU, Michigan State University—Rice Genome Annotation Project; NCBI, National Center for Biotechnology Information—Non-redundant protein sequences; Os ind., *Oryza sativa* subsp. *indica*; Os jap., *O. sativa* subsp. *japonica*; TEs, Transposable elements; Cell wall biogen./mod., Cell wall biogenesis/modification.

IR64												CT8556-37-2-3-1-M												Sim.							
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.		
01g08975	36 617 478– 36 620 325	-	4	542	Os01t082	RAP-DB	Protein kinase, core domain containing protein.	100.0	100.0	-	PF07714	GO:0004672	Protein kinases	01c6043g07	36 617 548– 36 620 325	-	yes	5	497	Os01t082	RAP-DB	Protein kinase, core domain containing protein.	91.0	100.0	SSF56112	-	-	Protein kinases	85.3		
01g08979	36 629 662– 36 630 843	-	1	393	LOC_Os08g13340.1	MSU	retrotransposon protein, putative, Ty1-copia subclass, expressed	100.0	97.2	SSF56672	PF07727	-	TEs																		
01g08980	36 632 152– 36 632 951	-	2	229	LOC_Os03g11850.1	MSU	retrotransposon protein, putative, Ty1-copia subclass, expressed	85.0	97.4	-	PF00098 PF14223	-	TEs																		
01g08981	36 634 296– 36 635 544	-	3	280	BAD31085.1	NCBI	hypothetical protein [Os jap.]	52.0	82.8	-	-	-	Hypothetical	01c6045g01	36 641 013– 36 642 989	-	yes	4	442	No hits	-	-	-	-	PF12435	-	-	Hypothetical			
01g08983	36 642 407– 36 644 752	-	3	557	No hits	-	-	-	-	-	PF12435	-	Hypothetical																		
01g08984	36 645 927– 36 646 166	+	1	79	BAD73498.1	NCBI	hypothetical protein [Os jap.]	86.0	92.6	-	-	-	Hypothetical																		
01g08985	36 649 089– 36 649 715	-	1	208	Os01t082	RAP-DB	Similar to RING-H2 finger protein ATL3C.	100.0	99.5	SSF57850	PF13639	-	Domain-containing	01c6047g01	36 649 543– 36 649 715	-	no	1	57	Os01t082	RAP-DB	Similar to RING-H2 finger protein ATL3C.	100.0	100.0	-	-	-	Domain-containing	26.9		
01g08987	36 657 543– 36 657 911	-	1	122	LOC_Os01g60740.1	MSU	LTP16 - Protease inhibitor/seed storage/LTP family protein precursor, expressed	100.0	100.0	SSF47699	PF00234	-	Proteases/protease inhibitors	01c6047g03	36 657 543– 36 657 911	-	yes	1	123	LOC_Os01g60740.1	MSU	LTP16 - Protease inhibitor/seed storage/LTP family protein precursor, expressed	100.0	100.0	SSF47699	PF00234	-	Proteases/protease inhibitors	100.0		
01g08988	36 658 761– 36 662 786	-	6	250	LOC_Os01g60750.1	MSU	transposon protein, putative, unclassified, expressed	50.0	56.3	-	-	-	TEs																		
01g08989	36 664 604– 36 664 861	+	1	85	OslR64_01g044510	EuGe	Conserved hypothetical protein	-	-	-	-	-	Hypothetical	01c6050g03	36 664 604– 36 664 861	+	yes	1	86	No hits	-	-	-	-	-	-	-	Unknown	98.8		
01g08991	36 672 799– 36 674 021	+	3	251	Os01t082	RAP-DB	Alpha-expansin OsEXPA2.	100.0	100.0	SSF50685	PF01357	-	Cell wall biogen./mod.	01c6050g05	36 672 799– 36 674 021	+	yes	3	303	Os01t082	RAP-DB	Alpha-expansin OsEXPA2.	100.0	75.5	SSF50685	PF03330	-	Cell wall biogen./mod.	76.5		
01g08992	36 675 933– 36 687 432	+	14	933	No hits	-	-	-	-	SSF49590	PF03330	-	Oxido-reductases	01c6050g06	36 675 933– 36 682 719	+	yes	8	372	EEC71719.1	NCBI	hypothetical protein Osl_04250 [Os ind.]	53.0	98.5	-	PF01283	GO:0003735	Translocation	37.1		
										SSF51197	-	-		01c6051g01	36 684 001– 36 687 432	+	yes	4	239	Os01t082	RAP-DB	Hypothetical conserved gene.	92.0	89.7	SSF51197	-	-	Oxido-reductases	25.5		
01g08994	36 691 625– 36 693 986	+	2	68	LOC_Os04g54520.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass	54.0	89.2	-	-	-	TEs																		
01g08995	36 694 599– 36 696 161	-	2	479	OslR64_01g044560	EuGe	Conserved hypothetical protein	-	-	-	-	-	Hypothetical																		
01g08996	36 699 106– 36 700 524	-	1	472	XP_01563	NCBI	transcription repressor OFP2 [Os jap.]	100.0	99.6	-	PF04844 PF13724	-	Transcription	01c6051g03	36 699 106– 36 700 525	-	yes	1	468	XP_01563	NCBI	transcription repressor OFP2 [Os jap.]	100.0	98.5	-	PF04844 PF13724	-	Transcription	98.3		
01g08997	36 701 586– 36 701 804	+	1	72	BAD87079.1	NCBI	hypothetical protein [Os jap.]	57.0	100.0	-	-	-	Hypothetical	01c6051g04	36 701 586– 36 701 804	+	yes	1	73	BAD87079.1	NCBI	hypothetical protein [Os jap.]	57.0	100.0	-	-	-	Hypothetical	100.0		
01g08998	36 702 780– 36 704 103	-	3	83	BAS74992.1	NCBI	Os01g0823550 [Os jap.]	64.0	100.0	-	-	-	Hypothetical																		
01g08999	36 706 685– 36 707 444	+	3	170	Os01t082	RAP-DB	Conserved hypothetical protein.	100.0	100.0	-	-	-	Hypothetical	01c6052g02	36 706 685– 36 707 444	+	yes	3	171	Os01t082	RAP-DB	Conserved hypothetical protein.	100.0	100.0	-	-	-	Hypothetical	100.0		

Supplementary Table S4.1 (continued)

IR64														CT8556-37-2-3-1-M														Sim.			
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms		Group		
01g09000	36 707 826- 36 709 139	-	1	437	Os01t082	RAP- 3700-01	Protein of unknown function DUF641, plant domain containing protein.	100.0	99.8	-	PF04859	-	Domain- containing	01c6052g03	36 707 826- 36 709 139	-	yes	1	438	Os01t082	RAP- 3700-01	Protein of unknown function DUF641, plant domain containing protein.	100.0	100.0	-	PF04859	-	Domain- containing	99.8		
01g09001	36 710 784- 36 711 197	+	1	137	XP_01563	NCBI	probable E3 ubiquitin-protein ligase ATL44 [Os jap.]	100.0	100.0	SSF57850	PF13639	-	Protein degradation																		
01g09004	36 716 900- 36 718 942	-	1	680	Os01t082	RAP- 3900-01	U-box E3ubiquitin ligase, Positive regulation of cold stress response	100.0	99.7	SSF48371	-	-	Protein degradation																		
01g09006	36 730 215- 36 731 744	-	2	339	OslR64_0	Eu	Putative Os02g0818201 protein	-	-	-	-	-	Hypothetical																		
01g09008	36 736 499- 36 736 906	+	1	135	BAB93191	NCBI	hypothetical protein [Os jap.]	97.0	97.7	-	-	-	Hypothetical																		
01g09010	36 741 808- 36 744 410	-	2	229	EEC71722	NCBI	hypothetical protein Osl_04259 [Os ind.]	93.0	100.0	-	-	-	Hypothetical																		
														01c6062g01	36 742 294- 36 744 260	+	yes	2	81	LOC_Os01	MSU	expressed protein	69.0	98.2	-	-	-	Hypothetical			
01g09014	36 752 940- 36 757 206	+	4	814	EEC71723	NCBI	hypothetical protein Osl_04261 [Os ind.]	100.0	91.1	SSF56112	-	-	Protein kinases	01c6062g04	36 749 874- 36 757 206	+	yes	8	1104	EEC71723	NCBI	hypothetical protein Osl_04261 [Os ind.]	67.0	95.3	SSF56112	-	-	Protein kinases	64.5		
01g09016	36 761 010- 36 766 480	-	3	145	No hits	-	-	-	-	-	-	-	Unknown	01c6065g01	36 761 014- 36 761 331	-	yes	1	104	PWS21294	NCBI	hypothetical protein DKP78_24345, partial [Enterococcus faecium]	62.0	95.5	-	-	-	Hypothetical	46.9		
														01c6065g02	36 761 935- 36 770 791	+	yes	3	697	Os01t082	RAP- 4800-01	Conserved hypothetical protein.	57.0	97.5	SSF81383	PF12937	GO:0005515	Domain- containing			
01g09017	36 769 298- 36 770 791	+	2	405	EAY76330	NCBI	hypothetical protein Osl_04263 [Os ind.]	96.0	100.0	-	-	-	Hypothetical																		
01g09018	36 772 392- 36 773 066	+	1	224	Os01t082	RAP- 4900-01	Protein of unknown function DUF688 domain containing protein.	100.0	100.0	-	PF05097	-	Domain- containing	01c6065g03	36 772 392- 36 772 483	+	no	1	30	Os01t082	RAP- 4900-01	Protein of unknown function DUF688 domain containing protein.	100.0	100.0	-	-	-	Hypothetical	13.4		
01g09021	36 779 945- 36 780 704	+	2	125	EAY76332	NCBI	hypothetical protein Osl_04265 [Os ind.]	100.0	100.0	-	PF07859	GO:0016787	Hypothetical																		
01g09022	36 782 158- 36 789 253	+	7	324	Os01t082	RAP- 5000-01	Lateral organ boundary domain (LBD)-containing protein, LBD transcription factor, Regulation of constitutive aerenchyma and lateral root formation	76.0	99.2	-	PF03195	-	Growth and development	01c6067g02	36 779 945- 36 789 253	+	yes	9	557	No hits	-	-	-	-	SSF53474	PF07859	PF03195	-	Growth and development	57.2	
01g09024	36 799 833- 36 800 060	+	1	75	BAB61207	NCBI	P0460E08.17 [Os jap.]	89.0	86.6	-	-	-	Hypothetical																		
01g09026	36 802 993- 36 804 954	-	2	540	Os01t082	RAP- 5500-01	Nodulin-like domain containing protein.	100.0	100.0	SSF10347	PF06813	-	Transporters	01c6069g02	36 802 993- 36 804 954	-	yes	3	513	Os01t082	RAP- 5500-01	Nodulin-like domain containing protein.	100.0	94.8	SSF10347	PF06813	-	Transporters	94.8		
01g09029	36 809 709- 36 818 616	-	11	1068	Os01t082	RAP- 5700-01	Similar to VHS2 protein (Fragment).	51.0	99.6	SSF48464	PF01453	-	Intracellular transport	01c6069g05	36 809 709- 36 812 870	-	yes	3	487	XP_01563	NCBI	EP1-like glycoprotein 3 [Os jap.]	100.0	94.2	SSF51110	PF01453	-	Intracellular transport	94.1		
														01c6069g06	36 814 583- 36 818 526	-	yes	9	568	LOC_Os01	MSU	VHS and GAT domain containing protein, expressed	100.0	99.8	-	PF03127	GO:0006886	Intracellular transport			
01g09030	36 821 200- 36 826 970	-	5	559	LOC_Os01	MSU	transmembrane amino acid transporter protein, putative, expressed	77.0	82.2	-	PF01490	-	Transporters	01c6069g07	36 819 782- 36 826 971	-	yes	6	494	LOC_Os01	MSU	transmembrane amino acid transporter protein, putative, expressed	94.0	90.9	-	PF01490	-	Transporters	70.6		
01g09031	36 829 947- 36 830 635	-	2	191	Os01t082	RAP- 5800-01	Amino acid transporter, transmembrane domain containing protein.	97.0	97.8	-	PF01490	-	Transporters	01c6069g08	36 829 947- 36 830 635	-	yes	2	192	Os01t082	RAP- 5800-01	Amino acid transporter, transmembrane domain containing protein.	97.0	98.9	-	PF01490	-	Transporters	99.0		
01g09032	36 832 913- 36 834 367	+	1	484	LOC_Os01	MSU	SAG20, putative, expressed	97.0	99.6	-	PF11204	-	Senescence	01c6069g09	36 832 913- 36 834 463	+	yes	1	517	LOC_Os01	MSU	SAG20, putative, expressed	100.0	99.8	-	PF04749	-	Senescence	89.7		
01g09033	36 836 301- 36 836 792	+	3	69	Os01t082	RAP- 6000-00	Heavy metal transport/ detoxification protein domain containing protein.	100.0	100.0	SSF55008	PF00403	GO:0030001 GO:0046872	Transporters	01c6069g10	36 836 301- 36 836 792	+	yes	3	70	Os01t082	RAP- 6000-00	Heavy metal transport/ detoxification protein domain containing protein.	100.0	100.0	SSF55008	PF00403	GO:0030001 GO:0046872	Transporters	100.0		
01g09035	36 848 500- 36 850 672	+	5	547	Q6B6R4.1	NCBI	RecName: Full=WRKY transcription factor WRKY24; Short=OsWRKY24 [Os ind.]	100.0	91.4	SSF11829	PF03106	GO:0003700 GO:0006355 GO:0043565	Transcription	01c6069g12	36 848 500- 36 850 673	+	yes	5	556	LOC_Os01	MSU	WRKY24, expressed	100.0	99.3	SSF11829	PF03106	GO:0003700 GO:0006355 GO:0043565	Transcription	90.9		

Supplementary Table S4.1 (continued)

IR64											CT8556-37-2-3-1-M											Sim.									
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Sim.		
01g09036	36 851 730- 36 852 065	+	1	111	BAS75020	NCBI	Oso1g0826500 [Os jap.]	100.0	100.0	-	-	-	Hypothetical	01c6069g13	36 851 730- 36 852 065	+	yes	1	112	Oso1t082 6500-00	RAP- DB	Hypothetical protein	100.0	100.0	-	-	-	Hypothetical	100.0		
01g09037	36 854 274- 36 856 891	-	2	175	BAD68849	NCBI	Epstein-Barr virus EBNA-1-like protein [Os jap.]	86.0	69.9	-	-	-	Epstein-Barr virus EBNA-1-like protein																		
01g09038	36 857 248- 36 857 982	-	1	244	BAD22247	NCBI	Epstein-Barr virus EBNA-1-like protein [Os jap.]	90.0	90.0	-	PF04569	-	Epstein-Barr virus EBNA-1-like protein TEs																		
01g09039	36 858 223- 36 859 064	-	2	95	LOC_Os07 g09930.1	MSU	transposon protein, putative, unclassified, expressed	97.0	91.3	-	-	-	TEs	01c6072g01	36 860 633- 36 861 663	+	yes	5	172	LOC_Os01 g61100.1	MSU	transposon protein, putative, unclassified, expressed	84.0	98.6	-	-	-	TEs	89.5		
01g09040	36 860 633- 36 861 615	+	5	155	LOC_Os01 g61100.1	MSU	transposon protein, putative, unclassified, expressed	91.0	97.9	-	-	-	TEs	01c6072g02	36 862 743- 36 873 681	-	yes	22	727	LOC_Os01 g61100.1	MSU	ulp1 protease family, C-terminal catalytic domain containing protein, expressed	88.0	80.3	SSF54001	-	-	Proteases/ protease inhibitors	78.4		
01g09041	36 864 298- 36 873 681	-	24	701	LOC_Os01 g61110.1	MSU	ulp1 protease family, C-terminal catalytic domain containing protein, expressed	89.0	74.7	SSF54001	-	-	Proteases/ protease inhibitors	01c6072g03	36 874 237- 36 874 752	-	yes	1	172	Oso1t082 6800-01	RAP- DB	Conserved hypothetical protein.	100.0	100.0	-	PF14009	-	Hypothetical	100.0		
01g09042	36 874 237- 36 874 752	-	1	171	Oso1t082 6800-01	RAP- DB	Conserved hypothetical protein.	100.0	100.0	-	PF14009	-	Hypothetical	01c6072g05	36 882 321- 36 887 637	-	yes	8	821	XP_02588 0804.1	NCBI	uncharacterized protein LOC4327520 [Os jap.]	99.0	99.8	SSF15950 1	PF11891 PF04187	-	Growth and development	66.5		
01g09043	36 877 113- 36 887 403	-	10	1081	Oso1t082 6900-01	RAP- DB	Domain of unknown function DUF399 domain containing protein.	67.0	99.7	SSF15950 1	PF11891 PF04187	-	Growth and development	01c6072g06	36 887 801- 36 888 890	+	yes	1	362	LOC_Os01 g61130.1	MSU	retrotransposon protein, putative, LINE subclass, expressed	96.0	95.7	SSF56219	-	-	TEs			
01g09044	36 888 935- 36 889 474	+	1	179	LOC_Os12 g43910.1	MSU	retrotransposon protein, putative, LINE subclass, expressed	97.0	91.9	SSF56672	PF00078	-	TEs	01c6072g07	36 890 266- 36 890 880	+	yes	1	205	LOC_Os01 g61140.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	97.1	-	PF13966	-	TEs			
01g09045	36 891 149- 36 895 492	-	6	603	LOC_Os01 g61150.1	MSU	phox domain-containing protein, putative, expressed	94.0	98.8	-	PF00787 PF09325	-	Domain- containing	01c6072g08	36 891 149- 36 895 492	-	yes	6	604	LOC_Os01 g61150.1	MSU	phox domain-containing protein, putative, expressed	94.0	99.3	-	PF00787 PF09325	-	Domain- containing	99.5		
01g09046	36 898 351- 36 900 536	+	5	567	Oso1t082 7300-01	RAP- DB	Putative laccase precursor, Abiotic stress response	100.0	99.8	SSF49503	-	-	Stress-related	01c6072g09	36 898 351- 36 900 536	+	yes	5	568	Oso1t082 7300-01	RAP- DB	Putative laccase precursor, Abiotic stress response	100.0	100.0	SSF49503	-	-	Stress-related	99.8		
01g09047	36 903 417- 36 904 085	-	1	222	LOC_Os01 g61170.1	MSU	prenylated rab acceptor, putative, expressed	100.0	84.7	-	PF03208	-	Intracellular transport	01c6072g10	36 903 417- 36 904 085	-	yes	1	223	LOC_Os01 g61170.1	MSU	prenylated rab acceptor, putative, expressed	100.0	85.6	-	PF03208	-	Intracellular transport	99.1		
01g09048	36 904 975- 36 906 931	+	3	601	Oso1t082 7500-01	RAP- DB	Exo70 exocyst complex subunit family protein.	100.0	90.5	-	PF03081	GO:0000145 GO:0006887	Intracellular transport	01c6072g11	36 904 975- 36 905 453	+	no	1	159	Oso1t082 7500-01	RAP- DB	Exo70 exocyst complex subunit family protein.	100.0	100.0	-	-	-	Intracellular transport	20.0		
01g09049	36 908 696- 36 910 312	+	1	538	Oso1t082 7600-01	RAP- DB	Exo70 exocyst complex subunit family protein.	99.0	100.0	-	PF03081	GO:0000145 GO:0006887	Intracellular transport	01c6073g01	36 905 584- 36 910 312	+	yes	3	997	OEL21590 1	NCBI	Exocyst complex component EXO70B1 [Dichanthelium oligosanthes]	100.0	78.0	-	PF03081	GO:0000145 GO:0006887	Intracellular transport	53.3		
01g09050	36 911 118- 36 912 673	-	3	293	LOC_Os01 g61200.1	MSU	GDSL-like lipase/acylhydrolase, putative, expressed	98.0	98.6	SSF52266	-	-	Lipases	01c6073g02	36 911 118- 36 912 673	-	yes	3	294	Oso1t082 7700-00	RAP- DB	Lipase, GDSL domain containing protein.	98.0	99.0	-	PF00657	GO:0016788	Lipases	99.7		
01g09051	36 916 587- 36 917 213	+	1	208	AM10778	NCBI	ABA receptor 8 [Os ind.]	100.0	100.0	SSF55961	PF10604	-	Hormone- related	01c6073g03	36 916 587- 36 917 216	+	yes	1	202	XP_01564 3625.1	NCBI	abscisic acid receptor PYL4 [Os jap.]	100.0	96.2	SSF55961	PF10604	-	Hormone- related	95.7		
01g09052	36 919 566- 36 923 771	-	4	282	BAB68103	NCBI	hypothetical protein [Os jap.]	53.0	69.3	-	-	-	Hypothetical																		
01g09053	36 926 304- 36 928 224	+	6	400	LOC_Os01 g61230.1	MSU	dihydroflavonol-4-reductase, putative, expressed	88.0	92.6	SSF51735	PF01370 PF16363	-	Biosynthetic processes	01c6075g01	36 926 448- 36 928 224	+	yes	6	353	LOC_Os01 g61230.1	MSU	dihydroflavonol-4-reductase, putative, expressed	100.0	92.6	SSF51735	PF16363 PF01370	-	Biosynthetic processes	88.0		
01g09054	36 929 254- 36 931 173	+	6	260	XP_01564 5271.1	NCBI	gamma-interferon-responsive lysosomal thiol protein [Os jap.]	100.0	79.6	-	PF03227	-	Protein degradation	01c6075g02	36 929 254- 36 931 173	+	yes	6	248	XP_01564 5271.1	NCBI	gamma-interferon-responsive lysosomal thiol protein [Os jap.]	100.0	100.0	-	PF03227	-	Protein degradation	79.3		
01g09055	36 932 330- 36 939 268	+	7	800	Oso1t082 8300-01	RAP- DB	Protein of unknown function DUF248, methyltransferase putative family protein.	63.0	99.7	SSF53335	-	-	Domain- containing	01c6075g03	36 932 330- 36 936 872	+	yes	5	512	Oso1t082 8300-01	RAP- DB	Protein of unknown function DUF248, methyltransferase putative family protein.	100.0	99.7	-	PF03141	GO:0008168	Domain- containing	62.9		
01g09056	36 940 648- 36 944 279	-	3	58	BAS75036	NCBI	hypothetical protein [Os jap.]	66.0	90.0	-	-	-	Hypothetical	01c6075g05	36 940 368- 36 941 452	-	yes	2	145	LOC_Os01 g61290.1	MSU	hypothetical protein	75.0	97.2	-	-	-	Hypothetical	22.8		
01g09057	36 948 019- 36 950 263	-	2	224	BAB68107	NCBI	hypothetical protein [Os jap.]	76.0	95.3	-	-	-	Hypothetical																		

