The University Of Sheffield.

Discovery and analysis of novel resistance genes in rice to the parasitic weed *Striga hermonthica*

By:

Emily G Beardon

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

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

September 2017

Acknowledgements

Firstly I would like to thank my supervisor Professor Julie Scholes for accepting me onto the project, and her continuous advice and support throughout my PhD, and my supervisor Dr. Mathias Lorieux for his time, advice and training. I would like to thank them both for reading my thesis and their guidance and suggestions for improvements. Thanks also to Professor Jon Slate for his help and advice with statistics.

I am very grateful to the UK Biotechnology and Biological Sciences Research Council (BBSRC) for their funding, and the White Rose Doctoral Training Partnership for their training. Thank you to Dr. Charles Lane at the Food and Environment Research Agency (Fera) for hosting me on my internship and giving me the opportunity to learn new skills and apply my knowledge outside of academia.

In lab C45 I would like to thank Nic White for her daily assistance in the smooth running of the lab, which allowed me to carry out my research efficiently. I would also like to say a huge thank you to my colleague and friend Alexis Moschopoulos, a former postdoc on the project, for his patience, help with experiments and endless science discussions from which I learnt a great deal. Likewise, a huge thank you to Peijun Zhang, a friend and lab member, for all her assistance, as well as James Bradley and numerous other members of the lab over the last few years.

I am very lucky to have so many friends, both in Sheffield and away, who have supported and encouraged me throughout my PhD, and kept me smiling even when things were tough. You know who you are, and thank you. Finally, I would like to thank my boyfriend Roshan, my brothers and sister and especially my Mum, for supporting me no matter what, and my late Dad who inspired and encouraged me to apply for University in the first place.

Declaration

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

Abstract

Striga species are obligate hemi-parasitic plants that infect the roots of maize, sorghum, millet and upland rice in sub-Saharan Africa, causing reductions in yield that range from ~ 30 % to total crop failure. Improving the yields of cereals in the presence of *Striga* is difficult as the parasite begins to negatively affect the growth and development of the host immediately after attachment. By the time *Striga* emerges above ground, the crop is often severely damaged. The use of resistant cultivars would provide a low cost and effective form of control that would target the parasite before the impacts on growth and development of the host crop can occur. The aim of this thesis was to identify novel *Striga hermonthica* resistance Quantitative Trait Loci (QTL) and underlying resistance genes in rice, and to functionally validate their role in providing resistance using comparative and functional genomic approaches.

A Recombinant Inbred Line population of rice derived from a cross between the resistant cultivar IR64 (*Oryza sativa* ssp. *indica*) and the susceptible cultivar Azucena (*O. sativa* ssp. *japonica*) was phenotyped for post-attachment resistance to *S. hermonthica* and a QTL analysis performed. A major QTL on chromosome 12 was identified. This QTL mapped to the same position as a previously identified QTL for *S. hermonthica* resistance in the temperate *O. sativa* spp *japonica* cultivar Nipponbare, suggesting that resistance to *S. hermonthica* in these two cultivars may be governed by the same (or similar) genes. Bioinformatics tools and gene prediction software were used to identify genes present within the IR64 QTL. These genes comprised transposable elements, expressed and hypothetical proteins and a cluster of genes predicted to encode cell surface receptor-like proteins (RLPs) annotated as orthologs of genes conferring resistance to *Verticillium* wilt in tomato. *Verticillium* wilts are xylem invading fungal pathogens with a remarkably similar infection strategy and lifestyle to *Striga* species, making these genes top candidates for *S. hermonthica* resistance genes. A cluster of highly similar RLP genes is also present in the Nipponbare *S. hermonthica* resistance QTL.

RNAi lines where suites of these RLP genes had been down-regulated in Nipponbare were phenotyped for *S. hermonthica* resistance, and gene expression examined by qPCR. Increased susceptibility was associated with suppression of multiple RLP genes, and could not be associated with suppression of a single gene. *Tos17* and T-DNA insertion lines targeting 4 of the genes independently showed no increase in susceptibility, suggesting that these genes did not underlie the resistance, that there was functional redundancy or that multiple RLP genes may act together to confer resistance to *S. hermonthica* in this cultivar.

Finally, the genetic diversity of the candidate resistance genes in the QTL were examined across a range of diverse rice cultivars to determine if it was possible to exploit their diversity to reduce the number of candidate resistance genes. Cultivars were phenotyped for resistance to *S. hermonthica* and a correlation analysis performed for each gene between *S. hermonthica* resistance and similarity of the gene to the Nipponbare allele. A good correlation was observed for a number of candidate genes, and a region of the QTL was identified that was more likely to be involved in *S. hermonthica* resistance.

These results compiled across experiments provide evidence for the involvement of the RLP resistance genes in contributing to the resistance to *S. hermonthica* seen in the rice cultivars IR64 and Nipponbare.

Abbreviations and Acronyms

ABA	abscisic acid
ALS	acetolactate synthase
AM	arbuscular mycorrhiza
ANOVA	analysis of variance
Avr	avirulence gene
BIL	backcross inbred line
CC	coiled-coil
CIAT	International Centre for Tropical Agriculture
CRIPSR	clustered regularly interspaced short palindromic repeats
CVN	copy number variant
dai	days after inoculation
das	days after sowing
DMBQ	2,6-dimethoxybenzoquinone
ET	ethylene
ETI	effector-triggered immunity
ETS	effector-triggered susceptibility
GOI	gene of interest
GWAS	genome-wide association study
HIF	haustorial inducing factor
HR	hypersensitive response
JA	jasmonic acid
LOD	logarithm of the odds
LRR	leucine rich repeat
MAMPs	microbe associated molecular patterns
miRNA	microRNA
MSU	Michigan State University
mya	million years ago
NAM	nested association mapping
NBS	nucleotide binding site
NCBI	The National Center for Biotechnology Information
NERICA	New Rice for Africa
NIAS	The National Institute of Agrobiological Science, Japan
NIL	near isogenic line
PAMPs	pathogen associated molecular patterns
PCR	polymerase chain reaction
PR proteins	pathogenesis-related proteins
PRR	pattern recognition receptor
PTI	PAMP-triggered immunity
qPCR	quantitative PCR
QTL	quantitative trait locus / loci
R gene	resistance gene
RALF	Rapid ALkalinization Factor
RdDM	RNA-directed DNA methylation
RIL	recombinant inbred line
RLP	receptor-like protein
RNAi	RNA interference

ROS	reactive oxygen species
SA	salicylic acid
SG3	*Striga gesnerioides* race 3
siRNAs	small interfering RNAs
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
wpi	weeks post inoculation

Table of contents

Chapter 1 ..... **16**
1.1 Introduction to parasitic plants **17**
1.2 The life cycle of *Striga* **19**
1.3 Host responses to *Striga* infection **20**
1.4 Current strategies for controlling *Striga* **22**
1.5 How do plants defend themselves against parasitic weeds? **25**
1.5.1 Pre-attachment resistance 25
1.5.2 Post-attachment resistance 26
1.5.3 The role of signalling molecules 28
1.6 The genetic basis of *Striga* resistance **29**
1.7 Genetic mapping to identify *Striga* resistance QTL **37**
1.8 Rice as a model system for studying *Striga* resistance **38**
1.9 Aims and Objectives of thesis **39**

Chapter 2 ..... **40**
2.1 Introduction **41**
2.1.1 Aim of Chapter 2 43
2.2 Materials and Methods **45**
2.2.1 Plant materials 45
2.2.2 Phenotyping NAM parental lines for post-attachment resistance to *S. hermonthica* 45
2.2.3 QTL mapping strategy and analysis 46
2.2.4 Generation of F1 progeny by cross pollination 47
2.2.5 The phenotype of resistance 50
2.2.6 Statistical analysis 50
2.3 Results **52**
2.3.1 Evaluation of post-attachment resistance of NAM parents to *Striga hermonthica* .. 52
2.3.2 The phenotype of resistance in IR64 52
2.3.3 Identification of a *Striga hermonthica* resistance QTL 55
2.3.4 The mode of inheritance of the resistance to *Striga hermonthica* in IR64 and Nipponbare 61
2.4 Discussion **63**
2.4.1 What is the phenotype of resistance in IR64? 63
2.4.2. Identification of a *S. hermonthica* resistance QTL 64
2.4.3 What is the genetic basis of resistance in IR64 to *S. hermonthica*? 66
2.4.4 Conclusions 67

Chapter 3 ..... **68**
3.1 Introduction **69**
3.1.1 Aim of Chapter 3 72
3.2 Materials and Methods **73**
3.2.1 Genome sequences 73
3.2.2 Comparison of the genome structure spanning the *Striga hermonthica* resistance QTL region in rice cultivars 73
3.2.3 Prediction of genes in within the IR64 *S. hermonthica* resistance QTL and comparison with genes in the Nipponbare *S. hermonthica* resistance QTL 74

3.2.4 Cloning and sequencing of IR64 candidate resistance genes 74
3.2.4.1 DNA extraction 74
3.2.4.2 PCR amplification of candidate *S. hermonthica* RLP genes in IR64 75
3.2.4.3 Cloning of candidate IR64 resistance genes into pDONR/Zeo 76
3.2.5 Amplification and sequencing of homologs of candidate *S. hermonthica* resistance genes from Azucena 78
3.2.6 Growth and collection of IR64 root material for measurement of the expression of candidate resistance genes 78
3.2.7 RNA extraction, cDNA synthesis and qPCR to measure the expression of candidate resistance genes in IR64 following infection with *S. hermonthica* 79
3.2.9 Analysis of qPCR data 80
3.3 Results 82
3.3.1 Comparison of the genome structure of the *S. hermonthica* resistance QTL between rice cultivars 82
3.3.2 The IR64 QTL contains homologs of candidate *S. hermonthica* resistance genes in Nipponbare 82
3.3.3 Koshihikari contains homologs of candidate *S. hermonthica* resistance genes in Nipponbare 94
3.3.4 Azucena shares 3 homologs of candidate *S. hermonthica* resistance genes with IR64 and Nipponbare 98
3.3.5 Comparison of candidate *S. hermonthica* resistance RLP genes between cultivars 100
3.3.6 Comparison of candidate *S. hermonthica* resistance gene in IR64 with RLP genes in different plant species 106
3.3.7 The phenotype of resistance during *S. hermonthica* infection on IR64 109
3.3.8 Gene expression analysis of candidate *S. hermonthica* resistance on infected and uninfected IR64 root tissue 109
3.4 Discussion 112
3.4.1 Conclusion 119
Chapter 4 121
4.1 Introduction 122
4.1.1 Aims of Chapter 4 125
4.2 Materials and Methods 127
4.2.1 Plant materials 127
4.2.2 Phenotyping RNAi lines for post-attachment resistance to *S. hermonthica* 127
4.2.3 Growth and collection of RNAi root material for analysis of gene expression by qPCR 129
4.2.4 Confirming the presence of the insertion in the target genes and phenotyping the insertion lines for post-attachment resistance to *S. hermonthica* 129
4.2.5 Statistical analysis 131
4.3 Results 132
4.3.1 Phenotyping RNAi lines for resistance to *S. hermonthica* and examination of RLP gene expression 132
4.3.2 The efficiency of gene suppression for RNAi constructs 136
4.3.3 Insertion lines exhibited no increase in susceptibility to *S. hermonthica* 137
4.3.4 Verification of gene knockouts for *Tos17* and T-DNA insertion lines 137
4.4 Discussion 143
4.4.1 The role of the RLP genes in *S. hermonthica* resistance 143

4.4.2 The suppression of RLP transcript in RNAi lines 145
4.4.3 Conclusions and future directions 147

Chapter 5 149
5.1 Introduction 150
5.1.1 Diversity of the genus *Oryza* 150
5.1.2 Aims of Chapter 5 154
5.2 Materials and Methods 155
5.2.1 Plant materials 155
5.2.2 Comparison of SNPs between cultivars and predictions of resistance and susceptibility to *S. hermonthica* 155
5.2.3 Phenotyping rice cultivars for post-attachment resistance to *S. hermonthica* 157
5.2.4 DNA extraction and quantification, and genotyping of cultivars by PCR amplification 158
5.2.5 PCR amplification and sequencing of full-length genes from rice cultivars 158
5.2.6 Statistical analysis 159
5.3 Results 161
5.3.1 The genomic structure of the *S. hermonthica* resistance QTL in diverse rice cultivars, and predictions of resistance and susceptibility 161
5.3.2 Evaluation of post-attachment resistance of rice cultivars to *Striga hermonthica* . 166
5.3.3 The relationship between *S. hermonthica* resistance and SNP similarity to Nipponbare 169
5.3.4 Verification of SNP data: genotyping by PCR amplification of marker genes in different rice cultivars 182
5.3.5 Amplification and sequencing of candidate *S. hermonthica* resistance genes from rice cultivars 187
5.4 Discussion 194
5.4.1 Can any of the candidate resistance genes be discounted as being involved in resistance to *S. hermonthica*? 194
5.4.2 Interpretation of the data in the light of the assumptions 197
5.4.3 Other levels of complexity: post transcriptional and post-translational modifications 199
5.4.4 Conclusions and future directions 200

Chapter 6 202
6.1 Introduction 203
6.2 Were novel QTL identified in rice for *S. hermonthica* resistance? 203
6.3 What is the evidence that the candidate RLP genes underlie resistance to *S. hermonthica*? 206
6.4 The chromosome 12 QTL: a complex locus 209
6.5 Conclusions and future perspectives 211

References 212

Appendix 229

List of figures

Figure		Page
1.1	The distribution of the three most destructive *Striga* species in Africa	19
1.2	The life cycle of *Striga*	21
1.3	Phenotype of rice cultivars Nipponbare (A) and Koshikihari (B) infected with *Striga hermonthica*. (C) QTL analysis for *S. hermonthica* resistance of the Nipponbare x Koshihikari BIL population	33
1.4	The resistance of the BC_4F_5 lines from the Nipponbare x Koshihikari BIL population showing recombination in and around the QTL region on chromosome 12	34
1.5	Comparison of the genome structure of rice varieties Nipponbare and Koshihikari within the *S. hermonthica* resistance QTL on chromosome 12	35
1.6	Comparison of a near isogenic line (A) and a recombinanat inbred line (B) mapping population	38
2.1	Development of a rice Nested Association Mapping population	44
2.2	A) Rice seeds germinating between sheets of filter paper before being transferred to rhizotrons. B) *S. hermonthica*-infected rice plants growing in rhizotrons, watered by an automatic watering system	49
2.3	Cross-pollination of rice	51
2.4	The dry biomass (A) mean number (B) and mean cumulative length (c) of *S. hermonthica* plants harvested from parents of a Nested Association Mapping population 21 dai with germinated *S. hermonthica*	53
2.5	The phenotype of resistance and susceptibility of IR64 and Azucena rice cultivars to *S. hermonthica*	54
2.6	Distribution of resistance to *S. hermonthica* 21 dai in the rice Recombinant Inbred Line IR64 x Azucena mapping population	56
2.7	Logarithm of odds (LOD) scores for *S. hermonthica* resistance across all rice chromosomes as determined by a one-way ANOVA method, on an IR64 x Azucena RIL population	57
2.8	Logarithm of odds (LOD) scores for *S. hermonthica* resistance across all rice chromosomes as determined by a Kruskall Wallis method	58
2.9	Major QTL for *S. hermonthica* resistance in *O. sativa* cultivar IR64, located on chromosome 12 with a peak at 32cM (Interval Mapping)	59
2.10	Resistance of 14 lines of the RIL population showing recombination in and around the QTL region	60

2.11 Evaluation of post-attachment resistance of Azucena, IR64, Nipponbare, Koshihikari and the F_1 hybrids to *S. hermonthica* 62

3.1 The Invasion Model of plant immunity to illustrate an attempted plant-invader symbiosis 71

3.2 Structural comparison of the *S. hermonthica* resistance QTL between rice cultivars IR64, Nipponbare and the susceptible cultivar Koshihikari 83

3.3 The number of genes and proportion of different gene families present in the IR64 and Nipponbare *S. hermonthica* resistance QTL between positions 5.7-6.7 Mbp 85

3.4 IR64 gene prediction models for genes showing homology to candidate *S. hermonthica* resistance genes in Nipponbare 88

3.5 Position of IR64 gene predictions on the PacBio contig for genes showing homology to candidate *S. hermonthica* resistance RLP genes in Nipponbare 89

3.6 Nucleic acid sequenced alignment from part of the IR64 homolog to Os12g10870 90

3.7 Amino acid alignment for 7 predicted IR64 homologs of candidate *S. hermonthica* resistance RLP genes in Nipponbare 91-92

3.8 Phylogenetic relationship of 13 RLP genes in the *S. hermonthica* resistance QTL for the rice cultivar Nipponbare, and the 7 identified homologs in the IR64 *S. hermonthica* resistance QTL 93

3.9 Koshihikari gene prediction models for 4 genes showing homology to candidate *S. hermonthica* resistance genes in Nipponbare 96

3.10 Amino acid alignment for 4 predicted Koshihikari homologs of candidate *S. hermonthica* resistance RLP genes in Nipponbare 97-98

3.11 Amplification of full-length candidate *S. hermonthica* resistance genes from IR64 and Azucena 99

3.12 Amino acid alignment of Os12g10870 from Nipponbare and the IR64 homolog 102

3.13 Amino acid alignment of Os12g11370 from Nipponbare and the IR64 and Koshihikari homologs 103

3.14 Multiple Amino acid alignment of Os12g11500 with IR64, Azucena and Koshihikari homologs 104 – 105

3.15 Amino acid alignment of IR64_h_Os12g11930 aligned to the end sequence of both Os12g11500 and Os12g11930 from Nipponbare 105

3.16 Phylogenetic relationship of RLP genes in the S. *hermonthica* resistance QTL from the rice cultivar IR64 and 5 homologous RLP genes 108

3.17 Time course of the phenotype of infection of *S. hermonthica* on the roots 110

resistant rice cultivar IR64 and the susceptible cultivar Azucena at 2, 4, 6, 8 and 10 dai

3.18 Time course of relative gene expression quantified by qPCR of candidate *S. hermonthica* resistance genes in *S. hermonthica* infected and uninfected root tissue of the resistant rice cultivar IR64 111

3.19 Domain structure of a typical receptor-like protein (RLP) (A) and comparison with a receptor-like kinase 113

4.1 RNAi–mediated gene silencing 124

4.2 Gateway pANDA vector for suppression of gene function in rice by RNAi 124

4.3 The dry biomass of *S. hermonthica* (A) and differential expression of RLP genes (B-E) for RNAi rice lines targeting the gene Os12g11500, and empty vector controls 133

4.4 The dry biomass of *S. hermonthica* (A) and differential expression of RLP genes (B-E) for RNAi rice lines targeting the gene Os12g11680 134

4.5 The dry biomass of *S. hermonthica* (A) and differential expression of RLP genes (B-C) for RNAi rice lines targeting the gene Os12g11370 135

4.6 The phenotype of resistance and susceptibility of rice cultivars Nipponbare, Koshihikari and the susceptible RNAi line JS8.14-19 138

4.7 The efficiency of gene silencing for RNAi constructs, determined by the relationship between the expression of the *gus* transgene derived from the linker region of the construct, and the expression of the primary gene target 139

4.8 The relationship between the sum of total RLP gene expression for each RNAi line and resistance (mg dry weight) to *S. hermonthica* (A) and expression of the *gus*-linker transcript from the pANDA vector (B) 140

4.9 *S. hermonthica* dry weight harvested from the roots of lines carrying insertion in the genes Os12g10870, Os12g1168, Os12g12120 (all *Tos17* insertions) or Os12g11370 (T-DNA insertion) 25 dai 141

4.10 Verification of gene knockout for insertion lines by PCR amplification of cDNA 142

4.11 Sequenced products of Os12g11370 T-DNA insertion line from cDNA 142

5.1 Classification of 3000 rice genomes into 5 varietal groups 152

5.2 Proposed origins for domesticated rice derived from phylogenetic analysis 153

5.3 Genomic structure of 112 rice accessions in the *S. hermonthica* resistance QTL (6.212 – 6.561 Mb) obtained from SNP sequencing data 163

5.4 The phenotype of resistance to *S. hermonthica* for 26 genetically diverse rice cultivars 167

5.5	The root systems of 4 rice cultivars, Curinga (tropical *japonica*, susceptible), Jhona349 (*aus*, intermediate), Firooz (*aromatic*, resistant) and Nipponbare (temperate *japonica*, resistant) 21 dai with *S. hermonthica*	168
5.6	The relationship between *S. hermonthica* dry biomass of 26 diverse rice cultivars and SNP similarity to the Nipponbare allele for 6 genes in the *S. hermonthica* resistance QTL, showing a significant negative correlation between SNP similarity and resistance	170
5.7	The relationship between *S. hermonthica* dry biomass of 26 diverse rice cultivars and SNP similarity to the Nipponbare allele for 9 genes across the *S. hermonthica* resistance QTL, showing no significant correlation between *S. hermonthica* resistance and similarity to Nipponbare	172-173
5.8	The relationship between *S. hermonthica* dry biomass of 26 diverse rice cultivars and SNP similarity to the Nipponbare allele for 4 genes in the *S. hermonthica* resistance QTL where IR64, and many other resistant cultivars, show poor SNP similarity to Nipponbare	174
5.9	Allelic diversity of 5 genes in 26 diverse *O. sativa* cultivars, based on SNP similarity to Nipponbare A: Os12g10870; B: Os12g11370; C: Os12g11500; D: Os12g11860; E: Os12g12010	176-180
5.10	Genotyping 27 rice cultivars using PCR markers for 5 candidate resistance genes across the *S. hermonthica* resistance QTL	183-186
5.11	Amplification of full-length candidate *S. hermonthica* resistance genes Os12g11370 (A) and Os12g11500 (B) from 8 rice cultivars	188
5.12	Amino acid alignment of the candidate *S. hermonthica* resistance RLP gene Os12g11370 from 4 rice cultivars, Nipponbare, Fanny, Jasmine85 and Koshihikari	190-191
5.13	Amino acid alignment of the candidate *S. hermonthica* resistance RLP gene Os12g11500 from 6 rice cultivars, Nipponbare, Darmali, Fanny, Koshihikari, Jhona349 and Azucena	192-193
6.1	An overview of the work carried out in this PhD, the findings, and the future directions for further research into resistance to *Striga* in rice	204

List of tables

Table		Page
1.1	Genes present in the Nipponbare *S. hermonthica* resistance QTL	36
2.1	Primer sequences for confirming successful cross pollination of F_1 progeny	50
2.2	Summary statistics of the QTL for *S. hermonthica* resistance in the rice IR64 x Azucena RIL population	55
3.1	Sequences of Gateway primers for amplification of full-length candidate RLP *S. hermonthica* resistance genes in IR64	77
3.2	Sequences of primers used for sequencing of RLP *S. hermonthica* candidate resistance genes in IR64	78
3.3	Sequences of primers designed for qPCR amplification of candidate *S. hermonthica* resistance genes in IR64	80
3.4	Genes within the IR64 *S. hermonthica* resistance QTL region as predicted by Fgenesh gene-finder	86- 87
3.5	Genes in Koshihikari for the equivalent region to the IR64 *S. hermonthica* resistance QTL, as predicted by Fgenesh gene-finder	95
3.6	Gene comparisons for across the *S. hermonthica* resistance QTL between the annotated Nipponbare genes and those predicted within the IR64 QTL. Koshihikari gene predictions are included where known	107
4.1	RNAi constructs produced for down-regulation of suites of RLP genes in Nipponbare	128
4.2	Primer sequences for confirmation of the presence of the *Tos17* / T-DNA insertions for 4 insertion lines	128
4.3	Primer sequences for qPCR amplification of candidate *S. hermonthica* RLP resistance genes	130
4.4	Primer sequences for testing whether gene transcripts were present for target genes in 4 insertion lines	131
5.1	The 26 rice cultivars selected for phenotyping for post attachment resistance to *S. hermonthica* and their genetic grouping	156
5.2	The 19 genes across the Nipponbare *S. hermonthica* resistance QTL examined in this study, and their annotation	157
5.3	Primer sequences used for genotyping rice cultivars for five candidate resistance genes in the *S. hermonthica* resistance QTL region	159
5.4	Primer sequences used for amplification of full-length genes across the *S. hermonthica* resistance QTL for a selection of rice cultivars	159

5.5	Similarity of SNPs to the Nipponbare allele for 19 genes across the *S. hermonthica* resistance QTL, in 26 diverse rice cultivars	164
5.6	Predictions of resistance or susceptibility to *S. hermonthica* for 26 genetically diverse rice cultivars	165
5.7	Summary statistics for Pearson's product-moment correlation analysis carried out between *S. hermonthica* dry weight and SNP similarity to Nipponbare for 19 genes across the *S. hermonthica* resistance QTL gene in 26 diverse rice cultivars	175
5.8	Similarity of SNPs to the Nipponbare allele for 19 genes across the *S. hermonthica* resistance QTL, in 26 diverse rice cultivars. Genes with > 80 % SNP similarity to Nipponbare are highlighted	181
5.9	Summary table of genotyping results for five genes in the Nipponbare *S. hermonthica* resistance QTL for the 26 rice cultivars	187

Chapter 1

General Introduction

1.1 Introduction to parasitic plants

Parasitism occurs in all Kingdoms of life, and is known to be a very successful life strategy. In plants, parasitism has evolved independently at least a dozen times, and 1 % of all angiosperms (c. 4000 species) are thought to be parasitic (Berner *et al.*, 1995; Westwood *et al.*, 2010). The dependence of parasitic plants on their host for survival varies greatly among genera. Holoparasites lack chlorophyll and are entirely dependent on their host for a source of carbon, while hemiparasites are photosynthetic and are thus able to fix at least some of their own carbon, although the extent to which they do this varies considerably between species (Westwood *et al.*, 2010). Hemi-parasites may be either facultative, living autotrophically and only parasitizing host plants when they are available, or obligate, requiring a host plant to complete their life cycle (Westwood *et al.*, 2010). Additionally, some parasitic plants, such as *Rhinanthus* and *Olax* species, have functional roots to extract water and nutrients from the soil. Others only have undeveloped roots (*Orobanche* species) or no roots at all (*Cuscuta* and the mistletoes) (Hibberd & Jeschke, 2001). Despite this diversity, all parasitic plants are defined by their ability to form haustoria (from the Latin *haurire*, meaning to drink), the specialised transfer organ through which the parasite obtains its water, minerals and carbohydrates (Hibberd & Jeschke 2001).

The family *Orobanchaceae* are the most species rich of all parasitic plant lineages, comprising 89 genera and ca. 2061 species, and are unique in that they represent the full range of parasitic ability, from facultative to obligate parasites (Bennett & Mathews, 2006; Westwood *et al.*, 2010). The majority are root parasites, including the genera *Striga*, *Orobanche* and *Alectra*, which are among the world's most destructive agricultural pests affecting cereals and legume crops (Bennett & Mathews, 2006; Parker, 2013). *Striga* species are particularly devastating, infecting the staple crops of sub-Saharan Africa (maize, sorghum, pearl millet, upland rice and cowpea), causing average yield losses of between 30 – 90 % and even complete crop failure (Ejeta, 2007). Over 40 % of cereal producing land in sub-Saharan Africa is infested with *Striga hermonthica* (Delile) Benth, and *Striga asiatica* (L.) Kuntze which are continuing to spread, affecting the lives of 100 million people and causing annual crop losses in excess of US $1 billion (Scholes & Press, 2008; Rodenburg *et al.*, 2010; Spallek *et al.*, 2013). The effects of the parasite are worse where soil fertility is poor, reducing the already low yields further (Ejeta & Gressel, 2007; Atera *et al.*, 2012). The poorest subsistence farmers are the most severely affected by *Striga*, which threatens food security and prevents a means to escape poverty. It is thus not surprising that *Striga* is now considered to be the most serious biotic threat to cereal production in sub-Saharan Africa (Scholes & Press, 2008).

Research into control of *Striga* in Africa started over 70 years ago, but success has been limited (Atera *et al.*, 2012), and our understanding of parasitic plant-host interactions has lagged behind that of insect and microbial pathogens (Scholes & Press 2008). It is also becoming increasingly obvious that the degree of *Striga* infestation in sub-Saharan Africa is worsening. Increasing population is partly to blame, putting pressure on land availability and forcing agriculture onto less suitable marginal lands with low soil fertility (Ejeta, 2007). The shift in cultivation to continuous cropping, particularly monocropping, with no or shortened fallow periods and low use of fertilisers due to cost, has not helped (Emechebe *et al.*, 2004). Use of contaminated cereal seed and frequent cultivation of susceptible crops that allow the build-up of parasite seeds in the soil are also responsible, often leading farmers to abandon fields. *Striga* infection on rice is increasing due to changing diets, crop profitability and an expanding area under cultivation (Rodenburg *et al.*, 2010). It is therefore clear that an effective control strategy is urgently needed that is both cost effective and easy to implement for resource poor farmers.