Supplementary Table S4.1 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#	AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#	AA	BLAST hit	DB.		Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	
01g09059	36 953 848- 36 954 579	-	1	243	XP_01564 3980.1	NCBI	protein G1-like7 [Os jap.]	54.0	100.0	-	PF04852	-	Growth and development	01c6077g02	36 953 744- 36 954 506	-	yes	2	170	XP_01564 3980.1	NCBI	protein G1-like7 [Os jap.]	96.0	100.0	-	PF04852	-	Growth and development	44.3	
01g09060	36 955 531- 36 956 266	+	3	87	EEC66640. 1	NCBI	hypothetical protein Osl_32897 [Os ind.]	90.0	98.7	-	-	-	Hypothetical																	
01g09061	36 959 941- 36 961 442	+	4	317	XP_02587 8013.1	NCBI	thioredoxin-like fold domain-containing protein MRL7 homolog, chloroplastic [Os jap.]	100.0	100.0	SSF52833	-	-	Growth and development	01c6079g01	36 959 941- 36 961 442	+	yes	4	285	Os01t082 9000-01	RAP-DB	Thioredoxin-like fold domain containing protein.	100.0	89.6	SSF52833	-	-	Growth and development	89.6	
01g09062	36 962 775- 36 964 637	-	2	426	Os01t082 9100-01	RAP-DB	Ankyrin domain containing protein.	100.0	99.8	SSF49354 SSF48403	PF00635 PF12796	-	Domain-containing	01c6079g02	36 962 775- 36 964 699	-	yes	2	437	Os01t082 9100-01	RAP-DB	Ankyrin domain containing protein.	90.0	99.5	SSF48403 SSF49354	PF12796 PF13857 PF00635	-	Domain-containing	90.7	
01g09063	36 966 795- 36 973 554	+	7	672	LOC_Os03 g06540.1	MSU	retrotransposon protein, putative, unclassified, expressed	74.0	96.1	-	PF14223	-	TEs																	
01g09064	36 974 652- 36 977 778	-	4	205	EAY76358. 1	NCBI	hypothetical protein Osl_04290 [Os ind.]	83.0	58.8	-	-	-	Hypothetical	01c6079g03	36 966 896- 36 977 211	-	yes	5	305	No hits	-	-		-	-	-	-	Unknown	14.3	
01g09065	36 979 431- 36 980 198	+	1	255	LOC_Os01 g61350.1	MSU	glutaredoxin, putative, expressed	100.0	99.6	SSF52833	-	-	Oxido-reductases																	
01g09067	36 983 482- 36 983 884	-	2	104	LOC_Os01 g61360.1	MSU	DEFL4 - Defensin and Defensin-like DEFL family, expressed	100.0	96.2	-	-	-	Stress-related																	
01g09068	36 984 738- 36 985 898	-	1	386	Os01t082 9600-01	RAP-DB	Protein of unknown function DUF688 domain containing protein.	100.0	99.5	-	PF05097	-	Domain-containing	01c6083g01	36 985 497- 36 986 613	-	no	2	220	Os01t082 9600-01	RAP-DB	Protein of unknown function DUF688 domain containing protein.	65.0	100.0	-	PF05097	-	Domain-containing	28.2	
01g09069	36 987 825- 36 989 015	+	1	396	LOC_Os01 g61380.1	MSU	lactate/malate dehydrogenase, putative, expressed	100.0	99.5	SSF56327 SSF51735	-	-	Respiration	01c6083g02	36 987 825- 36 989 015	+	yes	1	397	LOC_Os01 g61380.1	MSU	lactate/malate dehydrogenase, putative, expressed	100.0	100.0	SSF56327 SSF51735	-	-	Respiration	99.5	
01g09070	36 989 675- 36 992 270	-	5	716	Os01t082 9900-01	RAP-DB	Iron(III)-deoxymugineic acid transporter, Translocation of iron in reproductive organs and phloem in joints	100.0	94.4	-	PF03169	GO:0055085	Transporters	01c6083g03	36 989 675- 36 992 217	-	yes	5	699	Os01t082 9900-01	RAP-DB	Iron(III)-deoxymugineic acid transporter, Translocation of iron in reproductive organs and phloem in joints	100.0	94.7	-	PF03169	GO:0055085	Transporters	97.1	
01g09071	36 993 321- 36 995 984	-	2	544	LOC_Os01 g61400.1	MSU	sufB/sufD domain containing protein, putative, expressed	100.0	100.0	SSF10196 0	-	-	Domain-containing	01c6084g01	36 993 321- 36 995 984	-	yes	2	545	Os01t083 0000-01	RAP-DB	Similar to Plastid sufB (Fragment).	100.0	100.0	SSF10196 0	-	-	Domain-containing	100.0	
01g09072	36 996 564- 37 000 374	+	8	475	Os01t083 0100-01	RAP-DB	Pyridine nucleotide-disulphide oxidoreductase, NAD-binding region domain containing protein.	95.0	100.0	-	PF07992	GO:0016491 GO:0055114	Oxido-reductases	01c6084g02	36 996 564- 36 996 946	+	no	1	127	LOC_Os01 g61410.1	MSU	NADH-ubiquinone oxidoreductase, mitochondrial precursor, putative, expressed	100.0	100.0	-	PF07992	GO:0016491 GO:0055114	Oxido-reductases	95.6	
													01c6085g01	36 997 715- 37 000 374	+	yes	7	328	Os01t083 0100-01	RAP-DB	Pyridine nucleotide-disulphide oxidoreductase, NAD-binding region domain containing protein.	93.0	100.0	-	PF07992	GO:0016491 GO:0055114	Oxido-reductases			
01g09073	37 003 101- 37 004 786	-	1	561	Os01t083 0200-01	RAP-DB	Zinc finger, RING/FYVE/PHD-type domain containing protein.	100.0	100.0	SSF57850	PF13923	-	Domain-containing	01c6085g02	37 003 101- 37 004 786	-	yes	1	562	Os01t083 0200-01	RAP-DB	Zinc finger, RING/FYVE/PHD-type domain containing protein.	100.0	100.0	SSF57850	PF13923	-	Domain-containing	100.0	
01g09074	37 006 480- 37 010 860	-	4	76	LOC_Os01 g61430.1	MSU	HIT zinc finger domain containing protein, expressed	55.0	67.8	-	-	-	Domain-containing	01c6085g03	37 006 480- 37 007 802	-	yes	3	73	LOC_Os01 g61430.1	MSU	HIT zinc finger domain containing protein, expressed	53.0	69.1	-	-	Domain-containing	81.4		
													01c6085g04	37 009 162- 37 010 860	-	yes	3	87	CAB34747 06.1	NCBI	unnamed protein product [Digitaria exilis]	85.0	53.4	SSF14423 2	PF04438	-	Domain-containing			
01g09076	37 013 151- 37 014 915	-	2	394	Os01t083 0500-01	RAP-DB	20G-Fe(II) oxygenase domain containing protein.	100.0	99.7	-	PF03171	GO:0016491	Oxido-reductases	01c6088g01	37 013 151- 37 014 915	-	yes	2	395	Os01t083 0500-01	RAP-DB	20G-Fe(II) oxygenase domain containing protein.	100.0	100.0	SSF51197	-	-	Oxido-reductases	99.8	
01g09077	37 015 659- 37 020 018	-	6	936	LOC_Os12 g03160.1	MSU	transposon protein, putative, unclassified, expressed	87.0	80.3	-	PF10551 PF03108	-	TEs																	
01g09079	37 022 937- 37 024 081	+	2	232	No hits	-	-	-	-	-	-	-	Unknown	01c6088g02	37 023 021- 37 024 831	+	yes	3	220	LOC_Os01 g61450.1	MSU	expressed protein	52.0	96.0	-	-	Hypothetical	26.0		
01g09081	37 031 283- 37 038 031	+	7	650	LOC_Os01 g61460.1	MSU	leaf senescence related protein, putative, expressed	80.0	99.8	-	PF13839 PF14416	-	Senescence	01c6088g04	37 031 283- 37 034 775	+	yes	5	421	LOC_Os01 g61460.1	MSU	leaf senescence related protein, putative, expressed	96.0	99.5	-	PF14416 PF13839	-	Senescence	62.5	
01g09083	37 041 477- 37 042 043	-	1	188	LOC_Os01 g61470.1	MSU	zinc finger, C3HC4 type domain containing protein, expressed	100.0	98.4	SSF57850	PF13639	-	Domain-containing	01c6091g01	37 041 477- 37 042 046	-	yes	1	190	XP_01561 8570.1	NCBI	RING-H2 finger protein ATL66 [Os jap.]	100.0	100.0	SSF57850	PF13639	-	Domain-containing	98.4	
01g09086	37 052 191- 37 052 862	+	1	223	Os01t083 1000-01	RAP-DB	Basic helix-loop-helix (bHLH) transcription factor, Axillary meristem formation	96.0	99.5	SSF47459	PF00010	GO:0046983	Transcription	01c6091g03	37 052 191- 37 052 862	+	yes	1	224	Os01t083 1000-01	RAP-DB	Basic helix-loop-helix (bHLH) transcription factor, Axillary meristem formation	96.0	100.0	SSF47459	PF00010	GO:0046983	Transcription	99.1	

Supplementary Table S4.1 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	
01g09087	37 057 660- 37 068 704	+	11	1531	LOC_Os02 g13750.1	MSU	retrotransposon protein, putative, unclassified, expressed	81.0	84.5	-	PF00098 PF14223	-	TEs																	
01g09088	37 072 525- 37 075 451	+	4	281	Os01t083 1200-01	RAP- DB	BAG family protein, Control of innate immunity and broad- spectrum disease resistance	100.0	92.9	-	PF00240 PF02179	GO:0005515 GO:0051087	Stress-related	01c6099g01	37 072 525- 37 075 451	+	yes	4	282	Os01t083 1200-01	RAP- DB	BAG family protein, Control of innate immunity and broad- spectrum disease resistance	100.0	93.2	-	PF02179 PF00240	GO:0051087 GO:0005515	Stress-related	99.6	
01g09089	37 076 094- 37 078 320	-	3	497	LOC_Os01 g61510.1	MSU	ammonium transporter protein, putative, expressed	100.0	100.0	-	PF00909	GO:0008519 GO:0015696 GO:0016020	Transporters	01c6099g02	37 076 094- 37 078 320	-	yes	3	498	LOC_Os01 g61510.1	MSU	ammonium transporter protein, putative, expressed	100.0	100.0	-	PF00909	GO:0008519 GO:0015696 GO:0016020	Transporters	100.0	
01g09090	37 079 973- 37 080 416	-	1	147	LOC_Os01 g61520.1	MSU	expressed protein	60.0	93.3	-	-	-	Hypothetical																	
01g09091	37 081 956- 37 083 221	-	3	268	BAD81856 .1	NCBI	protein kinase-like [Os jap.]	84.0	71.4	SSF56112	-	-	Protein kinases																	
01g09092	37 084 054- 37 084 944	+	2	257	Os01t083 1800-00	RAP- DB	Conserved hypothetical protein.	93.0	80.5	-	-	-	Hypothetical	01c6101g01	37 084 054- 37 085 051	+	yes	1	321	Os01t083 1800-00	RAP- DB	Conserved hypothetical protein.	84.0	92.2	-	-	-	Hypothetical	65.4	
01g09093	37 085 711- 37 087 384	-	3	497	Os01t083 1900-02	RAP- DB	Similar to Ammonium transporter.	100.0	100.0	-	PF00909	GO:0008519 GO:0015696 GO:0016020	Transporters	01c6101g02	37 085 711- 37 086 458	-	yes	2	220	Os01t083 1900-02	RAP- DB	Similar to Ammonium transporter.	100.0	100.0	-	PF00909	GO:0008519 GO:0015696 GO:0016020	Transporters	96.4	
													01c6102g01	37 086 512- 37 087 384	-	no	2	260	Os01t083 1900-01	RAP- DB	Similar to Ammonium transporter.	100.0	100.0	-	PF00909	GO:0008519 GO:0015696 GO:0016020	Transporters			
01g09094	37 089 676- 37 092 414	-	5	252	XP_01562 4995.1	NCBI	probable phytol kinase 2, chloroplastic [Os jap.]	100.0	72.4	-	-	-	Protein kinases	01c6102g02	37 089 676- 37 092 423	-	yes	5	256	XP_01562 4995.1	NCBI	probable phytol kinase 2, chloroplastic [Os jap.]	100.0	73.7	-	-	-	Protein kinases	98.4	
01g09095	37 094 178- 37 095 956	+	3	379	LOC_Os01 g61570.1	MSU	GDSL-like lipase/acylhydrolase, putative, expressed	100.0	95.5	-	-	GO:0016788	Lipases	01c6102g03	37 094 178- 37 095 956	+	yes	3	365	LOC_Os01 g61570.1	MSU	GDSL-like lipase/acylhydrolase, putative, expressed	100.0	100.0	SSF52266	-	-	Lipases	95.5	
01g09096	37 097 213- 37 101 422	-	5	295	Os01t083 2200-01	RAP- DB	Similar to self-like protein.	78.0	100.0	-	-	-	Stress-related	01c6102g04	37 097 163- 37 101 422	-	yes	5	313	Os01t083 2200-01	RAP- DB	Similar to self-like protein.	73.0	100.0	SSF52833	PF10262	-	Stress-related	93.3	
01g08976	36 621 492- 36 621 809	+	1	105	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6043g08	36 623 760- 36 624 523	-	yes	2	70	No hits	-	-	-	-	-	-	-	Unknown		
01g08977	36 626 142- 36 626 686	+	2	116	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6043g09	36 626 142- 36 626 686	+	yes	2	117	No hits	-	-	-	-	-	-	-	Unknown	95.7	
01g08978	36 627 123- 36 627 643	-	2	126	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6043g10	36 627 123- 36 627 643	-	yes	2	127	No hits	-	-	-	-	-	-	-	Unknown	99.2	
01g08982	36 636 735- 36 641 209	-	3	139	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6045g02	36 643 431- 36 643 937	+	yes	1	169	No hits	-	-	-	-	-	-	-	Unknown		
														01c6045g03	36 644 278- 36 644 575	+	no	2	82	No hits	-	-	-	-	-	-	-	Unknown		
														01c6046g01	36 645 686- 36 647 020	+	yes	3	179	No hits	-	-	-	-	-	-	-	Unknown		
														01c6047g02	36 650 100- 36 654 126	+	yes	3	112	No hits	-	-	-	-	-	-	-	Unknown		
01g08986	36 655 264- 36 655 618	+	2	97	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6050g01	36 660 145- 36 661 690	+	yes	5	95	No hits	-	-	-	-	-	-	-	Unknown		
														01c6050g02	36 662 441- 36 662 788	+	yes	1	116	No hits	-	-	-	-	-	-	-	Unknown		
01g08990	36 665 879- 36 668 285	-	3	338	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6050g04	36 665 659- 36 668 231	-	yes	4	338	No hits	-	-	-	-	-	-	-	Unknown	34.0	
01g08993	36 689 035- 36 690 359	+	2	75	No hits	-	-	-	-	-	-	-	Unknown																	
														01c6051g02	36 689 035- 36 690 359	+	yes	2	76	No hits	-	-	-	-	-	-	-	Unknown	100.0	
														01c6052g01	36 704 712- 36 705 997	+	yes	3	113	No hits	-	-	-	-	-	-	-	Unknown		

Supplementary Table S4.1 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	
01g09002	36 713 499-	-	3	77	No hits	-	-	-	-	-	-	-	Unknown	01c6053g01	36 713 499-	-	yes	4	205	No hits	-	-	-	-	-	-	Unknown	37.8		
	36 714 840														36 715 817															
01g09003	36 715 734+	+	2	183	No hits	-	-	-	-	-	-	-	Unknown																	
	36 716 550																													
01g09005	36 719 426-	+	6	388	No hits	-	-	-	-	-	-	-	Unknown	01c6060g01	36 731 735-	+	yes	3	189	No hits	-	-	-	-	-	-	Unknown			
	36 724 495														36 733 682															
01g09007	36 733 368-	+	1	104	No hits	-	-	-	-	-	-	-	Unknown																	
	36 733 682																													
01g09009	36 738 534-	-	3	54	No hits	-	-	-	-	-	-	-	Unknown	01c6061g01	36 738 408-	+	yes	3	158	No hits	-	-	-	-	-	-	Unknown			
	36 740 528														36 740 340															
01g09011	36 747 594-	-	1	84	No hits	-	-	-	-	-	-	-	Unknown	01c6062g02	36 747 594-	-	yes	1	85	No hits	-	-	-	-	-	-	Unknown	100.0		
	36 747 848														36 747 848															
01g09012	36 748 549-	-	1	190	No hits	-	-	-	-	-	-	-	Unknown	01c6062g03	36 748 549-	-	yes	1	191	No hits	-	-	-	-	-	-	Unknown	98.4		
	36 749 121														36 749 121															
01g09013	36 750 163-	-	5	369	No hits	-	-	-	-	-	-	-	Unknown																	
	36 752 397																													
01g09015	36 758 869-	+	3	330	No hits	-	-	-	-	-	-	-	Unknown																	
	36 760 463																													
01g09019	36 773 999-	-	1	79	No hits	-	-	-	-	-	-	-	Unknown																	
	36 774 238																													
01g09020	36 777 152+	+	2	78	No hits	-	-	-	-	-	-	-	Unknown	01c6067g01	36 777 152-	+	yes	3	111	No hits	-	-	-	-	-	-	Unknown	70.0		
	36 778 560														36 778 560															
01g09023	36 792 449-	-	3	294	No hits	-	-	-	-	-	-	-	Unknown	01c6067g03	36 792 449-	-	yes	3	282	No hits	-	-	-	-	-	-	Unknown	94.2		
	36 794 021														36 794 021															
01g09025	36 800 946-	+	2	54	No hits	-	-	-	-	-	-	-	Unknown	01c6069g01	36 800 946-	+	yes	2	55	No hits	-	-	-	-	-	-	Unknown	100.0		
	36 801 193														36 801 193															
															01c6069g03	36 805 189-	+	yes	2	114	No hits	-	-	-	-	-	Unknown			
															36 806 668															
01g09027	36 806 595-	+	1	84	No hits	-	-	-	-	-	-	-	Unknown																	
	36 806 849																													
01g09028	36 807 960-	+	1	82	No hits	-	-	-	-	-	-	-	Unknown	01c6069g04	36 807 960-	+	yes	1	83	No hits	-	-	-	-	-	-	Unknown	98.8		
	36 808 208														36 808 208															
01g09034	36 839 522-	+	4	428	No hits	-	-	-	-	-	-	-	Unknown	01c6069g11	36 837 460-	-	yes	8	287	No hits	-	-	-	-	-	-	Unknown	1.6		
	36 843 727														36 845 968															
															01c6072g04	36 877 800-	+	yes	1	142	No hits	-	-	-	-	-	Unknown			
															36 878 225															
															01c6075g04	36 938 436-	+	yes	1	243	No hits	-	-	-	-	-	Unknown			
															36 939 180															
															01c6075g06	36 944 054-	-	yes	1	77	No hits	-	-	-	-	-	Unknown			
															36 944 285															
01g09058	36 952 657-	+	2	99	No hits	-	-	-	-	-	-	-	Unknown	01c6077g01	36 952 657-	+	yes	2	100	No hits	-	-	-	-	-	-	Unknown	100.0		
	36 953 322														36 953 322															
01g09066	36 981 644-	-	2	138	No hits	-	-	-	-	-	-	-	Unknown																	
	36 982 832																													
01g09075	37 011 617-	-	2	143	No hits	-	-	-	-	-	-	-	Unknown																	
	37 012 162																													
01g09078	37 021 974-	+	1	111	No hits	-	-	-	-	-	-	-	Unknown																	
	37 022 309																													
01g09080	37 026 727-	+	2	120	No hits	-	-	-	-	-	-	-	Unknown	01c6088g03	37 026 727-	+	yes	2	67	No hits	-	-	-	-	-	-	Unknown	52.1		
	37 030 254														37 030 081															
															01c6088g05	37 035 707-	-	yes	2	78	No hits	-	-	-	-	-	Unknown			
															37 036 531															
01g09082	37 038 590-	-	3	68	No hits	-	-	-	-	-	-	-	Unknown																	
	37 039 750																													

Supplementary Table S4.1 (continued)

IR64												CT8556-37-2-3-1-M																	
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.
01g09084	37 045 030- 37 045 290	+	1	86	No hits	-	-	-	-	-	-	-	Unknown																
														01c6091g02	37 043 135- 37 050 261	+	yes	7	198	No hits	-	-	-	-	-	-	-	Unknown	
01g09085	37 047 312- 37 047 530	-	1	72	No hits	-	-	-	-	-	-	-	Unknown																
01g09097	37 103 490- 37 106 686	-	3	174	No hits	-	-	-	-	-	-	-	Unknown																

Supplementary Table S4.2 Annotations of predicted genes on the sequences of the IR64 genome and CT8556-37-2-3-1-M contigs (> 2442 bp) within 2 cM distance from the peak of the *Striga asiatica*-resistance QTL, *qSaB5.1*. Genes at the peak of the QTL are highlighted. Pos., Start and stop positions of coding region (CDS); S., Strand orientation; #E, Number of exons; #AA, Number of amino acids in the translated sequence; DB., Database; Cov., Percentage query coverage of BLAST hit; Iden., Percentage identity of BLAST hit; SSF, Superfamily; PF, Pfam domains; GO, Gene Ontology; C., Is the gene complete?; Sim., Percentage similarity between amino acid sequences of IR64 and CT8556-37-2-3-1-M genes; RAP-DB, Rice Annotation Project Database; MSU, Michigan State University—Rice Genome Annotation Project; NCBI, National Center for Biotechnology Information—Non-redundant protein sequences; Os ind., *Oryza sativa* subsp. *indica*; Os jap., *O. sativa* subsp. *japonica*; TEs, Transposable elements; Cell wall biogen./mod., Cell wall biogenesis/modification.

IR64											CT8556-37-2-3-1-M																			
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.	
05g05765	23 030 654- 23 031 864	-	3	292	Os05t044 2700-01	RAP- DB	No apical meristem (NAM) protein domain containing protein.	100.0	93.2	SSF10194 1	PF02365	GO:0003677 GO:0006355	Growth and development	05c4580g01	23 030 654- 23 031 400	-	yes	2	161	Os05t044 2700-01	RAP- DB	No apical meristem (NAM) protein domain containing protein.	100.0	99.4	SSF10194 1	-	GO:0003677 GO:0006355	Growth and development	54.8	
05g05766	23 034 368- 23 035 181	+	2	222	AAT69608. 1	NCBI	hypothetical protein [Os jap.]	89.0	79.3	-	-	-	Hypothetical	05c4581g01	23 034 368- 23 038 025	+	yes	4	279	AAT69608. 1	NCBI	hypothetical protein [Os jap.]	71.0	81.4	-	-	-	Hypothetical		
05g05767	23 037 922- 23 038 422	+	1	166	LOC_Os05 g37090.1	MSU	hypothetical protein	100.0	95.8	-	-	-	Hypothetical																	
05g05770	23 050 539- 23 051 987	-	2	425	LOC_Os05 g29100.1	MSU	retrotransposon protein, putative, unclassified, expressed	72.0	78.2	-	PF04195	-	TEs	05c4583g01	23 050 539- 23 051 987	-	yes	2	426	LOC_Os05 g29100.1	MSU	retrotransposon protein, putative, unclassified, expressed	72.0	78.2	-	PF04195	-	TEs	100.0	
05g05771	23 054 089- 23 054 496	+	1	135	LOC_Os05 g37110.1	MSU	expressed protein	100.0	99.3	-	-	-	Hypothetical																	
05g05772	23 055 301- 23 060 231	-	12	617	LOC_Os05 g37120.1	MSU	sec23/Sec24 zinc finger family protein, expressed	100.0	99.4	-	PF04811	GO:0006886 GO:0006888	Intracellular transport																	
05g05773	23 063 436- 23 066 021	-	3	393	LOC_Os01 g64170.2	MSU	glycosyl hydrolases family 17, putative, expressed	93.0	73.5	-	PF00332	GO:0004553 GO:0005975	Biosynthetic processes	05c4584g01	23 063 436- 23 066 021	-	yes	3	394	Os01t086 0800-02	RAP- DB	Glycoside hydrolase, family 17 protein.	88.0	74.4	SSF51445	-	-	Biosynthetic processes	100.0	
05g05774	23 068 594- 23 069 040	-	1	148	Os05t044 3500-01	RAP- DB	Similar to Ferredoxin VI, chloroplast precursor (Fd VI).	100.0	100.0	SSF54292	PF00111	GO:0009055 GO:0051536	Growth and development	05c4584g02	23 068 594- 23 069 040	-	yes	1	149	Os05t044 3500-01	RAP- DB	Similar to Ferredoxin VI, chloroplast precursor (Fd VI).	100.0	100.0	SSF54292	PF00111	GO:0009055 GO:0051536	Growth and development	100.0	
05g05776	23 071 766- 23 074 056	-	4	342	Os05t044 3700-01	RAP- DB	Syntaxin 6, N-terminal domain containing protein.	96.0	99.7	SSF47661	-	GO:0016020 GO:0016192	Intracellular transport	05c4584g04	23 071 766- 23 074 056	-	yes	4	343	Os05t044 3700-01	RAP- DB	Syntaxin 6, N-terminal domain containing protein.	96.0	99.7	SSF47661	-	GO:0016020 GO:0016192	Intracellular transport	100.0	
05g05777	23 074 684- 23 078 195	-	8	527	Os05t044 3800-01	RAP- DB	Similar to Plastid division protein ftsZ1 precursor.	86.0	99.1	-	PF12327 PF00091	-	Growth and development	05c4584g05	23 074 684- 23 078 195	-	yes	8	528	Os05t044 3800-01	RAP- DB	Similar to Plastid division protein ftsZ1 precursor.	86.0	99.1	-	PF12327 PF00091	-	Growth and development	100.0	
05g05778	23 079 301- 23 084 923	+	11	1381	LOC_Os12 g41870.1	MSU	transposon protein, putative, unclassified, expressed	100.0	90.9	-	PF10551 PF10536	-	TEs																	
05g05779	23 086 677- 23 088 334	-	2	41	LOC_Os06 g14860.1	MSU	hypothetical protein	78.0	96.9	-	-	-	Hypothetical																	
05g05780	23 090 828- 23 095 909	+	13	560	XP_01563 9062.2	NCBI	transcription factor TGAL5 isoform X6 [Os jap.]	82.0	99.6	SSF57959	-	Transcription	05c4588g01	23 090 828- 23 101 728	+	yes	14	584	XP_01563 9062.2	NCBI	transcription factor TGAL5 isoform X6 [Os jap.]	79.0	99.6	-	PF00170 PF14144 GO:0006355 GO:0006351 GO:0043565	GO:0003700 Transcription	95.9			
05g05781	23 097 190- 23 105 699	-	8	692	AAV44105. 1	NCBI	unknown protein [Os jap.]	97.0	69.4	-	PF12776 PF13359	-	TEs																	
05g05783	23 107 877- 23 112 666	+	7	1410	LOC_Os03 g47600.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	82.4	SSF56672 SSF53098	PF17919 PF13456 PF00078 PF00665	-	TEs																	
05g05784	23 113 290- 23 117 209	-	4	1025	LOC_Os10 g22540.1	MSU	retrotransposon protein, putative, unclassified, expressed	84.0	87.5	-	-	-	TEs																	
05g05785	23 118 935- 23 119 528	-	1	197	LOC_Os05 g37580.1	MSU	retrotransposon protein, putative, unclassified, expressed	85.0	98.2	-	PF04195	-	TEs																	

Supplementary Table S4.2 (continued)

IR64													CT8556-37-2-3-1-M													Sim.				
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits		PF hits	GO terms	Group	
05g05786	23 120 467- 23 122 959	-	4	193	No hits	-	-	-	-	-	-	PF06882	-	Hypothetical																
05g05787	23 124 263- 23 126 074	+	2	391	Os05t044 4200-01	RAP- DB	Similar to T6/4.5 protein (WIP6 protein).	100.0	98.7	SSF57667	-	-	Transcription	05c4589g01	23 124 263- 23 125 875	+	no	2	325	EAY98224.	NCBI	hypothetical protein Osl_20137 [Os ind.]	100.0	100.0	SSF57667	-	-	-	Domain-containing Unknown	82.9
05g05789	23 130 060- 23 130 347	-	1	95	OslR64_0 5g025970	EuGe ne	Putative expressed protein	-	-	-	-	-	Hypothetical	05c4590g02	23 130 060- 23 130 347	-	yes	1	96	No hits	-	-	-	-	-	-	-	-	Unknown	94.7
05g05791	23 137 669- 23 140 503	+	10	331	LOC_Os05 g37200.1	MSU	solute carrier family 35 member F2, putative, expressed	100.0	92.7	SSF10348	-	-	Transporters	05c4590g04	23 137 669- 23 140 503	+	yes	10	332	LOC_Os05 g37200.1	MSU	solute carrier family 35 member F2, putative, expressed	100.0	92.7	-	PF06027	GO:0016021 GO:0022857 GO:0055085	Transporters	100.0	
05g05793	23 143 404- 23 148 651	+	12	407	No hits	-	-	-	-	-	-	PF06027 GO:0016021 GO:0022857 GO:0055085	Transporters	05c4591g01	23 143 925- 23 148 651	+	yes	11	378	XP_00665 5365.1	NCBI	PREDICTED: solute carrier family 35 member F1-like isoform X1 [Oryza brachyantha]	51.0	51.6	-	PF06027	GO:0016021 GO:0022857 GO:0055085	Transporters	91.9	
05g05794	23 149 144- 23 150 580	-	2	124	Os05t045 5900-01	RAP- DB	Pentatricopeptide repeat domain containing protein.	90.0	74.8	-	-	-	Translation																	
05g05795	23 151 665- 23 152 111	-	1	148	BAS94272. 1	NCBI	Os05g0444875 [Os jap.]	69.0	96.1	-	-	-	Hypothetical	05c4592g01	23 151 665- 23 152 111	-	yes	1	149	Os05t044 4875-00	RAP- DB	Hypothetical protein	96.0	97.2	-	-	-	Hypothetical	98.0	
05g05796	23 154 671- 23 155 042	-	1	123	LOC_Os11 g25430.1	MSU	retrotransposon protein, putative, unclassified	100.0	89.4	-	-	PF06882	-	TEs																
05g05802	23 170 991- 23 172 550	+	1	519	Os05t044 5100-01	RAP- DB	Cytochrome P450 family protein.	100.0	98.8	SSF48264	PF00067	GO:0005506 GO:0016705 GO:0020037 GO:0055114	Cytochrome P450	05c4595g01	23 170 991- 23 172 550	+	yes	1	520	Os05t044 5100-01	RAP- DB	Cytochrome P450 family protein.	100.0	99.0	SSF48264	PF00067	GO:0005506 GO:0016705 GO:0020037 GO:0055114	Cytochrome P450	99.8	
05g05803	23 178 063- 23 183 559	-	6	185	No hits	-	-	-	-	-	-	-	Unknown	05c4595g02	23 178 063- 23 178 273	-	yes	2	42	Os05t044 5200-00	RAP- DB	Hypothetical protein	93.0	64.4	-	-	-	Hypothetical	22.2	
05g05804	23 186 483- 23 187 370	-	2	150	BAC65029. 1	NCBI	hypothetical protein [Os jap.]	71.0	72.6	-	-	-	Hypothetical	05c4598g01	23 186 483- 23 187 370	-	yes	2	151	Os08t025 9550-00	RAP- DB	Hypothetical protein	71.0	72.6	-	-	-	Hypothetical	100.0	
05g05805	23 189 562- 23 191 574	+	4	180	Os05t044 5500-01	RAP- DB	Similar to Acidic ribosomal protein (Fragment).	73.0	94.2	-	-	PF00428	-	Translation	05c4600g01	23 189 562- 23 191 574	+	yes	4	181	LOC_Os05 g37330.1	MSU	60S acidic ribosomal protein, putative, expressed	63.0	98.2	-	PF00428	-	Translation	100.0
05g05806	23 192 936- 23 196 546	+	6	938	Os05t044 5800-00	RAP- DB	Conserved hypothetical protein.	76.0	99.9	-	-	-	Hypothetical																	
05g05807	23 198 087- 23 210 804	-	19	1918	Os05t044 5900-02	RAP- DB	5-methylcytosine DNA glycosylase/lyase	100.0	94.1	-	-	PF15628	-	DNA processes	05c4606g01	23 200 370- 23 201 426	-	yes	4	197	Os05t044 5900-02	RAP- DB	5-methylcytosine DNA glycosylase/lyase	93.0	88.0	-	-	-	DNA processes	54.3
														05c4607g01	23 202 822- 23 208 599	-	yes	7	859	Os05t044 5900-02	RAP- DB	5-methylcytosine DNA glycosylase/lyase	100.0	99.5	SSF48150	-	GO:0003824 GO:0006281	DNA processes		
05g05808	23 212 514- 23 213 807	-	2	155	Os01t097 5000-02	RAP- DB	Protein of unknown function DUF966 family protein.	63.0	54.9	-	-	-	Domain-containing	05c4610g02	23 212 514- 23 217 064	-	yes	5	223	No hits	-	-	-	-	-	-	-	Unknown	60.6	
05g05810	23 220 128- 23 224 466	+	6	299	TVU19730. 1	NCBI	hypothetical protein EJ805_35899, partial [Eragrostis curvula]	57.0	77.1	-	-	PF01125	GO:0005634	Hypothetical	05c4611g01	23 220 128- 23 224 466	+	yes	6	300	TVU19730. 1	NCBI	hypothetical protein EJ805_35899, partial [Eragrostis curvula]	57.0	77.1	-	PF01125	GO:0005634	Hypothetical	100.0
05g05811	23 225 196- 23 225 771	-	1	191	OslR64_0 5g026100	EuGe ne	Conserved hypothetical protein	-	-	-	-	-	Hypothetical	05c4611g02	23 225 196- 23 225 771	-	yes	1	192	No hits	-	-	-	-	-	-	Unknown	99.0		
05g05812	23 226 934- 23 233 083	+	3	1952	LOC_Os04 g09840.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	94.1	SSF53098	PF03732	SSF56672	PF13456 PF00665 PF00078 PF17917	TEs																
05g05813	23 233 712- 23 235 340	-	1	542	LOC_Os06 g32800.1	MSU	transposon protein, putative, unclassified, expressed	100.0	87.1	-	-	-	TEs																	
05g05815	23 238 183- 23 239 876	-	2	532	LOC_Os07 g31060.1	MSU	retrotransposon protein, putative, unclassified, expressed	98.0	87.8	-	-	PF04195	-	TEs																
05g05816	23 241 044- 23 244 935	+	5	1003	KAF29310 34.1	NCBI	hypothetical protein DAI22_05g179300 [Os jap.]	81.0	95.7	-	-	-	Hypothetical	05c4620g01	23 242 764- 23 244 935	+	yes	3	582	EEC79310. 1	NCBI	hypothetical protein Osl_20145 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	57.7	
05g05817	23 246 321- 23 256 945	-	20	1870	LOC_Os05 g37410.1	MSU	hhH-GPD superfamily base excision DNA repair protein, expressed	100.0	92.6	-	-	PF15628	-	DNA processes	05c4624g01	23 254 694- 23 256 945	-	no	4	623	EAY98235. 1	NCBI	hypothetical protein Osl_20146 [Os ind.]	100.0	84.3	-	-	-	Hypothetical	31.5
05g05818	23 259 352- 23 259 888	-	1	178	BAS94287. 1	NCBI	Os05g0446700 [Os jap.]	100.0	98.9	-	-	-	Hypothetical																	

Supplementary Table S4.2 (continued)

IR64														CT8556-37-2-3-1-M																	
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Sim.		
05g05819	23 262 135- 23 266 587	-	5	614	Os05t044 6800-03	RAP- DB	Similar to Arginyl-tRNA--protein transferase 1 (EC 2.3.2.8) (R-transferase 1) (Arginyltransferase 1) (Arginine-tRNA--protein transferase 1). Splice isoform ATE1-2.	91.0	99.8	-	PF04376 PF04377	GO:0004057 GO:0016598 GO:0004057 GO:0016598	Protein degradation	05c4624g04	23 262 788- 23 266 587	-	yes	5	712	Os05t044 6800-03	RAP- DB	Similar to Arginyl-tRNA--protein transferase 1 (EC 2.3.2.8) (R-transferase 1) (Arginyltransferase 1) (Arginine-tRNA--protein transferase 1). Splice isoform ATE1-2.	79.0	99.8	-	PF04377 PF04376	GO:0004057 GO:0016598 GO:0004057 GO:0016598	Protein degradation	79.5		
05g05821	23 270 469- 23 271 273	-	2	93	Os05t044 6800-03	RAP- DB	Similar to Arginyl-tRNA--protein transferase 1 (EC 2.3.2.8) (R-transferase 1) (Arginyltransferase 1) (Arginine-tRNA--protein transferase 1). Splice isoform ATE1-2.	61.0	87.7	-	-	-	Protein degradation																		
05g05822	23 275 129- 23 277 093	-	5	373	Os05t044 6900-01	RAP- DB	Starch-binding domain-containing protein, Starch biosynthesis	100.0	99.5	-	PF00686	GO:2001070	Biosynthetic processes	05c4625g01	23 275 129- 23 276 961	-	yes	5	330	Os05t044 6900-01	RAP- DB	Starch-binding domain-containing protein, Starch biosynthesis	100.0	99.4	-	PF00686	GO:2001070	Biosynthetic processes	88.2		
05g05823	23 280 280- 23 280 798	-	1	172	Os05t044 7000-01	RAP- DB	Similar to Pleckstrin homology domain-containing protein 1 (AtPH1).	100.0	99.4	SSF50729	PF00169	-	Domain- containing	05c4626g01	23 278 591- 23 280 565	-	yes	2	139	EEC79313. 1	NCBI	hypothetical protein Osl_20149 [Os ind.]	68.0	100.0	SSF50729	PF00169	-	Domain- containing	43.5		
05g05825	23 287 132- 23 294 709	-	9	834	Os05t044 7200-02	RAP- DB	Similar to AUX1-like protein.	60.0	99.8	-	PF01490	-	Hormone- related	05c4630g01	23 287 132- 23 291 370	-	yes	8	607	Os05t044 7200-02	RAP- DB	Similar to AUX1-like protein.	83.0	99.8	-	PF01490	-	Hormone- related	72.7		
05g05826	23 299 275- 23 299 979	+	2	139	LOC_Os05 g06880.1	MSU	transposon protein, putative, unclassified	58.0	58.8	-	-	-	TEs																		
05g05827	23 300 496- 23 303 737	+	2	1047	LOC_Os11 g47200.1	MSU	transposon protein, putative, unclassified, expressed	100.0	93.0	-	PF04195	-	TEs																		
05g05828	23 304 376- 23 310 528	-	4	1978	LOC_Os01 g45810.1	MSU	transposon protein, putative, unclassified, expressed	100.0	94.0	SSF53098	PF13456 SSF56672	-	TEs																		
05g05829	23 311 075- 23 312 051	+	3	108	BAD62524 .1	NCBI	hypothetical protein [Os jap.]	52.0	89.3	-	-	-	Hypothetical																		
05g05830	23 312 597- 23 312 941	+	1	114	OsIR64_0 5g026180	EuGe ne	Conserved hypothetical protein	-	-	-	-	-	Hypothetical																		
05g05831	23 315 700- 23 326 077	-	14	1419	XP_01563 9680.1	NCBI	mediator of RNA polymerase II transcription subunit 13 [Os jap.]	96.0	99.6	-	PF18296	-	Transcription	05c4635g01	23 316 904- 23 330 240	-	yes	21	1739	LOC_Os05 g37500.1	MSU	expressed protein	99.0	99.7	-	PF18296	-	Transcription	90.7		
05g05832	23 326 612- 23 334 281	-	15	633	EEC79315. 1	NCBI	hypothetical protein Osl_20152 [Os ind.]	65.0	100.0	-	PF11597	-	Transcription																		
05g05833	23 334 936- 23 335 834	+	2	174	BAS94301. 1	NCBI	Os05g0447680 [Os jap.]	59.0	99.0	-	-	-	Hypothetical																		
05g05834	23 337 422- 23 340 461	-	4	601	Os05t044 7700-01	RAP- DB	Tetratricopeptide-like helical domain containing protein.	80.0	98.1	SSF48452	-	GO:0005515	Domain- containing	05c4637g01	23 337 422- 23 339 194	-	yes	3	402	Os05t044 7700-01	RAP- DB	Tetratricopeptide-like helical domain containing protein.	88.0	97.5	-	PF14559	-	Domain- containing	66.7		
05g05835	23 343 723- 23 344 301	+	1	192	Os05t044 7800-00	RAP- DB	Similar to H0212B02.6 protein.	99.0	96.8	-	PF04669	-	Cell wall biogen./mod.	05c4638g03	23 343 723- 23 344 124	+	yes	1	134	Os05t044 7800-00	RAP- DB	Similar to H0212B02.6 protein.	99.0	98.5	-	-	-	Cell wall biogen./mod.	67.7		
05g05836	23 345 066- 23 345 305	-	1	79	LOC_Os04 g11160.1	MSU	retrotransposon protein, putative, Ty1-copia subclass, expressed	72.0	96.5	-	-	-	TEs																		
05g05839	23 352 176- 23 354 597	+	2	66	LOC_Os12 g38820.1	MSU	expressed protein	91.0	90.0	-	-	-	Hypothetical																		
05g05840	23 355 426- 23 355 939	-	2	154	EAY98244. 1	NCBI	hypothetical protein Osl_20154 [Os ind.]	75.0	100.0	-	-	-	Hypothetical	05c4641g01	23 355 426- 23 355 905	-	yes	1	160	EAY98244. 1	NCBI	hypothetical protein Osl_20154 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	71.6		
05g05842	23 359 433- 23 362 750	+	2	486	Os05t044 8300-01	RAP- DB	Similar to glycerol-3-phosphate acyltransferase 8.	100.0	99.8	SSF69593	PF01553	-	Biosynthetic processes																		
05g05843	23 365 144- 23 370 060	-	6	428	LOC_Os12 g26400.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	76.0	-	-	-	TEs																		
05g05844	23 372 864- 23 373 565	-	1	233	LOC_Os05 g37630.1	MSU	expressed protein	100.0	96.2	-	-	-	Hypothetical	05c4653g01	23 372 864- 23 373 382	-	yes	1	173	Os05t044 8650-00	RAP- DB	Hypothetical protein	100.0	96.5	-	-	-	Hypothetical	72.1		

Supplementary Table S4.2 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	
05g05846	23 379 027- 23 380 001	+	1	324	XP_01563 8626.1	NCBI	ethylene-responsive transcription factor ERF094 [Os jap.]	100.0	100.0	-	PF00847	GO:0003700 GO:0006355	Transcription	05c4653g03	23 379 027- 23 379 287	+	no	1	87	LOC_Os05 g37640.1	MSU	expressed protein	100.0	100.0	-	-	-	-	Hypothetical	26.9
														05c4655g02	23 382 992- 23 384 998	-	yes	3	106	EEC70779. 1	NCBI	hypothetical protein Osl_02213 [Os ind.]	73.0	53.8	-	-	-	-	Hypothetical	
05g05849	23 386 070- 23 389 455	-	4	334	No hits	-	-	-	-	-	-	-	Unknown																	
05g05850	23 390 211- 23 391 653	-	1	480	Os05t044 9200-01	RAP- DB	Transferase family protein.	100.0	99.4	SSF52777	-	-	Transferases	05c4655g03	23 386 070- 23 391 203	-	yes	6	582	EEC79316. 1	NCBI	hypothetical protein Osl_20158 [Os ind.]	56.0	99.7	SSF52777	-	-	-	Transferases	71.3
05g05851	23 391 941- 23 392 306	+	1	121	BAS94313. 1	NCBI	Os05g0449300 [Os jap.]	100.0	99.2	-	-	-	Hypothetical																	
05g05854	23 397 207- 23 397 767	+	1	186	LOC_Os05 g37670.1	MSU	retrotransposon protein, putative, unclassified, expressed	93.0	95.4	-	-	-	TEs	05c4658g02	23 397 207- 23 397 536	+	yes	1	110	LOC_Os05 g37670.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	98.2	-	-	-	-	TEs	58.1
05g05855	23 398 651- 23 399 361	-	2	208	LOC_Os05 g37680.1	MSU	transposon protein, putative, unclassified, expressed	75.0	94.9	-	-	-	TEs	05c4659g02	23 398 651- 23 399 348	-	yes	2	203	LOC_Os05 g37680.1	MSU	transposon protein, putative, unclassified, expressed	100.0	96.1	-	-	-	-	TEs	92.8
05g05858	23 411 759- 23 414 060	+	3	597	Os05t044 9500-01	RAP- DB	Component of the SCF E3 ubiquitin ligase complex, Jasmonate-regulated defense responses, Regulation of leaf senescence	100.0	99.5	SSF52047 SSF81383	PF18791 PF18511	-	Stress-related	05c4660g03	23 411 759- 23 414 060	+	yes	3	598	Os05t044 9500-01	RAP- DB	Component of the SCF E3 ubiquitin ligase complex, Jasmonate-regulated defense responses, Regulation of leaf senescence	100.0	99.5	SSF52047 SSF81383	PF18511 PF18791	-	-	Stress-related	100.0
05g05860	23 419 174- 23 423 216	+	10	694	LOC_Os05 g37700.1	MSU	periplasmic beta-glucosidase precursor, putative, expressed	87.0	99.7	SSF52279 SSF51445	-	-	Hormone- related	05c4660g05	23 419 174- 23 423 216	+	yes	11	669	LOC_Os05 g37700.1	MSU	periplasmic beta-glucosidase precursor, putative, expressed	87.0	95.5	-	PF01915 PF00933	GO:0004553 GO:0005975 related	Hormone- related	96.1	
05g05861	23 425 886- 23 426 284	-	1	132	EEE63919. 1	NCBI	hypothetical protein Osl_18744 [Os jap.]	100.0	98.5	-	-	-	Hypothetical	05c4660g06	23 425 886- 23 426 284	-	yes	1	133	EEE63919. 1	NCBI	hypothetical protein Osl_18744 [Os jap.]	100.0	98.5	-	-	-	-	Hypothetical	100.0
05g05862	23 427 402- 23 429 807	-	3	173	EEC79318. 1	NCBI	hypothetical protein Osl_20161 [Os ind.]	90.0	97.4	-	-	-	Hypothetical	05c4660g07	23 427 402- 23 429 807	-	yes	3	203	EEC79318. 1	NCBI	hypothetical protein Osl_20161 [Os ind.]	77.0	98.1	-	-	-	-	Hypothetical	82.2
05g05863	23 431 324- 23 431 716	-	1	130	EEC79318. 1	NCBI	hypothetical protein Osl_20161 [Os ind.]	82.0	96.2	-	-	-	Hypothetical																	
05g05864	23 432 058- 23 432 453	-	1	131	EEC79318. 1	NCBI	hypothetical protein Osl_20161 [Os ind.]	95.0	99.2	-	-	-	Hypothetical	05c4662g01	23 431 569- 23 432 655	-	yes	3	210	EEC79318. 1	NCBI	hypothetical protein Osl_20161 [Os ind.]	86.0	100.0	-	-	-	-	Hypothetical	59.3
05g05865	23 433 595- 23 434 644	-	2	314	LOC_Os05 g37730.1	MSU	MYB family transcription factor, putative, expressed	100.0	99.4	SSF46689	PF00249	-	Transcription	05c4662g02	23 433 595- 23 434 644	-	yes	2	310	LOC_Os05 g37730.1	MSU	MYB family transcription factor, putative, expressed	100.0	97.8	SSF46689	PF00249	-	-	Transcription	98.4
05g05866	23 443 216- 23 446 167	+	3	111	AAU90114. .1	NCBI	hypothetical protein [Os jap.]	73.0	96.3	-	-	-	Hypothetical	05c4662g03	23 443 216- 23 446 167	+	yes	3	112	AAU90114 .1	NCBI	hypothetical protein [Os jap.]	73.0	97.5	-	-	-	-	Hypothetical	99.1
05g05867	23 449 471- 23 449 803	+	1	110	LOC_Os05 g37740.1	MSU	retrotransposon protein, putative, unclassified	100.0	96.4	-	PF04195	-	TEs																	
05g05868	23 450 092- 23 451 023	+	2	223	LOC_Os05 g31120.1	MSU	retrotransposon protein, putative, unclassified, expressed	68.0	88.1	-	-	-	TEs																	
05g05870	23 455 134- 23 456 078	+	1	314	CAH68017. .1	NCBI	H0807C06-H0308C08.4 [Oryza sativa]	100.0	96.8	-	-	-	Hypothetical																	
05g05871	23 456 330- 23 460 895	+	2	1460	LOC_Os05 g07520.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	93.8	SSF56672 SSF53098 SSF54160 SSF57756 SSF50630	PF00098 PF08284 PF00078 PF17921 PF17919 PF03732	-	TEs																	
05g05872	23 461 243- 23 462 503	-	2	264	BAD72332. .1	NCBI	plant disease resistance polyprotein-like [Os jap.]	54.0	79.0	-	-	-	Stress-related																	
05g05873	23 463 750- 23 468 567	-	7	1332	BAC05657. 1	NCBI	putative retrotransposon Cinful-1 [Os jap.]	90.0	75.2	SSF56672	PF17917 PF00078	-	TEs	05c4668g01	23 467 365- 23 468 567	-	no	3	327	LOC_Os05 g37760.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	71.8	-	-	-	-	TEs	24.0
05g05874	23 472 076- 23 472 447	-	1	123	BAS94320. 1	NCBI	Os05g0450200 [Os jap.]	100.0	96.7	-	-	-	Hypothetical	05c4668g02	23 472 076- 23 472 447	-	yes	1	124	Os05t045 0200-00	RAP- DB	Hypothetical protein	100.0	96.7	-	-	-	-	Hypothetical	100.0