Not all *Striga* species pose a threat to agricultural production however. The *Striga* genus consists of over 40 species, only 11 of which are considered agricultural pests (Ejeta, 2007). Of these, *S. hermonthica*, *S. asiatica* and *S. gesnerioides* inflict the greatest economic damage (Mohamed *et al.*, 2007). *Striga hermonthica*, also known as the giant witchweed, is the best studied of these. As an obligate allogamous (out-crossing) species (Safa *et al.*, 1984) with high genetic diversity and tall stature, it causes the most severe crop losses (Spallek *et al.*, 2013). It is also the most widespread of *Striga* species, affecting much of western, central and eastern Africa. *Striga asiatica* is most commonly found in southern and eastern Africa, but is also present in parts of Asia, Australia and the United States. In contrast to other *Striga* species, *S. gesnerioides* (Willd.) Vatke infects dicotyledonous plants, in particular cowpea. It causes greatest damage in Western Africa, but it also occurs in the Arabian Peninsula, Asia and the United States (Mohamed *et al.*, 2001, 2007; Ejeta, 2007). Unlike *S. hermonthica*, both *S. asiatica* and *S. gesnerioides* are autogamous (inbreeding) species, and therefore show more distinct morphotypes and a higher degree of specialisation (Mohamed *et al.*, 2001; Botanga *et al.*, 2002). Figure 1.1 shows the distribution of *Striga* species throughout Africa.

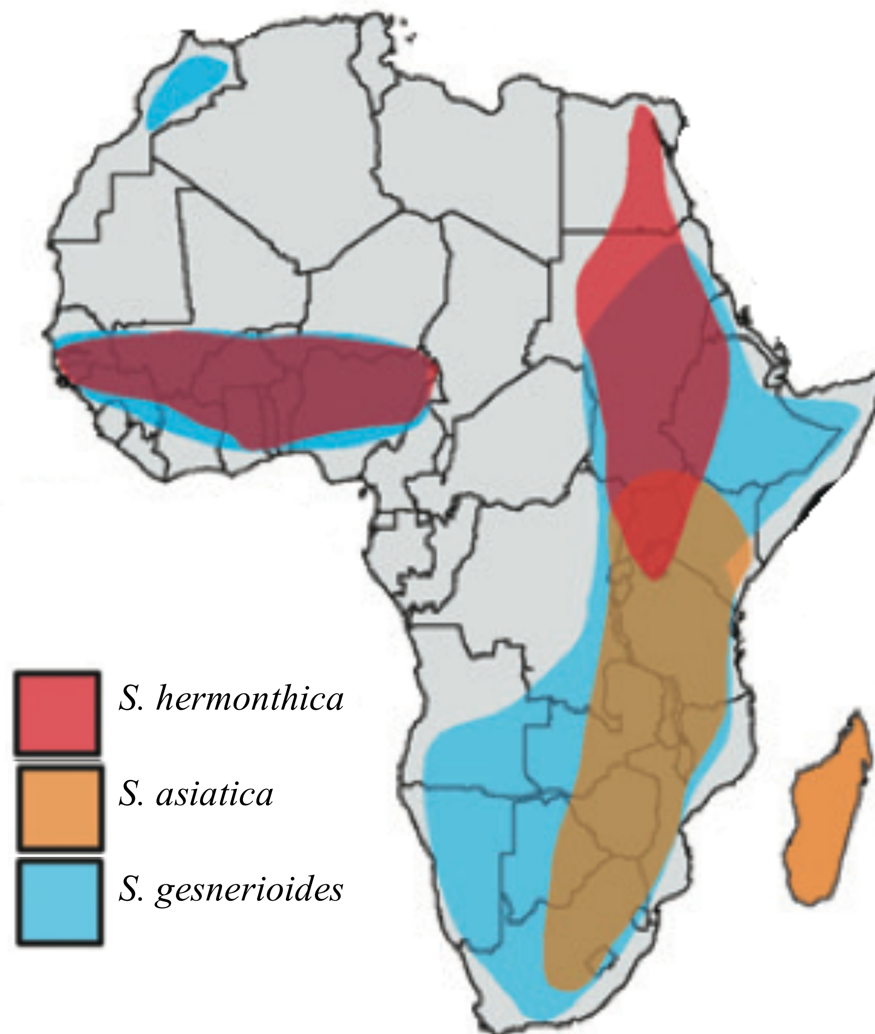


Figure 1.1 The distribution of the three most destructive *Striga* species in Africa. Figure from Spallek *et al* 2013.

1.2 The life cycle of *Striga*

The highly specialised life cycle of *Striga* means that it is very hard to control the parasite in the field; each stage is tightly coupled with the lifecycle of the host plant, requiring a complex exchange of signals for successful parasitism. Understanding the life cycle and biology of the parasite is essential for effective control, as the parasite can be targeted at different stages of development (Yoder & Scholes, 2010) (Figure 1.2). Before germination, *Striga* seeds require exposure to warm, humid conditions for one to two weeks (conditioning) to enter a state where they become sensitive to host-derived germination stimulants (Vallance, 1950; Yoder, 2001). Three different classes of compounds have been identified in the root exudates of host plants that will stimulate the seed germination of root parasitic plants: dihydrosorgoleone,

strigolactones and sesquiterpene lactones, the most potent of which are the strigolactones (Bouwmeester *et al.*, 2003; Yoneyama *et al.*, 2010). Strigolactones were first discovered as seed germination stimulants of parasitic plants but more recently were also identified as plant hormones that regulate shoot and root branching in plants (Gomez-Roldan *et al.*, 2008; Umehara *et al.*, 2008) and inducers of hyphal branching of arbuscular mycorrhizal (AM) fungi (Akiyama *et al.*, 2005). Multiple strigolactone receptors have also now been identified in *Striga* (Conn *et al.*, 2015; Toh *et al.*, 2015; Tsuchiya *et al.*, 2015).

Following exposure to strigolactones, *Striga* seeds germinate and a radicle begins to grow. The tip of the radicle produces hydrogen peroxide which activates the release of 2,6-dimethoxybenzoquinone (DMBQ) from host cell walls through the oxidative degradation of lignin (Keyes *et al.*, 2007). This acts as an haustorial inducing factor (HIF) causing swelling at the radicle tip and the production of haustorial hairs which aid attachment to the host roots, and an haustorium develops (Yoshida *et al.*, 2016). Cells under the haustorium divide to form a wedge which functions as a penetration peg, pushing through the host cortex and endodermis to establish direct xylem-xylem connections between the host and parasite (Dorr, 1997; Keyes *et al.*, 2001; Yoshida *et al.*, 2016). After successful penetration, the parasite can then extract water, minerals and nutrients from the host plant, produce its own cotyledons and stem, and grow towards the soil surface. Approximately 6 weeks after emergence the *Striga* plant flowers (Rich & Ejeta, 2007). Each *Striga* plant produces 10,000-200,000 tiny seeds approximately 0.2 - 0.5 mm in length, which are easily transported by wind, water, animal movement and contaminated crop seed and farm equipment (Berner *et al.*, 1994; Hearne, 2009). Seeds can remain viable for over 20 years in the soil, making eradication extremely difficult.

1.3 Host responses to *Striga* infection

The effects of *Striga* infection on the growth of the host plant occur as soon as a vascular connection is established, well before the parasite emerges above ground (Press & Stewart, 1987; Frost *et al.*, 1997). It is this phenomenon that has earned the parasite its name "*Striga*" (Latin for "witch"), referring to the seemingly bewitched effect on the crop (Rich & Ejeta, 2008). Early effects of *Striga* infection include reduced internode expansion and severe stunting, reduced tillering (in the case of rice) and decreased photosynthesis and transpiration (Press & Stewart, 1987; Frost *et al.*, 1997; Gurney *et al.*, 1999).

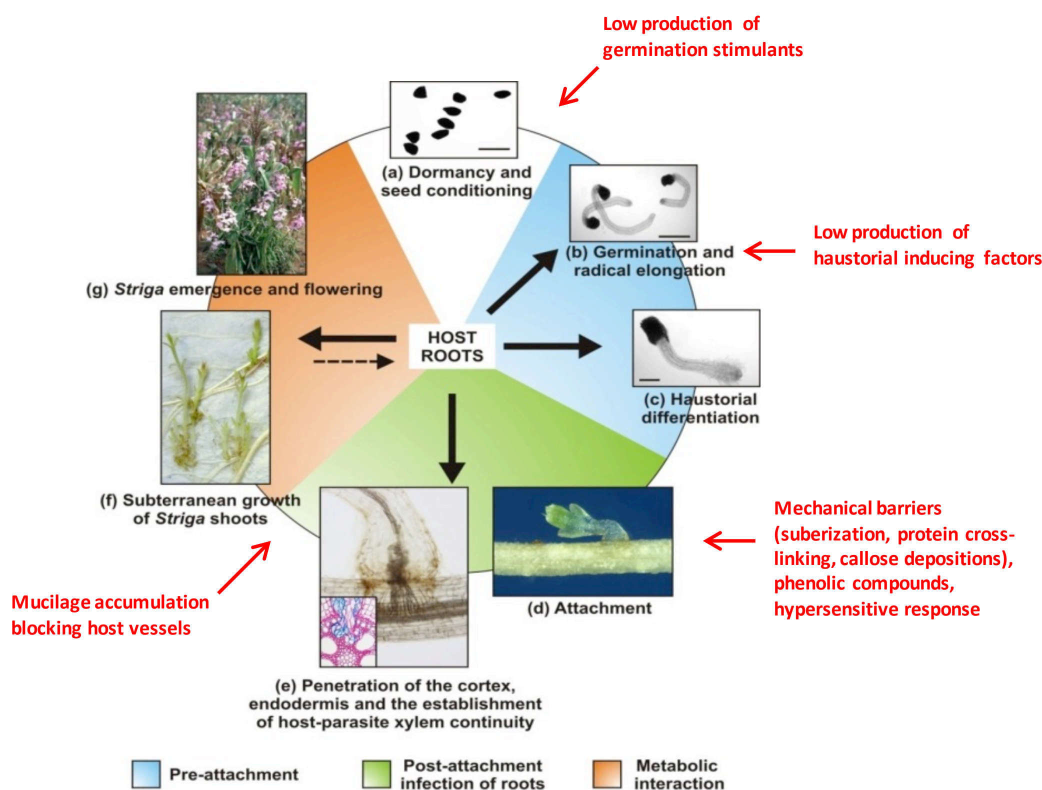


Figure 1.2 The life cycle of *Striga.* Red arrows indicate opportunities for host resistance. Adapted from Scholes and Press 2008.

Reductions in host biomass cannot be accounted for by an increase in parasite biomass alone (i.e. just due to acquisition of host resources), leading to the suggestion that *Striga* has a phytotoxic effect on the host, perhaps through the production of a toxin capable of being transported through the host, affecting plant metabolism (Press & Stewart, 1987; Rank *et al.*, 2004). However, a more recent hypothesis suggests that the severe stunting of the host plant may be hormone based, controlled by plant growth regulators rather than a toxin. Auxins and cytokinins are known to control shoot branching and tillering in plants (Leyser, 2003; Hayward *et al.*, 2009) and recently strigolactones have been shown to be involved as well. High levels of strigolactones inhibit tiller bud outgrowth (and thus number) in rice (Umehara *et al.*, 2008). Reduced tillering is also seen under *Striga* infection of rice (Jamil *et al.*, 2012), suggesting the effects of *Striga* infection on the growth of the host plant could be regulated at least in part by these hormones. As the parasite grows and its biomass increases it also acts as an additional sink, effectively competing for host carbon, inorganic solutes and water (Frost *et al.*, 1997; Gurney *et al.*, 1999). For some hosts, symptoms of infection may be characteristic of severe drought stress, and show increased root: shoot ratios compared to uninfected plants (Aflakpui et al. 2002). Not surprisingly, the time of attachment of the parasite to the host roots is also

important, with early attachment causing the greatest reductions in host biomass and grain yield, and larger parasite biomass (Gurney *et al.*, 1999). Indeed, the speed of emergence, rather than the parasite number, has been found to be the main determinant of severity of impact on host growth (Kaewchumnong & Price, 2008). Because the impacts of *Striga* infection are seen soon after attachment, effective control strategies must act early on, preventing attachment or killing the parasite before it can become properly established (Scholes & Press, 2008). However, many control strategies currently used do not act in this way.

1.4 Current strategies for controlling *Striga*

Cultural control strategies, such as nutrient fertilisation and / or intercropping, help to target the pre- attachment stages of parasite development. The improved soil fertility achieved through fertilisation is thought to reduce the amount of strigolactones produced by host plant roots, thereby alleviating the impacts of *Striga* infection due to reduced germination of *Striga* seeds in the soil (Jamil *et al.*, 2011a). In contrast, strigolactone production is markedly increased under phosphorus and nitrogen deficiency (Yoneyama *et al.*, 2007b,a). However, inorganic fertilisers are often expensive or unavailable, meaning that large scale fertilizer use is not a viable option for many farmers (Hearne, 2009; Atera *et al.*, 2012). Crop rotation and intercropping with leguminous trap crops such as cowpea, groundnut or soybean are less expensive alternatives (Gbehounou & Adango, 2003) and can help to alleviate the *Striga* problem. Oswald *et al.*, (2002) showed that productivity of maize could be increased when intercropped with a range species, of which cowpea and yellow gram were the most effective. Additional benefits of intercrops included increased shading, lower temperatures and higher humidity under the intercrop canopy, which reduced the number of *Striga* plants in these systems (Oswald *et al.*, 2002). Legume crops contribute to control by improving soil fertility through nitrogen fixation and by causing suicidal germination of *Striga* seeds. Indeed, germination stimulants present in the root exudates of legumes generally stimulate germination of *Striga* seeds. *Desmodium* species, which can be used as fodder crop for cattle, are particularly effective and produce large amounts of germination stimulants that will cause suicidal germination of *Striga* seeds. The fact that *Desmodium* is a perennial intercrop and is present even when the crop is not in the ground also means that it reduces the *Striga* seed bank more effectively over time than annual intercrops. In addition to improving the nitrogen content of the soil, and stimulating suicidal germination of seeds, novel flavonoids in the root exudates of *D. uncinatum* also inhibit subsequent radicle growth and development (Khan *et al.*, 2008; Hooper *et al.*, 2010). All these mechanisms together contribute to a dramatic suppression of *Striga* attachment to the host crop and increases in yield (Kifuko-Koech *et al.*,

2012; Midega *et al.*, 2013). However, willingness to invest in a non-food intercrop may be an issue for some subsistence farmers. In such cases an alternative legume, such as soybean, may be more appropriate (Hearne, 2009) although this is much less effective. Repeated use of intercrops is also required before significant improvements in grain yield are achieved, due to the high numbers and longevity of *Striga* seed that have accumulated in the soil (Atera *et al.*, 2012). Hand weeding is another control measure popular among farmers (Atera *et al.*, 2012). It has the potential to reduce the *Striga* seed bank in the soil if carried out before flowering (Hearne, 2009). However, it is labour intensive and can only be achieved once the parasite has emerged above ground and reached a manageable size. Unfortunately many of the harmful effects on host growth occur before this stage (Frost *et al.*, 1997).

One chemical control strategy for *Striga* involves growing herbicide resistant seeds, which is currently only used for maize. Tolerant maize germplasm harbours a natural mutation for imidazolinone resistance, which confers resistance against the imidazolinone herbicide (Kanampiu *et al.*, 2003). This mutation does not occur in dicotyledonous plants such as *Striga*, making it susceptible to the application of this herbicide. Maize seeds are treated with low doses of imidazolinone prior to planting, and the herbicide works by inhibiting the activity of acetolactate synthase (ALS), killing the parasite and forming a protective zone around the host plant roots. Herbicide coated maize seed is already used in East Africa under the name StrigAway (Kanampiu *et al.*, 2003; Rodenburg *et al.*, 2010). However, its effectiveness is variable and very dependent on environmental conditions such as the time of *Striga* germination and the distribution of rainfall; too much rain will wash the herbicide beneath the rhizosphere of the crop (Kanampiu *et al.*, 2003). Using seed based technologies for *Striga* control requires farmers to purchase treated seed annually, as non-treated seed shows no resistance (Hearne, 2009). This is unattractive for the majority of African farmers, who rely on seed produced on their own farm (Rodenburg *et al.*, 2010).

1.4.1 The use of resistant cultivars in controlling *Striga*

Because the impacts of *Striga* on host growth and development occur very early on, within a few days of attachment, an effective control strategy should prevent attachment or kill the parasite before it can become properly established. Removal of *Striga* from the host once it has emerged above ground does not reverse the damage it has already caused. The use of resistant cultivars has long been considered a sustainable and cost effective form of crop protection against *Striga*, especially when used in combination with other control measures that aim to reduce the *Striga* seed bank and improve soil fertility (Scholes & Press, 2008). Crop resistance would not only reduce *Striga* infection and yield losses, but would also reduce rates

of seed production, thereby reducing the seed bank in the soil and the chance of future infestations (Rodenburg *et al.*, 2006).

However, there are several aspects of *Striga* biology that pose a treat to the durability of host plant resistance. Firstly, the high genetic variability within and between different species and populations of *Striga*; *S. hermonthica* exhibits particularly high genetic variation due to its outbreeding mating system, leading to high levels of heterozygosity within each population (Safa *et al.*, 1984; Mohamed *et al.*, 2007). Secondly, each *Striga* plant is capable of producing 10,000 – 200,000 dust-like seeds that can remain viable in the soil for 20 years (Berner *et al.*, 1995). This creates a reservoir of genetic diversity within the soil, and a range of *Striga* genotypes capable of infecting a wide range of host species and cultivars. This high genetic variability means that no cultivar shows complete resistance to *Striga*; even the most resistant plants may still support one or two parasites. It is therefore unlikely that a single resistance gene would be useful in the field, as any new resistance is immediately confronted by the genetic diversity of the *Striga* seed bank (Mohamed *et al.*, 2007). Additionally, host-plant resistance is known to vary between different *Striga* species and ecotypes, which may also be affected by climate, and vary across different sites and years (Johnson *et al.*, 1997; Rodenburg *et al.*, 2017).

For resistance to be effective, it must be both broad spectrum against a range of *Striga* species and ecotypes, and must also be durable. This requires pyramiding multiple resistance genes with different modes of action into a single locally adapted cultivar. It also requires a better understanding of *Striga* diversity and virulence, and how this translates into host-parasite specificity. Improved knowledge of the molecular genetic basis of resistance with respect to parasite virulence will allow predictive breeding programs to be targeted for different regions to match appropriate host resistance against the prevailing parasite ecotype.

For rice, breeding for *Striga* resistance has received relatively little attention, perhaps because 75 % of the worlds rice production occurs in irrigated lowlands where *Striga* species are absent (Seck *et al.*, 2012). Parasitic weeds are only prevalent in rain-fed ecosystems, which in Africa accounts for around 72 % of total rice production (Rodenburg *et al.*, 2010). However, due to increased demand for rice in Africa, national, regional and international agencies are now placing a higher priority on the rice sector as a means to improve food security and economic growth (Balasubramanian *et al.*, 2007).

1.5 How do plants defend themselves against parasitic weeds?

Plants possess sophisticated defence mechanisms to defend themselves from attack from a wide variety of pests, pathogens and parasites (Jones & Dangl, 2006). Much of the work on plant immunity has been carried out on bacterial and fungal pathogens, but much less is known about defence against parasitic plants. However, several studies have shown that plants recognise and defend themselves against parasitic plants in a very similar manner to microbial pathogens. Different defence responses may operate at different stages of parasite infection and may vary between different host species and genotypes (Yoshida & Shirasu, 2009).

1.5.1 Pre-attachment resistance

Since *Striga* seeds require a chemical signal from the host to germinate, the first opportunity for resistance is to select cultivars which produce low amounts of germination stimulants (Vogler *et al.*, 1996). This form of resistance is best described in *Striga* resistant genotypes of sorghum (Haussmann *et al.*, 2001) and to a lesser extent rice (Jamil *et al.*, 2011b) but is also important in resistance to *Orobanche* (Rubiales *et al.*, 2003). In rice, NERICA cultivars exhibit a positive relationship between strigolactone production and the germination, attachment and emergence of *Striga* plants (Jamil *et al.*, 2011b). QTL mapping using a Recombinant Inbred Line (RIL) mapping population created from Bala, an *indica* rice cultivar, and Azucena, a tropical *japonica* cultivar, identified a major QTL on chromosome 1 termed qSLB1.1 for both the production of strigolactones and the germination of *S. hermonthica* seeds (Cardoso *et al.*, 2014). Two cytochrome P450 genes, SLB1 and SLB2, which showed high homology with the MAX1 strigolactone biosynthesis gene in *Arabidopsis*, were identified in the QTL for Azucena, the high strigolactone producer. These genes were missing in Bala, which produced lower strigolactone levels, but overexpression of either of these genes in Bala resulted in higher strigolactone production (Cardoso *et al.*, 2014). A recent study identified a gene in sorghum responsible for an alteration in the types of strigolactones present in root exudates, resulting in increased *Striga* resistance (Gobena *et al.*, 2017). Sorghum lines with mutations in the *lgs* allele produced drastically less of the dominant strigolactone 5-deoxystrigol, a potent stimulant of *Striga* germination, but this resulted in greater production of the strigolactone orobanchol. Interestingly, orobanchol does not stimulate germination of *Striga*, and plants mutant for the *lgs* allele did not differ greatly in the extent of AM colonisation or tiller number (Gobena *et al.*, 2017), indicating it may be possible to breed increased resistance to *Striga* without affecting the plants important associations with these fungi or alterations in the plants development.

The exudation of germination inhibitors and reduced production of HIFs from host plants are other pre-attachment forms of resistance (Rispail *et al.*, 2007). Gurney *et al.* (2003) showed that *Tripsacum dactyloides*, a wild relative of maize, had enhanced resistance against *S. hermonthica*, in part due to the low concentrations of HIFs produced. Parasites which did attach to *Tripsacum dactyloides* roots developed slowly with poor haustorial differentiation, and the formation of secondary haustoria was inhibited, suggesting a mobile signal is also produced in host roots which inhibits haustorial development (Gurney *et al.*, 2003).

1.5.2 Post-attachment resistance

Post attachment resistance to parasitic plants can occur in the root cortex, at the endodermis, or after vascular connection (Yoshida & Shirasu, 2009; Timko & Scholes, 2013). Typical defence responses are similar to those observed for microbial pathogens, and include suberin and callose depositions, increased production of phenolic compounds, lignifications of the endodermis, increases in peroxidase activity, expression of pathogenesis-related (PR) proteins and hypersensitive response (Mohamed *et al.*, 2003; Echevarría-Zomeño *et al.*, 2006; Pérez-de-Luque *et al.*, 2006a; Timko & Scholes, 2013).

Resistance to the parasitic plant *Orobanche cumana* in sunflower occurs at the cortex (Echevarría-Zomeño *et al.*, 2006). Here, suberization of host cell walls together with protein cross-linking, were observed in resistance interactions, preventing penetration by the parasite. In pea, accumulation of callose was also observed, as well as H_2O_2 and peroxidases (Pérez-de-Luque *et al.*, 2006a). Peroxidases are a group of enzymes involved in a variety of functions including lignification, suberization, and other stress responses, and together with H_2O_2, enable protein cross-linking and formation of papillae, thereby strengthening cell walls (Echevarría-Zomeño *et al.*, 2006). The presence of phenolic compounds in cortical cells was also observed during incompatible reactions. These are excreted into the apoplast, creating a toxic environment and effectively poisoning the parasite (Echevarría-Zomeño *et al.*, 2006). Plant defensins, which show antifungal properties against a broad range of fungi (Thomma *et al.*, 2002), are also known to provide resistance against parasitic plants. de Zélicourt *et al.*, (2007), showed that the sunflower (*Helianthis annuus*) defensin Ha-DEF1 was involved in resistance to *O. cumana*. Expression of Ha-DEF1 was induced upon *O. cumana* infection, while purified Ha-DEF1 induced browning symptoms at the radicle apex of *O. cumana* seedlings as a result of cell death. Anti-fungal activity of Ha-DEF1 was confirmed with bioassays on *Saccharomyces cerevisiae* and *Alternaria brassicicola*, which showed strong inhibition of growth and germ-tube development, respectively. However, no effect was seen on *S.*

hermonthica seedlings (de Zélicourt et al., 2007). This was the first account of defensins acting on plant cells.

Resistance to *S. hermonthica* in the resistant sorghum genotype Framida occurs at 3 different stages; the inhibition of haustorial development, reduced translocation of nutrients towards the haustoria, and the accumulation of phenolic substances around the central cylinder (Arnaud *et al.*, 1999), demonstrating resistance can act at different levels simultaneously. Yoshida & Shirasu (2009) also demonstrated multi-layers of incompatibility during resistance to *S. hermonthica* in non-host dicotyledonous plants. Four different non-host plants, *Arabidpsis*, cowpea, the legume *Lotus japonicas* and the hemiparasite *Phtheirospermum japonicum,* were infected with germinated *S. hermonthica* seeds to examine differences in incompatibility. *Arabidopsis* and cowpea exhibited resistance after vascular connections were established. Resistance in *Lotus japonicas* was due to mechanical barriers in the cortex, while resistance in the hemiparasite *Phtheirospermum japonicum* occurred very early, and in most cases *S. hermonthica* failed to form an attachment or penetrate the P. *japonicum* roots (Yoshida & Shirasu, 2009). Gurney et al. (2006) reported that resistance in the rice cultivar Nipponbare occurred at the endodermis, and no evidence of lignification was observed.

The study by Yoshida & Shirasu (2009) investigated the rate of infection and development of *S. hermonthica* on the resistant Nipponbare and the more susceptible rice cultivar Koshihikari. The frequency of different developmental stages of the parasite was recorded at two and four weeks post inoculation (wpi) of *S. hermonthica,* and calculated as a percentage of parasites that had penetrated host tissue. At 4 wpi significantly fewer parasites had developed to the six-leaf stage on Nipponbare compared to Koshihikari. However, at 2 wpi, the frequency of parasites that had developed vascular connections did not differ significantly between the cultivars, suggesting Nipponbare may also exhibit some form of post-vascular resistance (Yoshida & Shirasu, 2009). Resistance to *S. hermonthica* is very ecotype dependent, and therefore differences in resistance may be observed with different ecotypes. Post vascular resistance was seen in the vetch *Vicia sativa* when infected with *Orobache crenata.* This was shown to be due to the accumulation of mucilage, composed mainly of non-esterified pectins, which blocked host vessels, obstructing the parasites nutrient supply and causing browning and death of the tubercles (Pérez-de-Luque *et al.*, 2006b). In addition to the pre-attachment resistance in some rice (NERICA) cultivars described above, some NERICA cultivars also show very good post-attachment resistance to some *S. hermonthica* and *S. asiatica* ecotypes. In most cases this was characterised by the inability of the *Striga* to penetrate the endodermis. Some parasites were able to penetrate the endodermis and form a few vascular connections,

but these grew slowly and were associated with deposition of a dense staining material (Cissoko *et al.*, 2011). In most cases, this resistance was maintained when challenged by several ecotypes of both *Striga* species, indicating resistance was relatively broad spectrum. In cowpea, resistance to *S. gesnerioides* is often associated with a hypersensitive response at the site of parasite attachment (Li & Timko, 2009). This is characteristic of a gene-for-gene resistance response, where cultivars carrying the specific resistance (R) gene convey resistance against a specific race of the parasite (Scholes *et al.*, 2007) causing browning and subsequent death of the parasite within 3 or 4 days (Li & Timko, 2009; Huang *et al.*, 2012a). The gene responsible for this reaction has been identified in cowpea. This gene, termed RSG3-301, provides resistance against *S. gesnerioides* race 3 (SG3), and is the first resistance gene against *Striga* species to be cloned (Li & Timko, 2009). The hypersensitive-like response observed in cowpea resistance response is much less common in hosts exhibiting resistance to *Striga hermonthica* or *asiatica*, although it has been observed in some sorghum cultivars (Mohamed *et al.*, 2003).

1.5.3 The role of signalling molecules

Inducible defence responses in plants are regulated by a network of signalling molecules, of which salicylic acid (SA), jasmonic acid (JA), ethylene (ET) and abscisic acid (ABA) are thought to be the most important. Generally speaking, the SA signalling pathway is associated with defence against pathogens with a biotrophic lifestyle, while the JA/ET signalling pathway activates a set of defence responses targeting necrotrophic pathogens (Glazebrook, 2005). However, these pathways are known to interact extensively and mutually antagonise each other, allowing a plant to fine-tune its defence against different pathogens and modes of infection (Kunkel & Brooks, 2002). These pathways are also involved in defence against parasitic plants. Application of SA to the roots of red clover (*Trifolium pratense*) was found to reduce the number of *O. minor* parasites, while application of MeJA had no effect (Kusumoto *et al.*, 2007). The SA and JA signalling pathways are known to act antagonistically, although they may also act synergistically at lower concentrations (Mur *et al.*, 2006). Hiraoka & Sugimoto (2008) showed that infection of *S. hermonthica* on a susceptible sorghum cultivar induced JA-responsive genes and suppressed SA-responsive genes in host roots. In contrast, *Striga* infection on roots of the resistant cultivar Wad Ahmed induced both SA and JA-responsive genes to a lesser degree, suggesting both signalling pathways are involved in resistance. Application of SA to host roots reduced susceptibility of all cultivars (Hiraoka & Sugimoto, 2008), indicating this hormone may play a greater role in resistance against parasitic plants.

In rice, both the SA and JA pathways are induced on infection with *S. hermonthica*, although induction of the SA pathway occurs after that of JA (Mutuku *et al.*, 2015). Interestingly, mutants in JA biosynthesis were susceptible, although resistance was recovered by application of exogenous JA. In contrast, SA is not necessary, as SA-deficient *NahG* plants were also resistant; these mutants also showed induction in the JA pathway. However, plants containing a silencing construct to knockdown the WRKY45 transcription factor, which is a key regulator of the SA signalling pathway, were susceptible to *S. hermonthica*, and showed down-regulation in the JA pathway. Foliar application of JA recovered resistance. These results suggests that in rice, WRKY45 is involved in the regulation of both the SA and JA pathways (Mutuku *et al.*, 2015). The identification of the resistance genes will help clarify the defence mechanisms involved in this resistance.