Supplementary Table S4.2 (continued)

IR64											CT8556-37-2-3-1-M											Sim.							
Gene ID	Pos. (bp)	S.	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation		Cov.	Iden.	SSF hits	PF hits	GO terms	Group	
05g05875	23 481 424- 23 490 195	+	9	737	OslR64_0 5g026350	EuGe ne	Putative expressed protein	—	—	—	—	Hypothetical	05c4670g01	23 481 424- 23 481 831	+	yes	1	136	KAF09000	NCBI	hypothetical protein E2562_026798 [Oryza meyeriana var. granulata]	72.0	52.3	—	—	—	Hypothetical	35.6	
													05c4671g01	23 484 924- 23 486 513	+	yes	3	182	No hits	—	—	—	—	—	—	Unknown			
05g05877	23 499 537- 23 501 522	+	1	661	XP_02588 1091.1	NCBI	UPF0503 protein At3g09070, chloroplastic [Os jap.]	100.0	99.4	—	PF05340	—	Growth and development	05c4673g01	23 501 031- 23 501 481	+	yes	1	149	XP_00665	NCBI	PREDICTED: UPF0503 protein At3g09070, chloroplastic-like [Oryza brachyantha]	96.0	83.6	—	PF05340	—	Growth and development	19.7
05g05879	23 507 055- 23 507 726	-	2	172	No hits	—	—	—	—	—	—	Unknown	05c4673g03	23 507 055- 23 507 726	-	yes	2	158	KAF29471	NCBI	hypothetical protein DAI22_02g347700 [Os jap.]	50.0	89.9	—	—	—	Hypothetical	90.1	
05g05881	23 511 581- 23 513 713	-	3	350	LOC_Os01 g21620.1	MSU	expressed protein	57.0	51.8	SSF53901	—	GO:0016746	Hypothetical																
05g05882	23 515 940- 23 519 202	+	7	520	LOC_Os05 g37820.1	MSU	transporter, major facilitator family, putative, expressed	100.0	96.1	—	PF07690	GO:0022857 GO:0055085	Transporters	05c4678g01	23 516 109- 23 519 202	+	yes	7	434	Os05t045 1100-01	RAP- DB	Major facilitator superfamily protein.	93.0	96.4	—	PF07690	GO:0022857 GO:0055085	Transporters	79.2
05g05883	23 519 423- 23 519 617	-	1	64	OslR64_0 5g026420	EuGe ne	Conserved hypothetical protein	—	—	—	—	—	Hypothetical	05c4678g02	23 519 440- 23 519 620	-	yes	1	46	No hits	—	—	—	—	—	—	Unknown	44.8	
05g05884	23 520 214- 23 521 907	-	4	166	EAY98259 1	NCBI	hypothetical protein Osl_20166 [Os ind.]	100.0	100.0	—	—	—	Hypothetical	05c4678g03	23 520 214- 23 521 907	-	yes	4	167	EAY98259	NCBI	hypothetical protein Osl_20166 [Os ind.]	100.0	100.0	—	—	—	Hypothetical	100.0
05g05886	23 525 005- 23 527 456	-	3	63	EAY98260 1	NCBI	hypothetical protein Osl_20167 [Os ind.]	62.0	92.3	—	—	—	Hypothetical	05c4679g02	23 526 909- 23 527 456	-	yes	3	40	EAY98260	NCBI	hypothetical protein Osl_20167 [Os ind.]	92.0	100.0	—	—	—	Hypothetical	55.6
05g05887	23 527 714- 23 530 630	-	3	356	LOC_Os05 g37860.1	MSU	expressed protein	57.0	92.2	—	—	—	Hypothetical																
05g05889	23 532 064- 23 536 506	-	4	632	AAV59353 1	NCBI	unknown protein [Os jap.]	69.0	90.5	—	PF03108	—	TEs	05c4682g01	23 531 984- 23 536 506	-	yes	5	648	LOC_Os02 g26630.1	MSU	transposon protein, putative, unclassified, expressed	65.0	92.1	—	PF03108	—	TEs	73.3
													05c4683g01	23 538 297- 23 539 421	-	yes	3	80	Os05t045 2151-00	RAP- DB	Hypothetical protein.	65.0	54.2	—	—	—	Hypothetical		
05g05890	23 537 530- 23 541 362	+	7	592	AAV59354 1	NCBI	putative auxin-independent growth promoter [Os jap.]	96.0	99.3	—	PF10250	—	Hormone- related	05c4683g02	23 540 194- 23 542 220	+	yes	5	387	AAV59354	NCBI	putative auxin-independent growth promoter [Os jap.]	100.0	99.7	—	PF10250	—	Hormone- related	32.3
													05c4683g03	23 543 103- 23 546 804	+	yes	1	103	BAD30778 .1	NCBI	hypothetical protein [Os jap.]	69.0	70.4	—	—	—	Hypothetical		
05g05893	23 548 135- 23 549 222	+	2	273	KAF29310 66.1	NCBI	hypothetical protein DAI22_05g182401 [Os jap.]	74.0	98.5	—	—	—	Hypothetical	05c4687g01	23 553 844- 23 556 289	-	yes	3	55	AAT47030	NCBI	unknown protein [Os jap.]	76.0	95.1	—	—	—	Hypothetical	
05g05896	23 565 484- 23 568 100	+	2	244	Os05t045 2800-01	RAP- DB	Zinc finger, RING/FYVE/PHD- type domain containing protein.	100.0	100.0	—	PF12906	GO:0008270	Domain- containing	05c4687g02	23 561 700- 23 568 100	+	yes	4	311	Os05t045 2800-01	RAP- DB	Zinc finger, RING/FYVE/PHD- type domain containing protein.	73.0	97.8	SSF57850	—	—	Domain- containing	70.6
05g05897	23 570 681- 23 575 643	-	11	397	Os05t045 2900-02	RAP- DB	Similar to Phosphatidic acid phosphatase-like protein.	69.0	83.6	—	PF14360	—	Biosynthetic processes	05c4687g04	23 570 681- 23 575 643	-	yes	11	398	Os05t045 2900-02	RAP- DB	Similar to Phosphatidic acid phosphatase-like protein.	69.0	84.0	—	PF14360	—	Intracellular transport	99.5
05g05898	23 580 228- 23 585 173	-	7	469	LOC_Os06 g19150.1	MSU	retrotransposon protein, putative, unclassified, expressed	59.0	91.6	—	—	—	TEs	05c4688g01	23 580 161- 23 580 500	-	yes	1	113	EEC79324	NCBI	hypothetical protein Osl_20173 [Os ind.]	100.0	100.0	—	—	—	Hypothetical	18.1
05g05899	23 586 838- 23 587 262	-	2	123	EEC79325 1	NCBI	hypothetical protein Osl_20174 [Os ind.]	72.0	100.0	—	—	—	Hypothetical	05c4689g01	23 586 838- 23 587 996	-	yes	2	231	No hits	—	—	—	—	—	Unknown	64.7		
05g05900	23 587 744- 23 587 992	-	1	82	No hits	—	—	—	—	—	—	Unknown																	
05g05902	23 590 956- 23 591 954	+	2	210	AAT07605 1	NCBI	hypothetical protein [Os jap.]	98.0	60.2	—	—	—	Hypothetical																
05g05904	23 598 470- 23 599 186	+	1	238	OslR64_0 5g026550	EuGe ne	Conserved hypothetical protein	—	—	—	—	—	Hypothetical	05c4693g02	23 598 470- 23 599 153	+	yes	2	169	No hits	—	—	—	—	—	—	Unknown	61.8	
05g05905	23 600 291- 23 601 745	-	5	368	EAY98268 1	NCBI	hypothetical protein Osl_20175 [Os ind.]	100.0	99.7	—	—	—	Hypothetical	05c4693g03	23 600 291- 23 600 798	-	yes	3	106	LOC_Os05 g37930.2	MSU	expressed protein	100.0	100.0	—	—	—	Hypothetical	28.5
													05c4696g02	23 607 857- 23 610 425	-	yes	2	131	LOC_Os03 g23060.1	MSU	transposon protein, putative, unclassified, expressed	52.0	73.1	—	—	—	TEs		
05g05908	23 610 293- 23 610 756	+	3	98	LOC_Os05 g37940.1	MSU	transposon protein, putative, unclassified, expressed	100.0	93.9	—	—	—	TEs																
05g05913	23 623 387- 23 626 626	-	8	248	OEL36015 1	NCBI	hypothetical protein BAE44_0002966 [Dichantheium oligosanthes]	79.0	68.2	—	PF09778	—	Guanylylate cyclase	05c4698g02	23 623 387- 23 626 272	-	yes	7	204	LOC_Os05 g37950.1	MSU	guanylyl cyclase, putative, expressed	62.0	98.8	—	PF09778	—	Guanylylate cyclase	81.9

Supplementary Table S4.2 (continued)

IR64													CT8556-37-2-3-1-M																
Gene ID	Pos. (bp)	S.	#	AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#	AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.
05g05914	23 631 877- 23 633 284	-	3	147	Os011076 4800-01	RAP- DB	Indole-3-acetic acid (IAA)-amido synthetase, Disease resistance, Abiotic stress tolerance	57.0	85.7	-	PF03321	-	Hormone- related	05c4699g01	23 631 877- 23 633 284	-	yes	3	191	PQM3630 5.1	NCBI	putative indole-3-acetic acid- amido synthetase GH3.1 [Prunus yedoensis var. nudiflora]	62.0	50.4	-	PF03321	-	Hormone- related	52.6
05g05915	23 633 797- 23 635 205	-	2	49	LOC_Os02 g24370.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass	88.0	83.7	-	-	-	TEs	05c4699g02	23 636 399- 23 638 500	+	yes	4	172	LOC_Os05 g37970.2	MSU	universal stress protein domain containing protein, putative, expressed	100.0	100.0	SSF52402	PF00582	-	Stress-related	100.0
05g05916	23 636 399- 23 638 500	+	4	171	LOC_Os05 g37970.2	MSU	universal stress protein domain containing protein, putative, expressed	100.0	100.0	SSF52402	PF00582	-	Stress-related	05c4699g03	23 639 779- 23 640 120	+	yes	1	114	BAH93182 .1	NCBI	Os05g0453800 [Os jap.]	55.0	93.5	-	-	-	DNA processes	98.2
05g05917	23 639 779- 23 640 120	+	1	113	BAH93182 .1	NCBI	Os05g0453800 [Os jap.]	55.0	95.2	-	-	-	DNA processes	05c4700g01	23 642 970- 23 646 059	+	yes	2	128	No hits	-	-	-	-	-	-	Unknown	56.4	
05g05918	23 641 671- 23 646 059	+	7	225	Os05t045 3900-00	RAP- DB	Similar to DNA replication complex GINS protein PSF1.	60.0	84.1	SSF15857 3	-	-	DNA processes	05c4582g01	23 040 464- 23 041 533	-	yes	2	49	No hits	-	-	-	-	-	-	-	Unknown	89.6
05g05764	23 028 388- 23 029 815	-	2	310	No hits	-	-	-	-	-	-	-	Unknown	05c4582g02	23 044 079- 23 045 887	-	yes	3	74	No hits	-	-	-	-	-	-	-	Unknown	97.3
05g05768	23 040 614- 23 041 533	-	2	48	No hits	-	-	-	-	-	-	-	Unknown	05c4584g03	23 069 592- 23 069 953	+	yes	2	72	No hits	-	-	-	-	-	-	-	Unknown	61.1
05g05769	23 044 079- 23 045 887	-	3	73	No hits	-	-	-	-	-	-	-	Unknown	05c4590g01	23 126 723- 23 129 662	-	yes	4	73	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05775	23 069 592- 23 070 123	+	3	108	No hits	-	-	-	-	-	-	-	Unknown	05c4590g03	23 130 845- 23 133 895	+	yes	3	86	No hits	-	-	-	-	-	-	-	Unknown	52.4
05g05782	23 106 687- 23 107 495	-	2	245	No hits	-	-	-	-	-	-	-	Unknown	05c4590g05	23 142 443- 23 142 881	-	yes	1	146	No hits	-	-	-	-	-	-	-	Unknown	93.6
05g05788	23 126 723- 23 129 662	-	4	72	No hits	-	-	-	-	-	-	-	Unknown	05c4591g02	23 149 144- 23 149 661	-	yes	2	47	No hits	-	-	-	-	-	-	-	Unknown	99.0
05g05790	23 130 845- 23 135 421	+	4	105	No hits	-	-	-	-	-	-	-	Unknown	05c4592g02	23 156 236- 23 156 541	-	yes	1	102	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05792	23 142 414- 23 142 881	-	1	155	No hits	-	-	-	-	-	-	-	Unknown	05c4593g01	23 158 389- 23 158 685	+	yes	2	28	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05797	23 156 236- 23 156 541	-	1	101	No hits	-	-	-	-	-	-	-	Unknown	05c4593g02	23 161 417- 23 161 939	-	yes	2	22	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05798	23 156 946- 23 157 767	+	2	109	No hits	-	-	-	-	-	-	-	Unknown	05c4600g02	23 192 936- 23 193 145	+	yes	1	70	No hits	-	-	-	-	-	-	-	Unknown	99.0
05g05799	23 158 389- 23 158 685	+	2	27	No hits	-	-	-	-	-	-	-	Unknown	05c4610g01	23 210 513- 23 210 740	+	yes	1	76	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05800	23 161 417- 23 161 939	-	2	21	No hits	-	-	-	-	-	-	-	Unknown	05c4624g02	23 258 970- 23 259 725	-	yes	3	69	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05801	23 168 551- 23 168 964	+	1	137	No hits	-	-	-	-	-	-	-	Unknown	05c4624g03	23 260 981- 23 262 275	-	yes	4	93	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05809	23 217 601- 23 218 609	+	2	155	No hits	-	-	-	-	-	-	-	Unknown	05c4624g05	23 268 015- 23 269 187	+	yes	2	139	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05814	23 236 753- 23 237 868	-	3	189	No hits	-	-	-	-	-	-	-	Unknown	05c4624g06	23 269 863- 23 270 594	+	yes	2	37	No hits	-	-	-	-	-	-	-	Unknown	100.0

Supplementary Table S4.2 (continued)

IR64												CT8556-37-2-3-1-M												Sim.					
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.		Iden.	SSF hits	PF hits	GO terms	Group
05g05824	23 283 020- 23 285 126	-	2	267	No hits	-	-	-	-	-	-	-	Unknown	05c4638g01	23 339 287- 23 340 461	-	yes	2	105	No hits	-	-	-	-	-	-	-	Unknown	
														05c4638g02	23 342 305- 23 343 030	-	yes	3	98	No hits	-	-	-	-	-	-	-	Unknown	
05g05837	23 346 117- 23 346 368	+	1	83	No hits	-	-	-	-	-	-	-	Unknown																
05g05838	23 346 842- 23 347 898	+	2	235	No hits	-	-	-	-	-	-	-	Unknown																
														05c4639g01	23 347 562- 23 347 852	-	yes	1	97	No hits	-	-	-	-	-	-	-	Unknown	
05g05841	23 358 062- 23 358 397	+	1	111	No hits	-	-	-	-	-	-	-	Unknown																
05g05845	23 375 893- 23 378 052	+	2	75	No hits	-	-	-	-	-	-	-	Unknown																
														05c4653g02	23 375 936- 23 378 156	+	yes	2	57	No hits	-	-	-	-	-	-	-	Unknown	
05g05847	23 380 721- 23 381 059	-	1	112	No hits	-	-	-	-	-	-	-	Unknown																
														05c4655g01	23 380 914- 23 381 852	+	yes	2	65	No hits	-	-	-	-	-	-	-	Unknown	
05g05848	23 382 055- 23 384 210	+	3	221	No hits	-	-	-	-	-	-	-	Unknown																
05g05852	23 393 601- 23 394 562	+	3	170	No hits	-	-	-	-	-	-	-	Unknown																
05g05853	23 395 111- 23 396 167	+	2	123	No hits	-	-	-	-	-	-	-	Unknown	05c4658g01	23 395 111- 23 396 167	+	yes	2	124	No hits	-	-	-	-	-	-	-	Unknown	99.2
														05c4659g01	23 397 858- 23 398 273	+	yes	2	98	No hits	-	-	-	-	-	-	-	Unknown	
05g05856	23 401 021- 23 403 986	+	2	99	No hits	-	-	-	-	-	-	-	Unknown																
														05c4660g01	23 404 569- 23 405 367	+	yes	2	62	No hits	-	-	-	-	-	-	-	Unknown	
05g05857	23 406 409- 23 410 592	+	2	63	No hits	-	-	-	-	-	-	-	Unknown	05c4660g02	23 406 409- 23 410 592	+	yes	2	64	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05859	23 414 799- 23 417 104	+	2	68	No hits	-	-	-	-	-	-	-	Unknown	05c4660g04	23 414 799- 23 418 213	+	yes	3	124	No hits	-	-	-	-	-	-	-	Unknown	40.7
05g05869	23 451 602- 23 452 966	+	3	335	No hits	-	-	-	-	-	-	-	Unknown																
														05c4670g02	23 482 104- 23 482 537	-	yes	2	128	No hits	-	-	-	-	-	-	-	Unknown	
														05c4671g02	23 488 839- 23 490 195	+	yes	2	108	No hits	-	-	-	-	-	-	-	Unknown	
05g05876	23 492 602- 23 492 868	-	1	88	No hits	-	-	-	-	-	-	-	Unknown	05c4671g03	23 492 602- 23 492 868	-	yes	1	89	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05878	23 501 953- 23 502 267	-	1	104	No hits	-	-	-	-	-	-	-	Unknown	05c4673g02	23 501 953- 23 502 267	-	yes	1	105	No hits	-	-	-	-	-	-	-	Unknown	99.0
05g05880	23 508 052- 23 510 454	+	4	419	No hits	-	-	-	-	-	-	-	Unknown	05c4674g01	23 509 612- 23 510 499	+	yes	2	257	No hits	-	-	-	-	-	-	-	Unknown	51.8
05g05885	23 522 677- 23 524 200	-	4	250	No hits	-	-	-	-	-	-	-	Unknown	05c4678g04	23 522 677- 23 523 140	-	yes	2	107	No hits	-	-	-	-	-	-	-	Unknown	41.6
														05c4679g01	23 524 079- 23 525 466	+	yes	3	60	No hits	-	-	-	-	-	-	-	Unknown	
05g05888	23 531 291- 23 531 617	+	1	108	No hits	-	-	-	-	-	-	-	Unknown																
05g05891	23 541 816- 23 544 186	-	4	339	No hits	-	-	-	-	-	-	-	Unknown																

Supplementary Table S4.2 (continued)

IR64													CT8556-37-2-3-1-M													Sim.			
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits		PF hits	GO terms	Group
05g05892	23 544 987- 23 546 828	-	2	450	No hits	-	-	-	-	-	-	-	Unknown																
05g05894	23 550 530- 23 559 156	+	4	75	No hits	-	-	-	-	-	-	-	Unknown																
05g05895	23 561 700- 23 562 289	+	2	69	No hits	-	-	-	-	-	-	-	Unknown																
														05c4687g03	23 569 112- 23 569 778	+	yes	2	55	No hits	-	-	-	-	-	-	-	Unknown	
05g05901	23 588 337- 23 588 858	-	2	72	No hits	-	-	-	-	-	-	-	Unknown	05c4689g02	23 588 337- 23 588 858	-	yes	2	73	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05903	23 594 309- 23 595 707	-	2	134	No hits	-	-	-	-	-	-	-	Unknown																
														05c4693g01	23 594 908- 23 596 293	-	yes	2	82	No hits	-	-	-	-	-	-	-	Unknown	
05g05906	23 602 981- 23 606 699	-	3	340	No hits	-	-	-	-	-	-	-	Unknown																
														05c4696g01	23 603 348- 23 606 648	+	yes	3	184	No hits	-	-	-	-	-	-	-	Unknown	
05g05907	23 607 481- 23 608 488	+	3	32	No hits	-	-	-	-	-	-	-	Unknown																
05g05909	23 613 722- 23 613 964	+	1	80	No hits	-	-	-	-	-	-	-	Unknown																
05g05910	23 614 405- 23 614 656	-	1	83	No hits	-	-	-	-	-	-	-	Unknown	05c4697g01	23 614 405- 23 614 656	-	yes	1	84	No hits	-	-	-	-	-	-	-	Unknown	100.0
05g05911	23 615 433- 23 616 545	-	2	117	No hits	-	-	-	-	-	-	-	Unknown	05c4697g02	23 615 433- 23 616 545	-	yes	2	112	No hits	-	-	-	-	-	-	-	Unknown	93.2
05g05912	23 618 473- 23 621 917	+	4	248	No hits	-	-	-	-	-	-	-	Unknown	05c4698g01	23 620 538- 23 622 190	+	yes	4	111	No hits	-	-	-	-	-	-	-	Unknown	14.8
														05c4700g02	23 648 538- 23 648 789	+	yes	1	83	No hits	-	-	-	-	-	-	-	Unknown	
05g05919	23 649 245- 23 651 693	+	4	229	No hits	-	-	-	-	-	-	-	Unknown	05c4701g01	23 649 801- 23 651 693	+	yes	5	89	No hits	-	-	-	-	-	-	-	Unknown	31.0

Supplementary Table S4.3 Annotations of predicted genes on the sequences of the IR64 genome and CT8556-37-2-3-1-M contigs (> 2442 bp) within 2 cM distance from the peak of the *Striga asiatica*-resistance QTL, *qSaB11.1*. Genes at the peak of the QTL are highlighted. Pos., Start and stop positions of coding region (CDS); S., Strand orientation; #E, Number of exons; #AA, Number of amino acids in the translated sequence; DB., Database; Cov., Percentage query coverage of BLAST hit; Iden., Percentage identity of BLAST hit; SSF, Superfamily; PF, Pfam domains; GO, Gene Ontology; C., Is the gene complete?; Sim., Percentage similarity between amino acid sequences of IR64 and CT8556-37-2-3-1-M genes; RAP-DB, Rice Annotation Project Database; MSU, Michigan State University—Rice Genome Annotation Project; NCBI, National Center for Biotechnology Information—Non-redundant protein sequences; Os ind., *Oryza sativa* subsp. *indica*; Os jap., *O. sativa* subsp. *japonica*; TEs, Transposable elements; Cell wall biogen./mod., Cell wall biogenesis/modification.

IR64											CT8556-37-2-3-1-M																			
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.	
11g05079	19 735 754– 19 736 227	+	2	101	LOC_Os11 g30770.1	MSU	expressed protein	86.0	97.7	–	–	–	Hypothetical	11c3264g01	19 735 754– 19 736 227	+	yes	2	102	LOC_Os11 g30770.1	MSU	expressed protein	86.0	97.7	–	–	–	Hypothetical	100.0	
11g05083	19 742 678– 19 742 980	+	1	100	EEE52149. 1	NCBI	hypothetical protein Osj_33986 [Os jap.]	81.0	96.3	–	–	–	Hypothetical	11c3264g03	19 742 678– 19 742 980	+	yes	1	101	EEE52149. 1	NCBI	hypothetical protein Osj_33986 [Os jap.]	81.0	96.3	–	–	–	Hypothetical	100.0	
11g05084	19 744 008– 19 745 414	+	1	468	EEC68218. 1	NCBI	hypothetical protein Osl_36208 [Os ind.]	99.0	99.4	SSF48371	–	–	Hypothetical	11c3264g04	19 744 008– 19 752 571	+	yes	3	683	EEC68218. 1	NCBI	hypothetical protein Osl_36208 [Os ind.]	100.0	99.3	SSF48371	–	–	Domain- containing	67.6	
11g05086	19 749 291– 19 749 623	+	1	110	LOC_Os10 g06180.1	MSU	transposon protein, putative, unclassified	100.0	70.0	–	PF05754	–	TEs																	
11g05087	19 752 140– 19 752 499	+	1	119	LOC_Os11 g30790.1	MSU	expressed protein	100.0	98.3	–	–	–	Hypothetical																	
11g05088	19 753 436– 19 754 745	–	2	177	BAT14138. 1	NCBI	Osl1g0503650 [Os jap.]	64.0	87.7	–	–	–	Hypothetical																	
11g05089	19 755 563– 19 756 054	+	1	163	LOC_Os11 g30810.1	MSU	sulfotransferase domain containing protein, expressed	100.0	99.4	–	PF00685	GO:0008146	Sulfo- transferases	11c3265g01	19 755 563– 19 756 054	+	yes	1	164	Os11t050 3900-01	RAP- DB	Sulfotransferase family protein.	100.0	99.4	SSF52540	–	–	Sulfo- transferases	100.0	
11g05090	19 756 174– 19 757 188	+	2	227	LOC_Os11 g30810.1	MSU	sulfotransferase domain containing protein, expressed	86.0	85.7	SSF52540	–	–	Sulfo- transferases	11c3265g02	19 756 174– 19 757 188	+	yes	2	228	Os11t050 3900-01	RAP- DB	Sulfotransferase family protein.	86.0	85.7	–	PF00685	GO:0008146	Sulfo- transferases	99.6	
11g05091	19 758 164– 19 758 706	+	1	180	ABA93895 .1	NCBI	Flavonol sulfotransferase, putative [Os jap.]	100.0	98.3	SSF52540	–	–	Sulfo- transferases	11c3265g03	19 758 164– 19 758 706	+	yes	1	181	ABA93895 .1	NCBI	Flavonol sulfotransferase, putative [Os jap.]	100.0	98.3	–	PF00685	GO:0008146	Sulfo- transferases	100.0	
11g05092	19 759 033– 19 762 537	–	5	842	LOC_Os11 g30840.1	MSU	retrotransposon protein, putative, unclassified, expressed	93.0	96.4	SSF56672 SSF53098	PF17921 PF00078	–	TEs	11c3265g04	19 759 033– 19 762 537	–	yes	5	721	LOC_Os11 g30840.1	MSU	retrotransposon protein, putative, unclassified, expressed	92.0	85.3	SSF53098 SSF56672	PF17921 PF17919	–	TEs	81.4	
11g05093	19 765 041– 19 768 742	–	5	886	LOC_Os04 g05710.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	87.0	75.3	SSF54160 SSF53098 SSF56672	PF17921 PF17919 PF00078	–	TEs	11c3267g01	19 765 899– 19 770 310	–	yes	5	1190	LOC_Os07 g26220.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	99.0	82.8	SSF56672 SSF57756 SSF54160 SSF53098	PF17921 PF00078 PF03732 PF17919	–	TEs	} 88.2	
11g05094	19 769 297– 19 770 310	–	1	337	LOC_Os01 g34250.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	90.8	–	PF03732	–	TEs																	
11g05095	19 770 654– 19 771 847	–	2	374	LOC_Os02 g34440.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	80.4	SSF90257	–	–	TEs	11c3267g02	19 770 654– 19 771 847	–	yes	2	375	LOC_Os02 g34440.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	80.4	SSF90257	–	–	TEs	100.0	
11g05096	19 774 060– 19 776 465	–	4	388	No hits	–	–	–	–	–	–	–	Unknown	11c3268g01	19 775 392– 19 776 465	–	yes	2	208	LOC_Os11 g30880.1	MSU	retrotransposon protein, putative, unclassified, expressed	64.0	95.5	–	–	–	TEs	52.8	
11g05097	19 777 204– 19 777 557	+	1	117	LOC_Os11 g30910.1	MSU	sulfotransferase domain containing protein, expressed	91.0	62.6	–	PF00685	GO:0008146	Sulfo- transferases	11c3268g02	19 777 204– 19 777 557	+	yes	1	118	LOC_Os11 g30910.1	MSU	sulfotransferase domain containing protein, expressed	91.0	61.7	SSF52540	–	–	Sulfo- transferases	99.2	
11g05098	19 778 684– 19 780 561	+	5	148	ABA93901 .1	NCBI	Sulfotransferase domain containing protein [Os jap.]	86.0	72.7	–	–	–	Sulfo- transferases																	
11g05100	19 787 953– 19 788 243	+	2	71	EEE52155. 1	NCBI	hypothetical protein Osj_33998 [Os jap.]	93.0	93.9	–	–	–	Hypothetical																	
11g05101	19 789 185– 19 790 309	+	1	374	LOC_Os11 g30910.1	MSU	sulfotransferase domain containing protein, expressed	100.0	98.9	–	PF00685	GO:0008146	Sulfo- transferases																	
11g05102	19 791 690– 19 795 119	+	4	426	LOC_Os11 g30920.1	MSU	transposon protein, putative, Mariner sub-class	97.0	96.9	–	–	–	TEs	11c3270g01	19 789 350– 19 795 119	+	yes	5	749	LOC_Os11 g30920.1	MSU	transposon protein, putative, Mariner sub-class	55.0	96.6	–	PF00685	GO:0008146	TEs	55.5	

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.							
Gene ID	Pos. (bp)	S.	#A	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation		Cov.	I den.	SSF hits	PF hits	GO terms	Group	
11g05103	19 795 504- 19 800 232	-	6	422	No hits	-	-	-	SSF51445	-	-	Hypothetical	11c3270g02	19 795 504- 19 797 708	-	yes	3	153	EEC68181.	NCBI	hypothetical protein Osl_36136 [Os ind.]	68.0	71.2	-	-	-	Hypothetical	24.9	
													11c3272g01	19 799 734- 19 802 582	+	yes	3	294	No hits	-	-	-	-	SSF53756	-	-	Domain-containing		
11g05104	19 802 601- 19 802 921	+	1	106	BAC83728.	NCBI	HGWP repeat containing protein-like [Os jap.]	88.0	66.7	-	PF03578	-	Domain-containing																
11g05105	19 803 584- 19 804 153	+	1	189	BAD53599	NCBI	HGWP repeat containing protein-like [Os jap.]	98.0	66.8	-	PF03578	-	Domain-containing																
11g05107	19 808 125- 19 816 077	-	6	565	BAH94473	NCBI	Os09g0272800 [Os jap.]	90.0	63.1	-	PF07727 PF14223	-	TEs																
11g05108	19 816 682- 19 819 492	+	5	729	LOC_Os11g28230.1	MSU	retrotransposon protein, putative, unclassified, expressed	90.0	80.5	-	PF04195	-	TEs	11c3285g01	19 816 076- 19 819 461	+	yes	4	792	LOC_Os11g28230.1	MSU	retrotransposon protein, putative, unclassified, expressed	95.0	81.0	-	PF04195	-	TEs	81.4
11g05109	19 820 286- 19 824 090	-	5	925	LOC_Os11g43020.1	MSU	retrotransposon protein, putative, unclassified, expressed	98.0	77.5	SSF56672 SSF53098	PF17919 PF00665	-	TEs	11c3286g01	19 820 286- 19 824 090	-	yes	5	809	LOC_Os10g18360.1	MSU	retrotransposon protein, putative, unclassified, expressed	68.0	84.3	-	PF00665 PF13456	GO:0015074 GO:0003676	TEs	77.9
11g05110	19 824 269- 19 827 688	-	4	522	LOC_Os11g30990.1	MSU	retrotransposon protein, putative, unclassified, expressed	65.0	81.8	-	PF03732	-	TEs	11c3286g02	19 824 269- 19 827 688	-	yes	4	510	LOC_Os11g30990.1	MSU	retrotransposon protein, putative, unclassified, expressed	66.0	88.1	-	PF03732	-	TEs	85.9
11g05111	19 828 379- 19 830 222	-	3	475	LOC_Os06g50290.1	MSU	transposon protein, putative, Pong sub-class, expressed	96.0	83.4	-	PF14303	-	TEs	11c3286g03	19 828 379- 19 830 222	-	yes	3	476	LOC_Os06g50290.1	MSU	transposon protein, putative, Pong sub-class, expressed	96.0	83.4	-	PF14303	-	TEs	99.6
11g05112	19 830 838- 19 831 708	-	4	191	LOC_Os06g10740.1	MSU	transposon protein, putative, unclassified, expressed	69.0	71.9	-	PF05699	GO:0046983	TEs	11c3286g04	19 830 838- 19 831 708	-	yes	4	192	LOC_Os06g10740.1	MSU	transposon protein, putative, unclassified, expressed	69.0	71.9	-	PF05699	GO:0046983	TEs	100.0
11g05113	19 833 136- 19 833 468	-	1	110	LOC_Os10g06180.1	MSU	transposon protein, putative, unclassified	100.0	63.6	-	PF05754	-	TEs																
11g05114	19 834 401- 19 839 431	+	5	1022	LOC_Os10g09440.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	98.0	85.8	SSF56672 SSF57756	PF08284 PF03732	-	TEs																
11g05115	19 839 595- 19 841 277	+	1	560	ABA97812	NCBI	retrotransposon protein, putative, Ty3-gypsy subclass [Os jap.]	100.0	97.9	SSF53098 SSF54160	PF17921	-	TEs																
11g05116	19 841 944- 19 845 545	-	3	434	LOC_Os10g12140.1	MSU	transposon protein, putative, unclassified, expressed	68.0	98.0	SSF57667 PF02892	PF05754	-	TEs	11c3289g01	19 844 134- 19 845 545	-	yes	3	305	LOC_Os10g12140.1	MSU	transposon protein, putative, unclassified, expressed	99.0	97.7	-	PF02892	GO:0003677	TEs	68.5
11g05119	19 854 991- 19 863 453	+	9	1003	LOC_Os11g31050.1	MSU	retrotransposon protein, putative, unclassified, expressed	66.0	92.8	SSF56672 SSF50630	PF17919 PF08284	-	TEs	11c3289g03	19 856 449- 19 861 941	+	yes	6	743	LOC_Os11g31050.1	MSU	retrotransposon protein, putative, unclassified, expressed	64.0	98.8	SSF50630 SSF56672	PF08284 PF03732	-	TEs	51.3
11g05121	19 864 600- 19 864 968	+	1	122	Os11t050g6700-00	RAP-DB	Similar to ATPase, coupled to transmembrane movement of substances.	100.0	95.9	-	-	-	Transporters	11c3291g01	19 864 375- 19 864 968	+	yes	2	164	Os11t050g6700-00	RAP-DB	Similar to ATPase, coupled to transmembrane movement of substances.	69.0	98.2	-	-	-	Transporters	65.9
11g05122	19 865 658- 19 867 052	-	1	464	LOC_Os11g31060.1	MSU	IQ calmodulin-binding and BAG domain containing protein, putative, expressed	98.0	99.6	SSF63491	PF02179	GO:0051087	Stress-related	11c3291g02	19 865 658- 19 866 527	-	yes	1	289	LOC_Os11g31060.1	MSU	IQ calmodulin-binding and BAG domain containing protein, putative, expressed	97.0	99.3	-	-	-	Stress-related	67.9
													11c3293g01	19 866 970- 19 867 052	-	no	1	27	LOC_Os11g31060.1	MSU	IQ calmodulin-binding and BAG domain containing protein, putative, expressed	100.0	100.0	-	-	-	Stress-related		
11g05124	19 871 772- 19 872 375	-	2	122	LOC_Os11g31070.1	MSU	hypothetical protein	100.0	97.6	-	-	-	Hypothetical	11c3293g03	19 871 772- 19 872 375	-	yes	2	121	LOC_Os11g31070.1	MSU	hypothetical protein	100.0	96.8	-	-	-	Hypothetical	97.5
11g05128	19 897 278- 19 901 934	+	3	496	OsIR64_1g021040	Eu Gene	Omega-hydroxypalmitate O-feruloyl transferase	-	-	-	PF02458	GO:0016747	Cell wall biogen./mod.	11c3293g08	19 893 930- 19 900 327	+	yes	5	338	Os11t050g7200-00	RAP-DB	Similar to transferase.	50.0	94.7	-	PF02458	GO:0016747	Transferases	33.8
													11c3297g01	19 901 700- 19 901 936	-	yes	1	79	TVU25329.	NCBI	hypothetical protein [Eragrostis curvula]	86.0	80.6	-	-	-	Hypothetical		
11g05129	19 902 631- 19 908 741	+	2	2017	AAL58229.	NCBI	putative gag-pol precursor [Os jap.]	100.0	97.5	SSF53098 SSF56672	PF13456 PF00665	-	TEs																

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	
11g05130	19 909 060- 19 911 984	-	7	821	LOC_Os11 g44050.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	94.2	-	PF04195	-	TEs																	
11g05131	19 914 563- 19 920 670	-	3	787	XP_01561 6711.1	NCBI	DELLA protein RHT-1 [Os.jap.]	100.0	99.1	-	PF03514	-	Hormone- related	11c3297g02	19 914 563- 19 915 408	-	yes	2	182	LOC_Os11 g31100.1	MSU	gibberellin response modulator protein, putative, expressed	100.0	100.0	-	PF03514	-	Hormone- related	73.3	
														11c3298g01	19 919 292- 19 920 673	-	yes	2	411	LOC_Os11 g31100.1	MSU	gibberellin response modulator protein, putative, expressed	96.0	97.5	-	-	-	Hormone- related		
11g05132	19 923 363- 19 933 967	-	8	474	EEC68223. 1	NCBI	hypothetical protein Osl_36219 [Os ind.]	92.0	64.2	-	-	-	Hypothetical	11c3298g02	19 923 363- 19 933 975	-	yes	8	468	EEES2159. 1	NCBI	hypothetical protein Osl_34006 [Os.jap.]	92.0	69.4	-	-	-	Hypothetical	97.1	
														11c3298g03	19 934 955- 19 935 472	+	yes	2	98	Os11t050 7701-00	RAP- DB	Hypothetical protein	92.0	93.3	-	-	-	Hypothetical		
11g05133	19 934 381- 19 936 348	-	3	218	EAY81044. 1	NCBI	hypothetical protein Osl_36223 [Os ind.]	100.0	75.2	-	-	-	Hypothetical	11c3298g04	19 936 070- 19 936 348	-	yes	1	93	LOC_Os11 g31120.1	MSU	hypothetical protein	100.0	100.0	-	-	-	Hypothetical	41.7	
														11c3302g01	19 945 008- 19 946 833	-	yes	2	219	LOC_Os11 g31140.1	MSU	expressed protein	91.0	96.5	-	-	-	Hypothetical		
11g05139	19 952 481- 19 958 479	+	8	465	LOC_Os03 g32070.1	MSU	retrotransposon protein, putative, unclassified, expressed	80.0	57.7	-	-	-	TEs																	
														11c3303g03	19 961 951- 19 962 193	-	yes	1	81	KAF29109 70.1	NCBI	hypothetical protein DAI22_11g142700 [Os.jap.]	100.0	92.5	-	-	-	Hypothetical		
11g05140	19 962 149- 19 965 944	+	3	561	LOC_Os04 g51840.1	MSU	transposon protein, putative, CACTA, En/Spm sub-class, expressed	59.0	89.1	-	PF13960 PF13952	-	TEs																	
11g05142	19 968 748- 19 972 504	+	6	1036	LOC_Os05 g39570.1	MSU	transposon protein, putative, CACTA, En/Spm sub-class, expressed	100.0	90.5	SSF54001	-	-	TEs																	
11g05146	19 989 346- 19 991 516	-	5	302	Os11t050 8600-01	RAP- DB	Sugar transporter, TAL effector- mediated susceptibility to bacterial pathogen	100.0	99.0	-	PF03083	GO:0016021	Transporters	11c3306g03	19 989 346- 19 991 516	-	yes	5	303	Os11t050 8600-01	RAP- DB	Sugar transporter, TAL effector- mediated susceptibility to bacterial pathogen	100.0	99.3	-	PF03083	GO:0016021	Transporters	99.7	
11g05147	20 001 192	+	4	333	No hits	-	-	-	-	-	-	-	Unknown	11c3306g06	20 000 841- 20 001 192	+	yes	2	101	Os11t051 0100-00	RAP- DB	Hypothetical protein	85.0	100.0	-	-	-	Hypothetical	30.0	
11g05149	20 006 011- 20 009 022	+	2	289	EEC68231. 1	NCBI	hypothetical protein Osl_36234 [Os ind.]	54.0	89.1	-	PF13456	GO:0003676 GO:0004523	TEs	11c3306g08	20 006 011- 20 006 412	+	yes	1	134	No hits	-	-	-	-	-	-	-	Unknown	83.1	
														11c3306g09	20 008 663- 20 009 022	+	yes	1	120	EEC68231. 1	NCBI	hypothetical protein Osl_36234 [Os ind.]	100.0	100.0	SSF53098	-	-	Hypothetical		
11g05154	20 022 582- 20 023 721	+	2	214	LOC_Os04 g32900.1	MSU	transposon protein, putative, unclassified	65.0	50.0	-	-	-	TEs																	
11g05155	20 025 452- 20 033 421	+	10	1215	LOC_Os11 g31300.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	94.0	70.8	-	PF07197	-	TEs																	
11g05156	20 033 894- 20 036 102	-	2	666	LOC_Os01 g58710.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	92.0	91.8	-	PF00665 PF13456	GO:0015074 GO:0003676 GO:0004523	TEs																	
11g05157	20 038 268- 20 039 950	-	2	226	LOC_Os09 g07000.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	99.0	63.8	-	-	-	TEs																	
11g05158	20 041 340- 20 043 963	+	2	451	LOC_Os10 g18970.1	MSU	retrotransposon protein, putative, unclassified	69.0	97.4	-	PF03732	-	TEs																	
11g05159	20 044 641- 20 045 766	-	2	110	LOC_Os07 g28340.1	MSU	transposon protein, putative, unclassified, expressed	99.0	89.9	-	PF05754	-	TEs																	
11g05160	20 048 123- 20 048 536	-	1	137	KAF29109 73.1	NCBI	hypothetical protein DAI22_11g143000 [Os.jap.]	100.0	96.5	-	-	-	Hypothetical																	
11g05164	20 067 197- 20 069 102	+	3	300	Os11t051 2000-01	RAP- DB	No apical meristem (NAM) protein domain containing protein.	100.0	98.7	SSF10194 1	PF02365 GO:0006355	GO:0003677 development	Growth and development	11c3313g01	20 067 197- 20 069 102	+	yes	3	301	Os11t051 2000-01	RAP- DB	No apical meristem (NAM) protein domain containing protein.	100.0	97.0	SSF10194 1	PF02365 GO:0006355	GO:0003677 development	Growth and development	97.0	
11g05166	20 073 238- 20 075 324	+	3	277	Os11t051 2100-01	RAP- DB	Similar to No apical meristem protein, expressed.	100.0	98.9	SSF10194 1	PF02365 GO:0006355	GO:0003677 development	Growth and development	11c3313g03	20 073 238- 20 075 324	+	yes	3	278	Os11t051 2100-01	RAP- DB	Similar to No apical meristem protein, expressed.	100.0	97.1	SSF10194 1	PF02365 GO:0006355	GO:0003677 development	Growth and development	96.8	

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	
11g05195	20 219 844- 20 221 695	-	8	250	LOC_Os11g31540.1	MSU	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	100.0	99.6	SSF52058	PF08263	-	Hormone-related	11c3330g01	20 219 844- 20 221 695	-	yes	8	251	Os11t0514500-01	RAP-DB	Sorghum bicolor leucine-rich repeat-containing extracellular glycoprotein precursor.	100.0	100.0	SSF52058	PF08263	-	Hormone-related	99.6	
11g05196	20 225 386- 20 227 442	-	8	241	LOC_Os11g31550.1	MSU	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	98.0	79.0	SSF52058	PF08263 PF00560	-	Hormone-related	11c3331g01	20 225 386- 20 227 442	-	yes	8	242	LOC_Os11g31550.1	MSU	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	98.0	79.0	SSF52058	PF00560 PF08263	-	Hormone-related	100.0	
11g05197	20 231 250- 20 231 684	-	1	144	EEC68242.1	NCBI	hypothetical protein OsI_36257 [Os ind.]	100.0	99.3	-	-	-	Hypothetical	11c3331g02	20 231 250- 20 231 684	-	yes	1	145	EEC68242.1	NCBI	hypothetical protein OsI_36257 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	99.3	
11g05198	20 232 829- 20 234 915	-	6	187	LOC_Os11g31560.1	MSU	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	100.0	71.3	-	PF13855 PF00560	GO:0005515 GO:0005515	Hormone-related	11c3331g03	20 232 829- 20 234 915	-	yes	6	188	LOC_Os11g31560.1	MSU	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	100.0	71.3	-	PF00560 PF13855	GO:0005515 GO:0005515	Hormone-related	100.0	
11g05199	20 237 736- 20 240 172	-	11	424	Os11t0514800-01	RAP-DB	BRO1 domain domain containing protein.	100.0	98.3	-	PF03097	-	Intracellular transport	11c3331g04	20 237 736- 20 240 172	-	yes	11	402	Os11t0514800-01	RAP-DB	BRO1 domain domain containing protein.	100.0	93.6	-	-	-	Intracellular transport	93.9	
11g05201	20 252 505- 20 254 151	-	1	548	Os11t0515000-00	RAP-DB	Armadillo-like helical domain containing protein.	99.0	75.8	SSF48371	-	-	Domain-containing	11c3331g05	20 245 979- 20 254 184	-	yes	3	195	Os11t0515000-00	RAP-DB	Armadillo-like helical domain containing protein.	68.0	97.0	SSF48371	-	-	Domain-containing	23.6	
11g05203	20 262 211- 20 262 471	-	1	86	LOC_Os11g31610.1	MSU	expressed protein	85.0	93.2	-	-	-	Hypothetical																	
11g05206	20 274 418- 20 278 523	-	3	459	Os11t0515500-01	RAP-DB	Transport inhibitor response 1 (TIR1)-like protein, Target of siR109944 (small interfering RNA)	83.0	86.1	SSF52047	PF18511 PF18791	-	Hormone-related	11c3335g01	20 274 418- 20 278 523	-	yes	3	460	Os11t0515500-01	RAP-DB	Transport inhibitor response 1 (TIR1)-like protein, Target of siR109944 (small interfering RNA)	83.0	86.3	SSF52047	PF18791 PF18511	-	Hormone-related	99.6	
11g05207	20 284 476- 20 286 002	-	5	190	LOC_Os11g31630.1	MSU	expressed protein	94.0	89.4	-	-	-	Hypothetical	11c3335g02	20 284 476- 20 285 804	-	yes	5	125	LOC_Os11g31630.1	MSU	expressed protein	91.0	97.3	-	-	-	Hypothetical	64.7	
11g05208	20 292 372- 20 295 941	+12	488	LOC_Os11g31640.1	MSU	serine palmitoyltransferase 2, putative, expressed	100.0	100.0	-	PF00155	GO:0009058 GO:0030170	Biosynthetic processes	11c3336g01	20 292 372- 20 295 941	+12	yes	12	489	Os11t0516000-01	RAP-DB	Similar to Serine palmitoyltransferase (Fragment).	100.0	100.0	-	PF00155	GO:0009058 GO:0030170	Biosynthetic processes	100.0		
11g05209	20 297 549- 20 301 069	+4	185	ABA93933.1	NCBI	hypothetical protein LOC_Os11g31650 [Os jap.]	51.0	77.9	-	-	-	Hypothetical	11c3337g01	20 300 585- 20 301 069	+4	yes	2	76	Os11t0516100-01	RAP-DB	Ribosomal protein L31 domain containing protein.	91.0	100.0	SSF143800	-	-	Translation	40.5		
11g05210	20 302 027- 20 303 036	-3	245	KAF29110.14.1	NCBI	hypothetical protein DAI22_11g145650 [Os jap.]	100.0	85.9	-	-	-	Hypothetical																		
11g05211	20 304 579- 20 305 058	+1	159	LOC_Os11g31660.1	MSU	expressed protein	88.0	86.4	-	-	-	Hypothetical	11c3338g01	20 304 579- 20 305 058	+1	yes	1	160	LOC_Os11g31660.1	MSU	expressed protein	88.0	86.4	-	-	-	Hypothetical	100.0		
11g05212	20 305 395- 20 309 405	+8	447	EEE52178.1	NCBI	hypothetical protein OsJ_34044 [Os jap.]	92.0	100.0	-	-	-	Hypothetical	11c3339g01	20 305 481- 20 309 405	+8	yes	8	456	EEE52178.1	NCBI	hypothetical protein OsJ_34044 [Os jap.]	91.0	99.8	-	-	-	Hypothetical	96.3		
11g05213	20 311 969- 20 313 607	-3	428	LOC_Os11g31670.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	83.6	-	-	-	TEs	11c3339g02	20 311 969- 20 313 575	-3	yes	2	457	LOC_Os11g31670.1	MSU	retrotransposon protein, putative, unclassified, expressed	92.0	92.6	-	-	-	TEs	80.2		
														11c3339g03	20 314 822- 20 315 408	-2	yes	2	123	LOC_Os11g31680.1	MSU	retrotransposon protein, putative, unclassified, expressed	74.0	72.5	-	-	-	TEs		
11g05214	20 314 822- 20 339 211	-31	1873	LOC_Os11g31680.1	MSU	retrotransposon protein, putative, unclassified, expressed	58.0	90.6	SSF53098	PF07727 PF00665 PF13966	-	TEs	11c3340g01	20 316 005- 20 339 211	-28	yes	28	1708	No hits	-	-	-	-	-	SSF53098	PF13976 PF13966 PF00665	-	TEs	72.5	
														11c3340g02	20 342 867- 20 343 233	-3	yes	3	37	LOC_Os11g31700.1	MSU	expressed protein	75.0	100.0	-	-	-	Hypothetical		
11g05215	20 342 867- 20 359 729	-29	1057	EEC68252.1	NCBI	hypothetical protein OsI_36272 [Os ind.]	79.0	82.8	-	-	-	Hypothetical	11c3341g01	20 347 399- 20 349 812	-7	no	7	217	KAF29110.23.1	NCBI	hypothetical protein DAI22_11g146500 [Os jap.]	100.0	69.1	-	-	-	Hypothetical	73.5		
														11c3342g01	20 350 021- 20 359 519	-19	no	19	699	EEE52180.1	NCBI	hypothetical protein OsJ_34047 [Os jap.]	100.0	78.7	-	-	-		Hypothetical	
11g05216	20 365 912- 20 366 466	+1	184	BAT14201.1	NCBI	Os11g0518200 [Os jap.]	100.0	100.0	-	-	-	Hypothetical	11c3344g01	20 365 938- 20 366 273	-1	yes	1	112	KAF29110.25.1	NCBI	hypothetical protein DAI22_11g146650 [Os jap.]	100.0	100.0	-	-	-	Hypothetical			
														11c3345g01	20 366 771- 20 368 328	-4	yes	4	180	LOC_Os11g31710.1	MSU	expressed protein	100.0	82.1	-	-	-	Hypothetical		
														11c3349g01	20 372 690- 20 375 637	-4	yes	4	127	KAF29110.27.1	NCBI	hypothetical protein DAI22_11g146750 [Os jap.]	64.0	95.1	-	-	-	Hypothetical		