1.6 The genetic basis of *Striga* resistance

Although many different *Striga* resistance phenotypes have been observed, the underlying genetics governing these responses is not well understood. Many studies investigating the molecular basis of *Striga* resistance have shown it to be polygenic, controlled by both major and minor genes (Timko & Scholes, 2013). These may be recessive or quantitative in nature. In an investigation into the inheritance of low *Striga* germination stimulants in sorghum, Vogler *et al.* (1996) used one resistant line “SRN-39” and three susceptible lines to show that low stimulant production was controlled by a single, recessive allele. A similar investigation using recombinant inbred lines (RILs) of sorghum also found low production of germination stimulants to be inherited by a single recessive gene of major effect. However, low and high stimulant producing classes were not always distinct, suggesting some minor genes may also be involved (Haussmann *et al.*, 2001). Haussmann *et al*. (2004) used two RIL mapping populations to identify resistance quantitative trait loci (QTL) in sorghum. Resistant parental lines were characterised as low germination stimulant producing (IS9830) or expressing mechanical resistance (N13). Resistance was tested in the field over two years and five different sites across Mali and Kenya, and composite interval mapping used to detect eleven QTL in both populations, indicating resistance is not simply inherited. The QTL of most significance corresponded to the *lgs* locus, a major locus for low germination stimulants, but other QTL were also identified. Five QTL were shared between the two populations. Since these withstood across populations, years and environments, they are ideal candidates for marker-assisted selection (Haussmann *et al.*, 2004).

In another study, the production of low amounts of haustorial initiation factors in sorghum were found to be inherited as a single dominant gene, *Lhf*. This gene was only present in the wild sorghum species PQ434. Mapping populations created from crosses of high stimulant producing lines, together with genotyping using microsatellite markers, enabled this gene to be mapped within 19.3cM from the Xtxp358 marker on linkage group nine (Grenier *et al.*, 2007).

Resistance against *S. gesnerioides* in cowpea (*Vigna unguiculata*) is better understood. *S. gesnerioides* exhibits clear race structure, and seven races are currently known to exist (Li *et al.*, 2009). Resistance to these races in cowpea is thought to be conferred by a single dominant gene in most cases, and has been mapped to two linkage groups (Ouedraogo *et al.*, 2001). As mentioned above, the first resistance gene to *S. gesnerioides* in the cowpea cultivar B301 has now been cloned (Li & Timko, 2009). The protein product from this gene localises to the plasma membrane, and contains coiled-coil, nucleotide binding site and leucine rich repeat (CC-NBS-LRR) domains. A possible role as a guard molecule against parasite attack was suggested (Li & Timko, 2009).

In rice, mapping populations have also been used successfully to uncover the genetic basis of *Striga* resistance. Gurney *et al*. (2006) used Backcross Inbred Lines (BILs) of Nipponbare, which exhibits strong post-attachment resistance to *S. hermonthica*, and Kasalath which is more susceptible, in order to identify the genomic regions governing post-attachment resistance. Composite interval mapping was used to locate seven resistance QTL on chromosomes 1, 4, 5, 6, 7, 8 and 12. Four of these were confirmed in a second screen, and those on chromosome 4 and 12 appeared to be the most important. Although the mean resistance of the mapping population was intermediate between the parents, the slight bimodality in the distribution of resistance observed indicates the segregation of some genes of major effect (Gurney *et al.*, 2006). Interestingly, a large effect QTL on chromosome 4 was due to a resistant allele in Kasalath rather than in Nipponbare, which suggests a phenotype even more resistant than Nipponbare could be created by combining these into the same cultivar. In order to establish whether this Kasalath derived resistance QTL also conferred resistance to *S. hermonthica* in a different genetic background, Swarbrick *et al.* (2009) carried out a further QTL analysis using a Koshihikari-Kasalath Backcross Inbred Line mapping population. Three QTL conferring resistance to *Striga* were identified, one of which came from the susceptible Koshihikari genotype. The location of the large-effect QTL in the Koshihikari-Kasalath population overlapped with the QTL detected previously, verifying its role in resistance and narrowing down its position in the genome (Swarbrick *et al.*, 2009).

The QTL identified by Gurney *et al.*, (2006) were very interesting as there was a suggestion that resistance may be due to a few genes of major effect. However, because the rice cultivar Kasalath also possessed significant resistance it was difficult to resolve the QTL further. In order to investigate the Nipponbare resistance QTL in more detail, a second study was carried out at the University of Sheffield using a BIL population derived from a cross between Nipponbare and Koshihikari. Koshihikari is a temperate *japonica* cultivar that is susceptible to the *S. hermonthica* ecotype used in the mapping studies, so better resolution is achieved between lines when phenotyping for resistance. This study revealed one major QTL on chromosome 12 (Nipponbare) between positions 4.0 - 8.0 Mbp, with a peak at 6.74 Mbp (Figure 1.3) (Scholes *et al*., unpublished data). This QTL region explains 55 % of variation for *S. hermonthica* resistance, and with a LOD score of greater than 25 strongly suggests that the resistance is due to a single gene or to a few genes of major effect. Such a major QTL is surprising given the genetic diversity of the parasite, and indicates the presence of a very important resistance gene or genes, supporting the conclusion of Gurney *et al.*, (2006).

It can be seen from Figure 1.3 that the QTL is broad and the area contains many hundreds of genes. Lack of recombination in this area meant it was impossible to refine the QTL using lines from the original population, therefore a series of BC_4F_5 lines were produced at the University of Sheffield and NIAS, Japan, by backcrossing a small selection of resistant lines to Koshihikari and selfing to the F_5 generation. These lines were then phenotyped for *S. hermonthica* resistance. Phenotyping the BC_4F_5 lines refined the QTL position to between 5.7 - 6.7 Mbp (Figure 1.4). Within the QTL there was a 350 kb region that is not present in the same region of Koshihikari genome (Figure 1.5) (Scholes *et al*., unpublished data). It is therefore possible that resistance could be provided within this Nipponbare specific region, although the QTL extended beyond it on both sides (Scholes, personal communication).

The *S. hermonthica* resistance QTL in Nipponbare contains 131 genes, over half of which are annotated as transposons or retrotransposons (see http://rice.plantbiology.msu.edu/ for annotations). Also within this region is a cluster of 13 receptor-like proteins (RLPs) annotated as homologs of genes providing resistance to *Verticillium* wilt in tomato (Table 1.1). *Verticillium* wilts are soil borne, xylem-invading fungal plant pathogens with a broad host range, infecting over 200 host species (Fradin & Thomma, 2006). In a similar manner to *Striga* species, *Verticillium* wilts respond to host root exudates, which stimulate germination of microsclerotia in the soil. These penetrate the root, crossing the endodermis to enter the xylem, where conidia are produced and transported around the plant in the water stream (Fradin &

Thomma, 2006). The strikingly similar infection strategy of this parasite with *Striga* species, together with the presence of a cluster of *Verticillium* wilt homologs in the Nipponbare *Striga* resistance QTL, makes these genes excellent candidates for *Striga* resistance. Additionally, these RLP genes also belong to the same gene class as a gene recently identified in tomato (*CuRe1*) that provides increased resistance to the stem holoparisitic plant *Cuscuta reflexa*. Overexpression of the cell surface RLP gene *CuRe1* into susceptible *S. pennellii* plants resulted in hypersensitive response symptoms and improved resistance to *C. reflexa* (Hegenauer *et al.*, 2016). The expression of these candidate resistance genes in Nipponbare was investigated under *S. hermonthica* infection using quantitative PCR (qPCR). Expression of four of these genes (Os12g11370, Os12g11660, Os12g11680 and Os12g11720) was upregulated in Nipponbare, but not Koshihikari, in response to *S. hermonthica* infection, although this does not necessarily imply an involvement in resistance (Scholes *et al*., unpublished data).

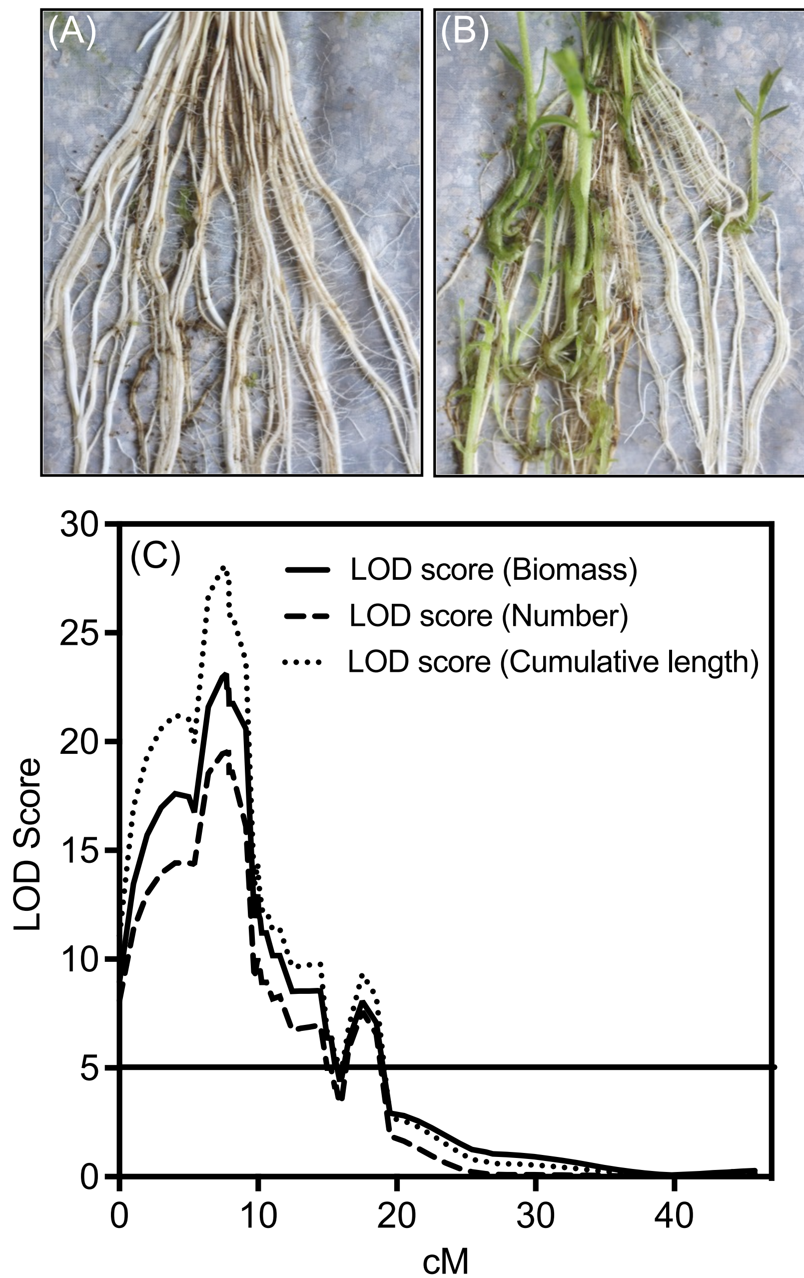


Figure 1.3 Phenotype of rice cultivars Nipponbare (a) and Koshikihari (b) infected with *Striga hermonthica* (Kibos ecotype). (c) Genome-wide quantitative trait locus (QTL) for *Striga hermonthica* resistance traits (Striga biomass, number and cumulative length per rice plant) on chromosome 12 in the Nipponbare/Koshihikari //Koshihikari BIL population. A single QTL on Chromosome 12 at 7.7cM / 6.75 Mbp explains over 55 % of the variation in *Striga* resistance (*Striga* biomass or number) (Scholes *et al.*, unpublished data).

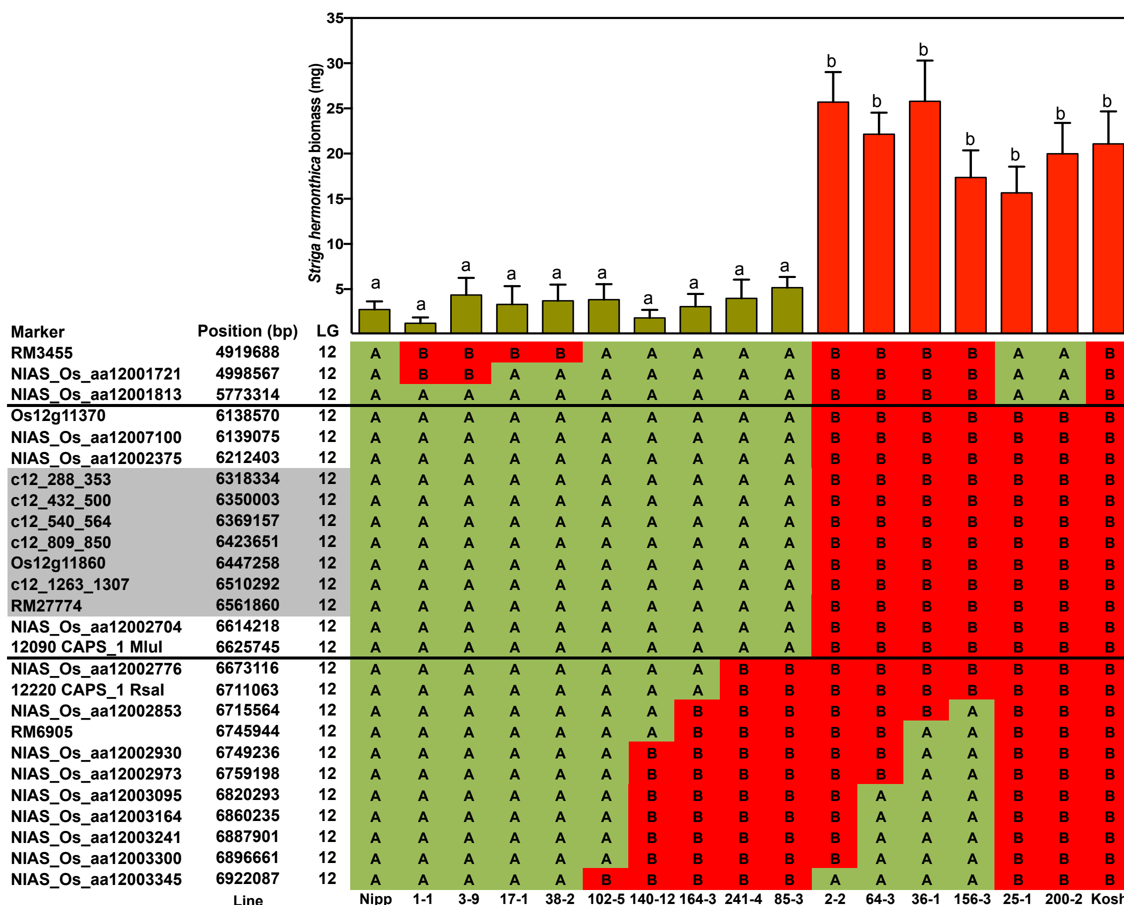


Figure 1.4 The resistance of the BC_4F_5 lines from the Nipponbare x Koshihikari BIL population showing recombination in and around the QTL region on chromosome 12. Top panel: Dry biomass of *S. hermonthica* 21 days after infection with germinated *S. hermonthica* seed collected from Kibos in Western Kenya in 2013. Values are means ± SE where N=4. Lower panel: Recombination breakpoints for each line are shown in relation to chromosomal position, where A = Nipponbare genotype and B = Koshihikari genotype. Letters above bars indicate significant differences between lines ($p < 0.05$). (Scholes *et al.*, unpublished data).

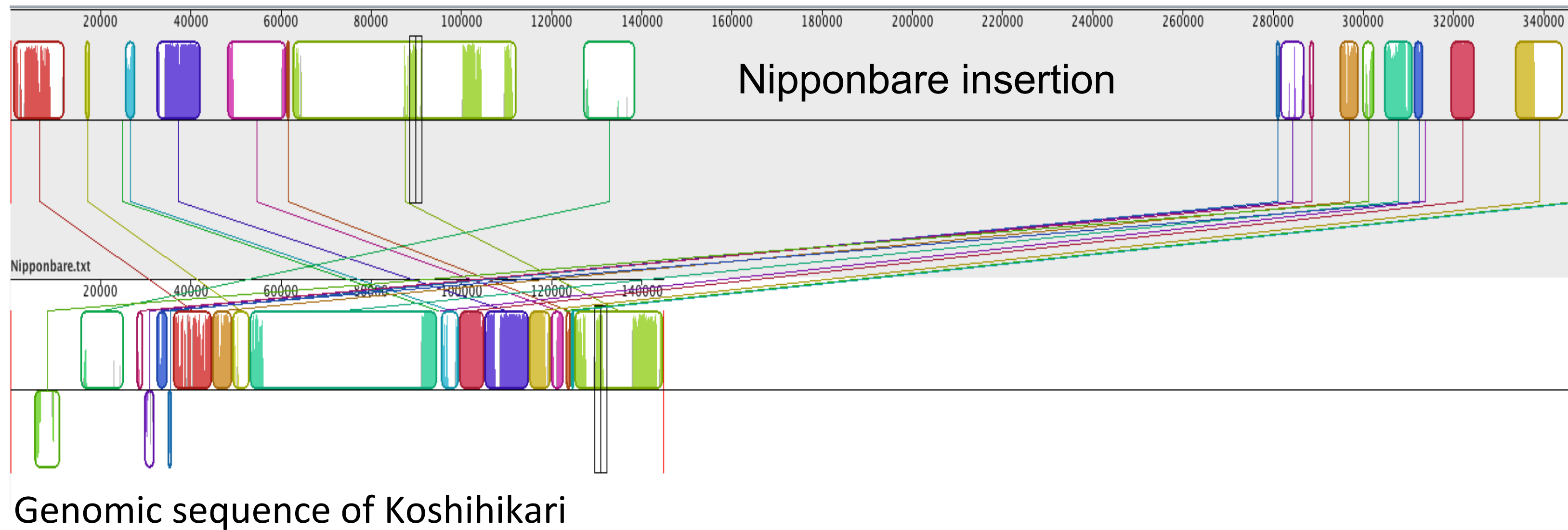


Figure 1.5 Comparison of the genome structure of rice varieties Nipponbare and Koshihikari within the *S. hermonthica* QTL region on chromosome 12. Numbers refer to the sequence length in bp. Coloured blocks surround a region of the genome that aligns to part of the genome being compared that is homologous and free from genomic rearrangement. Lines link regions of homology. Blocks below the centre line indicate regions of inverse orientation relative to the top sequence. Areas outside blocks do not align, and so contain sequences specific to that genome or lacking homology between genomes. (Scholes *et al*., unpublished data).

Table 1.1 Genes present in the Nipponbare *S. hermonthica* resistance QTL. Transposable elements and expressed proteins have been omitted. aa = amino acid. Source: MSU (http://rice.plantbiology.msu.edu/) and International Rice Genome Sequencing Project (IRGSP) (http://rapdb.dna.affrc.go.jp).

Genes in Nipponbare	Annotation	Length (aa)
LOC_Os12g10850.1	hhH-GPD superfamily base excision DNA repair protein, putative, expressed	475
LOC_Os12g10870.1	verticillium wilt disease resistance protein, putative, expressed	1016
LOC_Os12g10930.1	NL0E, putative, expressed	749
LOC_Os12g11370.1	verticillium wilt disease resistance protein, putative, expressed	1015
LOC_Os12g11500.1 IRGSP	resistance protein SlVe1 precursor, putative, expressed	1013
LOC_Os12g11510.1	hcr2-0B, putative, expressed	829
LOC_Os12g11660.1	RALFL45 - Rapid ALkalinization Factor RALF family protein precursor, expressed	*99*
LOC_Os12g11680.1 IRGSP	verticillium wilt disease resistance protein precursor, putative, expressed	*1000*
LOC_Os12g11720.1	verticillium wilt disease resistance protein precursor, putative, expressed	*1020*
LOC_Os12g11860.1	verticillium wilt disease resistance protein precursor, putative, expressed	*1006*
LOC_Os12g11930.1	disease resistance protein SlVe2 precursor, putative, expressed	*1016*
LOC_Os12g11940.1	disease resistance family protein, putative, expressed	*855*
LOC_Os12g12000.1	RALFL46 - Rapid ALkalinization Factor RALF family protein precursor, expressed	97
LOC_Os12g12010.1	verticillium wilt disease resistance protein precursor, putative, expressed	1000
LOC_Os12g12120.1	verticillium wilt disease resistance protein precursor, putative, expressed	1006
LOC_Os12g12130.1	verticillium wilt disease resistance protein, putative, expressed	1026

1.7 Genetic mapping to identify *Striga* resistance QTL

It is clear from many studies discussed that the use of mapping populations has been extremely important in the discovery of *Striga* resistance QTL thus far. Several different types of mapping population exist, but most are originally produced from two genetically distinct parents (Morrell *et al.*, 2012). BILs, also referred to as Near Isogenic Lines (NILs), are generated by repeated backcrossing of the F_1 progeny to one of the original parents, so that each line contains a small number of introgressed fragments from the donor genome in an otherwise homogenous genetic background (Fernie & Klee, 2011). All lines together represent the whole genome of the donor parent in an overlapping manner (see Figure 1.6) (Ali *et al.*, 2010). Recombinant Inbred Lines (RILs) are created by repeated selfing (self-fertilising) of the F_1 progeny until homozygosity is reached. Here, each individual contains multiple introgressed fragments, and on average the genome of each parent is represented equally throughout the population (Mauricio, 2001; Fernie & Klee, 2011). Figure 1.6 shows a comparison between a NIL an RIL mapping population. Nested Association Mapping (NAM) populations are a relatively new design of mapping population which involves crossing multiple lines to a single reference parent (Morrell *et al.*, 2012), to create many RILs. Because of the diversity of the parents, NAM represents a cost effective and powerful strategy for genome-wide fine mapping by sequencing only the founder parents and genotyping the RILs with the same genetic markers (Yu *et al.*, 2008).

In addition to mapping populations, the availability of annotated genome sequences and advances in genomic technologies generally have opened doors for studying a whole host of candidate genes for complex traits by taking advantage of the natural genetic diversity and historical recombination events found in diverse cultivars, using a technique known as association mapping (Zhu *et al.*, 2008). Association mapping has the benefit of increased mapping resolution at reduced research time, and can be used to identify novel and superior alleles (Zhu *et al.*, 2008; Morrell *et al.*, 2012). Association studies use single nucleotide polymorphisms (SNPs) to search for statistical associations between phenotype and genotype (Morrell *et al.*, 2012), are now a promising approach in genetic mapping, and can be used to fine map genes within a QTL (Huang *et al.*, 2010; Han & Huang, 2013). There are two approaches that may be taken: (1) candidate-gene association mapping, which targets selected candidate genes and identifies SNPs within these genes and between lines; while (2) genome-wide association mapping searches for associations between genotype and phenotype at a whole-genome level. For example, Huang et al., (2010) carried out genome-wide association studies (GWAS) of 14 agronomic traits using 3.6 million SNPs across 517 rice landraces. The loci

identified explained on average 36 % of phenotypic variation. The peaks of these loci were often near known genes, suggesting GWAS can also be used to aid gene identification (Huang *et al.*, 2010).

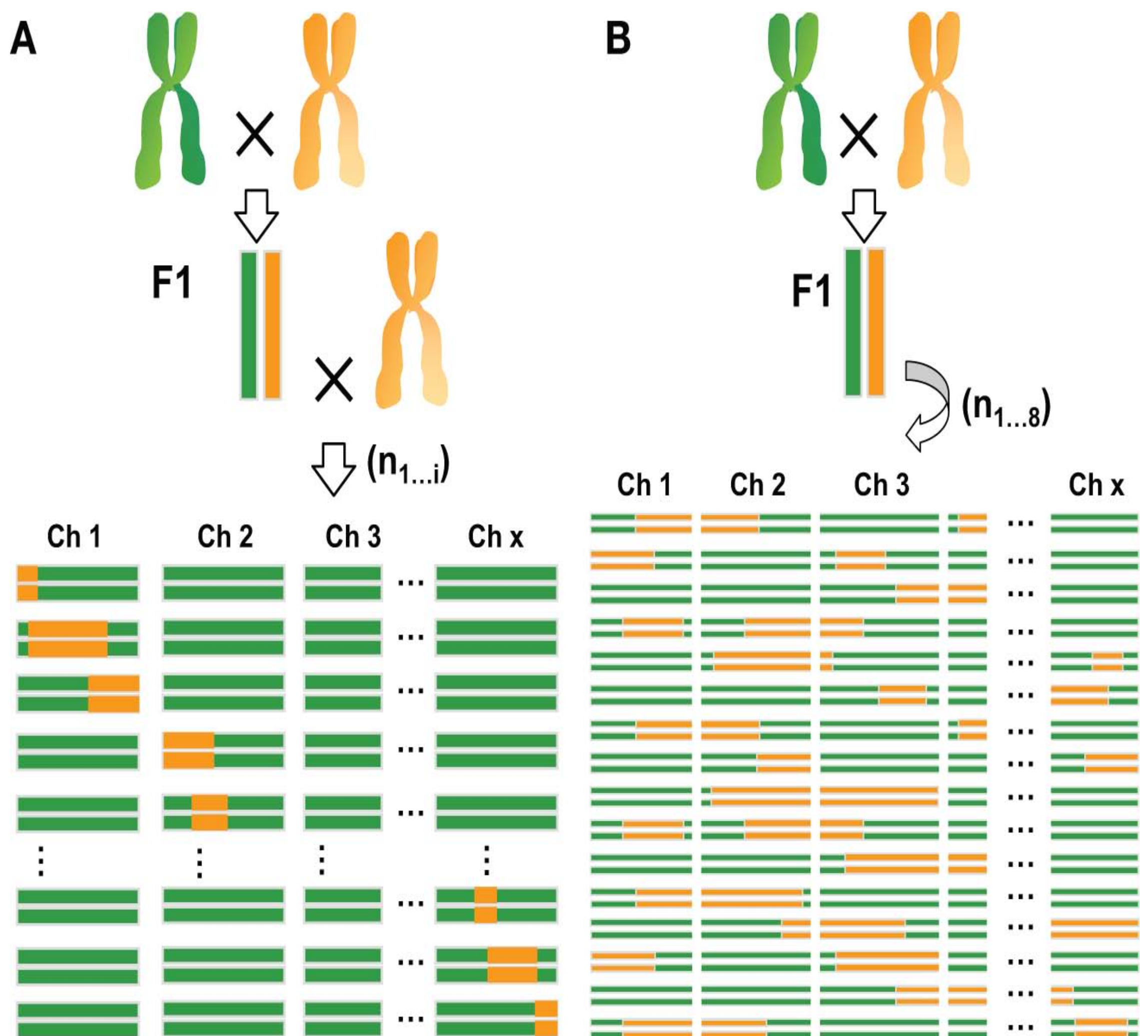


Figure 1.6 Comparison of a near isogenic line (A) and a recombinanat inbred line (B) mapping population. Figure from Fernie and Klee 2011.

1.8 Rice as a model system for studying *Striga* resistance

Although more commonly associated with Asia, rice has been cultivated in Africa for centuries (Nwanze *et al.*, 2006). Increased population and changes in consumer preferences have lead to growing demand for rice as a food source, and rice is now grown and consumed in 38 countries in sub-Saharan Africa (Balasubramanian *et al.*, 2007). Using rice to better understand

the molecular genetic basis of resistance to *Striga* has several advantages. Firstly it is considered the model monocot for molecular genetic studies; it's small genome size of 389 Mb has been accurately sequenced (Matsumoto *et al.*, 2005) and shows a high degree of synteny with other cereals (Bolot *et al.*, 2009), suggesting improved knowledge of the genetics underlying important agronomic traits may be applied to other cereals. In addition, rice possesses enormous within-species diversity (Garris *et al.*, 2005; Zhu & Ge, 2005; Schatz *et al.*, 2014). The re-sequencing of 3000 rice genomes from all known sub-groups, including both cultivated and wild accessions, has opened the door to a better understanding of the genetic diversity within this species (Xu *et al.*, 2011; The 3000 Rice Genomes Project, 2014), and the availability of mapping populations and wild rice species is a valuable source of genetic material which can be exploited for improved breeding purposes (Ali *et al.*, 2010; Fragoso *et al.*, 2017). In addition, insertion mutant germplasm such as the *Tos17* retrotransposon lines exists for rice (Hirochika, 2001; Sallaud *et al.*, 2004), which can contribute to functional genomics. Rice therefore is an ideal model cereal for investigating the genetic basis of resistance to *Striga*.

1.9 Aims and Objectives of thesis

The aim of this thesis is to identify novel QTL and candidate resistance genes in rice and functionally validate their role in providing resistance to *S. hermonthica* (Kibos isolate) using comparative and functional genomic approaches. Specific aims are:

1) To discover novel QTL for post-attachment resistance in rice to *S. hermonthica* using a rice mapping population (Chapter 2).

2) To identify candidate resistance genes within the IR64 QTL region and test the hypothesis that the candidate resistance genes are the same or similar to those identified in the Nipponbare resistance QTL (Chapter 3).

3) To determine whether the candidate receptor-like protein genes in Nipponbare underlie resistance to *S. hermonthica* (Chapter 4).
4) To examine the diversity of candidate resistance genes in the QTL across a range of genetically diverse rice cultivars to test the hypothesis that their diversity can help identify genes or combinations of genes underlying resistance to *S. hermonthica* (Chapter 5).