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M																				
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.		
11g05219	20 384 531- 20 388 979	-	4	404	LOC_Os12 g26190.1	MSU	retrotransposon protein, putative, unclassified, expressed	96.0	76.6	-	PF13976 PF07727	-	TEs																		
11g05220	20 390 333- 20 391 484	-	1	383	LOC_Os11 g31720.1	MSU	retrotransposon protein, putative, unclassified, expressed	99.0	96.8	-	PF14223	-	TEs																		
11g05221	20 392 848- 20 394 964	+	3	245	AAV44030. 1	NCBI	putative polyprotein [Os jap.]	52.0	74.8	-	-	-	Hypothetical																		
11g05222	20 397 327- 20 398 232	+	1	301	LOC_Os09 g16340.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass	100.0	91.5	-	-	-	TEs																		
11g05223	20 398 485- 20 404 346	+	3	1707	LOC_Os10 g10460.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	87.0	97.7	SSF53098 SSF56672 SSF50630 SSF57756	PF08284 PF00078 PF17921 PF17917 PF03732 PF00098	-	TEs																		
11g05225	20 410 663- 20 411 169	-	1	168	BAT14203. 1	NCBI	Osl1g0518701 [Os jap.]	100.0	100.0	-	-	-	Hypothetical	11c3364g01	20 410 961- 20 411 169	-	no	1	69	Os11t051 8701-00	RAP- DB	Hypothetical protein	100.0	100.0	-	-	-	Hypothetical	41.1		
11g05228	20 423 446- 20 435 063	-	12	1789	LOC_Os01 g41620.1	MSU	transposon protein, putative, unclassified, expressed	89.0	95.0	-	PF10536 PF10551 PF03108	-	TEs																		
11g05229	20 435 438- 20 438 280	-	7	368	No hits	-	-	-	-	SSF56112	-	-	Protein kinases	11c3366g01	20 426 391- 20 437 651	-	yes	6	242	ABA93944 .1	NCBI	transposon protein, putative, Mutator sub-class [Os jap.]	83.0	80.0	SSF56112	-	-	TEs	29.8		
11g05230	20 438 565- 20 463 016	+	31	1378	EEC68255. 1	NCBI	hypothetical protein Osl_36277 [Os ind.]	55.0	86.2	-	-	-	Hypothetical	11c3366g02	20 438 565- 20 455 095	+	yes	24	1031	EEC68255. 1	NCBI	hypothetical protein Osl_36277 [Os ind.]	77.0	88.9	-	-	-	Hypothetical	60.4		
														11c3367g01	20 456 261- 20 456 392	-	yes	1	44	LOC_Os11 g31780.1	MSU	expressed protein	100.0	100.0	-	-	-	Hypothetical			
11g05231	20 463 139- 20 466 653	+	8	417	LOC_Os11 g31790.1	MSU	hypothetical protein	94.0	99.7	-	PF03004	-	Hypothetical	11c3367g02	20 463 139- 20 463 809	+	yes	3	155	LOC_Os11 g31790.1	MSU	hypothetical protein	94.0	99.3	-	-	-	Hypothetical			
														11c3368g01	20 464 268- 20 466 653	+	yes	4	179	Os02t058 3500-00	RAP- DB	Transposase, Pta/En/Spm, plant domain containing protein.	87.0	100.0	-	-	-	Domain- containing	77.0		
11g05232	20 467 458- 20 473 607	-	8	427	LOC_Os11 g31800.1	MSU	expressed protein	88.0	99.5	-	-	-	Hypothetical	11c3368g02	20 466 990- 20 474 508	-	yes	10	544	Os11t051 9200-00	RAP- DB	Hypothetical protein	69.0	99.7	-	-	-	Hypothetical	68.1		
11g05234	20 478 032- 20 486 598	+	3	102	BAD19291 .1	NCBI	putative BRUSHY1 [Os jap.]	55.0	73.2	-	-	-	Cell division																		
11g05236	20 497 169- 20 501 154	-	9	219	EEC68256. 1	NCBI	hypothetical protein Osl_36279 [Os ind.]	90.0	100.0	-	-	-	Hypothetical	11c3374g01	20 497 169- 20 501 154	-	yes	9	220	EEC68256. 1	NCBI	hypothetical protein Osl_36279 [Os ind.]	90.0	100.0	-	-	-	Hypothetical	100.0		
11g05237	20 502 796- 20 509 291	-	16	573	Os11t051 9500-00	RAP- DB	Conserved hypothetical protein.	96.0	82.9	-	-	-	Hypothetical	11c3374g02	20 502 796- 20 509 291	-	yes	16	574	Os11t051 9500-00	RAP- DB	Conserved hypothetical protein.	96.0	82.9	-	-	-	Hypothetical	100.0		
11g05241	20 517 221- 20 522 427	+	5	927	LOC_Os12 g01790.1	MSU	retrotransposon protein, putative, unclassified, expressed	65.0	82.5	-	PF00098 PF13976 PF14223 PF00083 PF13961	-	TEs																		
11g05242	20 524 139- 20 525 239	-	3	231	BAD19888 .1	NCBI	root cap protein 1-like [Os jap.]	84.0	68.7	-	-	-	Transporters																		
11g05243	20 525 463- 20 527 918	+	6	604	LOC_Os12 g01790.1	MSU	retrotransposon protein, putative, unclassified, expressed	99.0	69.5	SSF56672 SSF53098	PF07727	-	TEs																		
11g05244	20 528 931- 20 537 450	+	7	469	Os11t052 2000-01	RAP- DB	XY4 protein (Fragment).	55.0	95.8	SSF53254	PF00300	-	Biosynthetic processes	11c3385g01	20 530 119- 20 535 211	+	yes	3	151	XP_03470 7592.1	NCBI	uncharacterized protein LOC117930913 [Vitis riparia]	55.0	59.0	SSF53254	-	-	Domain- containing	16.7		
11g05245	20 540 565- 20 546 691	-	12	1216	LOC_Os11 g31890.1	MSU	NLI interacting factor-like phosphatase, putative, expressed	99.0	93.9	SSF56784 SSF52113	PF03031	-	Phosphatases	11c3388g01	20 540 565- 20 546 947	-	yes	13	1010	Os11t052 1900-02	RAP- DB	Similar to NLI interacting factor- like phosphatase family protein, expressed.	99.0	97.1	SSF56784 SSF52113	PF03031	-	Phosphatases	81.4		
11g05247	20 551 337- 20 553 005	-	3	148	Os11t052 1500-01	RAP- DB	Similar to Acyl carrier protein, chloroplast precursor (ACP) (ACP05) (Clone 29C08).	100.0	99.3	SSF47336	PF00550	-	Biosynthetic processes	11c3389g02	20 551 337- 20 553 005	-	yes	3	149	Os11t052 1500-01	RAP- DB	Similar to Acyl carrier protein, chloroplast precursor (ACP) (ACP05) (Clone 29C08).	100.0	99.3	SSF47336	PF00550	-	Intracellular transport	98.7		

Supplementary Table S4.3 (continued)

IR64														CT8556-37-2-3-1-M																
Gene ID	Pos. (bp)	S.	C.	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.		
11g05249	20 555 772- 20 556 050	+	1	92	LOC_Os11 g31930.1	MSU	expressed protein	100.0	98.9	-	-	-	Hypothetical	11c3389g04	20 555 772- 20 556 271	+	yes	2	125	EAY81106. 1	NCBI	hypothetical protein Osl_36285 [Os ind.]	71.0	100.0	-	-	-	Hypothetical	67.2	
11g05250	20 557 081- 20 559 500	+	5	381	LOC_Os11 g31940.1	MSU	GDSL-like lipase/acylhydrolase, putative, expressed	100.0	92.1	-	PF00657	GO:0016788	Lipases	11c3389g05	20 557 081- 20 559 500	+	yes	5	382	LOC_Os11 g31940.1	MSU	GDSL-like lipase/acylhydrolase, putative, expressed	100.0	92.3	SSF52266	-	-	-	Lipases	99.7
11g05252	20 561 911- 20 562 330	+	1	139	BAT14218. 1	NCBI	Os11g0520900 [Os jap.]	100.0	98.6	-	-	-	Hypothetical																	
11g05253	20 567 441- 20 575 388	+	17	838	XP_01561 7893.1	NCBI	coiled-coil domain-containing protein SCD2 isoform X1 [Os jap.]	100.0	96.5	-	-	-	Cell division	11c3391g01	20 567 848- 20 575 388	+	yes	17	684	EEC68259. 1	NCBI	hypothetical protein Osl_36288 [Os ind.]	94.0	96.9	-	-	-	Hypothetical	76.8	
11g05255	20 582 363- 20 585 358	-	3	527	LOC_Os01 g41516.1	MSU	retrotransposon protein, putative, Ty1-copia subclass, expressed	80.0	88.0	SSF56672	PF07727	-	TEs																	
11g05256	20 586 567- 20 587 678	-	2	334	LOC_Os11 g27650.1	MSU	retrotransposon protein, putative, Ty1-copia subclass, expressed	100.0	91.6	-	PF00098	PF14223	TEs																	
11g05257	20 587 857- 20 588 141	+	1	94	BAD10262 .1	NCBI	hypothetical protein [Os jap.]	100.0	88.3	-	-	-	Hypothetical	11c3398g01	20 587 857- 20 588 141	+	yes	1	95	BAD10262 .1	NCBI	hypothetical protein [Os jap.]	100.0	88.3	-	-	-	Hypothetical	100.0	
11g05259	20 611 349- 20 611 603	-	1	84	LOC_Os05 g41660.1	MSU	ICE-like protease p20 domain containing protein, putative, expressed	70.0	58.3	-	-	-	Proteases/ protease inhibitors	11c3402g01	20 611 349- 20 611 603	-	yes	1	85	LOC_Os05 g41660.1	MSU	ICE-like protease p20 domain containing protein, putative, expressed	70.0	56.7	-	-	-	Proteases/ protease inhibitors	97.6	
11g05260	20 618 097- 20 622 787	+	11	517	Os11t052 2900-01	RAP- DB	Peptidase S10, serine carboxypeptidase family protein.	82.0	92.1	-	PF00450	GO:0004185 GO:0006508	Proteases/ protease inhibitors	11c3402g02	20 618 097- 20 622 787	+	yes	11	518	LOC_Os11 g31980.1	MSU	OSCP63 - Putative Serine Carboxypeptidase homologue, expressed	82.0	92.1	SSF53474	-	-	Proteases/ protease inhibitors	99.6	
11g05262	20 625 145- 20 625 453	-	1	102	EAY81111. 1	NCBI	hypothetical protein Osl_36291 [Os ind.]	100.0	97.1	-	-	-	Hypothetical	11c3402g04	20 625 145- 20 625 453	-	yes	1	103	EAY81111. 1	NCBI	hypothetical protein Osl_36291 [Os ind.]	100.0	99.0	-	-	-	Hypothetical	98.0	
11g05264	20 627 049- 20 627 806	-	2	48	No hits	-	-	-	-	-	-	-	Unknown	11c3402g06	20 627 049- 20 630 793	-	yes	3	93	LOC_Os02 g50750.1	MSU	retrotransposon protein, putative, unclassified, expressed	57.0	63.5	-	-	-	TEs	52.2	
11g05265	20 628 803- 20 635 186	+	6	782	LOC_Os11 g31990.1	MSU	retrotransposon protein, putative, unclassified, expressed	85.0	75.7	-	PF03732	-	TEs																	
11g05266	20 635 674- 20 638 143	+	3	618	LOC_Os05 g47450.1	MSU	retrotransposon protein, putative, unclassified, expressed	89.0	86.3	SSF56672	PF00665	SSF53098	PF00078	PF17919	PF13456	PF04195	TEs													
11g05267	20 638 994- 20 642 025	-	9	740	LOC_Os06 g44740.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	76.3	-	-	-	TEs																	
11g05268	20 645 454- 20 647 567	+	4	179	No hits	-	-	-	-	-	-	-	Unknown	11c3417g01	20 645 454- 20 646 095	+	yes	2	89	LOC_Os04 g20290.1	MSU	retrotransposon protein, putative, unclassified, expressed	55.0	75.0	-	-	-	TEs	46.9	
11g05269	20 653 222- 20 653 641	+	1	139	EAY81112. 1	NCBI	hypothetical protein Osl_36292 [Os ind.]	100.0	97.1	-	-	-	Hypothetical	11c3418g01	20 653 222- 20 653 641	+	yes	1	140	EAY81112. 1	NCBI	hypothetical protein Osl_36292 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	97.1	
11g05270	20 655 238- 20 655 926	-	3	75	LOC_Os11 g32020.1	MSU	expressed protein	80.0	81.8	-	-	-	Hypothetical	11c3420g01	20 655 319- 20 655 926	-	yes	3	49	LOC_Os11 g32020.1	MSU	expressed protein	100.0	79.6	-	-	-	Hypothetical	64.0	
11g05271	20 658 580- 20 661 137	+	3	316	LOC_Os11 g32030.1	MSU	sex determination protein tasselseed-2, putative, expressed	100.0	95.9	SSF51735	PF13561	-	Oxido- reductases	11c3420g02	20 658 580- 20 660 549	+	no	3	109	LOC_Os11 g32030.1	MSU	sex determination protein tasselseed-2, putative, expressed	100.0	100.0	SSF51735	-	-	Oxido- reductases	58.9	
													11c3421g01	20 660 904- 20 661 137	+	yes	1	78	LOC_Os11 g32030.1	MSU	sex determination protein tasselseed-2, putative, expressed	100.0	100.0	SSF51735	PF13561	-	Oxido- reductases			
11g05273	20 664 714- 20 668 723	-	5	1218	LOC_Os09 g37820.1	MSU	transposon protein, putative, unclassified, expressed	99.0	92.4	-	PF09322	PF03108	PF10551	TEs																
11g05274	20 670 004- 20 673 466	+	4	303	EEE52191. 1	NCBI	hypothetical protein Osj_34068 [Os jap.]	50.0	86.3	-	-	-	Hypothetical	11c3421g03	20 670 004- 20 673 466	+	yes	4	304	EEE52191. 1	NCBI	hypothetical protein Osj_34068 [Os jap.]	50.0	86.3	-	-	-	Hypothetical	98.0	
11g05275	20 673 786- 20 675 523	-	2	278	EAZ18539. 1	NCBI	hypothetical protein Osj_34069 [Os jap.]	64.0	99.4	-	-	-	Hypothetical	11c3421g04	20 673 786- 20 674 358	-	yes	1	191	EEC68263. 1	NCBI	hypothetical protein Osl_36294 [Os ind.]	63.0	100.0	-	-	-	Hypothetical	43.8	

Supplementary Table S4.3 (continued)

IR64													CT8556-37-2-3-1-M													Sim.			
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits		PF hits	GO terms	Group
11g05276	20 676 542- 20 678 501	-	3	155	LOC_Os11g32050.1	MSU	expressed protein	100.0	96.8	-	-	-	Hypothetical	11c3424g01	20 676 706- 20 678 501	-	yes	3	155	EEC68263.1	NCBI	hypothetical protein Osl_36294 [Os ind.]	100.0	97.4	-	-	-	Hypothetical	94.3
11g05277	20 679 432- 20 687 242	-	8	1005	LOC_Os08g31010.1	MSU	retrotransposon protein, putative, unclassified, expressed	78.0	83.2	SSF56672	PF17917 PF03732	-	TEs																
11g05278	20 689 689- 20 691 055	+	2	125	XP_02587.7466.1	NCBI	TPD1 protein homolog 1-like [Os jap.]	100.0	100.0	-	-	-	Growth and development	11c3431g02	20 689 689- 20 691 055	+	yes	2	126	XP_02587.7466.1	NCBI	TPD1 protein homolog 1-like [Os jap.]	100.0	100.0	-	-	-	Growth and development	100.0
11g05279	20 691 710- 20 692 242	-	2	129	EAY81116.1	NCBI	hypothetical protein Osl_36296 [Os ind.]	100.0	99.2	-	-	-	Hypothetical	11c3431g03	20 691 710- 20 692 242	-	yes	2	130	EAY81116.1	NCBI	hypothetical protein Osl_36296 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	99.2
11g05280	20 694 066- 20 695 271	+	2	369	Os03t0193901-00	RAP-DB	Conserved hypothetical protein.	87.0	75.1	-	-	-	Hypothetical	11c3431g04	20 694 066- 20 695 400	+	yes	2	170	No hits	-	-	-	-	-	-	-	Unknown	24.0
11g05281	20 695 932- 20 697 251	-	2	108	LOC_Os01g39030.1	MSU	expressed protein	52.0	63.3	-	-	-	Hypothetical	11c3431g05	20 695 932- 20 697 251	-	yes	2	109	LOC_Os01g39030.1	MSU	expressed protein	52.0	63.3	-	-	-	Hypothetical	97.2
11g05283	20 711 519- 20 714 524	+	4	585	Os11t0523700-01	RAP-DB	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation	90.0	99.2	SSF47459	PF00010	GO:0046983	Transcription	11c3433g01	20 711 519- 20 712 376	+	no	1	286	Os11t0523700-01	RAP-DB	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation	79.0	98.2	-	-	-	Transcription	90.4
													11c3434g01	20 712 536- 20 714 524	+	yes	4	247	Os11t0523700-01	RAP-DB	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation	100.0	100.0	SSF47459	PF00010	GO:0046983	Transcription		
11g05286	20 725 543- 20 731 027	-	14	853	AAG43286.2	NCBI	putative auxin response factor 1 [Os ind.]	100.0	99.2	SSF101936	PF06507 PF02362	-	Hormone-related	11c3434g03	20 725 543- 20 731 274	-	yes	13	831	A2ZET6.1	NCBI	RecName: Full=Auxin response factor 23; AltName: Full=OsARF1 [Os ind.]	100.0	100.0	SSF542776	PF06507 PF02309	-	Hormone-related	96.6
11g05289	20 736 863- 20 743 527	-	8	593	LOC_Os11g40940.1	MSU	retrotransposon protein, putative, unclassified, expressed	70.0	88.9	SSF57756	-	GO:0003676 GO:0008270	TEs																
11g05292	20 759 126- 20 760 494	+	2	177	LOC_Os01g04680.1	MSU	retrotransposon protein, putative, unclassified, expressed	75.0	67.7	-	PF13966	-	TEs																
														11c3437g03	20 759 966- 20 761 291	-	yes	2	113	BAF18922.2	NCBI	Os06g0186300 [Os jap.] -> Leucine-rich repeat domain containing protein.	79.0	58.2	-	-	-	Domain-containing	
11g05293	20 761 521- 20 762 228	+	2	122	LOC_Os12g21530.1	MSU	retrotransposon, putative, centromere-specific	99.0	95.0	-	-	-	TEs																
11g05294	20 763 707- 20 765 850	+	3	111	LOC_Os09g03870.1	MSU	retrotransposon protein, putative, unclassified, expressed	88.0	74.1	-	-	-	TEs																
11g05295	20 769 974- 20 775 748	+	9	275	XP_015615353.1	NCBI	chromophore lyase CRL, chloroplast [Os jap.]	100.0	100.0	-	PF06206	GO:0016829 GO:0017009	Cell division	11c3437g04	20 769 974- 20 775 748	+	yes	9	276	XP_015615353.1	NCBI	chromophore lyase CRL, chloroplast [Os jap.]	100.0	100.0	-	PF06206	GO:0016829 GO:0017009	Cell division	100.0
11g05296	20 780 165- 20 788 525	-	11	1116	LOC_Os11g32170.1	MSU	stripe rust resistance protein Yr10, putative, expressed	80.0	99.3	SSF51101	PF01419 SSF52540	PF00931 PF18052	Stress-related	11c3437g06	20 780 165- 20 788 525	-	yes	12	1046	LOC_Os11g32170.1	MSU	stripe rust resistance protein Yr10, putative, expressed	87.0	85.8	SSF51101	PF01419 SSF52540	PF00931 PF18052	Stress-related	91.2
11g05297	20 789 003- 20 796 349	-	10	747	LOC_Os11g32180.1	MSU	retrotransposon protein, putative, unclassified, expressed	84.0	72.2	SSF53098	PF17919 SSF56672	-	TEs	11c3437g07	20 789 003- 20 789 757	-	yes	2	98	LOC_Os06g25190.1	MSU	retrotransposon protein, putative, unclassified	98.0	63.0	-	-	-	TEs	74.2
														11c3440g01	20 792 226- 20 794 462	-	yes	5	386	LOC_Os11g32180.1	MSU	retrotransposon protein, putative, unclassified, expressed	98.0	65.6	SSF56672	-	-	TEs	
														11c3441g01	20 795 407- 20 796 349	-	yes	3	111	LOC_Os06g25190.1	MSU	retrotransposon protein, putative, unclassified	100.0	66.1	-	-	-	TEs	
11g05298	20 800 321- 20 803 217	+	3	97	LOC_Os02g25990.1	MSU	retrotransposon, putative, centromere-specific	71.0	57.3	-	-	-	TEs	11c3441g02	20 800 321- 20 803 217	+	yes	3	98	LOC_Os02g25990.1	MSU	retrotransposon, putative, centromere-specific	71.0	57.3	-	-	-	TEs	100.0
11g05299	20 803 480- 20 803 761	-	1	93	LOC_Os11g32190.1	MSU	retrotransposon, putative, centromere-specific	100.0	94.6	-	-	-	TEs	11c3441g03	20 803 480- 20 803 761	-	yes	1	94	LOC_Os11g32190.1	MSU	retrotransposon, putative, centromere-specific	100.0	94.6	-	-	-	TEs	100.0