Chapter 2

Identification of QTL underlying resistance to *Striga hermonthica* (Kibos ecotype) in a rice Recombinant Inbred Line (RIL) population derived from a cross between IR64 (*O. sativa* ssp. *indica*) and Azucena (*O. sativa* ssp. *japonica*)

2.1 Introduction

Rice has been cultivated in Africa for centuries (Nwanze et al., 2006). Over the last 40 years, cultivation and consumption of rice in West and sub-Saharan Africa has increased dramatically (Balasubramanian *et al.*, 2007) and is now becoming one of the most important crops both in terms of consumption and cash income (Atera & Itoh, 2011). Cultivation of the West African species *Oryza glaberrima* is being replaced by the Asian species *Oryza sativa*, which possesses more favourable agronomic traits such as reduced lodging and higher yield, although it is less well adapted to the African climate (Jones *et al.*, 1997; Semagn *et al.*, 2007). In sub-Saharan Africa, the majority of rice is grown by poor subsistence farmers with limited access to expensive chemical inputs, meaning rice production here has the lowest yield in the world (Nwanze *et al.*, 2006). Weed competition is a serious constraint, accounting for yield losses in excess of 2.2 million tons per year, at an estimated value of $1.45 billion (Rodenburg & Johnson, 2009). Among the most damaging weeds of the semi-arid regions are *Striga* species; obligate root hemiparasites which infect staple cereal crops and cause severe reductions in growth and development of the host plant (Parker & Riches, 1993). *Striga* thrives where soil fertility is low, typically affecting the poorest subsistence farmers most severely, threatening food security and preventing a means to escape poverty (Rodenburg *et al.*, 2010).

Currently the most popular control measures with farmers are hand-weeding, intercropping and crop rotation to improve soil fertility (Atera *et al.*, 2012). Other control measures include the use of fertilisers and herbicide resistant seed, however these are often unaffordable for subsistence farmers (Hearne, 2009; Rodenburg *et al.*, 2010). The impacts of *Striga* on host growth and development are seen within a few days of attachment, well before the parasite emerges above ground (Press & Stewart, 1987; Frost *et al.*, 1997). Therefore an effective control strategy should act early on, preventing attachment or killing the parasite before it can become properly established. The use of genetically improved cultivars exhibiting a good level of resistance to *Striga* offers a cost-effective alternative which could be used effectively as part of an integrated control program aimed at improving soil fertility and reducing the *Striga* soil seed bank (Hearne, 2009; Yoder & Scholes, 2010; Atera et al., 2012).

Although *Striga* resistance is known to exist in rice germplasm, very few genotypes show complete resistance or immunity, even the most resistant cultivars can still support one or two *Striga* plants. This poses a threat to the durability of resistance, especially given the high level of genetic diversity in *Striga* populations and the huge numbers of seed produced by one *Striga* plant alone (Rodenburg & Bastiaans, 2011). The success of a parasite on a host plant also depends on the interaction between the host genotype and the parasite population or

ecotype (Huang *et al.*, 2012b). As on outbreeding parasite, *S. hermonthica* shows particularly high levels of genetic diversity (Safa *et al.*, 1984). Host resistance may also vary under different climates, as well as between different sites and years (Rodenburg *et al.*, 2017). For resistance to be broad spectrum and durable, it is therefore desirable to combine multiple resistance genes for both pre and post-attachment resistance, or different mechanisms of post attachment resistance, into a single cultivar.

Genetic mapping studies have identified a number of QTL for resistance to *S. hermonthica* in cereals. In sorghum, several *S. hermonthica* resistant QTL were mapped in two different mapping populations (Haussmann *et al.*, 2004). Resistant parental lines exhibited mechanical resistance (N13) or low production of germination stimulants (IS9830). Field trials were conducted over two years in Mali and Kenya, and composite interval mapping used to detect 11 different resistant QTL in each population, with 5 QTL common between the two populations. A highly significant QTL x environment interaction was observed at each test location (Haussmann *et al.*, 2004). Mapping studies have also identified examples of both pre and post-attachment resistance to *Striga* in rice (Harahap *et al.*, 1993; Gurney *et al.*, 2006; Cissoko *et al.*, 2011; Jamil *et al.*, 2011; Atera *et al.*, 2015; Samejima *et al.*, 2016). Pre-attachment resistance is associated with lower production of germination stimulants in host root exudates (Jamil *et al.*, 2011b) while post-attachment resistance is associated with intense necrosis at the attachment site and a failure of the parasite to penetrate the root endodermis (Gurney *et al.*, 2006; Cissoko *et al.*, 2011). Gurney *et al* (2006) identified seven putative *S. hermonthica* resistance QTL from a BIL population developed from a cross between Nipponbare, an *O. sativa ssp*. temperate *japonica* cultivar that is resistant to *S. hermonthica*, and Kasalath, an *indica* cultivar that is more susceptible. The QTL on chromosome 12 (Nipponbare allele) and on chromosome 4 (Kasalath allele) explained the greatest percent of phenotypic variance in *S. hermonthica* resistance. A further QTL analysis in a Koshihikari-Kasalath BIL confirmed the QTL on chromosome 4 and verified its role in a different genetic background (Swarbrick *et al.*, 2009). The Nipponbare QTL was also mapped in a second population; a BIL population created from a cross between Nipponbare (resistant to *S. hermonthica*) and Koshihikari (temperate *japonica* and more susceptible), also identified the same QTL on chromosome 12 (Scholes, personal communication) (section 1.6).

A rice Nested Association Mapping (NAM) population has recently been produced, consisting of a series of 10 RILs, each representing a genetically different tropical *japonica* parent, but all crossed to the same *indica* reference parent, IR64 (Fragoso *et al.*, 2017). This population therefore represents broad genetic diversity in rice (Fragoso *et al.*, 2017), and is ideal for

uncovering new sources of *S. hermonthica* resistance. Approximately 300 seeds from the F_1 progeny of each cross were selected for single-seed descent to the F_7 generation to produce just under 200 homozygous RILs for each population (see Figure 2.1) (Fragoso *et al.*, 2017). A RIL population contains multiple introgressed fragments, and on average the genome of each parent is represented equally throughout the genome. This allows for the testing of epistasis and polygenic resistance (Mauricio, 2001; Fernie & Klee, 2011).

2.1.1 Aim of Chapter 2

The aim of this chapter is to discover novel QTL for post-attachment resistance in rice to S. *hermonthica* (Kibos isolate) using a RIL population (from the NAM population) and to test the hypothesis that the broad genetic diversity of the *O. sativa japonica* and *indica* parents will allow the discovery of novel *S. hermonthica* -resistance QTL. The specific objectives are to:

1) Phenotype the parental lines of the rice NAM population for post-attachment resistance to *S. hermonthica* to identify an appropriate RIL population (where the parental lines show different *S. hermonthica* -resistance/susceptibility phenotypes).
2) Phenotype lines of the RIL population for post-attachment resistance to *S. hermonthica* and carry out an analysis to identify *S. hermonthica* -resistance QTL.
3) Determine the mode of inheritance of the resistance by phenotyping F_1 progeny derived from a cross between the two parental lines of the RIL population.
4) Characterise the resistance phenotype of the resistant parent of the RIL population at a microscopic level.

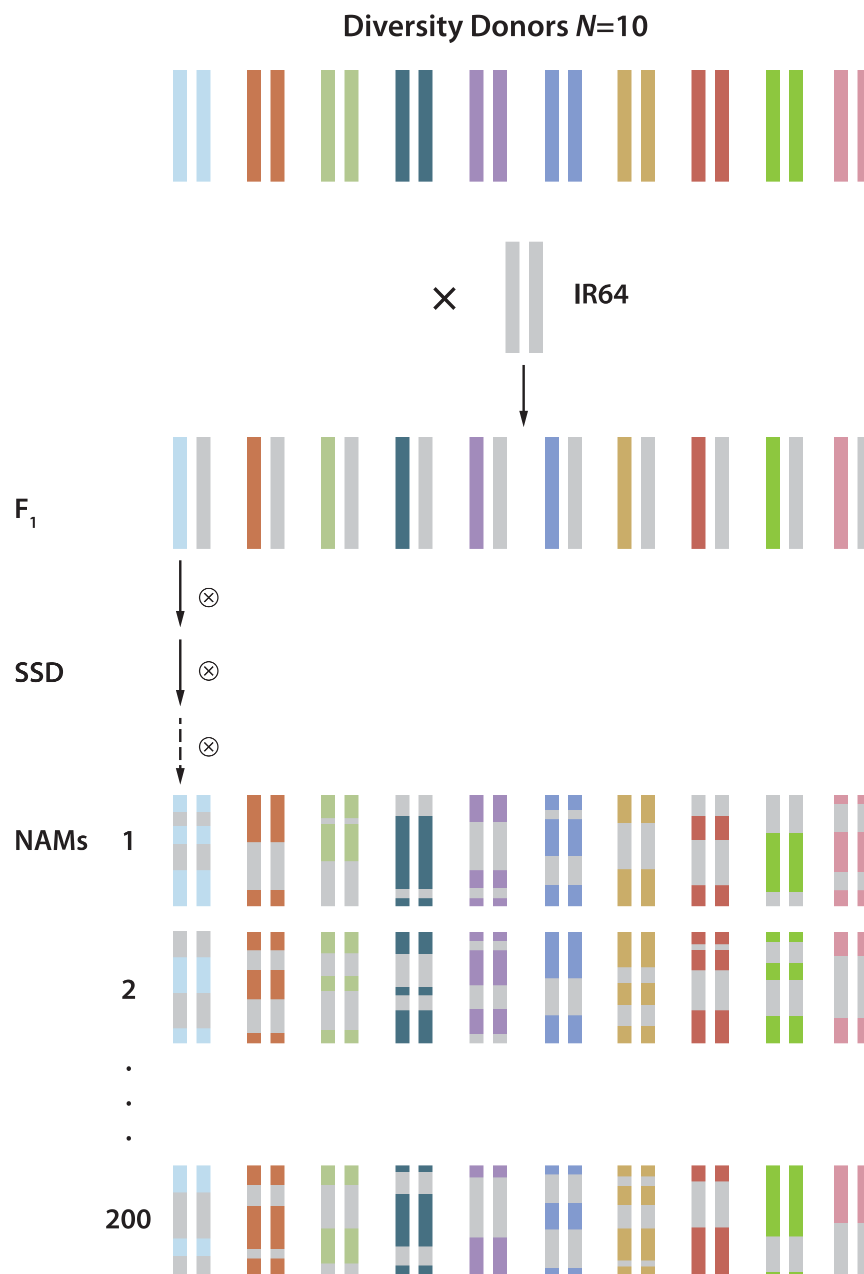


Figure 2.1. Development of a rice Nested Association Mapping population. The 10 founder parents are *O. sativa ssp.* tropical *japonica*, all crossed to the common parent IR64 (*O. sativa ssp. indica*). Each RIL population is comprised of nearly 200 individuals derived from each F1 cross, followed by single seed descent to F_7. The greater NAM population consists of 1,879 lines in total. Diagram from Fragoso *et al.*, 2017.

2.2 Materials and Methods

2.2.1 Plant materials

Rice seeds of parental genotypes of the NAM population, consisting of IR64 (Oryza sativa ssp. *indica*) and 9 *Oryza sativa* ssp. tropical *japonica* genotypes (CT9998-41-12-M-4-1, CT10037-56-6-M-M-1, CT10005-12-1-M-4-1, ITA 164-1, CT10006-7-2-M-2-1, CT8556-37-1-3-1-M-1, CT10035-26-4-2-M-1, CT10035-42-4-4-M-1, CT10045-5-5-M-1-1), were supplied by Dr. Mathias Lorieux at the International Centre for Tropical Agriculture (CIAT), Colombia. However, since the creation of this population, two of the lines (CT8556-37-1-3-1-M-1 and CT10035-26-4-2-M-1) are now known to cluster closer to the *indica* rather than the *japonica* subspecies (Fragoso *et al.*, 2017). Parental and F_1 rice seeds derived from a cross between IR64 and Azucena were also supplied by Dr. Mathias Lorieux. Nipponbare and Koshihikari seeds (both *Oryza sativa ssp.* temperate *japonica*) were supplied by Dr. Hori Kiyosumi at the National Institute for Agrobiological Sciences (NIAS), Japan. *S. hermonthica* seeds were collected from plants parasitising maize in Kibos, Western Kenya, in 2009 or 2013.

2.2.2 Phenotyping NAM parental lines for post-attachment resistance to *S. hermonthica*

Rice seeds were germinated between 2 sheets of moistened, glass fibre filter paper (GF/A, Whatman™, Buckinghamshire), held between 2 blocks of wet horticultural rockwool (Groden®, Vital) and incubated at 30 °C (Figure 2.2). After 7 days, seedlings were transferred to rhizotrons, consisting of a 25 x 25 x 2cm perspex chamber containing vermiculite medium (William Sinclair, Horticulture Ltd, Gainsborough, UK) a strip of rockwool, and a sheet of 100 µm mesh (Plastok Group, Birkenhead, UK) placed on top. Roots of the rice seedlings grew down over the mesh, and holes at the top and bottom of the rhizotrons allowed for shoot growth and water drainage, respectively (Cissoko *et al.*, 2011). Roots were kept in the dark by covering the rhizotrons with aluminium foil. Plants were grown in a controlled walk-in chamber (Figure 2.2) with a day/night temperature of 28 ˚C /24 ˚C, 60 % relative humidity and a 12 h photoperiod at an irradiance of 500 μmol s^{-1} m^{-2}. Each plant was watered 4 times a day with a total volume of 27 ml of 40 % Long Ashton solution (Hewitt, 1966) containing a 2 mM ammonium nitrate as the nitrogen source, using an automatic watering system.

S. *hermonthica* seeds were surfaced sterilised with 10 % bleach for 5-10 min, washed thoroughly with distilled water, and gently spread over moistened filter paper (GF/A, Whatman™, Buckinghamshire) in Petri dishes. These were incubated in the dark at 30 °C for 13 days. Eighteen hours before infection of rice seedlings, *S. hermonthica* seeds were germinated by adding 2 ml of 0.1 ppm GR24, an artificial germination stimulant, to ensure consistent attachment to the rice roots and avoid any differences in pre-attachment resistance that might

exist between different cultivars. The percentage germination of *S. hermonthica* seeds was recorded before infection. Fifteen days after sowing (das) the rice seeds, seedlings were inoculated with 15 mg of germinated *S. hermonthica* seeds by pipetting them along the roots and then carefully positioning them within 5 mm of the roots using a small, soft paintbrush.
Five rice plants of each parental line were infected with *S. hermonthica* seeds collected from plants parasitising maize in Kibos, Western Kenya, in 2009. To compare the virulence of this *S. hermonthica* seed batch with that collected from the same location in 2013, an additional 5 plants of IR64 and Azucena were also infected with the 2013 *S. hermonthica* seed batch. Twenty-one days after inoculation (dai) with *S. hermonthica* seeds images of the root systems of rice plants were acquired by placing an open rhizotron onto the surface of a flatbed scanner (Canon CanoScan 9000F) and scanning at high resolution. *S. hermonthica* plants were then harvested from rice roots, placed in Petri dishes and photographed using a Canon EOS500D digital camera. They were then dried at 30 °C for 5 days and then weighed to obtain the total dry biomass of *S. hermonthica* on each rice plant. The number and length of parasites was also determined from Petri dish images using Image J software (http://imagej.nih.gov/ij/). *S. hermonthica* plants that were smaller than 3.5 mm in length were not considered successful attachments and were discounted from measurements.

2.2.3 QTL mapping strategy and analysis

The IR64 x Azucena RIL population, which consists of 184 homozygous lines, was chosen for QTL mapping of *S. hermonthica* resistance. This population was genotyped by sequencing at the University of Yale to identify 44,500 high quality SNP markers. Sequencing reads mapped onto the Nipponbare reference sequence (Nipponbare reference MSU v.7.0). The computer program MapDisto (Lorieux, 2012) was used to impute missing data, correct for genotyping errors and construct a genetic map. Different algorithms were run within MapDisto to achieve this. LB impute was used to correct for false homozygosity and impute missing genotypes (Fragoso *et al.*, 2016), and BP-Impute used for breakpoint imputation. These were combined with final imputation using the R/qtl 'argmax' function. Further genotyping corrections were carried out using the 'Color genotypes' tool in MapDisto. Redundant SNP markers from non-recombining regions were filtered out. The above analysis was performed by Dr Mathias Lorieux, CIAT, Colombia and collaborators at the University of Yale. The filtered SNP data was made available for the mapping of *S. hermonthica* resistance QTL.

Phenotyping large numbers of plants for resistance / susceptibility to *Striga* is time consuming; therefore a two-step approach was implemented to reduce the number required for accurate QTL detection. In step 1, 44 lines were chosen at random for phenotyping. Four rice plants of

each line were grown, as well as an additional 4 of each parent. These were inoculated with *S. hermonthica* and scored for resistance as described in section 2.2.2, making a total of 184 plants. As it was not possible to infect this many plants at one time point, plants were split into 3 batches of approximately 62 plants each. The 4 replicates were distributed over the 3 batches to help calibrate for variation that may occur between batches. Batches 1 and 2 were separated by 5 days, and batches 2 and 3 were separated by 2 days. A linear mixed model was run on the *S. hermonthica* dry biomass data to determine how much variation was due to variation between batches compared to variation between genotypes, with batch and genotype fitted as random effects. As variation between batches was detected, data were normalises by Z-scores before running the QTL analysis. Z-scores were calculated for each line by finding the difference in mean *S. hermonthica* biomass between each line and the mean biomass of the population, and then dividing this value by the standard deviation of the population. QTL analysis was carried out in the computer program MapDisto (Lorieux, 2012). Linkage groups were first determined by mapping SNPs on the physical map (Nipponbare reference MSU v.7.0). A one-way ANOVA was implemented to search for trait-to-marker associations using the built-in F-test feature in MapDisto. The more robust Kruskal-Wallis method was also run for comparison. In addition, the R/qtl interface was used to perform an additional QTL analysis by selecting the interval mapping method (Lander & Botstein, 1989).

On detection of a QTL, an additional 20 lines were selected that showed recombination within the QTL region. These were phenotyped for *S. hermonthica* resistance as before. Four replicates were again used for each line, spread over 2 batches which were separated by 2 days. Data from both phenotyping experiments were then pooled and normalised using Z-scores to account for variation between replicates and batches. Data normalisation and scaling was carried out by Dr. Mathias Lorieux, CIAT, Colombia. Pooled data were used in subsequent QTL analyses, which was carried out as described above.

2.2.4 Generation of F1 progeny by cross pollination

In order to determine the mode of inheritance of resistance in IR64 and Nipponbare, the resistant cultivars were each crossed with two susceptible cultivars, Azucena and Koshihikari. In addition, a cross was also made between IR64 and Nipponbare, and between Azucena and Koshihikari. The F_1 generations from each cross were phenotyped for resistance to *S. hermonthica* (Kibos ecotype).

In order to carry out the crosses, the outer leaves of rice panicles were cut away from the base to expose all spikelets, before flowering (Figure 2.3). The top of each spikelet in the panicle

was cut off at a slight angle just above the anthers using scissors. Anthers were removed from each spikelet using a glass syringe attached to a vacuum. A second panicle from a different cultivar was chosen for the pollen donor. The two panicles were tied together, covered in a glassine bag, which was closed with a paperclip, and staked. Panicles were flicked several times around midday for the next 2 - 3 days to increase the chance of pollination. Once seeds had developed and panicles had turned yellow, panicles were cut from the plant and dried at 40 °C for one week. Seeds were stored at 4 °C until use. To test that cross pollination of parental cultivars had taken place, rice seedlings where genotyped by PCR using polymorphic simple sequence repeat (SSR) markers and a primer in the Os12g11370 gene that had previously been shown to amplify different product sizes for the cultivars crossed (Table 2.1). Thermo Scientific Phire Plant Direct PCR kit was used for genotyping, which was carried out before plants were infected with *S. hermonthica*. A small piece of leaf tip approximately 2 mm in diameter was taken from each seedling, and placed in 20 µl of Dilution Buffer. A leaf tip was also removed from each parental plant. The leaf sample was then crushed against the side of the tube with a sterile pipette tip until the solution turned green. After spinning down, 1 µl of the supernatant was used as a DNA template in a 20 µl PCR reaction. Each 20 µl PCR reaction consisted of 10 µl 2 x Phire Plant PCR Buffer, 1 µl of each primer at a concentration of 10 µM (to give a final concentration of 0.5 µM in the reaction), 0.4 µl Phire Hot Start II DNA polymerase, 1 µl DNA template and 6.6 µl H_2O. The PCR cycling program was: 5 min at 98 °C (initial denaturation); then 40 cycles of 5 s at 98 °C (denaturation), 5 s at 62 - 65 °C according to individual primers (annealing), and 20 s at 72 °C (extension); followed by a final extension of 1 min at 72 °C. 5 µl of 5 X DNA Loading Buffer, blue (Bioline) was added to each sample, and 12 µl of the mix loaded onto a 1.5 – 2 % agarose gel, according to product size. F_1 samples were run next to parental genotypes until good separation of bands was visible.

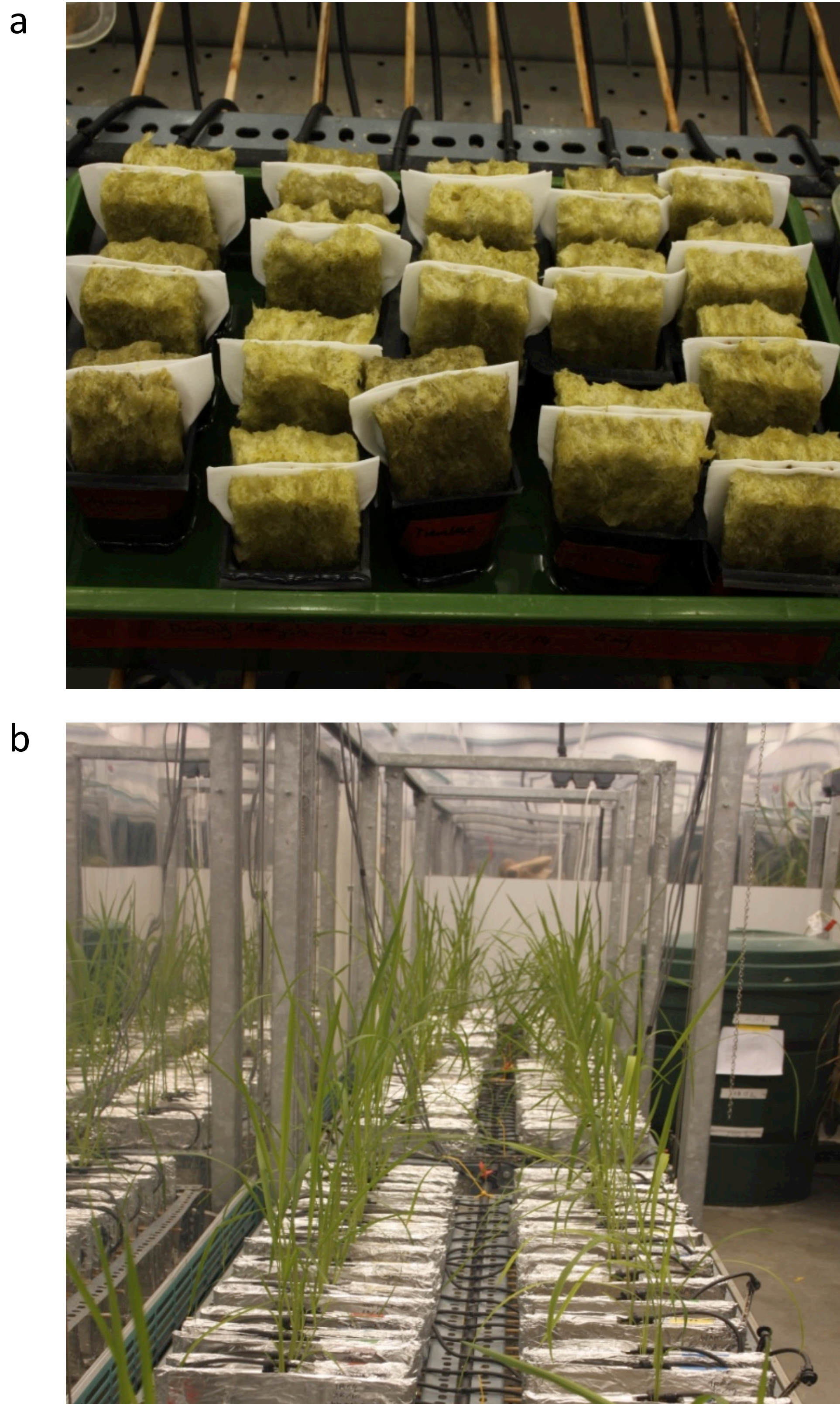


Figure 2.2. a) Rice seeds germinating between sheets of filter paper before being transferred to rhizotrons. b) *S. hermonthica*-infected rice plants growing in rhizotrons, watered by an automatic watering system.

Table 2.1. Primer sequences for confirming successful cross pollination of F_1 progeny.

Primer name	Forward primer sequence (5’-3’)	Reverse primer sequence (5’-3’)	Cross
11370 genotype	AATTCGCTCACTAGGATTGAGCTT	AAGATCTGCGGAGGCACCTT	Nip x Kosh IR64 x Kosh Azucena x Kosh
RM101	GTGAATGGTCAAGTGACTTAGGTGGC	ACACAACATGTTCCCTCCCATGC	IR64 x Azucena
RM164	TCTTGCCCGTCACTGCAGATATCC	GCAGCCCTAATGCTACAATTCTTC	Nip x Azucena

2.2.5 The phenotype of resistance

The phenotype of resistance or susceptibility on IR64 and Azucena was investigated by taking images of *S. hermonthica* at the site of attachment on rice roots. Images were taken at 21 dai using a Leica stereomicroscope (Leica MZ FLIII). To obtain a cross-sectional view of susceptible and resistant host responses, a selection of attachments were cut and fixed in Carnoy’s fixative (4 : 1, 100 % ethanol : acetic acid) by vacuum infiltration for 20 min. Samples were then washed twice in 100 % ethanol for 30 min, before being infiltrated in a solution of Technovit 1: 100 % ethanol (1:1) (Technovit [(R)] 7100) for 2 h. Samples were then placed in 100 % Technovit 1 for 15 min, then transferred to fresh Technovit 1 for a further 3 d. Each sample was then embedded in resin mould by mixing Technovit 1 and Hardener 2 in a 15 : 1 ratio according to manufacturer’s instructions. After mounting onto histoblocks, 5 µm sections were cut through the attachment using a Leica microtome (RM2145), placed on a microscope slide (Super Premium Microscope Slides, Gurr [(R)]) in distilled water, and dried on a hot plate at 65 °C. Sections were stained for 2 min in 1 % Toluidine blue at 65 °C, then washed for 5 min in distilled water. Sections were mounted permanently in Dep-Pex mounting medium (Gurr [(R)], PROLABO [(R)]) and photographs taken with an Olympus (BX51) microscope.

2.2.6 Statistical analysis

The statistical package R, version 3.3.0 (http://www.r-project.org) was used for analysis of S. *hermonthica* biomass and number data of the NAM parental lines and the F_1 progeny. For the NAM parental lines, a one-way ANOVA was carried out to assess significant differences between genotypes, followed by a Tukey Multiple Comparison test to identify where the significant differences lie. Two-sample t-tests were carried out on the F_1 biomass data. Tests were carried out on log_{10} transformed data, to adjust for non-normal distribution.

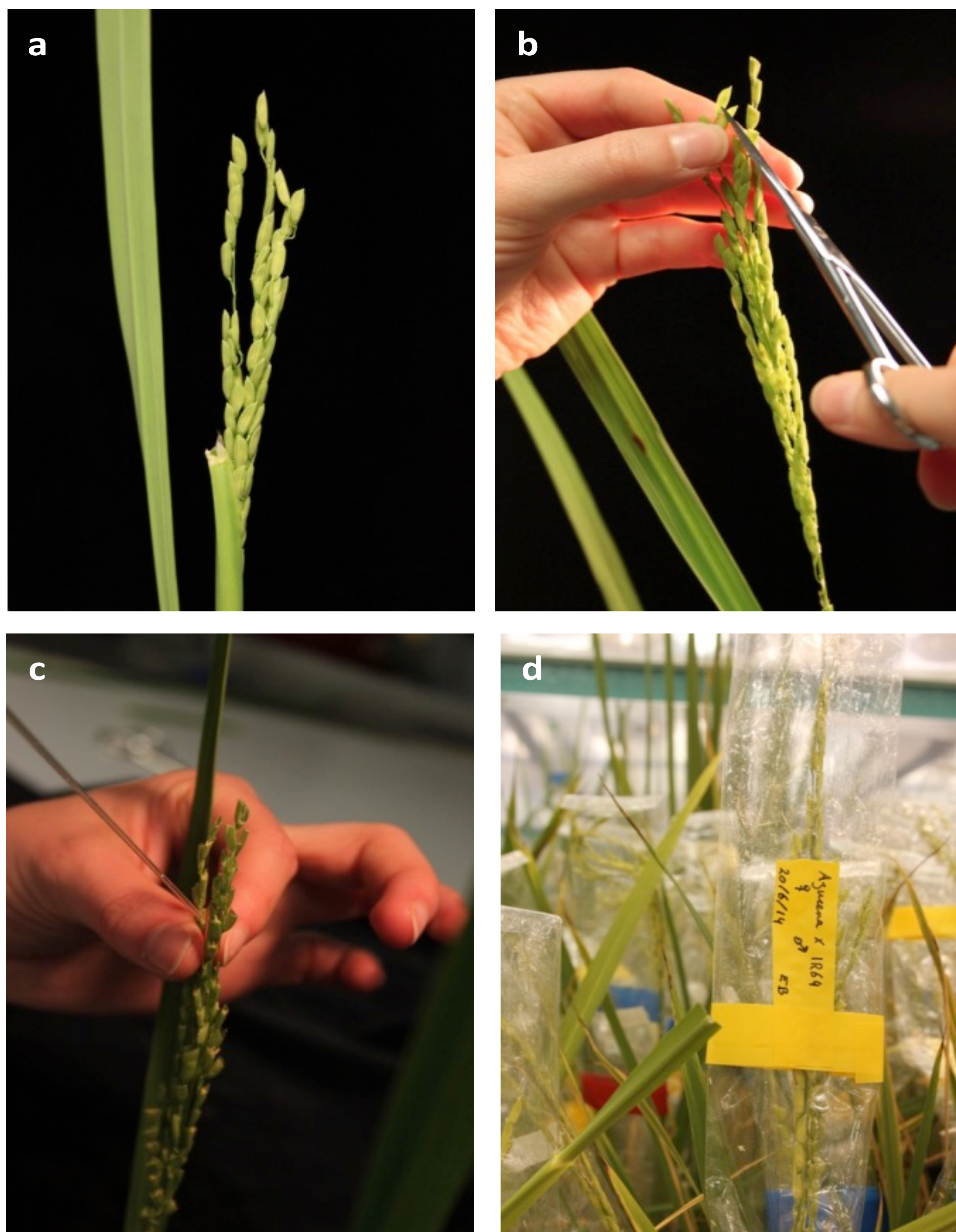


Figure 2.3. Cross-pollination of rice. a) panicle before removal of outer leaves. b) spikelets were cut at an angle above the anthers. c) removal of anthers by vacuum. d) panicles bagged together in a glassine bag.

2.3 Results

2.3.1 Evaluation of post-attachment resistance of NAM parents to *Striga hermonthica*

In order to identify an appropriate rice mapping population for detection of *S. hermonthica* resistance QTL, 11 parents of a NAM population were phenotyped for post-attachment *S. hermonthica* resistance (Figure 2.4). There was a highly significant difference between genotypes for *S. hermonthica* dry biomass (ANOVA: F = 24.3, d.f. = 12, 54, p < 0.001), number (ANOVA: F = 12.4, d.f. = 12, 54, p < 0.001) and cumulative length (ANOVA: F = 18.1, d.f. = 12, 54, p < 0.001). There was no significant difference in *S. hermonthica* dry biomass, number or cumulative length between plants infected with *S. hermonthica* seed collected in 2009 compared to those collected in 2013. The *S. hermonthica* biomass collected from these lines fell broadly into two categories: resistant or susceptible. The IR64 parent showed good resistance to *S. hermonthica*, while Azucena was the most susceptible. NAM parents CT10006, ITA164 and CT10035-26 were all susceptible to *S. hermonthica*, with a dry biomass of between 10.9 mg and 14.2 mg. All other cultivars exhibited a good level of resistance, which did not differ significantly from IR64 (Tukey's honesty significant differences p < 0.05). There was also a highly significant difference in the number of *S. hermonthica* seedlings between genotypes (ANOVA: F = 12.4, d.f. =12, 54, p < 0.001). Mean number of *S. hermonthica* seedlings per genotype, and mean cumulative length of *S. hermonthica* seedlings, showed similar patterns to *S. hermonthica* biomass, although for mean number the difference between resistance and susceptibility was less well defined (Figure 2.4). The RIL population from a cross between the most susceptible cultivar, Azucena, and IR64 was chosen for further study.

2.3.2 The phenotype of resistance in IR64

Transverse sections of IR64 and Azucena roots at the site of *S. hermonthica* attachments 21 dai revealed the phenotype of these cultivars to be resistant and susceptible respectively **(**Figure 2.5). The cultivar Azucena was very susceptible with many *S. hermonthica* individuals on each host root system (Figure 2.5 a), whereas very few parasites attached and grew on the root systems of IR64 (Figure 2.5 b). The resistance response of IR64 was characterised by necrosis at the site of attachment (Figure 2.5 b inset). Transverse sections through the root and parasite haustorium revealed that the parasite had penetrated the root cortex, endodermis and made vascular connections with the host xylem in the susceptible cultivar Azucena (Figure 2.5 c). In contrast, in the resistant cultivar IR64, the parasite penetrated the cortex but failed to penetrate the endodermis and thus could not form connections to the host xylem vessels.

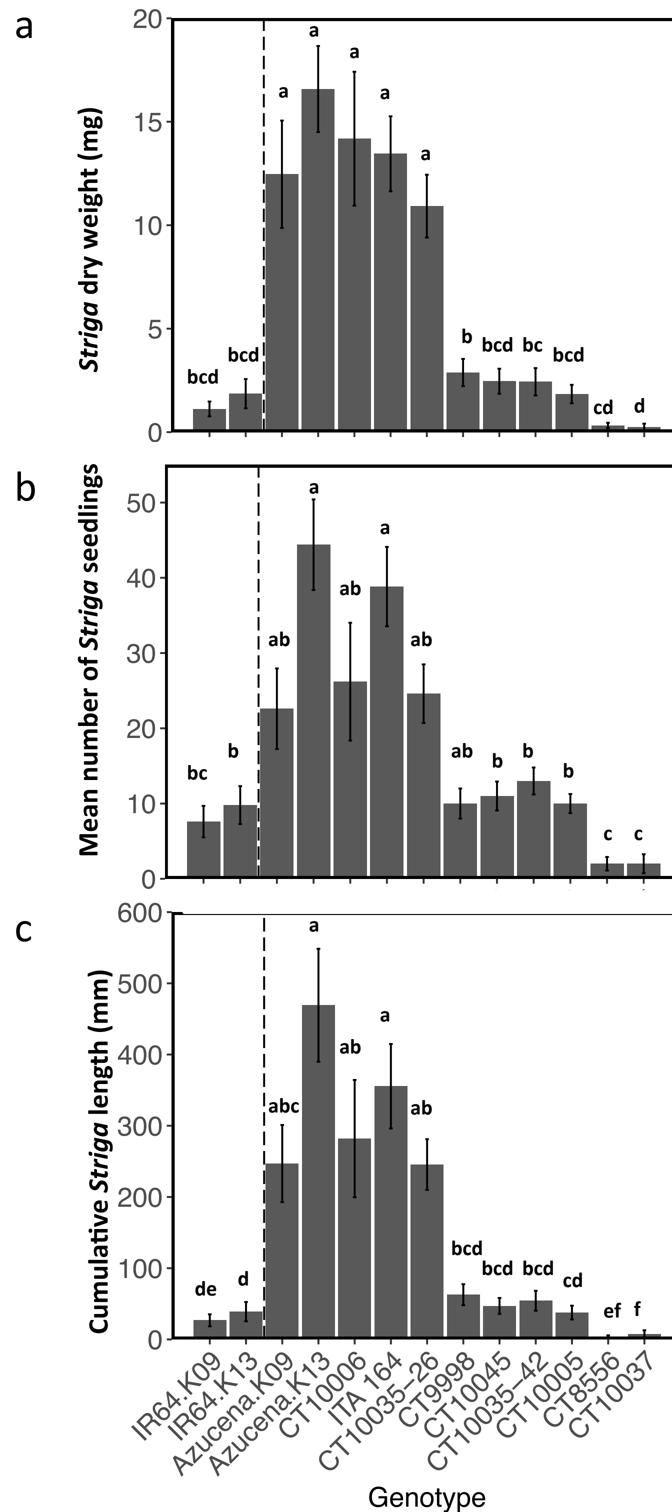


Figure 2.4. The dry biomass (a) mean number (b) and mean cumulative length (c) of *S. hermonthica* plants harvested from parents of a Nested Association Mapping population 21 dai with germinated *Striga hermonthica* seed collected from Kibos in Western Kenya. IR64 and Azucena cultivars were infected with *S. hermonthica* seed collected in both 2009 (K09) and 2013 (K13). All other genotypes were infected with Kibos 2009 seed. Values are means ± SE, where n = 5. There was a highly significant effect of genotype on *S. hermonthica* biomass, number and cumulative length (ANOVA $p < 0.001$). Tukey's honesty significant differences ($p < 0.05$) are represented by letters above each bar: different letter codes represent significant differences.

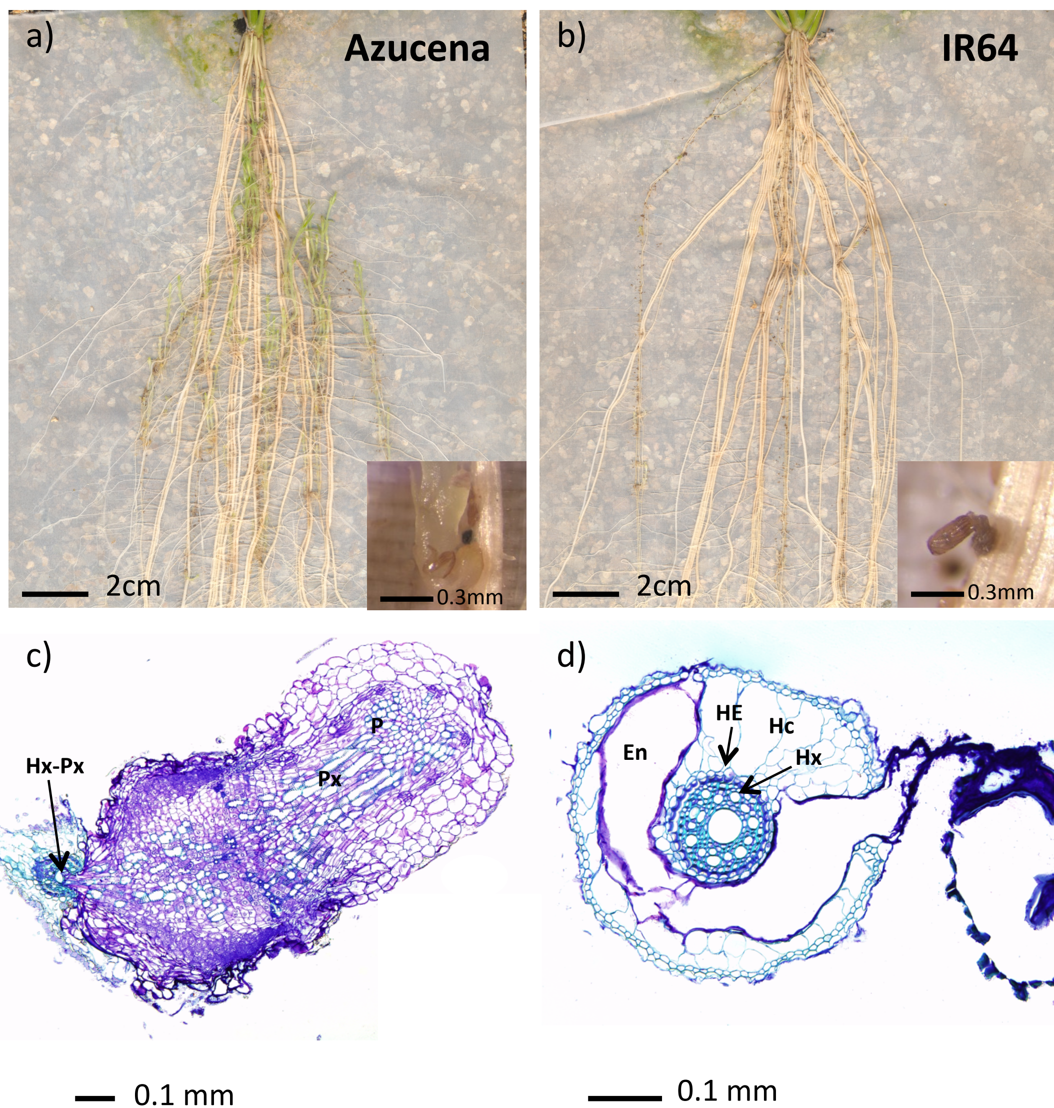


Figure 2.5. The phenotype of resistance and susceptibility of IR64 and Azucena rice cultivars to *S. hermonthica*, (Kibos ecotype). a and b) Root systems of *S. hermonthica* infected plants 21 days after infection. b) The resistant reaction was associated with necrosis around the site of parasite attachment (inset). c-d) Transverse sections through the attachment and host root. c) The parasite has successfully penetrated the host cortex and endodermis and formed fully established xylem-xylem connections. d) Resistance in IR64 is shown by failure of the parasite to breach the endodermis and attach to the host xylem vessels, instead growing through the cortex and around the vascular core. Hx-Px, host-parasite xylem; En, Parasite endophyte; HE, host endodermis, Hc, host cortex, Hx, host xylem vessels.

2.3.3 Identification of a *Striga hermonthica* resistance QTL

S. hermonthica dry weight from the 64 RILs phenotyped ranged from very resistant to very susceptible (Figure 2.6). Azucena was the most susceptible with a mean biomass of 40.1 mg *S. hermonthica*. Twelve of the RILs were more resistant than IR64, with a mean biomass of less than 1.75 mg. Most lines had intermediate resistance between the IR64 and Azucena parents.

S. hermonthica dry weight was normalised using z-scores and these scores were used as trait data for QTL analyses. A number of different QTL analyses were performed (Table 2.2). A single major QTL was detected on chromosome 12 between positions 5.7 – 6.7 Mb with all analyses conducted (Table 2.2, Figures 2.7 - 2.10). No other QTL were detected on chromosomes 1 - 11. The one-way ANOVA produced a LOD score of 14.07 (Figure 2.7, Table 2.2), and the percentage of variance explained by the QTL was 67.9%. The same QTL was also detected by the more robust Kruskal-Wallis method with almost as high significance (LOD = 12.37) (Figure 2.8, Table 2.2), and by interval mapping (LOD = 14.34) (Figure 2.9, Table 2.2). A lack of recombination between positions 5759093 – 6719392bp (Figure 2.10) meant it was impossible to define the QTL further.

Table 2.2. Summary statistics of the QTL for *S. hermonthica* resistance in the rice IR64 x Azucena RIL population. QTL analysis was performed using 3 different methods.

Method	Chromosome	Marker	Position (cM)	LOD	% R^2	F / H statistic	-log 10(p)
One-way ANOVA	12	Chr12_5759093 – Chr12_6719392	32.0	14.07	67.9	116.48 (F)	14.47
Kruskal-Wallis	12	Chr12_5759093 – Chr12_6719392	31.99	12.37	-	56.97 (H)	13.36
Interval Mapping	12	Chr12_5672456- Chr12_6863703	32.0	14.34	-	-	-

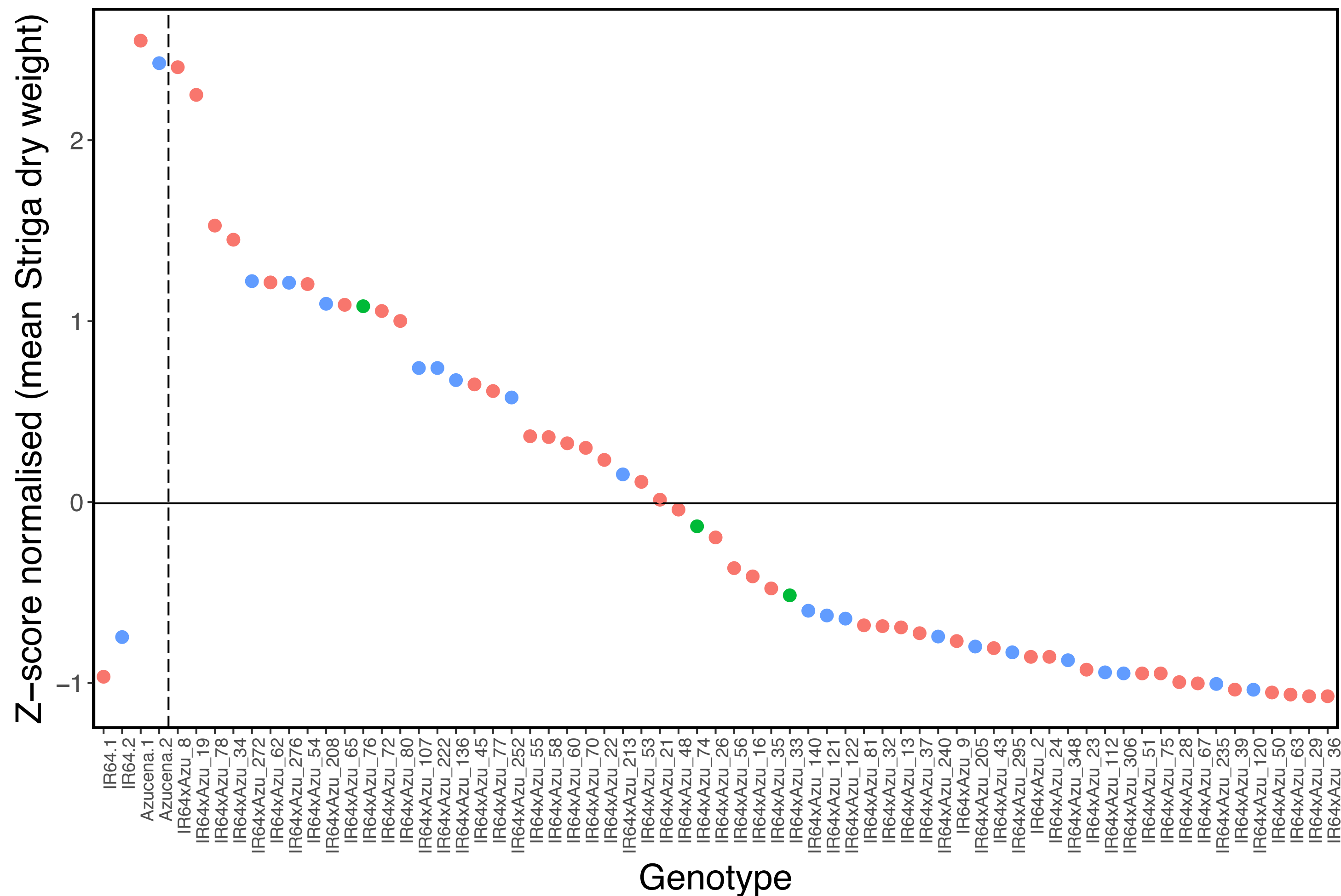


Figure 2.6. Distribution of resistance to *Striga hermonthica* (Kibos isolate) 21 dai in the rice Recombinant Inbred Line IR64 x Azucena mapping population. Forty-four randomly chosen lines (shown in red) where phenotyped initially. Further lines (shown in blue) showing recombination within the QTL region where then phenotyped. Green dots are lines phenotyped in both experiments. Parents of the mapping population are IR64 (*indica*) and Azucena (*japonica*). Data shown are z-score values calculated from mean *S. hermonthica* dry weight, n=4.

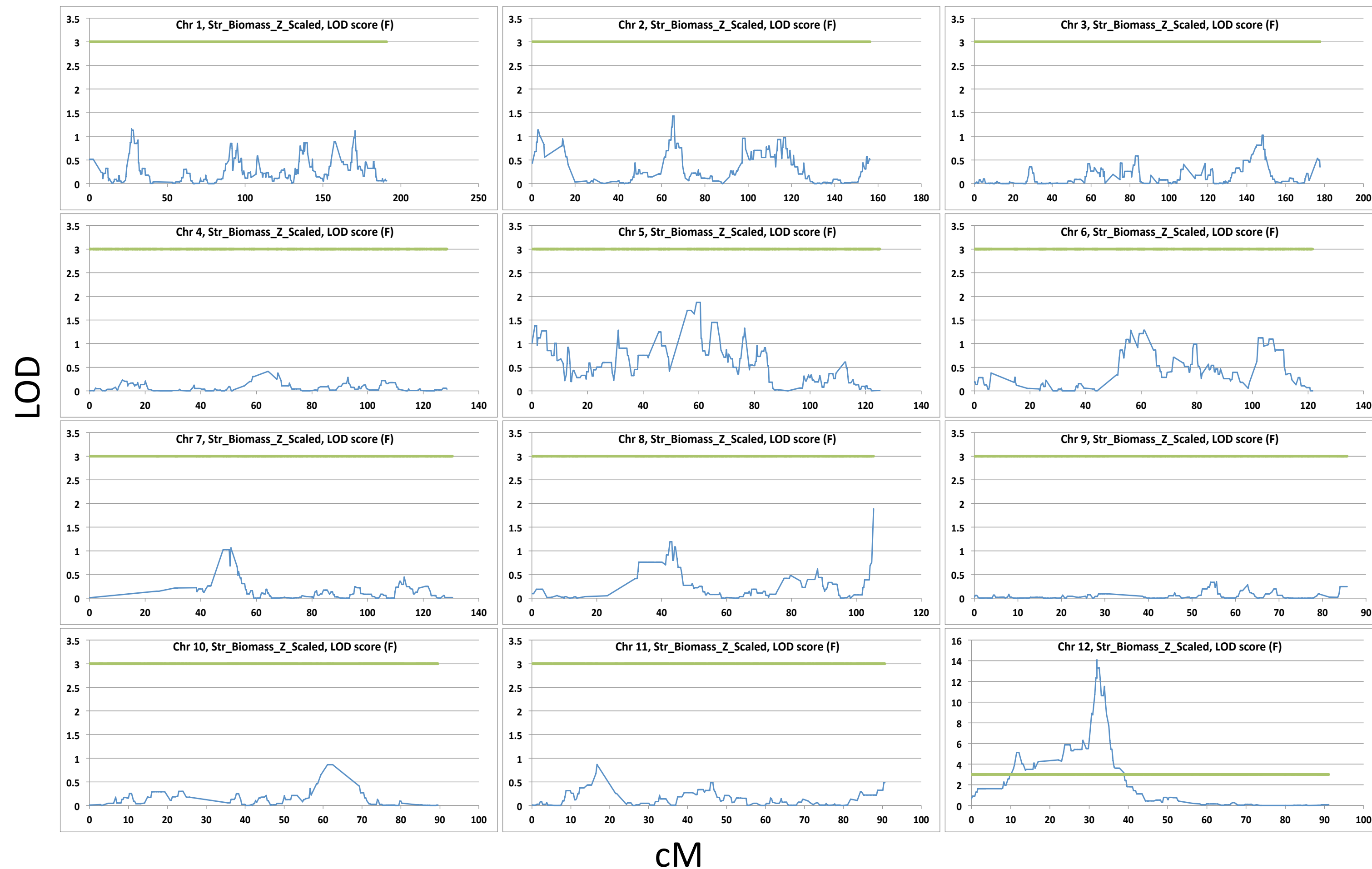


Figure 2.7. Logarithm of odds (LOD) scores for *S. hermonthica* resistance across all rice chromosomes as determined by a one-way ANOVA method. QTL mapping analysis was carried out using 64 recombinant inbred lines (RILs) from the IR64 x Azucena rice mapping population. Green horizontal lines represent threshold levels of statistical significance (LOD scores of 3).

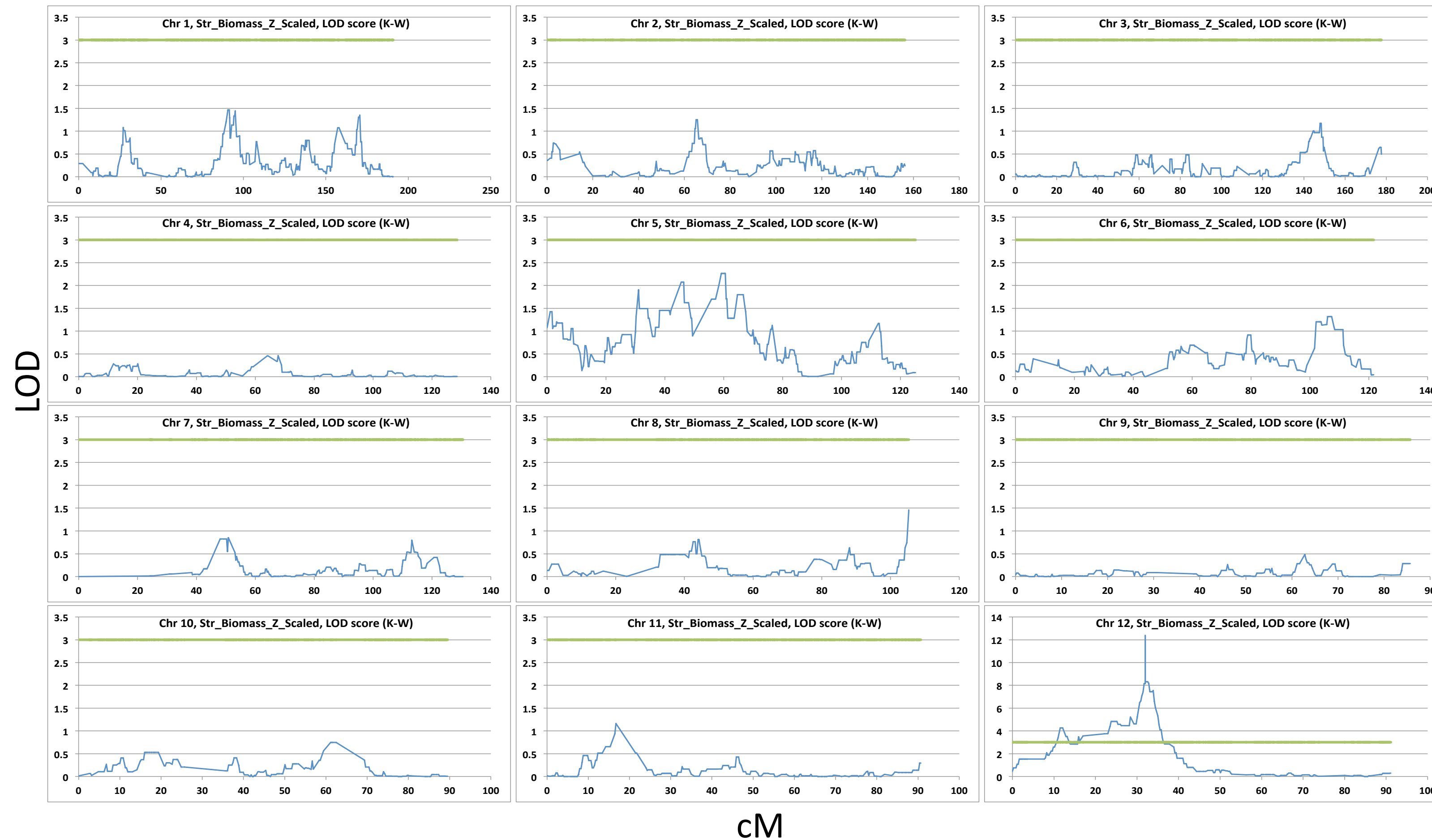


Figure 2.8. Logarithm of odds (LOD) scores for *S. hermonthica* resistance across all rice chromosomes as determined by a Kruskall Wallis method. QTL mapping analysis was carried out using 64 recombinant inbred lines (RILs) from the IR64 x Azucena rice mapping population. Green horizontal lines represent threshold levels of statistical significance (LOD scores of 3).