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.							
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group
11g05300	20 804 850- 20 806 187	+	2	185	LOC_Os04	MSU	retrotransposon protein, putative, Ty3-gypsy subclass	91.0	67.8	-	-	-	TEs																
11g05302	20 811 345- 20 812 068	+	2	220	BAD28103	NCBI	hypothetical protein [Os jap.]	79.0	52.9	-	PF05754	-	Hypothetical																
11g05303	20 812 785- 20 813 798	-	1	337	LOC_Os11	MSU	retrotransposon, putative, centromere-specific, expressed	100.0	97.6	SSF53098	-	-	TEs	11c3447g01	20 812 785- 20 813 798	-	yes	1	338	LOC_Os11	MSU	retrotransposon, putative, centromere-specific, expressed	100.0	97.6	-	PF00665	GO:0015074	TEs	100.0
11g05305	20 817 650- 20 824 395	-	13	992	Os11t052	RAP-DB	Similar to Jacalin-like lectin domain containing protein, expressed.	62.0	95.5	SSF52540 SSF51101	PF00931 PF18052 PF01419	-	Stress-related	11c3447g03	20 817 650- 20 823 813	-	yes	13	799	Os11t052	RAP-DB	Similar to Jacalin-like lectin domain containing protein, expressed.	77.0	95.5	SSF51101 SSF52540	PF00931 PF01419	-	Stress-related	95.3
														11c3448g01	20 823 950- 20 824 395	-	no	1	148	LOC_Os11	MSU	jacalin-like lectin domain containing protein, expressed	100.0	96.6	-	PF18052	-	Stress-related	
11g05307	20 831 238- 20 832 806	-	2	202	LOC_Os01	MSU	retrotransposon protein, putative, unclassified, expressed	95.0	68.2	-	-	-	TEs																
11g05308	20 834 053- 20 834 850	+	1	265	LOC_Os08	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	96.2	-	-	-	TEs																
11g05309	20 835 123- 20 836 913	+	1	596	LOC_Os08	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	99.0	93.1	SSF56672 SSF57756 SSF50630	PF03732 PF00098 PF08284	-	TEs																
11g05310	20 837 867- 20 838 352	-	1	161	AAN09855	NCBI	hypothetical protein [Os jap.]	100.0	69.5	-	-	-	Hypothetical																
11g05311	20 839 666- 20 842 554	+	2	786	CAH65969	NCBI	H0820C10.2 [Oryza sativa]	99.0	97.9	SSF54160 SSF53098 SSF56672	PF17917 PF17921	-	TEs																
11g05312	20 843 661- 20 844 290	-	1	209	Os07t015	RAP-DB	Similar to Calcineurin B-like protein.	100.0	63.2	-	-	-	Stress-related																
11g05313	20 845 647- 20 847 191	+	2	191	AAK13114	NCBI	Putative retroelement pol polyprotein [Os jap.]	93.0	74.2	-	-	-	TEs																
11g05314	20 847 995- 20 848 234	-	1	79	LOC_Os08	MSU	retrotransposon protein, putative, Ty3-gypsy subclass	57.0	80.0	-	-	-	TEs																
11g05315	20 849 633- 20 851 589	-	3	108	LOC_Os10	MSU	transposon protein, putative, unclassified, expressed	82.0	84.6	-	-	-	TEs	11c3458g01	20 849 633- 20 851 589	-	yes	3	109	LOC_Os10	MSU	transposon protein, putative, unclassified, expressed	82.0	84.6	-	-	-	TEs	100.0
11g05316	20 853 424- 20 853 900	-	1	158	No hits	-	-	-	-	-	PF05754	-	Hypothetical	11c3458g02	20 853 424- 20 853 900	-	yes	1	159	No hits	-	-	-	-	-	PF05754	-	Domain- containing Hypothetical	100.0
11g05318	20 863 733- 20 865 351	-	2	184	EAY81124	NCBI	hypothetical protein Osl_36305 [Os ind.]	73.0	99.3	-	-	-	Hypothetical	11c3461g02	20 863 733- 20 865 351	-	yes	2	185	EAY81124	NCBI	hypothetical protein Osl_36305 [Os ind.]	73.0	99.3	-	-	-	Hypothetical	100.0
11g05319	20 866 136- 20 869 474	+	2	490	Os11t052	RAP-DB	Similar to Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450- LIA1) (Obtusifoliol 14-alpha demethylase) (Fragment).	100.0	99.6	SSF48264	PF00067	GO:0005506 GO:0016705 GO:0020037 GO:0055114	Cytochrome P450	11c3461g03	20 866 136- 20 869 474	+	yes	2	491	Os11t052	RAP-DB	Similar to Cytochrome P450 51 (EC 1.14.13.70) (CYPLI) (P450- LIA1) (Obtusifoliol 14-alpha demethylase) (Fragment).	100.0	99.6	SSF48264	PF00067	GO:0005506 GO:0016705 GO:0020037 GO:0055114	Cytochrome P450	100.0
11g05320	20 871 275- 20 871 514	+	1	79	LOC_Os08	MSU	retrotransposon protein, putative, Ty3-gypsy subclass	57.0	84.4	-	-	-	TEs																
11g05321	20 872 407- 20 872 832	+	1	141	LOC_Os08	MSU	retrotransposon protein, putative, unclassified, expressed	88.0	66.1	-	-	-	TEs																
11g05322	20 873 540- 20 877 967	-	1	1475	LOC_Os01	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	100.0	96.7	SSF54160 SSF53098 SSF57756 SSF50630 SSF56672	PF17921 PF00078 PF17917 PF08284 PF00098 PF03732	-	TEs																
11g05323	20 878 362- 20 879 036	-	1	224	LOC_Os08	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	95.0	93.9	-	-	-	TEs																
11g05324	20 881 400- 20 881 825	+	1	141	LOC_Os08	MSU	retrotransposon protein, putative, unclassified, expressed	88.0	66.1	-	-	-	TEs																

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.								
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	
11g05325	20 883 374- 20 886 368	+	7	668	LOC_Os07 g24560.1	MSU	retrotransposon protein, putative, unclassified, expressed	99.0	70.8	—	PF04195	—	TEs	11c3468g01	20 883 374- 20 886 368	+	yes	7	669	LOC_Os07 g24560.1	MSU	retrotransposon protein, putative, unclassified, expressed	99.0	70.8	—	PF04195	—	TEs	100.0	
11g05326	20 886 577- 20 892 174	-	8	1272	ABA96243 .2	NCBI	retrotransposon protein, putative, unclassified [Os jap.]	84.0	66.3	SSF56672 SSF53098	PF00078 PF00665 PF03732	—	TEs	11c3468g02	20 886 577- 20 890 097	-	yes	5	624	LOC_Os04 g19770.1	MSU	retrotransposon protein, putative, unclassified	83.0	67.9	SSF53098 SSF56672	PF00078 PF00665	—	TEs	87.2	
														11c3469g01	20 890 186- 20 892 174	-	yes	3	521	LOC_Os09 g07870.1	MSU	retrotransposon protein, putative, unclassified, expressed	98.0	77.9	—	PF03732	—	TEs		
11g05327	20 893 925- 20 896 281	-	3	597	AAT07580. 1	NCBI	hypothetical protein [Os jap.]	50.0	69.9	—	—	—	Hypothetical	11c3470g01	20 895 442- 20 896 236	-	yes	1	265	LOC_Os05 g48910.1	MSU	retrotransposon protein, putative, unclassified	98.0	96.2	—	—	—	TEs	42.7	
11g05329	20 901 583- 20 910 419	+	26	919	LOC_Os11 g32260.1	MSU	lysosomal alpha-mannosidase precursor, putative, expressed	100.0	88.4	—	PF01074 PF07748 PF09261	GO:0004559 GO:0006013 GO:0004559 GO:0006013	Biosynthetic processes	11c3471g01	20 901 625- 20 910 419	+	yes	26	906	LOC_Os11 g32260.1	MSU	lysosomal alpha-mannosidase precursor, putative, expressed	100.0	88.3	SSF74650 SSF88688 SSF88713	—	—	Biosynthetic processes	98.5	
11g05330	20 912 284- 20 918 637	+	7	658	LOC_Os11 g32270.1	MSU	N-rich protein, putative, expressed	100.0	95.9	—	PF10539	—	Stress-related	11c3471g02	20 912 284- 20 918 637	+	yes	7	659	LOC_Os11 g32270.1	MSU	N-rich protein, putative, expressed	100.0	95.9	—	PF10539	—	Stress-related	100.0	
11g05331	20 919 810- 20 923 908	-	5	273	XP_01561 5649.1	NCBI	N-alpha-acetyltransferase MAK3 [Os jap.]	72.0	100.0	SSF55729	—	—	Growth and development	11c3471g03	20 919 810- 20 923 908	-	yes	5	274	LOC_Os11 g32280.1	MSU	acetyltransferase, GNAT family, putative, expressed	72.0	100.0	—	PF00583	GO:0008080	Growth and development	100.0	
11g05332	20 924 901- 20 925 146	-	1	81	Os021077 0800-01	RAP- DB	NADH/NADPH-dependent nitrate reductase	84.0	79.4	SSF52343	—	—	Nutrient assimilation	11c3471g04	20 924 901- 20 925 146	-	yes	1	82	Os021077 0800-01	RAP- DB	NADH/NADPH-dependent nitrate reductase	84.0	79.4	—	PF00175	GO:0016491	Nutrient assimilation	100.0	
11g05333	20 925 875- 20 926 834	-	1	319	LOC_Os11 g32290.1	MSU	3-5 exonuclease eri-1, putative, expressed	100.0	100.0	SSF53098	PF06839 PF00929	—	Translation																	
11g05334	20 927 642- 20 928 971	+	3	239	EEC68275. 1	NCBI	hypothetical protein Osl_36314 [Os ind.]	72.0	100.0	—	—	—	Hypothetical																	
11g05335	20 929 268- 20 930 865	-	2	212	EEE68685. 1	NCBI	hypothetical protein Osl_27313 [Os jap.]	58.0	96.7	—	—	—	Hypothetical	11c3474g01	20 928 897- 20 930 865	-	yes	3	197	EEE68685. 1	NCBI	hypothetical protein Osl_27313 [Os jap.]	62.0	96.7	—	—	—	Hypothetical	79.3	
11g05336	20 931 451- 20 931 843	-	1	130	LOC_Os11 g32310.1	MSU	expressed protein	100.0	97.7	—	PF06839	GO:0008270	Domain- containing																	
11g05337	20 935 932- 20 936 306	+	1	124	XP_01561 7522.1	NCBI	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplast [Os jap.]	94.0	97.4	—	PF12046	—	Growth and development																	
11g05338	20 937 050- 20 938 117	+	3	141	XP_02582 8557.1	NCBI	protein COFACTOR ASSEMBLY OF COMPLEX C SUBUNIT B CCB1, chloroplast [Panicum hallii]	100.0	99.3	—	PF12046	—	Growth and development	11c3476g01	20 936 139- 20 938 117	+	yes	4	197	Os11t052 6200-01	RAP- DB	Similar to predicted protein.	100.0	100.0	—	PF12046	—	Hypothetical	71.9	
11g05339	20 938 978- 20 940 366	+	2	158	No hits	—	—	—	—	—	—	—	Unknown	11c3476g02	20 938 978- 20 940 478	+	yes	2	165	LOC_Os11 g32330.1	MSU	retrotransposon protein, putative, unclassified	70.0	63.8	—	—	—	TEs	45.3	
11g05340	20 941 035- 20 942 723	+	3	125	LOC_Os04 g28800.1	MSU	retrotransposon protein, putative, unclassified	53.0	62.7	—	—	—	TEs	11c3476g03	20 941 035- 20 942 723	+	yes	3	126	LOC_Os04 g28800.1	MSU	retrotransposon protein, putative, unclassified	53.0	62.7	—	—	—	TEs	100.0	
11g05342	20 948 514- 20 951 349	-	5	547	No hits	—	—	—	—	—	PF12435	—	Hypothetical	11c3476g05	20 948 514- 20 951 349	-	yes	5	548	No hits	—	—	—	—	—	PF12435	—	Unknown	100.0	
11g05343	20 953 039- 20 953 482	+	1	147	BAD26378 .1	NCBI	hypothetical protein [Os jap.]	84.0	63.0	—	—	—	Hypothetical																	
11g05344	20 955 316- 20 958 782	-	6	371	LOC_Os11 g32360.1	MSU	expressed protein	100.0	99.3	SSF81383	—	GO:0005515	Hypothetical	11c3477g01	20 955 316- 20 958 782	-	yes	6	372	LOC_Os11 g32360.1	MSU	expressed protein	100.0	99.3	SSF81383	—	GO:0005515	Hypothetical	100.0	
11g05346	20 964 185- 20 964 397	-	1	70	BAD53449 .1	NCBI	hypothetical protein [Os jap.]	100.0	68.6	—	PF05754	—	Hypothetical																	
11g05347	20 972 678- 20 973 358	+	1	226	EAY81137. 1	NCBI	hypothetical protein Osl_36320 [Os ind.]	100.0	99.6	—	—	—	Hypothetical	11c3480g01	20 972 678- 20 973 358	+	yes	1	227	EAY81137. 1	NCBI	hypothetical protein Osl_36320 [Os ind.]	100.0	99.6	—	—	—	Hypothetical	100.0	
11g05348	20 975 587- 20 975 877	+	1	96	KAB81154 15.1	NCBI	hypothetical protein EE612_055867, partial [Oryza sativa]	82.0	100.0	—	—	—	Hypothetical	11c3480g02	20 975 587- 20 975 877	+	yes	1	97	KAB81154 15.1	NCBI	hypothetical protein EE612_055867, partial [Oryza sativa]	82.0	100.0	—	—	—	Hypothetical	100.0	
11g05349	20 977 419- 20 978 728	-	3	235	EAZ18562. 1	NCBI	hypothetical protein Osl_34090 [Os jap.]	100.0	98.7	—	—	—	Hypothetical	11c3480g03	20 977 419- 20 978 728	-	yes	3	236	EAZ18562. 1	NCBI	hypothetical protein Osl_34090 [Os jap.]	100.0	98.7	—	—	—	Hypothetical	100.0	
11g05350	20 979 766- 20 983 990	-	6	551	OslR64_1 1g022070	Eu Gene	FBD domain-containing protein	—	—	SSF52058 SSF81383	PF08387 PF00646	—	Domain- containing	11c3480g04	20 979 766- 20 983 990	-	yes	6	552	EEC68279. 1	NCBI	hypothetical protein Osl_36323 [Os ind.]	93.0	93.8	SSF81383 SSF52058	PF08387 PF00646	—	Domain- containing	100.0	

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.									
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group		
11g05351	20 984 364- 20 986 109	+	5	152	BAD36687	NCBI	hypothetical protein [Os jap.]	55.0	63.9	-	-	-	Hypothetical																		
11g05352	20 986 546- 20 992 458	-	3	1759	LOC_Os10 g10460.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	84.0	99.2	SSF53098 SSF57756 SSF56672 SSF50630	PF00098 PF17917 PF00078 PF03732 PF17921 PF08284	-	TEs																		
11g05353	20 992 713- 20 993 660	-	1	315	LOC_Os10 g10460.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	94.0	98.0	-	-	-	TEs																		
11g05354	20 994 378- 20 995 566	+	3	97	BAD36687	NCBI	hypothetical protein [Os jap.]	81.0	62.0	-	-	-	Hypothetical																		
11g05355	20 996 003- 21 001 915	-	3	1759	LOC_Os10 g10460.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	84.0	99.2	SSF53098 SSF57756 SSF56672 SSF50630	PF00098 PF17917 PF00078 PF03732 PF17921 PF08284	-	TEs																		
11g05356	21 002 170- 21 003 117	-	1	315	LOC_Os10 g10460.1	MSU	retrotransposon protein, putative, Ty3-gypsy subclass, expressed	94.0	98.0	-	-	-	TEs																		
11g05357	21 003 835- 21 005 023	+	3	97	BAD36687	NCBI	hypothetical protein [Os jap.]	81.0	62.0	-	-	-	Hypothetical																		
11g05359	21 009 003- 21 009 341	-	1	112	OslR64_1 g022090	Eu Gene	Conserved hypothetical protein	-	-	-	-	-	Hypothetical	11c3481g02	21 009 003- 21 009 341	-	yes	1	113	No hits	-	-	-	-	-	-	-	Unknown	100.0		
11g05360	21 010 146- 21 010 515	-	2	16	OslR64_1 g022090	Eu Gene	Conserved hypothetical protein	-	-	-	-	-	Hypothetical	11c3481g03	21 010 146- 21 010 515	-	yes	2	17	No hits	-	-	-	-	-	-	-	Unknown	100.0		
11g05362	21 013 737- 21 019 071	+	4	1206	EAY81143	NCBI	hypothetical protein Osl_36325 [Os ind.]	71.0	100.0	SSF51110 SSF56112	PF00069 PF01453	-	Protein kinases	11c3482g01	21 013 737- 21 015 281	+	no	2	474	EAY81143	NCBI	hypothetical protein Osl_36325 [Os ind.]	100.0	91.7	SSF51110	-	-	-	Domain- containing Protein kinases	75.0	
														11c3483g01	21 015 626- 21 018 638	+	no	4	432	EAY81143	NCBI	hypothetical protein Osl_36325 [Os ind.]	51.0	100.0	SSF56112	-	-	-			
11g05364	21 022 489- 21 022 797	-	1	102	EAY81144	NCBI	hypothetical protein Osl_36326 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	11c3484g02	21 022 489- 21 022 797	-	yes	1	103	EAY81144	NCBI	hypothetical protein Osl_36326 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	100.0		
11g05366	21 028 502- 21 034 731	+	5	1019	LOC_Os11 g32470.1	MSU	NEF1, putative, expressed	90.0	98.6	-	-	-	Growth and development	11c3486g01	21 028 944- 21 034 731	+	yes	5	873	EEE52205	NCBI	hypothetical protein Osl_34095 [Os jap.]	99.0	95.7	-	-	-	Hypothetical	85.6		
11g05367	21 035 717- 21 037 946	+	4	238	LOC_Os11 g32480.1	MSU	TPR repeat region family protein, expressed	78.0	97.8	SSF10991	PF03937	-	Domain- containing	11c3487g01	21 035 918- 21 037 946	+	yes	4	172	EAY81147	NCBI	hypothetical protein Osl_36329 [Os ind.]	100.0	99.4	SSF10991	PF03937	-	Domain- containing	71.9		
11g05368	21 038 577- 21 042 872	-	15	474	Osl11t052 8400-01	RAP- DB	Similar to aminopeptidase.	99.0	94.9	-	PF12576	-	Proteases/ protease inhibitors	11c3487g02	21 038 577- 21 042 872	-	yes	15	475	EEC68283	NCBI	hypothetical protein Osl_36330 [Os ind.]	99.0	95.3	-	PF12576	-	Hypothetical	100.0		
11g05369	21 044 494- 21 045 736	-	4	151	Osl11t052 8500-01	RAP- DB	Similar to Rubredoxin 1 (Rd-1).	100.0	100.0	SSF57802	-	-	Oxido- reductases																		
11g05370	21 046 656- 21 047 496	-	2	238	ABA94032	NCBI	GH3 auxin-responsive promoter family protein [Os jap.]	100.0	99.2	-	PF03321	-	Hormone- related																		
11g05371	21 050 911- 21 052 471	-	3	469	Osl11t052 8700-02	RAP- DB	Indole-3-acetic acid (IAA)-amido synthetase	100.0	99.6	-	PF03321	-	Hormone- related	11c3488g01	21 051 442- 21 052 261	-	no	3	223	Osl11t052 8700-01	RAP- DB	Indole-3-acetic acid (IAA)-amido synthetase, Plant architecture establishment, Drought tolerance	100.0	99.6	-	PF03321	-	Hormone- related	47.6		
11g05373	21 060 217- 21 064 199	+	2	122	LOC_Os11 g32530.1	MSU	retrotransposon protein, putative, unclassified, expressed	100.0	70.5	-	-	-	TEs	11c3489g01	21 061 649- 21 064 199	+	yes	3	134	LOC_Os11 g32530.1	MSU	retrotransposon protein, putative, unclassified, expressed	53.0	93.0	-	-	-	TEs	75.4		
11g05375	21 069 327- 21 070 857	+	4	227	No hits	-	-	-	-	PF03254	GO:0008107	Cell wall biogen./mod. GO:0042546																			

Supplementary Table S4.3 (continued)