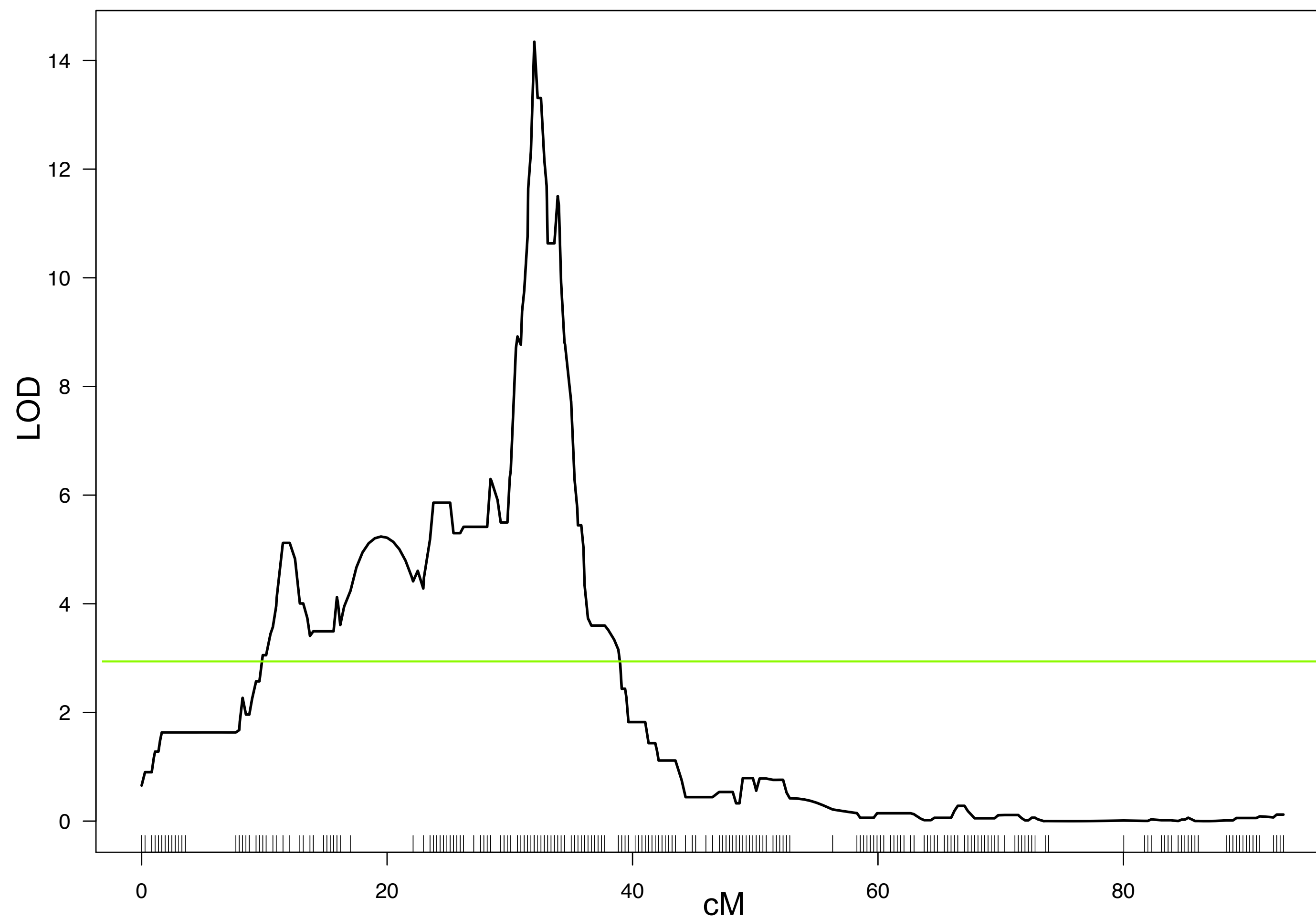
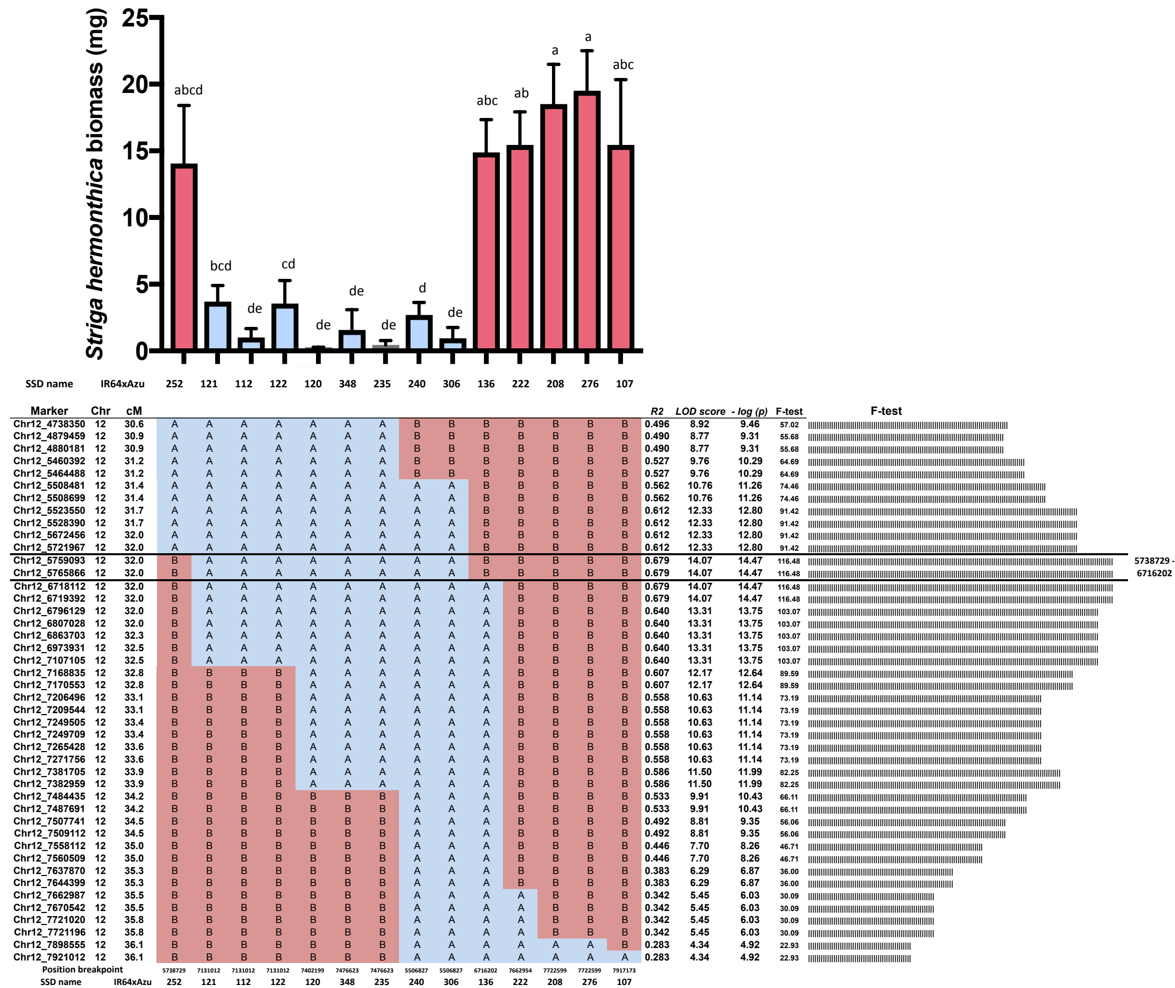


Figure 2.9. Major QTL for *S. hermonthica* resistance in *O. sativa* cultivar IR64, located on chromosome 12 with a peak at 32cM. (Interval mapping, LOD = 13.34). Green horizontal line represents threshold level of statistical significance (LOD scores of 3).



Marker	Chr	cM	252	121	112	122	120	348	235	240	306	136	222	208	276	107	R2	LOD score	- log (p)	F-test
Chr12_4738350	12	30.6	A	A	A	A	A	A	A	B	B	B	B	B	B	B	0.496	8.92	9.46	57.02
Chr12_4879459	12	30.9	A	A	A	A	A	A	A	B	B	B	B	B	B	B	0.490	8.77	9.31	55.68
Chr12_4880181	12	30.9	A	A	A	A	A	A	A	B	B	B	B	B	B	B	0.490	8.77	9.31	55.68
Chr12_5460392	12	31.2	A	A	A	A	A	A	A	B	B	B	B	B	B	B	0.527	9.76	10.29	64.69
Chr12_5464488	12	31.2	A	A	A	A	A	A	A	B	B	B	B	B	B	B	0.527	9.76	10.29	64.69
Chr12_5508481	12	31.4	A	A	A	A	A	A	A	A	A	B	B	B	B	B	0.562	10.76	11.26	74.46
Chr12_5508699	12	31.4	A	A	A	A	A	A	A	A	A	B	B	B	B	B	0.562	10.76	11.26	74.46
Chr12_5523550	12	31.7	A	A	A	A	A	A	A	A	A	B	B	B	B	B	0.612	12.33	12.80	91.42
Chr12_5528390	12	31.7	A	A	A	A	A	A	A	A	A	B	B	B	B	B	0.612	12.33	12.80	91.42
Chr12_5672456	12	32.0	A	A	A	A	A	A	A	A	A	B	B	B	B	B	0.612	12.33	12.80	91.42
Chr12_5721967	12	32.0	A	A	A	A	A	A	A	A	A	B	B	B	B	B	0.612	12.33	12.80	91.42
Chr12_5759093	12	32.0	B	A	A	A	A	A	A	A	A	B	B	B	B	B	0.679	14.07	14.47	116.48
Chr12_5765866	12	32.0	B	A	A	A	A	A	A	A	A	B	B	B	B	B	0.679	14.07	14.47	116.48
Chr12_6718112	12	32.0	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.679	14.07	14.47	116.48
Chr12_6719392	12	32.0	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.679	14.07	14.47	116.48
Chr12_6796129	12	32.0	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.640	13.31	13.75	103.07
Chr12_6807028	12	32.0	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.640	13.31	13.75	103.07
Chr12_6863703	12	32.3	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.640	13.31	13.75	103.07
Chr12_6973931	12	32.5	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.640	13.31	13.75	103.07
Chr12_7107105	12	32.5	B	A	A	A	A	A	A	A	A	A	B	B	B	B	0.640	13.31	13.75	103.07
Chr12_7168835	12	32.8	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.607	12.17	12.64	89.59
Chr12_7170553	12	32.8	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.607	12.17	12.64	89.59
Chr12_7206496	12	33.1	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.558	10.63	11.14	73.19
Chr12_7209544	12	33.1	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.558	10.63	11.14	73.19
Chr12_7249505	12	33.4	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.558	10.63	11.14	73.19
Chr12_7249709	12	33.4	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.558	10.63	11.14	73.19
Chr12_7265428	12	33.6	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.558	10.63	11.14	73.19
Chr12_7271756	12	33.6	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.558	10.63	11.14	73.19
Chr12_7381705	12	33.9	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.586	11.50	11.99	82.25
Chr12_7382959	12	33.9	B	B	B	B	A	A	A	A	A	A	B	B	B	B	0.586	11.50	11.99	82.25
Chr12_7484435	12	34.2	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.533	9.91	10.43	66.11
Chr12_7487691	12	34.2	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.533	9.91	10.43	66.11
Chr12_7507741	12	34.5	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.492	8.81	9.35	56.06
Chr12_7509112	12	34.5	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.492	8.81	9.35	56.06
Chr12_7558112	12	35.0	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.446	7.70	8.26	46.71
Chr12_7560509	12	35.0	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.446	7.70	8.26	46.71
Chr12_7637870	12	35.3	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.383	6.29	6.87	36.00
Chr12_7644399	12	35.3	B	B	B	B	B	B	B	A	A	A	B	B	B	B	0.383	6.29	6.87	36.00
Chr12_7662987	12	35.5	B	B	B	B	B	B	B	A	A	A	A	B	B	B	0.342	5.45	6.03	30.09
Chr12_7670542	12	35.5	B	B	B	B	B	B	B	A	A	A	A	B	B	B	0.342	5.45	6.03	30.09
Chr12_7721020	12	35.8	B	B	B	B	B	B	B	A	A	A	A	B	B	B	0.342	5.45	6.03	30.09
Chr12_7721196	12	35.8	B	B	B	B	B	B	B	A	A	A	A	B	B	B	0.342	5.45	6.03	30.09
Chr12_7898555	12	36.1	B	B	B	B	B	B	B	A	A	A	A	A	A	B	0.283	4.34	4.92	22.93
Chr12_7921012	12	36.1	B	B	B	B	B	B	B	A	A	A	A	A	A	A	0.283	4.34	4.92	22.93
Position breakpoint			5738729	7131012	7131012	7131012	7402199	7476623	7476623	5506827	5506827	6716202	7662954	7722599	7722599	7917173				
SSD name		IR64xAzu	252	121	112	122	120	348	235	240	306	136	222	208	276	107				

Figure 2.10. Resistance of 14 lines of the RIL population showing recombination in and around the QTL region. Top panel: Dry biomass of *Striga hermonthica* 21 dai with germinated *S. hermonthica* seed collected from Kibos in Western Kenya 2013. Values are means ± SE, where n = 4. Bottom panel: Recombination breakpoints for each line are shown in relation to base pair position, where A=IR64 genotype and B=Azucena genotype. LOD scores and F-test values for each marker position are shown in the columns on the right. QTL region is marked by 2 horizontal lines.

2.3.4 The mode of inheritance of the resistance to *Striga hermonthica* in IR64 and Nipponbare

To determine the mode of inheritance of resistance in IR64 and Nipponbare, these cultivars were both crossed with two susceptible cultivars, Azucena and Koshihikari, and phenotyped for resistance to *S. hermonthica.* Mean *S. hermonthica* dry weight on F_1 plants from all crosses was approximately midway between the parents, indicating co-dominance (Figure 2.11). The difference in *S. hermonthica* dry weight between genotypes was highly significant; *S. hermonthica* dry weight from the IR64 x Azucena cross was significantly different from both IR64 (two sample t-test: $t = 4.2$, d.f. = 10, $p < 0.01$) and Azucena ($t = 2.9$, d.f, = 10, $p < 0.05$). Similarly, *S. hermonthica* dry weight from the IR64 x Koshihikari cross was significantly different from both IR64 ($t = 4.0$, d.f. = 25, $p < 0.001$) and Koshihikari ($t = 2.1$, d.f. = 26, $p < 0.05$). There was also a significant difference in *S. hermonthica* dry weight for the Nipponbare crosses. There was a significant difference in *S. hermonthica* dry weight harvested from Nipponbare x Azucena F_1 plants when compared to Nipponbare ($t = 3.3$, d.f. = 26, $p < 0.01$) and Azucena ($t = 9.1$, d.f. = 27, $p <0.001$) parents. *S. hermonthica* dry weight harvested from Nipponbare x Koshihikari F1 plants was significantly different to Nipponbare ($t = 6.3$, d.f. = 25, $p <0.001$) and Koshihikari ($t = 2.5$, d.f. = 26, $p < 0.05$).

To test if the resistance in IR64 and Nipponbare is likely to be due the presence of the same resistance gene or genes, F_1 plants were produced from a cross between these two cultivars and phenotyped for *S. hermonthica* resistance. There was no significant difference in *S. hermonthica* dry biomass between genotypes; all genotypes were highly resistant to the parasite (Figure 2.11). Azucena and Koshihikari plants were also crossed with each other to test for resistance interactions between susceptible cultivars. There was no significant difference in *S. hermonthica* dry biomass between the F_1 plants and their susceptible parents (Figure 2.11).

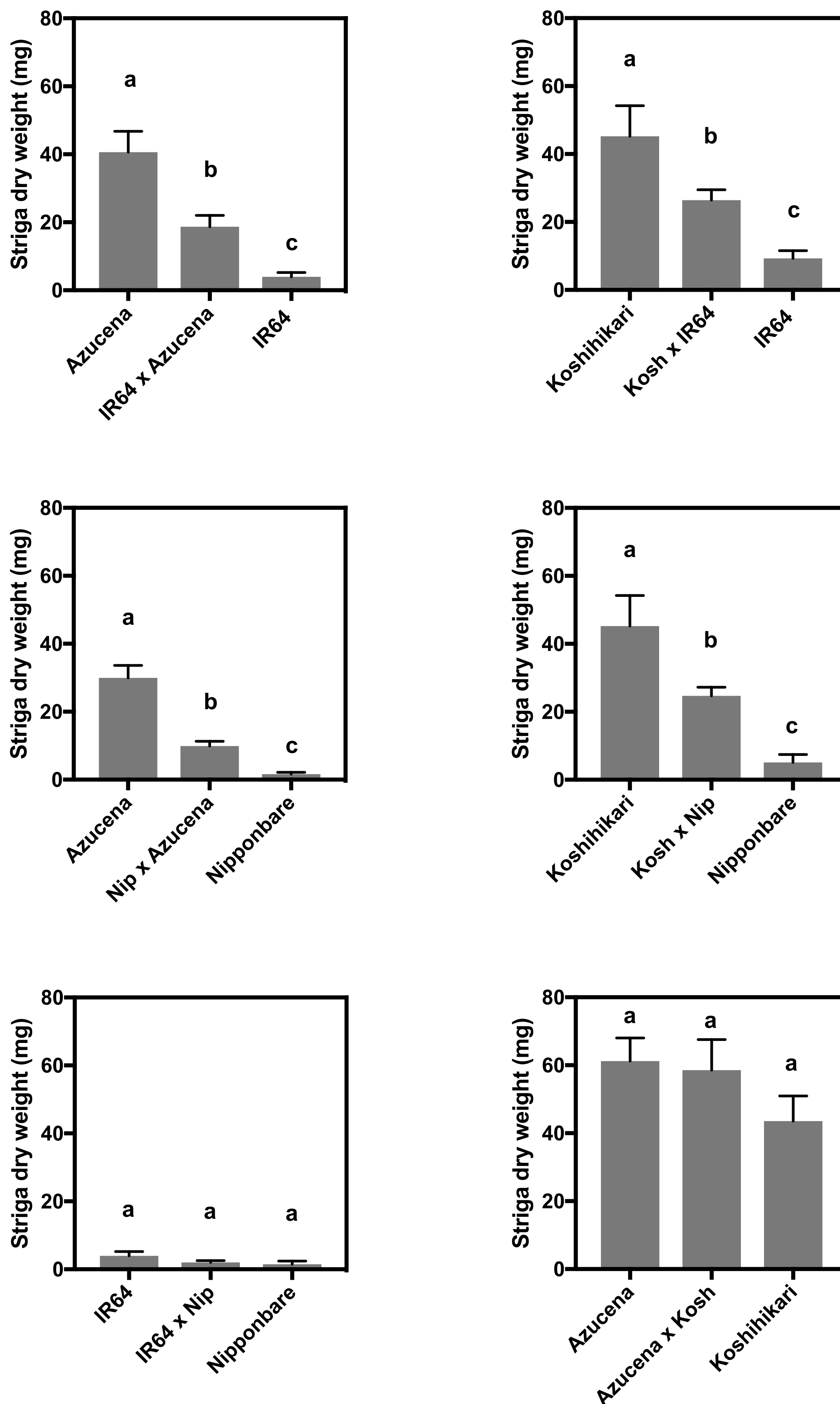


Figure 2.11. Evaluation of post-attachment resistance of Azucena, IR64, Nipponbare, Koshihikari and the F_1 hybrids to *Striga hermonthica*. Data shown are *S. hermonthica* mean dry weight 28 dai ± SE. Significant differences (p<0.05) are represented by letters above each bar: different letter codes represent significant differences.

2.4 Discussion

Breeding broad spectrum, durable resistance to *Striga* species in cereal crops is likely to require pyramiding of multiple resistance genes for both pre and post-attachment resistance, or different mechanisms of post attachment resistance (Scholes & Press, 2008; Timko and Scholes 2013). The aim of this study was to identify novel QTL for resistance to *S. hermonthica* (Kibos ecotype). Initially ten tropical *japonica* parents of a NAM population and the common parent IR64, were phenotyped for resistance to *S. hermonthica* to enable selection of a RIL mapping population where the parental lines showed genetic differences for the trait of interest (Mauricio, 2001; Semagn *et al.*, 2006), in this case *S. hermonthica* resistance. The *O. sativa indica* cultivar IR64 showed strong resistance to *S. hermonthica* (Kibos isolate), as did six other parental lines. Four of the parental genotypes were highly susceptible to *S. hermonthica*. The most susceptible genotype was Azucena thus the IR64 (*indica*) x Azucena (*japonica*) RIL population was selected for phenotyping and analysis. *O sativa indica* and *japonica* subspecies are estimated to have diverged 0.4 million years ago (mya) (Zhu & Ge, 2005), therefore using a population derived from an *indica* x *japonica* cross guarantees high levels of polymorphism in the population, and so a good coverage of SNP markers (Fragoso *et al.*, 2017). In addition, a resistant *indica* parent was expected to harbour novel resistance genes not present in *japonica* cultivars.

2.4.1 What is the phenotype of resistance in IR64?

The phenotype of resistance in IR64 was associated with necrosis around the site of the parasite attachment. Cross sections through the root most commonly revealed that the parasite penetrated the root epidermis and cortex, but was halted at the endodermis. The haustorium did not develop or differentiate properly. Occasionally *S. hermonthica* plants did make a few vascular connections but these parasites were small and grew slowly. Interestingly, these phenotypes of resistance in IR64 are similar to those observed in some other rice cultivars. For example, Gurney *et al.* (2006) showed that the parasites attached to the host roots but within 24 h necrosis was visible at the site of attachment. In addition the parasite rarely breached the endodermis to form xylem-xylem connections, and often grew through the cortex, around the vascular core and out of the root. Similarly, resistant NERICA lines (NERICA 1 and NERICA 10) showed the same phenotypes in response to *S. hermonthica* infection; in most cases the parasite failed to breach the endodermis. However, in both Nipponbare and the NERICA cultivars some vascular connections were made but dense staining material was observed and parasites remained small (Cissoko *et al.*, 2011). In the resistance interaction between *Orobanche aegyptiaca* and the vetch *Vicia atropurpurea*

cultivar Popany, the parasite was also halted at the endodermis. This was coupled with a red-brown secretion in the apoplast of the host-parasite interface and around the vascular cylinders (Goldwasser *et al.*, 2000). This inability of the parasite to penetrate the endodermis therefore appears to be common form of post-attachment resistance to parasitic plants, although it is not the only resistance phenotype and the underlying mechanism governing this response is still unknown. Arrest of parasite development can occur at later stages of infection in some interactions. The wild relative of maize, *Tripsacum dactyloides*, exhibited resistance to *S. hermonthica* after vascular continuity was established (Gurney *et al.*, 2003). This was also observed for non-hosts *Arabidopsis* and cowpea when infected with *S. hermonthica* (Yoshida and Shirasu 2009). However when the non-host and hemiparasite *Phtheirospermum japonicum* was infected with *S. hermonthica*, resistance occurred much earlier; *S. hermonthica* radicles failed to penetrate the roots and instead continued to elongate (Yoshida & Shirasu, 2009).

2.4.2. Identification of a *S. hermonthica* resistance QTL

In this study a RIL population derived from a cross between the resistant cultivar IR64 (*indica*) and the susceptible cultivar Azucena (*tropical japonica*) was used for identifying *S. hermonthica* resistance QTL. Due to the very high map coverage (one SNP every 8 Kb) a novel two-step phenotyping strategy was taken to see if it was possible to speed up detection of any QTL present. A small subset of the population (44 lines) was randomly chosen for phenotyping resistance to *S. hermonthica* and an initial QTL analysis carried out to see if any QTL could be detected. Unexpectedly, this immediately revealed the presence of a QTL on chromosome 12 in a similar position to that detected previously in the Nipponbare/Koshihikari//Koshihikari BIL population (Scholes, personal communication). The high-resolution marker data was then used to identify other lines within the RIL population possessing recombination within the QTL region. These were phenotyped for *S. hermonthica* resistance, which increased the resolution of the initial QTL, producing a LOD score of 14.34 without having to phenotype the whole population. Thus the 2-step strategy enabled the discovery of the QTL within 3 months compared to 3 years for the Nipponbare / Koshihikari BIL population (Scholes, personal communication). It should be noted, however, that this strategy only works for strong effect QTL, which can be detected using small population sizes.

The *S. hermonthica* resistance QTL in IR64 was located between positions 5.7 – 6.7 Mbp on chromosome 12. This QTL region is quite large due to lack of recombination in the area, which is possibly due to structural differences between the IR64 and Azucena genomes. The local recombination rate between 5.7 - 7.2 Mbp was very low, between 5 - 10 fold less than the

whole genome average. SNP coverage was also low in the region due to sequence alignment issues (Lorieux, personal communication). Nevertheless, the importance of this QTL for *S. hermonthica* resistance is evident from the highly significant LOD scores and percentage of variance explained by the QTL (% R^2 = 67.9) which suggests that the resistance is due to a major gene, or a small number of tightly linked genes of major effect. No other QTL were detected on other chromosomes, which is in accordance with this hypothesis.

The *S. hermonthica* resistance QTL on chromosome 12 identified in this population overlaps that of the Nipponbare QTL detected previously (section 1.6 and Gurney *et al.*, 2006). In both cases, *S. hermonthica* seeds were collected from plants parasitizing maize in Kibos, western Kenya, although the earlier study used seed collected in 1997, whereas seed collected in 2013 was used in the present study. The identification of a QTL in the same position is extremely interesting as the resistance in IR64 was expected to be different due to the divergence of *indica* and *japonica* subspecies. This result suggests that the same or similar genes may underlie the resistance in both Nipponbare and IR64, or if several genes are involved, they could overlap between the two cultivars. It also suggests that these genes may be very important for resistance to *S. hermonthica* in rice. A detailed analysis of the genes present within the IR64 QTL region and a comparison with the Nipponbare *S. hermonthica* - resistance QTL is reported in Chapter 3.

The success of a parasite on a host plant depends on both genotype x environment interactions and the interaction between the host genotype and the particular parasite population or ecotype (Huang *et al.*, 2012b). As on outbreeding parasite, *S. hermonthica* shows particularly high levels of genetic diversity (Safa *et al.*, 1984). Other studies investigating the resistance of IR64 and Azucena found both of these cultivars were susceptible to *S. hermonthica*. When challenged with *S. hermonthica* seed collected from maize in Kibos in 1997, IR64 plants from IRRI were susceptible (Gurney *et al.*, 2006). In contrast, Cissoko (2012) found that IR64 plants from CIAT, Colombia, showed good resistance to the same ecotype of *S. hermonthica* as was also found in the present study. Therefore it is likely that IR64 cultivar from IRRI differs genetically from the CIAT genotype used here. In a pot experiment, IR64 plants from IRRI were susceptible to *S. hermonthica* seed collected from sorghum in Mali in 2003. Azucena plants were more resistant to *S. hermonthica* than IR64 in this study (Kaewchumnong & Price, 2008). This difference in resistance might also be explained by the different *S. hermonthica* ecotype used. As rice plants were grown in pots in this experiment; differences may also be due to different levels of pre-attachment resistance between cultivars, due to different levels of germination stimulant produced.

2.4.3 What is the genetic basis of resistance in IR64 to *S. hermonthica*?

The F_1 progeny of the cross between IR64 and Azucena had intermediate resistance to *S. hermonthica* when compared to the parental lines, indicating that resistance is co-dominant. The fact that the same result was seen in the F_1 progeny of the cross between Nipponbare and Koshihikari and that the F_1 progeny of a cross between Nipponbare and IR64 were resistant suggests that the same gene(s) may underlie the resistance in both cultivars. The *Cf* genes in tomato which provide resistance against the foliar fungal pathogen *C. fulvum* (Hammond-Kosack, 1994) are also thought to be co-dominant. These authors speculated that if the *Cf* resistance genes encoded proteins with a receptor or surveillance function, then greater perception of the invading pathogen would be achieved by greater concentrations of the receptor with respect to the concentration of avirulence gene product. The signal to the plant to activate a defence response is therefore the outcome of this interaction, and is proportional to the plants ability to detect the invading organism (receptor concentration) and the pathogens ability to repress the response mounted (*Avr* concentration) (Hammond-Kosack, 1994). The *Cf* genes are now known to encode cell surface Receptor Like Proteins (RLP) (Piedras *et al.*, 2000; Fradin *et al.*, 2009; Liu *et al.*, 2015). Given that the Nipponbare QTL region also contains a cluster of RLP genes with homology to *Verticillium* wilt RLP genes, it is tempting to speculate that a similar mechanism may be acting in resistance to *S. hermonthica*. In addition, a gene in tomato that confers increased resistance to the shoot parasitic plant *C.* reflexa (*CuRe1*) has also been shown to encode a RLP (Hegenauer et al., 2016).

This incomplete dominance of the resistance identified in IR64 and Nipponbare against the Kibos ecotype of *S. hermonthica* is in contrast to resistance identified against other parasitic plants. For example, three independent dominant genes were shown to provide resistance to *S. gesnerioides* in three different cowpea genotypes (Atokple *et al.*, 1995). *S. gesnerioide*s exhibits clear race structure, and seven races have so far been identified. Resistance genes for these have been mapped to two linkage groups for a variety of cowpea cultivars (Ouedraogo *et al.*, 2001), with race-specific resistance thought to be conferred by a single dominant gene in most cases (Li *et al.*, 2009). Resistance to the root holoparasite *O. cumana* is also dominant and monogenically inherited. In this case resistance to five races of *O. cumana* is provided by five independent genes termed *Or1-Or5,* respectively (Molinero-Ruiz *et al.*, 2006). In contrast, resistance to *S. hermonthica* is not thought to be race specific (it is an obligate outbreeder) thus it is interesting to speculate that cell surface immune receptors may detect conserved pathogen associated molecular patterns (PAMPs) thus providing broad spectrum resistance against the genetically diverse *S. hermonthica* seed bank.

2.4.4 Conclusions

This chapter has identified a major QTL for post-attachment *S. hermonthica* resistance in an IR64 x Azucena RIL mapping population. A major QTL was detected on chromosome 12 which mapped to the same position as the QTL previously identified in Nipponbare, despite the ancestral divergence of these two subspecies. Further examination of the IR64 QTL region is required to identify candidate *S. hermonthica* resistance genes and compare them with genes present in the Nipponbare QTL. Comparison of gene sequences from rice genotypes susceptible to *S. hermonthica* may also help identify or narrow down candidate resistance genes.

Chapter 3

Identification of candidate *Striga hermonthica* resistance genes in the rice cultivar IR64

3.1 Introduction

The work in Chapter 2 describes the mapping of a major QTL on chromosome 12 for *S. hermonthica* resistance in the *indica* rice cultivar IR64. The position of this QTL directly overlays that of the *S. hermonthica* resistance QTL previously detected in the temperate *japonica* cultivar Nipponbare. However, it is not known whether any of the genes within the QTL are shared between the two cultivars. The Nipponbare QTL contains a cluster of 13 genes annotated as orthologs of disease resistance genes that are currently top candidates for *S. hermonthica* resistance genes (see section 1.6). These are predicted to encode RLPs that localise to the plasma membrane. It remains to be determined whether the QTL in IR64 also contains these orthologs, or whether there are alternative or additional genes in IR64 that could provide *S. hermonthica* resistance by a different resistance mechanism. The genes in the IR64 QTL must first be determined in order to identify possible resistance gene candidates. This will enable their functional validation and a better understanding of the different mechanisms that exist within plants to defend themselves against attack by parasitic weeds.

Plants come under attack from a diverse range of pests, pathogens and parasites. Unlike animals, plants lack mobile defender cells and instead rely on the innate immunity of individual cells to defend themselves (Jones & Dangl, 2006). Plant defence strategies are highly sophisticated, consisting of multiple layers which can be constitutively expressed or induced upon pathogen attack, and may be activated locally or systemically (Dangl & Jones, 2001). The defence mounted depends on the lifestyle and infection strategy of the attacker, and results in activation of a set of defence-related genes regulated by a complex signalling network of phytohormones (Pieterse *et al.*, 2009). The outcome of infection is determined by the plants ability to mount a timely and appropriate defence response.

Successful recognition of the invader is key to successful induction of defence. Several models describing pathogen recognition and plant immune responses have been suggested. The gene-for-gene hypothesis states that a plant possessing a single dominant resistance (R) gene is resistant towards a pathogen possessing a corresponding dominant avirulence gene (*Avr*) (Flor, 1971) with resistance originally thought to be conferred by direct ligand-gene interaction (Keen, 1990; Thomma *et al.*, 2011). However, it is now known that many R proteins do not interact directly with their receptor (Luderer *et al.*, 2001; Selote & Kachroo, 2010) leading to the suggestion that some R proteins monitor the target of the effector, instead detecting changes in “pathogen-induced modified self”; the guard hypothesis (Jones & Dangl, 2006; Dodds & Rathjen, 2010). The zig-zag model, introduced by Jones & Dangl, (2006), postulated a two-branched immune system in plants. Firstly, trans-membrane pattern recognition

receptors (PRRs) at the cell surface recognise slowly evolving, conserved microbial features termed microbe or pathogen associated molecular patterns (MAMPs/PAMPs) such as flagellin or chitin, leading to the activation of basal resistance, or PAMP-triggered immunity (PTI). To overcome this, pathogens release fast evolving and highly specific effector molecules into the host cell, resulting in effector-triggered susceptibility (ETS). These may be recognised by intracellular nucleotide-binding (NB) leucine-rich repeat (LRR) resistance proteins, leading to the activation of a defence response and effector-triggered immunity (ETI). An evolutionary arms race drives pathogens to evolve new effectors to suppress ETI, and in turn the plant comes under pressure to evolve new methods for their detection (Jones & Dangl, 2006).

Although the PTI and ETI model brings together the gene-for-gene model with the recognition of more general pathogen elicitors, it does not conform to all aspects of plant immunity. For example, some R proteins have properties more consistent with PRR receptors and are not always pathogen specific (Fradin *et al.*, 2009; de Jonge *et al.*, 2012; Lozano-Torres *et al.*, 2012). For example in red current tomato (*Solanum pimpinellifolium*), the RLP protein *Cf-2* mediates dual disease resistance to both the leaf mould fungus *C. fulvum* and also the root parasitic nematode *Globodera rostochiensis* (Lozano-Torres *et al.*, 2012). Conversely, some PAMPs are only recognised by a very narrow range of host plants (Thomma *et al.*, 2011). Additionally, perception of effectors does not always occur inside host cells (Cook *et al.*, 2015). Therefore an alternative view of the plant immune system was proposed by Cook *et al.* (2015) termed the Invasion Model (Figure 3.1), which encompasses a continuum between PTI and ETI. Broadly, plants detect invaders either directly or indirectly by means of any type of receptor that is able to effectively portray the presence of an invader and convert the signal into an appropriate defence response (Thomma et al., 2011; Cook et al., 2015).

Despite recent progress, very little is known about the genes underlying resistance to parasitic plants. However, a race-specific gene-for-gene resistance mechanism is known to occur between the dicotyledonous cowpea *V. unguiculata* and *S. gesnerioides*. The R gene RSG3-301 in the cowpea cultivar B301, which encodes a protein containing coiled-coil, nucleotide binding site and leucine rich repeat (CC-NBS-LRR) domains and which localises to the plasma membrane, provides resistance against *S. gesnerioides* race 3. Although the parasite was able to penetrate the host cortex, it could not cross the endodermis and thus failed to make vascular connections (Li & Timko, 2009). More recently, a cell-surface receptor-like protein (RLP) was shown to provide increased resistance to the stem holoparasite *C. reflexa* in tomato. Resistant tomato (*Solanum lycopersicum*) plants produced reactive oxygen species (ROS) and ethylene in response to extracts of six different *Cuscuta* species, suggesting the resistant host

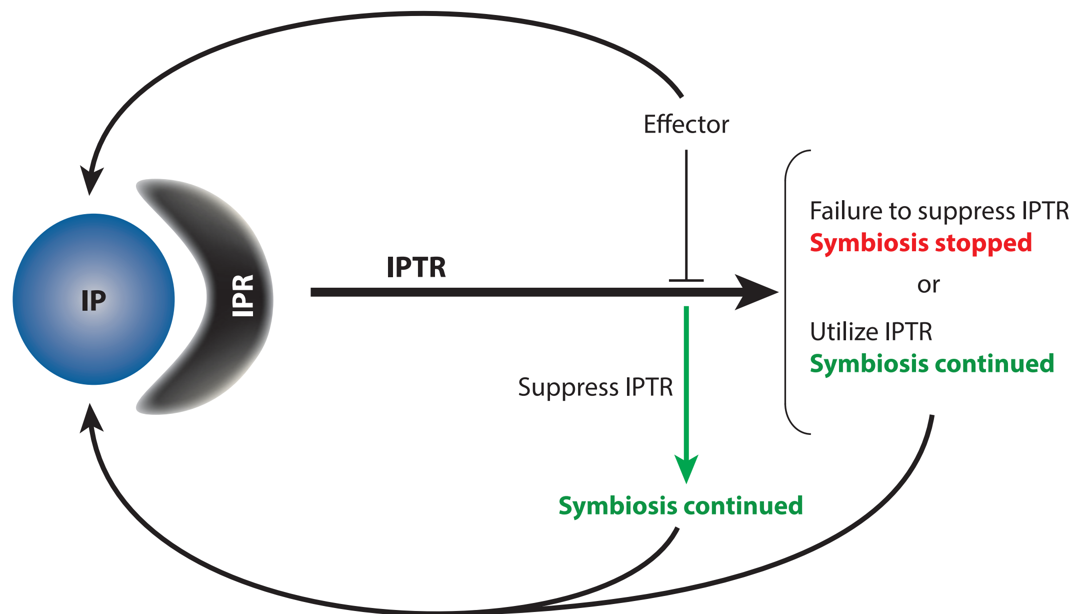


Figure 3.1. The Invasion Model of plant immunity proposed by Cook *et al* 2015, to illustrate an attempted plant-invader symbiosis. Upon attempted symbiosis, the invading organism releases Invasion patterns (IP) which are perceived by plant Invasion Pattern Receptors (IPR), eliciting an IP-triggered response (IPTR). The invader may use effectors to facilitate the interaction, but failure of the invader to manipulate IPTR leads to a halt in symbiosis. Potentially, IPTR may be utilised (e.g. by necrotrophs) or suppressed (e.g. by biotrophs) and symbiosis is continued. Multiple recognition events and attempts to influence IPTR eventually leads to a termination or continuation of the symbiosis. Figure from Cook *et al.*, 2015.

is able to detect a common factor present in all species (Hegenauer *et al.*, 2016). Heterologous expression of five tomato RLP genes in tobacco (*Nicotiana tabacum)*, a species unable to detect the *Cuscuta* factor, identified one gene which resulted in an oxidative burst and ethylene production in response to *C. reflexa* extracts. This gene, termed *CuRe1*, was absent from susceptible *S. pennellii* plants. Stable transformation of *CuRe1* into both *S. pennellii* and *N. tabacum* resulted in increased resistance to *C. reflexa*, although this resistance was incomplete (Hegenauer *et al.*, 2016).

No resistance genes against *Striga* species have yet been identified in rice. However, the QTL for *S. hermonthica* resistance in the temperate *japonica* Nipponbare contains a cluster of RLP genes annotated as homologs of *Verticillium* wilt disease resistance proteins in tomato. The highly similar infection strategy and lifecycle of these fungi compared to *Striga* species makes these genes excellent candidates for *S. hermonthica* resistance in this cultivar. The identification of a QTL for *S. hermonthica* resistance in the rice cultivar IR64 (Chapter 2), which maps to the same position on chromosome 12 as the QTL for *S. hermonthica* resistance in

Nipponbare, strongly suggests that the same, or very similar, resistance genes may exist between IR64 and Nipponbare in this region.

3.1.1 Aim of Chapter 3

The aim of this chapter is to identify candidate resistance genes within the IR64 QTL and test the hypothesis that the candidate resistance genes are the same or similar to those identified in the Nipponbare resistance QTL. Specific aims are:

1) To compare the genome structure of the IR64 and Nipponbare QTL regions.
2) To identify genes within the IR64 QTL region by using gene prediction software to annotate the IR64 QTL.
3) To compare similarity of the predicted genes with those of Nipponbare and the *S. hermonthica* susceptible cultivars Koshihikari and Azucena, where possible.
4) To clone and sequence the IR64 candidate resistance genes.
5) The examine gene expression by qPCR of candidate resistance genes in roots of IR64, in response to infection with *S. hermonthica*.

3.2 Materials and Methods

3.2.1 Genome sequences

The Nipponbare reference genome was downloaded from the MSU website http://rice.plantbiology.msu.edu/. In addition, some gene sequences were also obtained from the Nipponbare reference sequence from the IRGSP website (http:// rgp.dna.affrc.go.jp/IRGSP/index.html) where gene prediction algorithms predicted different start and stop sites for specific genes. Genome sequence from the rice cultivar Koshihikari (*O. sativa ssp.* temperate *japonica*), that spanned most of the equivalent *S. hermonthica* resistance- QTL region of both Nipponbare and IR64 (approximately 144,000 bp in length) was provided by Dr. Kiyosumi Hori (National Institute of Agrobiological Sciences (NIAS), Japan) from the sequencing of two Bac clones (J0090-E04 and J0090-F01). To examine genes not covered by this Bac sequence, the Koshihikari chromosome 12 genome sequence was downloaded from the National Centre for Biotechnology Information (NCBI, GenBank: DG000036.1). IR64 Illumina genome sequence was downloaded from http://schatzlab.cshl.edu/data/rice/. The sequence information for this cultivar has not been assembled into chromosomes, and currently exists as a series of over 36,000 contigs with no annotation. In addition, a longer PacBio sequence (2.07 Mb in length) covering the IR64 QTL region was kindly provided by the Schatz Laboratory (for confidential use in this project).

3.2.2 Comparison of the genome structure spanning the *Striga hermonthica* resistance QTL region in rice cultivars

Genome sequence from the Nipponbare and IR64 S*. hermonthica* resistance QTL regions and the equivalent region from Koshihikari were compared using Mauve Multiple Genome Alignment software http://darlinglab.org/mauve/mauve.html. Sequences were aligned with Progressive Mauve using default parameters. This software allows large-scale evolutionary events such inversions and rearrangements to be visualised and compared between closely related genomes, showing medium to high amounts of genome rearrangement. The Nipponbare sequence from 5.7 - 6.7 Mbp was aligned with the equivalent region in IR64 (approximately 526,000 bp). Both the IR64 and Nipponbare sequences were also compared to the equivalent sequence from the *S. hermonthica* susceptible cultivar Koshihikari. As the Koshihikari sequence for chromosome 12 obtained from NCBI is not high quality it was not possible to use, therefore the higher quality Illumina Bac sequence from NIAS was used for this comparison. However this sequence does not cover the entire QTL region.

3.2.3 Prediction of genes in within the IR64 *S. hermonthica* resistance QTL and comparison with genes in the Nipponbare *S. hermonthica* resistance QTL

In order to identify all genes within the IR64 QTL, the QTL region was selected from the PacBio sequence, and Fgenesh gene-finder (Softberry, Inc.) program used to predict gene models (Solovyev *et al.*, 2006). Organism specific gene-finder parameters were set to monocot plants. The amino acid sequences of all proteins predicted by Fgenesh were then blasted against the rice reference sequence using the blasp BLAST search on the MSU website using default settings (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml) to obtain gene annotations. In addition, gene prediction was also carried out for the available Koshihikari Bac sequence in the same way, to allow comparison of this region between *S. hermonthica* resistant and susceptible cultivars. To test the accuracy of Fgenesh gene predictions, the sequences of Nipponbare QTL region was also analysed with Fgenesh gene-finder in the same way and candidate genes compared back to the Nipponbare reference sequence. Recently, a genome annotation for the IR64 PacBio genome sequence also became available from the Schatz laboratory. This annotation was produced using the Maker2 genome annotation pipeline, and the coordinates of predicted genes checked against the Fgenesh predictions.

Nucleotide and amino acid sequences of all predicted genes within the IR64 QTL region and the equivalent region in Koshihikari (excluding transposable elements) were aligned with the Nipponbare homologs using the Create Alignment tool with highest accuracy settings in CLC bio Main Workbench version 7.0.3. The Create Pairwise Comparison function was then used to calculate the number of gaps, differences, identities and % identity between the alignments. Phylogenetic trees were constructed from alignments using the Neighbor Joining method, and bootstrapping analysis set to 1000 replicates.

3.2.4 Cloning and sequencing of IR64 candidate resistance genes

3.2.4.1 DNA extraction

DNA extraction was carried out according to the Qiagen DNeasy Plant Mini Kit. Approximately 80 - 100 mg of leaf tissue was used for each extraction. Leaf tissue was placed in a 2 ml Eppendorf tube containing a sterile ball bearing and frozen in liquid nitrogen, before being disrupted in a TissueLyser (Qiagen) for 45 s at 25 Hz until a fine powder was obtained. Following tissue disruption, 400 µl buffer AP1 and 4 µl RNase A was added to each tube, tubes were vortexed and then incubated for 10 min at 65 °C in a water bath. Tubes were inverted 2 - 3 times during incubation. Following incubation, 130 µl Buffer P3 was added and mixed briefly, then transferred to ice for a further 5 min. Samples were then centrifuged at 20,000 x g for 5

min. The lysate was removed and pipetted into a QIAshredder Mini spin column placed in a 2 ml collection tube, and centrifuged for 2 min at 20,000 x g. The flow-through was carefully transferred to a new tube so as not to disturb the pellet, and 1.5 x volume Buffer AW1 was added and mixed by pipetting. Six-hundred and fifty µl of the mixture was transferred to a DNeasy Mini spin column placed in a 2 ml collection tube, then centrifuged for 1 min at 6000 x g. The flow-through was discarded, and the step repeated with the remaining sample. After discarding flow-through again, 500 µl Buffer AW2 was added, and samples centrifuged for 1 min at 6000 x g. Flow-through was discarded and a further 500 µl Buffer AW2 was added, this time followed by centrifugation at 20,000 x g for 2 min. The spin column was carefully removed from the collection tube and placed in a new, sterile microcentrifuge tube. Thirty µl nuclease-free water was pipetted onto the centre of the spin column and incubated at room temperature for 5 min. Samples were centrifuged for 1 min at 6000 x g. The flow-through was pipetted back into the spin column to concentrate the sample, and centrifuged again at 6000 x g for 1 min. The quality and quantity of DNA in the samples was measured using a Nanodrop spectrophotometer (ND-8000, Thermo Scientific).

3.2.4.2 PCR amplification of candidate *S. hermonthica* RLP genes in IR64

Gateway primers were designed to PCR amplify predicted genes from IR64 sequences. For primer design, the 25 bp *att*B site of the primers required for Gateway cloning was followed by 18 - 25 bp of the gene-specific sequence homologous to the beginning (forward primer) and end (reverse primer) sequence of each gene. Potential hairpin formation and self-annealing was checked using the oligonucleotide properties calculator OligoCalc (http:// biotools.nubic.northwestern.edu/OligoCalc.html), and specificity confirmed with Primer BLAST (NCBI) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers and product sizes are shown in Table 3.1. Genes were amplified by PCR using the highly accurate Phusion Hot Start Flex DNA polymerase (New England BioLabs). A 25 µl reaction was carried out for each gene. Nuclease-free water was used as a no-template control. The reaction consisted of 5 µl 5 x Phusion HF Buffer, 0.5 µl 10 mM dNTPs (40 mM total), 1.25 µl of each 10 µM primer, 0.25 µl Phusion Hot Start Flex DNA polymerase, 5 µl DNA template (10 ng /µl) and 11.75 µl nuclease-free water. The PCR cycling conditions were: 30 s at 98 °C (initial denaturation); then 40 cycles of 10 s at 98 °C (denaturation), 30 s annealing at 72 °C for the first 15 cycles, then decreasing by 1 °C until 57 °C and remaining at this temperature for the remaining 25 cycles, 2 min at 72 °C (extension); followed by a final extension of 10 min at 72 °C. Five µl of each product was mixed with 5 x DNA Loading Buffer (Bioline) and run on a 1.6 % gel to check product sizes.

PCR purification of the amplified genes was carried out according to the QIAquick PCR Purification Kit (Qiagen). Five volumes of Buffer PB were added to the PCR reaction and mixed, and 10 µl 3 M sodium acetate (pH 5.0) added if the mixture turned orange or purple. The mixture was added to a QIAquick column and centrifuged for 60 s at 17,900 x g. The flow-through was discarded, and samples washed with 750 µl Buffer PE by centrifuging for 60 s at the same speed, followed by a further 60 s to remove residual buffer. QIAquick columns were placed in a clean, sterile microcentrifuge tubes, and DNA eluted in 30 µl nuclease-free water by centrifuging at 17,900 x g for 60 sec. The quality and quantity of DNA in the samples was measured using a Nanodrop spectrophotometer.

3.2.4.3 Cloning of candidate IR64 resistance genes into pDONR/Zeo

A BP recombination reaction was carried out for each IR64 gene amplified to create a Gateway entry clone. For each gene to be cloned, 1 µl purified PCR product (50 ng) was added to a tube containing 0.5 µl pDONR/Zeo, 1 µl BP clonase II and 2.5 µl nuclease-free water, to make a total volume of 5 µl. Samples were incubated at 25 °C overnight. The BP reaction mix was used for transformation into Library Efficiency® DH5α™ chemically competent *Escherichia coli* cells; 2.5 µl each BP reaction was added to 25 µl competent cells, mixed gently and incubated for 30 min on ice. Cells were then heat-shocked for 30 s at 42 °C in a water bath, and immediately returned to ice. Under sterile conditions, 125 µl of S.O.C. medium was added to each sample and incubated for 1 h at 37 °C while shaking. 100 µl was then spread onto pre-warmed low salt LB plates containing zeocin and incubated at 37 °C overnight. Low salt LB medium (500 ml): Tryptone 5 g, NaCl 2.5 g, Yeast Extract 2.5 g, Bacto agar 7.5 g; autoclaved and cooled to around 55 °C, then 250 µl zeocin added and mixed before pouring onto plates.

To test for successful transformation, colony PCRs were carried out using pDONR specific M13 primers for a selection of colonies (forward primer: CGCCAGGGTTTTCCCAGTCACGAC; reverse primer: TCACACAGGAAACAGCTATGAC) using the Qiagen *Taq* PCR Master Mix Kit. Each 10 µl PCR reaction consisted of 5 µl *Taq* PCR Master Mix, 0.2 µl of each primer (5 µM each) and 4.8 µl nuclease-free water. For the DNA template, a pipette tip was used to touch a colony, and then placed directly in the PCR mix. The PCR cycling program was: 3 min at 94 °C (initial denaturation); then 35 cycles of 30 s at 94 °C (denaturation), 30 s at 60 °C (annealing), and 3 min 30 s at 72 °C (extension); followed by a final extension of 10 min at 72 °C. 2.5 µl of 5X DNA Loading Buffer, blue (Bioline) was added to each sample, and the mix loaded onto a 1 % agarose gel. Overnight cultures were carried out for successfully transformed colonies (5 ml low salt LB medium containing zeocin, 37 °C, shaking). A miniprep was then carried out to isolate the plasmids using the QIAprep Spin Miniprep Kit (Qiagen). Overnight cultures were

spun down at 6,800 x g for 3 min to pellet the bacteria. Cells were re-suspended in 250 µl Buffer P1, and 250 µl Buffer P2 added and mixed thoroughly to start the lysis reaction. To neutralize the reaction, 350 µl Buffer N3 was added and mixed, and samples centrifuged for 10 min at 17,900 x g. The supernatant was added to QIAprep spin columns, centrifuged for 60 s at 17,900 x g, and the flow through discarded. Columns were washed with 500 µl Buffer PB and then again with 750 µl Buffer PE, and centrifuged for 60 s at 17,900 x g each time. Flow through was discarded and samples centrifuged for a further 60 s to remove residual wash buffer. DNA was eluted in 30 µl nuclease-free water. Samples from 2 colonies for each gene cloned were sequenced at GATC-biotech (Cologne, Germany) for Lightrun Sanger sequencing using both the vector M13 primers and a series of primers designed against specific regions of each gene (Table 3.2). Sequences were returned to the University of Sheffield for assembly.

Table 3.1 Sequences of Gateway primers for amplification of full-length candidate Receptor-Like Protein (RLP) *S. hermonthica* resistance genes in IR64. The first column lists the gene ID of the closest Nipponbare homolog**.**

Closest homolog to Nipponbare	Primer name	Gateway primer sequence (5' – 3')	Amplicon length (bp)
Os12g10870	10870_GW_F 10870_GW_R	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCGTCGTTCTCCAAGAGAGTC GGGGACCACTTTGTACAAGAAAGCTGGGTTCAGCGTTGCTTTCTCATGCACC	3048
Os12g10930	IR64 10930 GW F IR64 10930 GW R	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCATCGTCCTCCATGAGAGT GGGGACCACTTTGTACAAGAAAGCTGGGTTCAGTGTTCCTTTCTCATGTGGC	3045
Os12g11370	>11370_GW_F >11370_GW_R	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCGTCGTCCACCAAGAGGC GGGGACCACTTTGTACAAGAAAGCTGGGTTCATCGCTGTTTTTTCATATGCCTTCC	3045
Os12g11500	11500_RNAi_9_GW_F IR64 11500 GW F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCATCGTCCACCAAGAGAG GGGGACCACTTTGTACAAGAAAGCTGGGTTTACGTTTGGAACTGTATATTACTTGT	2511
Os12g11510	Os12g11510.1_GW_OX_CDS_F1 Os12g11510.1_GW_OX_CDS_R1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCGATCTGCTTACCATCTGAT GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAAGGATTTCCCCAGAAAGCT	2733
Os12g12000	IR64 12000 GW F IR64 12000 GW R	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCGAGAAGAGCGAGCG GGGGACCACTTTGTACAAGAAAGCTGGGTCTA TGG TAT TGG TGT GGG GCA	285
Os12g12010	Os12g12010.1_GW_OX_CDS_F1 Os12g11680.1_GW_OX_CDS_R1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCCTCCTCCATGAGAGTTG GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAGCTTGCTGATTCCTTTTGTTG	3000

Table 3.2 Sequences of primers used for sequencing of Receptor-Like Protein (RLP) *S. hermonthica* candidate resistance genes in IR64.

Closest homolog	Primer name	Primer sequence (5' – 3')
Os12g10870	11870_SEQ1 _R	GAAGCTACCTAACGAAGAGGG
	11870_SEQ1 _F	CTTCACAGGTATGATTCCGAGC
	11870_SEQ2 _R	AGGAGTCACTAATCTGATTGTTTCC
	11870_SEQ2 _F	AATTTGATTGACGGAAAGATACCC
Os12g10930	IR64 10930 seq1 R	CCTGTGAAGTTTGTGAGACTG
	IR64 10930 seq1 F	TTCTCACAGGCAAGTAGTCTG
	IR64 10930 seq2 F	AAGGTTGTGTACTTGAGGCAC
	IR64 10930 seq2 R	GCAACTAGAGATCTGGGTATC
Os12g11370	11370_genotype_F1	AATTCGCTCACTAGGATTGAGCTT
	11370_genotype_R1	AAGATCTGCGGAGGCACCTT
	11370 qPCR F	GGATCCTTCATATACAGTTGATGG
	11370 qPCR R	GAAGTTATTTGAGGCCATATCGG
Os12g11500	11500_SEQ1_R	ATTCTCCAAGCTAGTGTCCTGTGAG
	11500_SEQ1_F2	ACTGTTCTTCAGCTATCC
	11500_SEQ2_R2	CGATGAGTTGCAGTTTCC
	11500_SEQ2_F	TACTCGACTTACCTTGGTGAAACTG
Os12g11510	11510_SEQ1_R2	CGATGAGTTGCAGTTTCC
	11510_SEQ1_F	CTCCCCAGCCTTAGTGTTCTTC
	11510_SEQ2_R2	CCAATGTCAAGGATCTCC
	11510_SEQ2_F	ACCGGCGAGTTACCTGATAATATC
Os12g12010	112010_SEQ1_R	AGAGAAATTAGGCAGATTTCCAGAGAT
	12010_SEQ1_F	CAGCTTTCAAACAACAACTTCGAA
	IR64 12010 seq2 R2	TGAGAGGCATAGATGAGAACTG
	11680_SEQ2_F2	GATCTGACCCTTTACTTCCAC
All genes	M13 F	CGCCAGGGTTTTCCCAGTCACGAC
	M13 R	TCACACAGGAAACAGCTATGAC

3.2.5 Amplification and sequencing of homologs of candidate *S. hermonthica* resistance genes from Azucena

The genome sequence for Azucena is not available. Therefore, in order to determine whether candidate *S. hermonthica* resistance genes that had been identified in IR64 were also present in Azucena, PCR amplification was carried out on Azucena DNA for each candidate gene, as described in section 3.2.4. As the sequences for Azucena were unknown, the primers used were those designed against the IR64 gene sequences. Where these primer sequences differed between the IR64 and Nipponbare for a particular gene, both the IR64 and the Nipponbare primer sets were used. Amplified products were sequenced directly from PCR products as described in section 3.2.4.

3.2.6 Growth and collection of IR64 root material for measurement of the expression of candidate resistance genes

To examine the expression of the 7 candidate resistance genes identified in IR64, *S. hermonthica*-infected and uninfected root tissue was collected at 3 different time points. IR64 rice plants were grown in rhizotrons and infected with *S. hermonthica* as described in section 2.2.2. Twenty-four plants were infected with germinated *S. hermonthica* seeds, while another

24 were left uninfected. At 2, 4 and 10 dai, infected roots were cut from each plant, washed gently in distilled water to remove any *S. hermonthica* seeds, dried on tissue paper and immediately frozen in liquid nitrogen. Uninfected control root tissue was collected and treated in the same way. Tissue from 2 rice plants was combined to make one biological replicate. Tissue for four biological was harvested at each time point.

3.2.7 RNA extraction, cDNA synthesis and qPCR to measure the expression of candidate resistance genes in IR64 following infection with *S. hermonthica*

RNA extraction of root tissue was carried out using the Qiagen RNeasy Plant Mini Kit. Approximately 100 mg of root tissue was used per sample. Root tissue was ground in liquid nitrogen with a pestle and 600 µl Buffer RLT, containing β-mercaptoethanol, was added to each sample. Further tissue disruption was carried out by shaking for 60 s at 25 hz in a TissueLyser (Qiagen). The lysate was then added to a QIAshredder column and centrifuged at 20,000 x g for 2 min. The supernatant was carefully removed and immediately mixed with 275 µl ethanol, before being transferred to an RNeasy spin column and centrifuged at 8000 x g for 15 s. 700 µl Buffer RW1 was added to each sample, inverted to wash the column and then centrifuged for another 15 s. Samples were then washed twice with 500 µl Buffer RPE, and centrifuged at 8000 x g for 15 s and 2 min respectively. Columns were centrifuge for another 60 s to remove residual buffer, and RNA eluted in 40 µl RNAase-free water by centrifuging for a final 60 s at 8,000 x g. The quality and quantity of RNA was assessed using a Nanodrop spectrophotometer (ND-8000, Thermo Scientific) by measuring the ratio of absorbance at 260 : 280 and 260 : 230. In addition, 500 ng of RNA was run on a 1 % agarose gel to assess its integrity. Only RNA samples with a 260 / 280 ratio of > 1.8 were used for cDNA synthesis.

Synthesis of cDNA was carried out using the Maxima First strand cDNA synthesis Kit for qPCR with dsDNase (Thermo Scientific) following the manufacturer's instructions. For each reaction, 1 µg RNA was combined with 1 µl 10 x dsDNase Buffer and 1 µl dsDNase in a total volume of 10 µl, and mixed gently. Samples were incubated for 10 min at 37 °C, chilled on ice, briefly centrifuged. After chilling again, 4 µl 5 x Reaction Mix, 2 µl Maxima Enzyme Mix and 4 µl nuclease-free water was added to each sample and gently mixed and centrifuged briefly. Samples were incubated for 10 min at 25 °C, followed by 30 min at 50 °C, and finally 5 min at 85 °C to terminate the reaction. To test for RNA contamination of synthesised cDNA, a PCR reaction was carried out using primers for the eEF-1α gene (elongation factor 1-alpha Os03g08020) (forward primer: ATGGGTAAGGAGAAGACGCACATC; reverse primer: TCATTTCTTCTTGGCGGCAGC) and products run on a 1.5 % agarose gel. Amplification with these primers gives different product sizes for genomic DNA (1939 bp) and cDNA (1344 bp).

qPCR was carried out using the Bioline SensiMix TM SYBR No-ROX Kit using a Corbett Research RG-6000 machine. Each 10 µl reaction consisted of 5 µl 2 x Sensimix ™ SYBR No-ROX, 1 µl primers (2.5 µM) and 4 µl cDNA (1 in 12 dilution). Cycling conditions were: 10 min at 95 °C (initial denaturation), then 40 cycles of 15 s at 95 °C (denaturation), 15 s at 60 °C (annealing), and 15 s at 72 °C (extension). After the 40 cycles a melt curve is performed. Temperature increased from 61 °C – 91 °C at 1 °C intervals for 5 s at each step (Corbett Research RG-6000). Two technical replicates were run for each sample, and 3 biological replicates were run for each time point and treatment. For each gene of interest, the PCR reactions were carried out for all time points for both infected and uninfected tissue in the same run. Presenilin (Os01g16930) was used as a reference gene as the expression of this gene does not alter during *S. hermonthica* infection (Swarbrick *et al.*, 2008). A negative (nuclease-free water) and positive (genomic DNA) control was included in each run. Primer sequences used in the reactions are shown in Table 3.3. Primers were designed to have a TM of ~60 °C, be approximately 20 - 25 bp in length and have product sizes of between 50 – 150 bp. Primers contained a GC clamp of 2 at the 3' end where possible. Potential hairpin formation and self-annealing was checked using the oligonucleotide properties calculator OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html), and specificity confirmed with Primer BLAST (NCBI) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and PCR followed by agarose gel electrophoresis.