IR64													CT8556-37-2-3-1-M													Sim.					
Gene ID	Pos. (bp)	S.	#	AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#	AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits		PF hits	GO terms	Group		
11g05376	21 071 388- 21 075 563	-	4	298	OslR64_1 1g022190	Eu Gene	Bisdemethoxycurcumin synthase	-	-	SSF53901	-	GO:0016746	Stress-related	11c3493g01	21 071 388- 21 072 337	-	yes	3	257	LOC_Os11 g32540.1	MSU	chalcone synthase, putative, expressed	69.0	83.6	-	PF02797 PF00195	-	Stress-related	} 99.1		
														11c3495g01	21 075 366- 21 077 269	-	yes	2	129	TVU24636. 1	NCBI	hypothetical protein EJ805_27085, partial [Eragrostis curvula]	51.0	59.7	-	PF00195	-	Stress-related			
11g05377	21 078 506- 21 079 446	-	3	123	EEE52209. 1	NCBI	hypothetical protein Osl_34103 [Os jap.]	50.0	90.2	-	-	-	Hypothetical	11c3495g02	21 078 506- 21 080 078	-	yes	2	108	EEE52209. 1	NCBI	hypothetical protein Osl_34103 [Os jap.]	95.0	90.2	-	-	-	Hypothetical	59.3		
11g05378	21 080 000- 21 081 011	+	4	145	AAX96857. 1	NCBI	hypothetical protein [Os jap.]	98.0	92.3	-	-	-	Hypothetical																		
11g05379	21 081 851- 21 083 324	-	6	205	LOC_Os11 g32570.1	MSU	expressed protein	90.0	62.4	-	-	-	Hypothetical	11c3496g01	21 081 024- 21 083 324	-	yes	7	185	LOC_Os11 g32570.1	MSU	expressed protein	82.0	58.1	-	-	-	Hypothetical	78.8		
11g05382	21 090 688- 21 092 169	-	2	413	Osl11t052 9500-00	RAP- DB	Polyketide synthase, type III domain containing protein.	100.0	99.0	SSF53901	-	GO:0016746	Stress-related	11c3497g02	21 090 688- 21 092 169	-	yes	2	414	Osl11t052 9500-00	RAP- DB	Polyketide synthase, type III domain containing protein.	100.0	99.0	SSF53901	-	GO:0016746	Stress-related	100.0		
11g05383	21 094 313- 21 095 667	+	6	213	LOC_Os11 g32590.1	MSU	transposon protein, putative, Mariner sub-class, expressed	97.0	68.9	-	-	-	TEs	11c3497g03	21 094 313- 21 094 879	+	no	3	126	LOC_Os11 g32590.1	MSU	transposon protein, putative, Mariner sub-class, expressed	83.0	96.2	-	-	-	TEs	59.2		
11g05384	21 096 040- 21 096 599	-	2	140	EAY81154. 1	NCBI	hypothetical protein Osl_36337 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	11c3498g01	21 096 040- 21 096 599	-	yes	2	141	EAY81154. 1	NCBI	hypothetical protein Osl_36337 [Os ind.]	100.0	100.0	-	-	-	Hypothetical	100.0		
11g05385	21 100 156- 21 102 097	+	2	401	LOC_Os11 g32610.1	MSU	chalcone and stilbene synthases, putative, expressed	100.0	99.8	-	PF02797 PF00195	-	Stress-related	11c3498g02	21 100 156- 21 101 692	+	no	2	205	LOC_Os11 g32610.1	MSU	chalcone and stilbene synthases, putative, expressed	91.0	100.0	-	PF00195	-	Stress-related	47.9		
11g05386	21 106 437- 21 107 943	-	2	375	XP_01561 6272.2	NCBI	bisdemethoxycurcumin synthase-like [Os jap.]	100.0	98.1	-	PF02797 PF00195	-	Stress-related	11c3499g01	21 106 402- 21 111 027	-	yes	2	376	XP_01561 6272.2	NCBI	bisdemethoxycurcumin synthase-like [Os jap.]	82.0	98.4	-	PF00195 PF02797	-	Stress-related	83.9		
11g05387	21 110 859- 21 111 257	+	1	132	BAT14277. 1	NCBI	Osl1g0530000 [Os jap.]	100.0	97.0	-	-	-	Hypothetical																		
11g05388	21 112 570- 21 112 956	+	1	128	OslR64_1 1g022260	Eu Gene	expressed protein	-	-	-	-	-	Hypothetical	11c3501g01	21 112 602- 21 114 309	+	yes	2	155	No hits	-	-	-	-	-	-	-	Unknown	} 95.8		
11g05389	21 113 792- 21 114 309	+	2	62	No hits	-	-	-	-	-	-	-	Unknown																		
11g05080	19 737 413- 19 739 630	-	3	135	No hits	-	-	-	-	-	-	-	Unknown																		
														11c3264g02	19 738 274- 19 742 198	+	yes	3	123	No hits	-	-	-	-	-	-	-	Unknown			
11g05081	19 739 952- 19 740 428	+	2	35	No hits	-	-	-	-	-	-	-	Unknown																		
11g05082	19 741 859- 19 742 137	-	1	92	No hits	-	-	-	-	-	-	-	Unknown																		
11g05085	19 748 304- 19 748 943	+	2	47	No hits	-	-	-	-	-	-	-	Unknown																		
														11c3264g05	19 753 486- 19 753 945	+	yes	2	77	No hits	-	-	-	-	-	-	-	Unknown			
11g05099	19 782 772- 19 786 264	-	3	83	No hits	-	-	-	-	-	-	-	Unknown	11c3269g01	19 780 760- 19 782 894	-	yes	2	58	No hits	-	-	-	-	-	-	-	Unknown	} 39.0		
														11c3269g02	19 785 370- 19 786 281	+	yes	2	126	No hits	-	-	-	-	-	-	-	Unknown			
														11c3269g03	19 788 023- 19 788 420	-	yes	1	132	No hits	-	-	-	-	-	-	-	Unknown			
11g05106	19 807 185- 19 807 382	+	1	65	No hits	-	-	-	-	-	-	-	Unknown																		
11g05117	19 849 320- 19 850 144	+	3	74	No hits	-	-	-	-	-	-	-	Unknown																		
11g05118	19 851 311- 19 852 403	-	2	77	No hits	-	-	-	-	-	-	-	Unknown	11c3289g02	19 850 661- 19 852 403	-	yes	2	86	No hits	-	-	-	-	-	-	-	Unknown	69.3		
11g05120	19 863 607- 19 864 502	-	3	249	No hits	-	-	-	-	-	-	-	Unknown																		
11g05123	19 867 610- 19 870 667	-	2	111	No hits	-	-	-	-	-	-	-	Unknown	11c3293g02	19 867 610- 19 871 198	-	yes	2	112	No hits	-	-	-	-	-	-	-	Unknown	} 81.1		
														11c3293g04	19 875 293- 19 875 679	-	yes	1	118	No hits	-	-	-	-	-	-	-	Unknown			

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.							
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group
11g05125	19 874 812- 19 879 844	+	6	355	No hits	-	-	-	-	-	-	-	Unknown	11c3293g05	19 876 247- 19 879 844	+	yes	5	142	No hits	-	-	-	-	-	-	-	Unknown	28.7
11g05126	19 883 535- 19 894 564	-	4	85	No hits	-	-	-	-	-	-	-	Unknown	11c3293g06	19 886 932- 19 888 138	-	yes	2	33	No hits	-	-	-	-	-	-	-	Unknown	
														11c3293g07	19 889 501- 19 891 632	+	yes	2	57	No hits	-	-	-	-	-	-	-	Unknown	
11g05127	19 895 055- 19 895 258	-	1	67	No hits	-	-	-	-	-	-	-	Unknown																
11g05134	19 936 930- 19 938 583	+	2	267	No hits	-	-	-	-	-	-	-	Unknown																
11g05135	19 942 080- 19 942 439	+	1	119	No hits	-	-	-	-	-	-	-	Unknown																
11g05136	19 943 010- 19 943 243	-	1	77	No hits	-	-	-	-	-	-	-	Unknown																
11g05137	19 943 647- 19 947 009	+	4	391	No hits	-	-	-	-	-	-	-	Unknown																
11g05138	19 949 920- 19 951 213	-	2	38	No hits	-	-	-	-	-	-	-	Unknown	11c3303g01	19 949 920- 19 951 213	-	yes	2	39	No hits	-	-	-	-	-	-	-	Unknown	100.0
														11c3303g02	19 957 256- 19 958 479	+	yes	3	131	No hits	-	-	-	-	-	-	-	Unknown	
11g05141	19 967 083- 19 967 583	-	1	166	No hits	-	-	-	-	-	-	-	Unknown																
11g05143	19 974 844- 19 977 879	-	4	263	No hits	-	-	-	-	-	-	-	Unknown	11c3305g01	19 974 844- 19 977 879	-	yes	5	276	No hits	-	-	-	-	-	-	-	Unknown	74.6
11g05144	19 982 257- 19 986 035	-	6	166	No hits	-	-	-	-	-	-	-	Unknown	11c3306g01	19 982 257- 19 986 035	-	yes	6	151	No hits	-	-	-	-	-	-	-	Unknown	86.1
11g05145	19 988 138- 19 988 422	+	1	94	No hits	-	-	-	-	-	-	-	Unknown	11c3306g02	19 988 138- 19 988 422	+	yes	1	95	No hits	-	-	-	-	-	-	-	Unknown	98.9
														11c3306g04	19 995 207- 19 995 821	+	yes	2	53	No hits	-	-	-	-	-	-	-	Unknown	
														11c3306g05	19 996 980- 19 998 668	-	yes	2	143	No hits	-	-	-	-	-	-	-	Unknown	
11g05148	20 001 967- 20 002 470	+	1	167	No hits	-	-	-	-	-	-	-	Unknown	11c3306g07	20 001 967- 20 002 470	+	yes	1	168	No hits	-	-	-	-	-	-	-	Unknown	97.6
11g05150	20 009 209- 20 011 370	-	2	66	No hits	-	-	-	-	-	-	-	Unknown																
11g05151	20 014 025- 20 015 765	-	2	254	No hits	-	-	-	-	-	-	-	Unknown	11c3307g01	20 015 319- 20 015 765	-	yes	1	149	No hits	-	-	-	-	-	-	-	Unknown	53.2
11g05152	20 018 138- 20 019 574	+	5	156	No hits	-	-	-	-	-	-	-	Unknown																
11g05153	20 020 357- 20 020 626	+	1	89	No hits	-	-	-	-	-	-	-	Unknown																
11g05161	20 051 348- 20 054 753	+	2	182	No hits	-	-	-	-	-	-	-	Unknown																
11g05162	20 057 879- 20 058 319	-	1	146	No hits	-	-	-	-	-	-	-	Unknown																
11g05163	20 061 218- 20 065 683	+	3	64	No hits	-	-	-	-	-	-	-	Unknown																
														11c3313g02	20 070 817- 20 072 054	+	yes	2	129	No hits	-	-	-	-	-	-	-	Unknown	
11g05165	20 071 452- 20 072 007	+	2	60	No hits	-	-	-	-	-	-	-	Unknown																
														11c3313g06	20 083 987- 20 084 827	-	yes	3	45	No hits	-	-	-	-	-	-	-	Unknown	
11g05176	20 121 939- 20 122 876	-	2	35	No hits	-	-	-	-	-	-	-	Unknown																

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M											Sim.										
Gene ID	Pos. (bp)	S.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.		Annotation	Cov.	I den.	SSF hits	PF hits	GO terms	Group			
														11c3431g01	20 688 193-	+	yes	1	43	No hits	-	-						Unknown				
11g05282	20 699 157-	-	4	226	No hits	-	-	-	-	-	-	-	Unknown	11c3431g06	20 688 321	-	yes	3	103	No hits	-	-						Unknown	37.6			
	20 706 381														20 699 157-																	
11g05284	20 716 186-	+	2	23	No hits	-	-	-	-	-	-	-	Unknown		20 701 673																	
	20 716 337																															
11g05285	20 724 505-	+	1	93	No hits	-	-	-	-	-	-	-	Unknown	11c3434g02	20 724 505-	+	yes	1	94	No hits	-	-						Unknown	100.0			
	20 724 786														20 724 786																	
11g05287	20 731 381-	+	2	191	No hits	-	-	-	-	-	-	-	Unknown																			
	20 735 541																															
11g05288	20 736 266-	+	1	162	No hits	-	-	-	-	-	-	-	Unknown	11c3435g01	20 736 266-	+	yes	2	216	No hits	-	-						Unknown	41.9			
	20 736 754														20 738 078																	
11g05290	20 750 220-	+	1	136	No hits	-	-	-	-	-	-	-	Unknown	11c3435g02	20 738 078-	+	yes	1	137	No hits	-	-						Unknown	100.0			
	20 750 630														20 738 078																	
11g05291	20 750 988-	-	2	42	No hits	-	-	-	-	-	-	-	Unknown	11c3437g01	20 753 919-	-	yes	2	45	No hits	-	-						Unknown	29.1			
	20 754 206														20 754 206																	
														11c3437g02	20 757 909-	+	yes	2	83	No hits	-	-						Unknown				
															20 759 210																	
														11c3437g05	20 776 321-	+	yes	1	99	No hits	-	-						Unknown				
															20 776 617																	
11g05301	20 806 964-	-	1	122	No hits	-	-	-	-	-	-	-	Unknown																			
	20 807 332																															
11g05304	20 815 450-	+	2	54	No hits	-	-	-	-	-	-	-	Unknown	11c3447g02	20 815 450-	+	yes	2	55	No hits	-	-						Unknown	100.0			
	20 815 767														20 815 767																	
11g05306	20 827 773-	-	2	29	No hits	-	-	-	-	-	-	-	Unknown																			
	20 828 901																															
11g05317	20 858 234-	+	1	93	No hits	-	-	-	-	-	-	-	Unknown	11c3461g01	20 858 234-	+	yes	1	94	No hits	-	-						Unknown	100.0			
	20 858 515														20 858 515																	
11g05328	20 899 540-	+	1	121	No hits	-	-	-	-	-	-	-	Unknown																			
	20 899 905																															
11g05341	20 945 224-	-	4	195	No hits	-	-	-	-	-	-	-	Unknown	11c3476g04	20 945 224-	-	yes	4	196	No hits	-	-						Unknown	100.0			
	20 947 826														20 947 826																	
														11c3476g06	20 952 234-	+	yes	2	41	No hits	-	-						Unknown				
11g05345	20 960 865-	+	3	54	No hits	-	-	-	-	-	-	-	Unknown	11c3477g02	20 952 514																	
	20 963 196														20 960 865-																	
														11c3479g01	20 963 196																	
															20 971 093-																	
11g05358	21 005 460-	-	4	404	No hits	-	-	-	-	-	-	-	Unknown		20 972 184																	
	21 008 237																															
														11c3481g01	21 006 804-	+	yes	3	207	No hits	-	-						Unknown				
															21 008 314																	
11g05361	21 011 271-	-	2	40	No hits	-	-	-	-	-	-	-	Unknown																			
	21 011 633																															
11g05363	21 019 846-	+	3	190	No hits	-	-	-	-	-	-	-	Unknown	11c3484g01	21 019 846-	+	yes	3	191	No hits	-	-						Unknown	100.0			
	21 020 659														21 020 659																	
11g05365	21 026 147-	-	3	222	No hits	-	-	-	-	-	-	-	Unknown	11c3485g01	21 026 147-	-	yes	2	226	No hits	-	-						Unknown	83.8			
	21 027 310														21 026 987																	
														11c3485g02	21 028 432-	-	yes	1	108	No hits	-	-						Unknown				
															21 028 755																	
11g05372	21 059 498-	+	2	24	No hits	-	-	-	-	-	-	-	Unknown	11c3488g02	21 059 498-	+	yes	2	25	No hits	-	-						Unknown	100.0			
	21 059 639														21 059 639																	
11g05374	21 068 115-	-	1	190	No hits	-	-	-	-	-	-	-	Unknown																			
	21 068 687																															
11g05380	21 085 697-	+	1	103	No hits	-	-	-	-	-	-	-	Unknown	11c3496g02	21 085 697-	+	yes	1	104	No hits	-	-						Unknown	100.0			
	21 086 008														21 086 008																	
11g05381	21 086 891-	-	1	90	No hits	-	-	-	-	-	-	-	Unknown																			
	21 087 163																															

Supplementary Table S4.3 (continued)

IR64											CT8556-37-2-3-1-M																		
Gene ID	Pos. (bp)	S. #E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Gene ID	Pos. (bp)	S.	C.	#E	#AA	BLAST hit	DB.	Annotation	Cov.	Iden.	SSF hits	PF hits	GO terms	Group	Sim.	
11g05390	21 117 927-	-	2	65	No hits	-	-	-	-	-	-	Unknown	11c3497g01	21 087 116-	+	yes	3	84	No hits	-	-	-	-	-	-	-	-	Unknown	
	21 118 554													21 089 568															
11g05391	21 122 419-	-	2	31	No hits	-	-	-	-	-	-	Unknown	11c3502g01	21 117 927-	-	yes	3	125	No hits	-	-	-	-	-	-	-	Unknown	49.2	
	21 123 859													21 123 859															
													11c3502g02	21 125 637-	-	yes	3	70	No hits	-	-	-	-	-	-	-	Unknown		
														21 127 066															

Supplementary Table S5.1 Details of the *Oryza glaberrima* genotypes used.

Name *	Code †	Country of origin	IITA TOG accession number	IRRI IRGC accession number	ORSTOM ID
6	EM	Nigeria	5307	96726	—
7	KH	Nigeria	5314	—	—
8	DT	Nigeria	5321	—	—
10	KI	Nigeria	5326	—	—
12	GQ	Nigeria	5390	96740	—
13	LL	Nigeria	5418	—	—
15	KS	Nigeria	5424	—	—
17	HK	Nigeria	5429	112564	—
22	MA	Nigeria	5453	—	—
28	MD	Nigeria	5486	86752	—
30	ME	Nigeria	5494	—	—
32	MF	Nigeria	5500	—	—
36	KT	Nigeria	5556	—	—
41	HA	Ghana	5620	86764	—
43	GL	Nigeria	5666	—	—
44	MH	Nigeria	5672	—	—
46	DV	Nigeria	5674	96790	—
49	—	Nigeria	5681	96793	—
56	MT	Nigeria	5814	—	—
57	KV	Liberia	5815	86785	—
60	—	Nigeria	5882	96809	—
61	HL	Liberia	5885	112568	—
62	KL	Liberia	5887	86789	—
63	HP	Liberia	5919	86790	—
64	KD	Liberia	5923	86791	—
65	MV	Nigeria	5953	—	—
66	NA	Nigeria	5969	—	—
80	NB	Guinea	6203	86806	—
81	NC	Guinea	6205	—	—
82	ND	Zimbabwe	6206	96841	—
83	NE	Zimbabwe	6207	—	—
86	—	—	—	—	—
87	NF	Burkina Faso	6220	112577	—
89	NG	Mali	6231	96854	—
94	NH	Liberia	6308	—	—
97	NI	Liberia	6356	115598	—
100	IN	Liberia	6603	—	—
102	LE	Liberia	6688	—	—
103	LF	Liberia	6698	—	—
105	LG	Sierra Leone	6943	—	—
106	LH	Sierra Leone	6951	—	—
108	LK	Sierra Leone	7047	—	—
109	HE	Mali	7106	—	—
110	EA	Mali	7108	—	—
111	GV	Nigeria	5286	—	—
112	FF	Nigeria	5400	96741	—

Supplementary Table S5.1 (continued)

Name *	Code †	Country of origin	IITA TOG accession number	IRRI IRGC accession number	ORSTOM ID
113	IA	Nigeria	5439	—	—
114	FT	Mali	7115	103512	OR09461
115	GM	Nigeria	5464	—	—
116	FQ	Nigeria	5533	—	—
117	FB	Nigeria	5566	—	—
118	KE	Ghana	5591	—	—
119	ID	Nigeria	5639	—	—
121	GR	Senegal	7131	103439	OR09809
122	FL	Senegal	7132	—	—
124	LA	Senegal	7134	—	—
127	IC	Liberia	5775	—	—
129	FR	Nigeria	5997	—	—
130	LM	Sierra Leone	7420	—	—
132	LD	Mali	MG12	103544	OR09517
133	II	Mali	RAM131	—	—
134	LN	Mali	RAM137	—	—
136	IR	Mali	RAM48	—	—
137	KB	Mali	RAM55	—	—
138	IE	Mali	RAM77	—	—
142	—	Senegal	CG14	96717	OR09814
150	EC	Côte d'Ivoire	7206	103490	OR09271
151	ED	Guinea	14367	—	—
154	GE	Guinea	12444	104194	OR11161
155	LP	Mali	7213	103538	OR09511
156	GP	Mali	7214	103539	OR09512
158	HM	Senegal	7172 7173	103461 96893	OR10025
159	FP	Mali	7219	—	—
160	EE	Mali	7235	103549	OR09522
161	LQ	Côte d'Ivoire	10434	—	—
162	LR	Chad	7255	—	—
163	EF	Nigeria	12086	—	—
164	EG	Nigeria	12160	—	—
165	FN	Nigeria	12188	—	—
166	KQ	Nigeria	12249	—	—
168	FC	Cameroon	7273	—	—
169	HC	Cameroon	7274	—	—
172	EH	Burkina Faso	7291	104589 103469	—
178	HV	Ghana	7402	86826	—
179	FE	Ghana	7406	—	—
181	EP	Burkina Faso	7451	—	—
182	HR	Burkina Faso	7455	—	—
183	FG	Burkina Faso	7456	—	—
184	HT	Côte d'Ivoire	7554	—	—
185	EN	Nigeria	7993	—	—
186	GS	Nigeria	8049	—	—
187	EI	Gambia	8527	—	—

Supplementary Table S5.1 (continued)

Name *	Code †	Country of origin	IITA TOG accession number	IRRI IRGC accession number	ORSTOM ID
188	GK	Gambia	8537	—	—
190	HF	Gambia	8545	—	—
191	HD	Côte d'Ivoire	9524	—	—
200	—	—	—	—	—
205	LT	Guinea Bissau	12366	—	—
206	FM	Ghana	12372	—	—
208	EK	Tanzania	12387	—	—
209	IQ	Cameroon	12388	—	—
212	EV	Guinea	12399	—	—
213	HI	Guinea	12401	—	—
216	KP	Guinea	12411	—	—
217	LV	Guinea	12414	—	—
225	MB	Guinea	12439	104190	OR11157
231	—	Guinea	13645	—	—
234	EQ	Guinea	13708	—	—
238	IB	Guinea	14093	—	—
239	FD	Liberia	14116	—	—
243	GN	Zimbabwe	14184	—	—
253	MC	Guinea	14359	105021 115592	OR11188
256	GA	Guinea	14361	—	—
258	GF	Guinea	14364	105052	OR11195
264	IP	Guinea	14373	105043	OR11217
274	LC	Guinea	14602	105007	OR11172
276	ES	Guinea	14606	105051	OR11185
280	GI	Guinea	14610	105028	OR11198
297	GH	Côte d'Ivoire	7190 7191	103475 104573	OR09256
400	FH	Côte d'Ivoire	7190	103475 104573	OR09256
401	KA	Côte d'Ivoire	7300	103476	OR09257
402	GD	Côte d'Ivoire	7193	103478	OR09259
403	FV	Côte d'Ivoire	7194	103479	OR09260
404	HG	Côte d'Ivoire	7195	103480	OR09261
405	HS	Côte d'Ivoire	7196	103481	OR09262
407	—	Côte d'Ivoire	7198	103483	OR09264
408	HH	Côte d'Ivoire	7199	103484	OR09265
410	ET	Côte d'Ivoire	7201	103486	OR09267
411	EL	Côte d'Ivoire	7202	103487	OR09268
412	GT	Côte d'Ivoire	7203	103488	OR09269
415	IT	Côte d'Ivoire	7210	103492	OR09273
416	KK	Côte d'Ivoire	7211	103493	OR09274
418	IL	Côte d'Ivoire	12361	104035	OR09284
423	KC	Tanzania	—	104018	OR09347
425	IK	Tanzania	—	104020	OR09349
426	HB	Cameroon	—	103593	OR09400
427	IM	Cameroon	7266	103594	OR09401
428	MG	Cameroon	7268	103596	OR09404
429	KM	Cameroon	7273 7274	103598 104582 104583	OR09406

Supplementary Table S5.1 (continued)

Name*	Code†	Country of origin	IITA TOG accession number	IRRI IRGC accession number	ORSTOM ID
430	GB	Cameroon	—	104048	OR09407
431	FS	Cameroon	—	104049	OR09408
438-2	—	Mali	—	103499	OR09448_S
443-2	MI	Mali	7124	103523	OR09472_B
448	MK	Mali	—	—	OR09481
457	MM	Mali	—	—	OR09533
459	MN	Chad	—	104041	OR09666
460-1	—	Chad	7259	103586 104578 104588	OR09670
461	MP	Chad	7260 7261	103587 104579 104580	OR09671
463	LB	Chad	—	104045	OR09674
474-2	GG	Senegal	7147	103448	OR09847_G
475	IV	Senegal	—	—	OR09849
478	HQ	Senegal	—	—	OR09878
486	MQ	Senegal	—	—	OR09986
487	FI	Senegal	7167 7168	103459	OR09992
488	IG	Senegal	—	—	OR10006
489	FA	Senegal	—	—	OR10022
491	KN	Senegal	—	—	OR10038
492	IS	Senegal	—	—	OR10043
495	KF	Senegal	—	103570	OR10061
496	FK	Senegal	—	—	OR10067
498	MR	Senegal	—	—	OR10155
502	MS	Guinea	—	104174	OR11128
505	KG	Guinea	—	104185	OR11152
506-2	ER	Guinea	12436	104187	OR11154

* Names of genotypes from Orjuela *et al.* (2014).

† Codes of genotypes from Cubry *et al.* (2018).

IITA, The International Institute of Tropical Agriculture, Nigeria.

IRRI, The International Rice Research Institute, the Philippines.

ORSTOM, Office de la Recherche Scientifique et Technique Outre-mer (now Institut de Recherche pour le Développement, IRD, France).