Table 3.3. Sequences of primers designed for qPCR amplification of candidate *S. hermonthica* resistance genes in IR64. Presenilin (Os01g16930) was used as a reference gene.

Closest homolog	Primer name	Primer sequence (5' – 3')	Amplicon length (bp)
Os12g10870	10870 F 10870 R	GATCCTTCATATACAGTTGACAGG GCCATGTCAGCAATTCTGAGC	65
Os12g10930	IR64 10930 qPCR F1 IR64 10930 qPCR R1	CACCATCCCTGGGACTATCGG GGTGGAACTGGCCCCGTAAG	84
Os12g11370	11370 F 11370 R	GGATCCTTCATATACAGTTGATGG GAAGTTATTTGAGGCCATATCGG	79
Os12g11500	Os12g11500_F2 Os12g11500_R2	GTAGTACCACCGGTTCGG ATGACTGGAAGGCAGTGG	107
Os12g11510	IR64 11510 qPCR F3 IR64 11510 qPCR R3	GTGGGCATGGAAGACCTTGA AGGGTGAGATCCAATACTCGT	82
Os12g12000	Os12g12000-F1 Os12g12000-R1	GGAAGACGAAGATCAGATTGG TCCCCTCGCTCTTTATACTG	99
Os12g12010	12010 F2 12010 R2	TACTAGAAGTGTCTGGGTTGG GGAAGCTGGTAAGCGTCCG	119
Os01g16930	Os01g16930-F1 Os01g16930-R1	TAGAGCAGGAGGATGATTCG CACCAACATCCCTCATTTCC	54

3.2.9 Analysis of qPCR data

Relative quantification of gene expression was calculated according to the Comparative Ct method described in Pfaffl (2004). Raw data consisted of amplification efficiency (E) and take-

off (TO) cycle number for each technical replicate. Values for two technical replicates were averaged to give values for each biological replicate (sample). The mean amplification efficiency was calculated for the gene of interest (GOI) at each time point by averaging the biological replicates. For each biological replicate, E was raised to the power of the difference between the minimum take-off (defined as the replicate with the lowest take off value for a particular gene) and the sample take-off value of each replicate for that gene. The same procedure is followed for the reference gene in this case presenilin (Os01g16930). Values for the target gene were then divided by those of the reference gene. The equation for this calculation is shown below:

$$\text{Ratio} = \frac{(E_{target})^{\text{target (min. TO } - \text{ sample TO)}}}{(E_{ref})^{\text{ref (min. TO } - \text{ sample TO)}}}$$

E = amplification efficiency, target = target gene; ref = reference gene; and TO = take-off cycle.

Finally, all values were divided by the mean of the uninfected tissue for each time point and each gene of interest, to determine any the fold difference in gene expression between infected vs uninfected roots. For each gene, two-sample t-tests were carried out between infected and uninfected root tissue at each time point (2, 4 and 10 dai with *S. hermonthica*).

3.3 Results

3.3.1 Comparison of the genome structure of the *S. hermonthica* resistance QTL between rice cultivars

Mauve Multiple Alignment software was used to compare the genomic structure of the Nipponbare and IR64 QTL regions, and also the equivalent region for the *S. hermonthica* susceptible cultivar Koshihikari. Large-scale inversions and rearrangements across the QTL region between the three cultivars are shown in Figure 3.2. The sequence of the Nipponbare QTL was first aligned against the equivalent region in Koshihikari. It must be noted that the Koshihikari Bac sequence was used here which does not cover the entire QTL region, starting around 372 kb downstream of the Nipponbare sequence. Small regions of homology were observed between Nipponbare and Koshihikari. Additionally, a region of around 193 kb in length was observed in Nipponbare that was absent from the Koshihikari genome sequence (Figure 3.2a). At approximately 526 kb in length, the IR64 QTL region is 401 kb shorter than the Nipponbare QTL region of 927 kb. Despite this, many regions in the IR64 QTL show homology with parts of the Nipponbare QTL (Figure 3.2b). Some regions are rearranged, and 4 small regions have undergone inversion. The IR64 QTL was then compared to Koshihikari. Large regions of homology were observed with IR64 and Koshihikari (Figure 3.2c), and these showed very few rearrangements in the order of the sequence. In contrast to the alignment between Koshihikari and Nipponbare, no region was observed that was specific to IR64 and lacking in Koshihikari.

3.3.2 The IR64 QTL contains homologs of candidate *S. hermonthica* resistance genes in Nipponbare

Fgenesh gene-finder programme was used to predict gene models from the IR64 QTL sequence (PacBio) on chromosome 12. This revealed the presence of 76 genes. The amino acid sequences for the predicted genes were blasted against the rice MSU database to find the closest homolog and gene annotation. Of the 76 gene predictions, 11 were annotated as transposons, 31 as retrotransposons and 11 as expressed proteins. Five genes were not identified. Seven genes were predicted which shared closest homology to the candidate *S. hermonthica* resistance RLP genes in Nipponbare (Figure 3.3). A homolog to both the DNA repair protein (Os12g10850) and the Rapid Alkalization Factor family protein (Os12g12000) were also predicted for IR64. The different genes families found within the IR64 QTL are shown in Figure 3.3. All the genes except the transposons and retrotransposons are listed in Table 3.4 together with the annotation of their closest homolog. To distinguish the IR64 homologs from the Nipponbare genes, they will be named IR64_h_ + the name of the Nipponbare gene with the closest homology, from here on.

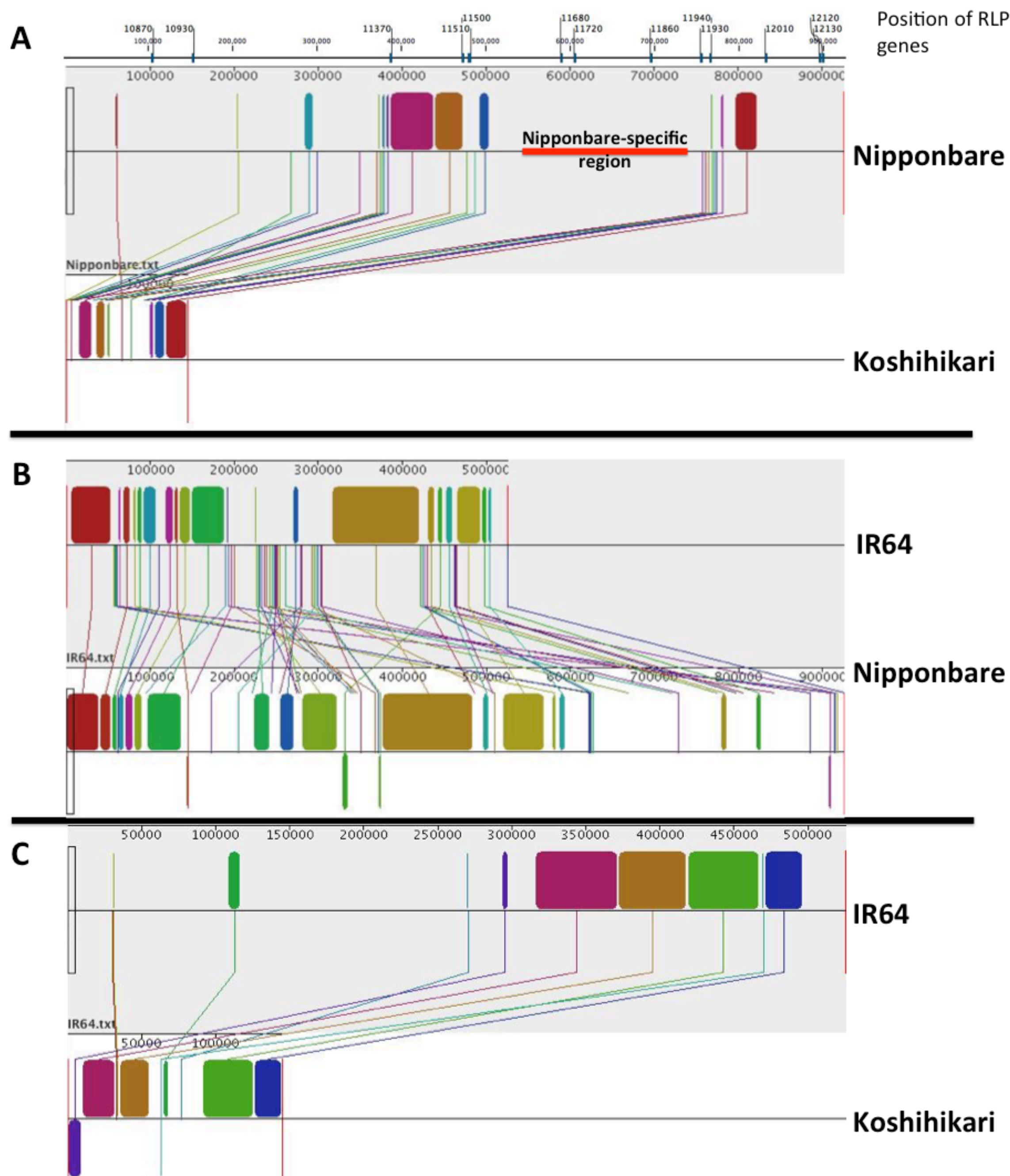
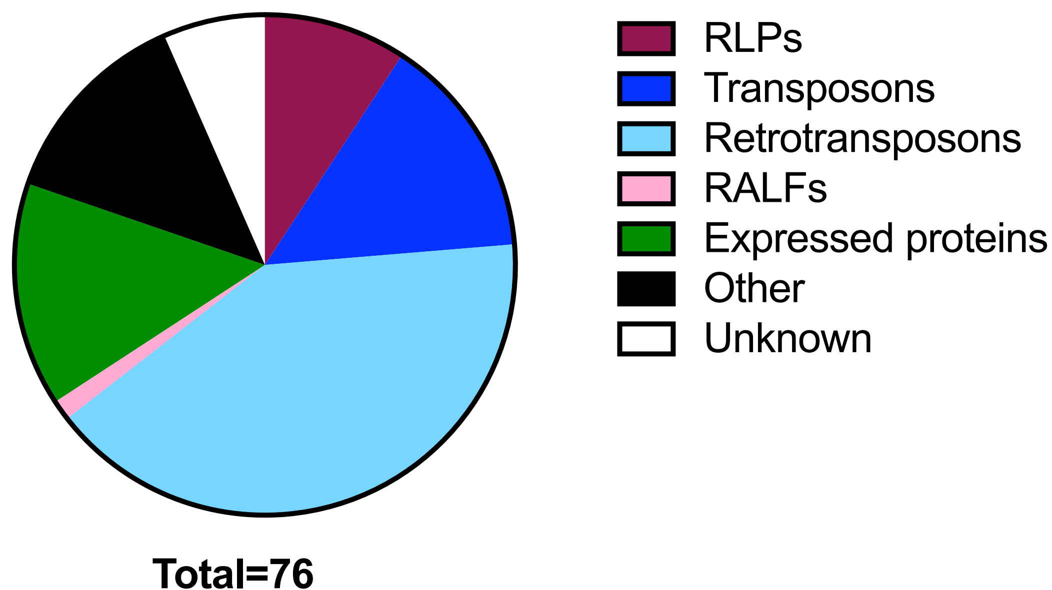


Figure 3.2. Structural comparison of the *Striga hermonthica* resistance QTL between rice cultivars IR64, Nipponbare and the susceptible cultivar Koshihikari. The position of the RLP genes on the Nipponbare sequence is indicated at the top, and numbers above each annotated gene refers to their Os12g number (MSU). Numbers above genome sequence for each cultivar refer to its length in bp and does not correlate with bp position on the genome. Coloured blocks indicate regions that are homologous and free from genomic rearrangement between cultivars. Lines between blocks link these regions of homology. Blocks below the centre line indicate regions of inverse orientation relative to the first (top) sequence being compared. Areas outside blocks did not align, and contain sequences specific to that genome or lacking homology between genomes. Genome alignments were carried out using Mauve Multiple Genome Alignment software using Progressive Mauve default parameters http://darlinglab.org/mauve/mauve.html.

Seven IR64 genes were predicted that shared closest homology to the candidate *S. hermonthica* resistance RLP genes in Nipponbare. These shared closest homology to Os12g10870, Os12g10930, Os12g11370, Os12g11500, Os12g11930, Os12g11510 and Os12g12010. The gene models for these genes and their coordinates on the PacBio contig are shown in Figure 3.4. Only one of these genes, IR64_h_Os12g10930, was predicted to have a single intron, all others were predicted to be encoded by a single exon (Figure 3.4). This is in contrast to Os12g10930 in Nipponbare, which contains 2 introns. Nipponbare Os12g11510 contains a single intron, whereas only 1 exon was predicted for the IR64 homolog for this gene. The predicted sequence for IR64_h_Os12g11500 translated to a protein of 818 amino acids in length, compared to 1013 in Nipponbare. A short protein 208 amino acids in length with closest homology to Os12g11930 (which forms a duplicated pair with Os12g11500 in Nipponbare) was also predicted for IR64. The position of these RLP gene homologs in relation to each other, and their orientation on the IR64 PacBio contig, are shown in Figure 3.5.

A comparison was carried out between the publically available Illumina and PacBio sequences for all these genes identified in IR64, by blasting the predicted gene sequences against the Illumina sequence. All gene sequences were identical between the two, with the exception of the IR64_h_ Os12g10870. The PacBio sequence for this gene had one nucleotide missing towards the end of the gene, which resulted in an early stop codon. To confirm the correct sequences from our own rice material, the RLP gene homologs in IR64 were cloned from genomic DNA, sequenced, assembled and aligned against the both the Illumina and the PacBio IR64 sequences (Figure 3.6). The sequences IR64_h_Os12g10870 matched the Illumina sequence, confirming a sequencing error in the PacBio version of this gene. This is shown is Figure 3.6. All other genes sequenced were identical to both the Illumina and PacBio sequences.

The amino acid sequences of the seven IR64 RLP homologs of candidate *S. hermonthica* resistance genes were aligned together to examine their similarity (Figure 3.7). Many of the genes showed good sequence similarity over much of the alignment; with the exception of IR64_h_Os12g11930, % identity between sequences ranged between 52.5 – 78.6 %. The greatest difference in sequences occurred at the predicted start site, with only Os12g10870, IR64_h_Os12g11370 and IR64_h_Os12g11510 predicted to contain signal peptides. The shorter sequence (818 amino acids) of IR64_h_Os12g11500 with respect to the other proteins meant that this protein is missing a transmembrane domain (Figure 3.7). In contrast, IR64_h_Os12g11930 contains a transmembrane domain but is missing a signal peptide and most of the LRR region present in the other 6 sequences (Figure 3.7).



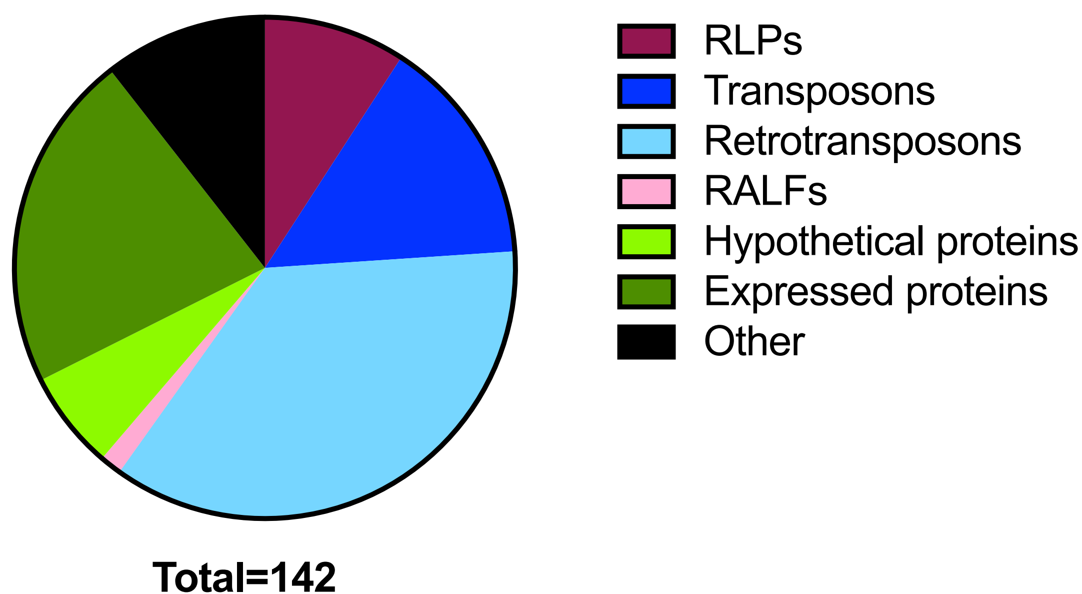


Figure 3.3. The number of genes and proportion of different gene families present in the IR64 (top) and Nipponbare (bottom) *S. hermonthica* resistance QTL between positions 5.7-6.7 Mbp. IR64 genes were predicted using Fgenesh gene-finder (Softberry, Inc) using PacBio sequence obtained from the Schatz lab (confidential). Annotation of IR64 genes was obtained by BLAST search against the Nipponbare reference genome on the MSU rice database.

Table 3.4 Genes within the IR64 *S. hermonthica* resistance QTL region as predicted by Fgenesh gene-finder (Softberry, Inc) using the IR64 PacBio contig sequence obtained from the Schatz lab (confidential). Genes encoding transposable elements have been omitted. Annotation was obtained by blasting the IR64 sequences against the Nipponbare reference genome (MSU). % identity to Nipponbare refers to the query coverage for each BLAST hit. Genes highlighted are homologs of candidate resistance RLP genes in the Nipponbare *S. hermonthica* resistance QTL.

Gene	Position on contig	exons	amino acids	chain	BLAST hit (MSU)	Annotation	Query Coverage	% Identity to Nipponbare
>FGENESH:174	1190669 - 1191103	1	144	-	LOC_Os10g10664.1	protein|hypothetical protein	83.33	99.17
>FGENESH:176	1197183 - 1197563	1	126	-	LOC_Os06g38460.1	protein|expressed protein	100	96.83
>FGENESH:180	1226216 - 1226653	1	145	+	LOC_Os12g10510.1	protein|expressed protein	100	99.31
>FGENESH:181	1226998 - 1230393	7	201	-	LOC_Os12g10520.2	protein|OsMADS33 - MADS-box family gene with MIKCc type-box, expressed	100	100
>FGENESH:183	1254705 - 1258356	6	151	+	LOC_Os12g10560.1	protein|clathrin adaptor complex small chain domain containing protein, expressed	100	99.34
>FGENESH:187	1278376 - 1278657	1	93	+	LOC_Os03g45010.1	protein|expressed protein	59.14	90.91
>FGENESH:189	1296477 - 1297364	1	295	+	LOC_Os12g10630.1	protein|ZF-HD protein dimerisation region containing protein, expressed	82.37	90.65
>FGENESH:190	1305976 - 1311010	10	402	-	LOC_Os12g10640.1	protein|uncharacterized protein ycf45, putative, expressed	92.54	99.73
>FGENESH:191	1315149 - 1317143	4	189	+	LOC_Os12g10650.1	protein|expressed protein	100	100
>FGENESH:192	1320933 - 1323441	2	214	+	LOC_Os12g10660.1	protein|B-box zinc finger family protein, putative, expressed	100	98.13
>FGENESH:193	1324179 - 1329337	5	1174	-	LOC_Os12g10670.1	protein|AAA-type ATPase family protein, putative, expressed	100	99.66
>FGENESH:195	1347745 - 1348477	2	132	-	LOC_Os12g10690.1	protein|expressed protein	100	92.42
>FGENESH:196	1352113 - 1359203	13	1227	+	LOC_Os12g10700.1	protein|expressed protein	100	96.74
>FGENESH:197	1367562 - 1371971	1	1469	-	LOC_Os12g10710.1	protein|NB-ARC domain containing protein, expressed	96.32	98.73
>FGENESH:198	1377212 - 1380083	8	196	+	LOC_Os12g10720.1	protein|glutathione S-transferase, putative, expressed	100	91.59
>FGENESH:199	1381329 - 1384120	9	212	+	LOC_Os12g10730.2	protein|glutathione S-transferase, putative, expressed	100	99.06
>FGENESH:201	1392019 - 1401841	19	854	+	LOC_Os12g10740.1	protein|leucine-rich repeat family protein, putative, expressed	100	81.97
>FGENESH:202	1404498 - 1405079	2	173	-	LOC_Os12g10750.1	protein|ARGOS, putative, expressed	72.83	95.24
>FGENESH:203	1411861 - 1412640	1	259	+	LOC_Os12g43564.1	protein|expressed protein	68.34	94.92
>FGENESH:204	1415680 - 1416227	2	172	-	LOC_Os12g10760.1	protein|expressed protein	100	100
>FGENESH:210	1461777 - 1465201	5	407	+	LOC_Os12g10810.1	protein|expressed protein	55.28	87.76
>FGENESH:214	1485289 - 1491393	8	670	+	LOC_Os12g10810.1	protein|expressed protein	39.4	84.47
>FGENESH:218	1514154 - 1515437	2	197	-	LOC_Os08g44290.1	protein|RNA recognition motif containing protein, putative, expressed	14.72	75.86
>FGENESH:219	1516435 - 1519039	7	465	+	LOC_Os12g10850.1	protein|hhH-GPD superfamily base excision DNA repair protein, putative, expressed	100	96.2
>FGENESH:220	1526291 - 1523245	1	1015	-	LOC_Os12g10870.1	protein|verticillium wilt disease resistance protein, putative, expressed	100	98.82
>FGENESH:221	1534068 - 1534274	1	68	-	LOC_Os12g10880.1	protein|expressed protein	100	83.82
>FGENESH:224	1562067 - 1564892	7	390	+	LOC_Os12g10910.1	protein|A/G-specific adenine DNA glycosylase, putative, expressed	47.44	78.6
>FGENESH:225	1569945 - 1572974	2	978	-	LOC_Os12g10930	genomic|NLOF, putative, expressed	64.59	88.72
>FGENESH:235	1631609 - 1633053	3	281	-	LOC_Os09g30060.1	protein|expressed protein	39.86	82.14
>FGENESH:242	1689663 - 1692707	1	1014	-	LOC_Os12g11370.1	protein|verticillium wilt disease resistance protein, putative, expressed	100	100
>FGENESH:245	1710511 - 1711172	2	87	-	LOC_Os12g11420.1	protein|expressed protein	58.62	82.35
>FGENESH:246	1712262 - 1712591	1	109	-	LOC_Os12g11400.1	protein|hypothetical protein	100	98.17

Table 3.4 continued

Gene	Position on contig	exons	amino acids	chain	BLAST hit (MSU)	Annotation	Query Coverage	% Identity to Nipponbare
>FGENESH:192	1320933 - 1323441	2	214	+	LOC_Os12g10660.1	protein|B-box zinc finger family protein, putative, expressed	100	98.13
>FGENESH:193	1324179 - 1329337	5	1174	-	LOC_Os12g10670.1	protein|AAA-type ATPase family protein, putative, expressed	100	99.66
>FGENESH:195	1347745 - 1348477	2	132	-	LOC_Os12g10690.1	protein|expressed protein	100	92.42
>FGENESH:196	1352113 - 1359203	13	1227	+	LOC_Os12g10700.1	protein|expressed protein	100	96.74
>FGENESH:197	1367562 - 1371971	1	1469	-	LOC_Os12g10710.1	protein|NB-ARC domain containing protein, expressed	96.32	98.73
>FGENESH:198	1377212 - 1380083	8	196	+	LOC_Os12g10720.1	protein|glutathione S-transferase, putative, expressed	100	91.59
>FGENESH:199	1381329 - 1384120	9	212	+	LOC_Os12g10730.2	protein|glutathione S-transferase, putative, expressed	100	99.06
>FGENESH:201	1392019 - 1401841	19	854	+	LOC_Os12g10740.1	protein|leucine-rich repeat family protein, putative, expressed	100	81.97
>FGENESH:202	1404498 - 1405079	2	173	-	LOC_Os12g10750.1	protein|ARGOS, putative, expressed	72.83	95.24
>FGENESH:203	1411861 - 1412640	1	259	+	LOC_Os12g43564.1	protein|expressed protein	68.34	94.92
>FGENESH:204	1415680 - 1416227	2	172	-	LOC_Os12g10760.1	protein|expressed protein	100	100
>FGENESH:210	1461777 - 1465201	5	407	+	LOC_Os12g10810.1	protein|expressed protein	55.28	87.76
>FGENESH:214	1485289 - 1491393	8	670	+	LOC_Os12g10810.1	protein|expressed protein	39.4	84.47
>FGENESH:218	1514154 - 1515437	2	197	-	LOC_Os08g44290.1	protein|RNA recognition motif containing protein, putative, expressed	14.72	75.86
>FGENESH:219	1516435 - 1519039	7	465	+	LOC_Os12g10850.1	protein|hhH-GPD superfamily base excision DNA repair protein, putative, expressed	100	96.2
>FGENESH:220	1526291 - 1523245	1	1015		LOC_Os12g10870.1	protein|verticillium wilt disease resistance protein, putative, expressed	100	98.82
>FGENESH:221	1534068 - 1534274	1	68	-	LOC_Os12g10880.1	protein|expressed protein	100	83.82
>FGENESH:224	1562067 - 1564892	7	390	+	LOC_Os12g10910.1	protein|A/G-specific adenine DNA glycosylase, putative, expressed	47.44	78.6
>FGENESH:225	1569945 - 1572974	2	978	-	LOC_Os12g10930	genomic|NL0E, putative, expressed	64.59	88.72
>FGENESH:235	1631609 - 1633053	3	281	-	LOC_Os09g30060.1	protein|expressed protein	39.86	82.14
>FGENESH:242	1689663 - 1692707	1	1014	-	LOC_Os12g11370.1	protein|verticillium wilt disease resistance protein, putative, expressed	100	100
>FGENESH:245	1710511 - 1711172	2	87	-	LOC_Os12g11420.1	protein|expressed protein	58.62	82.35
>FGENESH:246	1712262 - 1712591	1	109	-	LOC_Os12g11400.1	protein|hypothetical protein	100	98.17
>FGENESH:248	1715403 - 1715813	1	136	-	LOC_Os05g20010.1	protein|expressed protein	19.12	96.15
>FGENESH:256	1768306 - 1770759	1	817	+	LOC_Os12g11500.1	protein|resistance protein SlVe1 precursor, putative, expressed	96.57	98.86
>FGENESH:257	1779773 - 1780396	1	207	+	LOC_Os12g11930.1	protein|disease resistance protein SlVe2 precursor, putative, expressed	100	98.07
>FGENESH:258	1784556 - 1787288	1	910	-	LOC_Os12g11510	genomic|hcr2-0B, putative, expressed	54.15	98.31
>FGENESH:265	1839294 - 1839503	1	69	-	LOC_Os12g11980.1	protein|expressed protein	100	88.41
>FGENESH:266	1841865 - 1842071	1	68	-	LOC_Os12g11990.1	protein|expressed protein	100	97.06
>FGENESH:269	1864564 - 1864848	1	94	+	LOC_Os12g12000.1	protein|RALFL46 - Rapid ALkalinization Factor RALF family protein precursor, expressed	100	86.46
>FGENESH:270	1870560 - 1873493	1	977	+	LOC_Os12g12010.1	protein|verticillium wilt disease resistance protein precursor, putative, expressed	100	97.85
>FGENESH:271	1877169 - 1879931	3	853	-	LOC_Os11g46980.1	protein|receptor-like protein kinase 2 precursor, putative, expressed	87.57	72.36
>FGENESH:272	1891789 - 1893496	2	150	+	LOC_Os12g12170.1	protein|cytochrome b5-like Heme/Steroid binding domain containing protein, expressed	100	98.67
>FGENESH:278	1938377 - 1942685	10	819	-	LOC_Os12g12260.2	protein|diacylglycerol kinase 1, putative, expressed	80.22	99.85
>FGENESH:279	1957145 - 1958810	2	527	-	LOC_Os12g12290.1	protein|exostosin family domain containing protein, expressed	100	99.43
>FGENESH:280	1962401 - 1964889	8	328	-	LOC_Os12g12300.1	protein|EDM2, putative, expressed	100	94.2

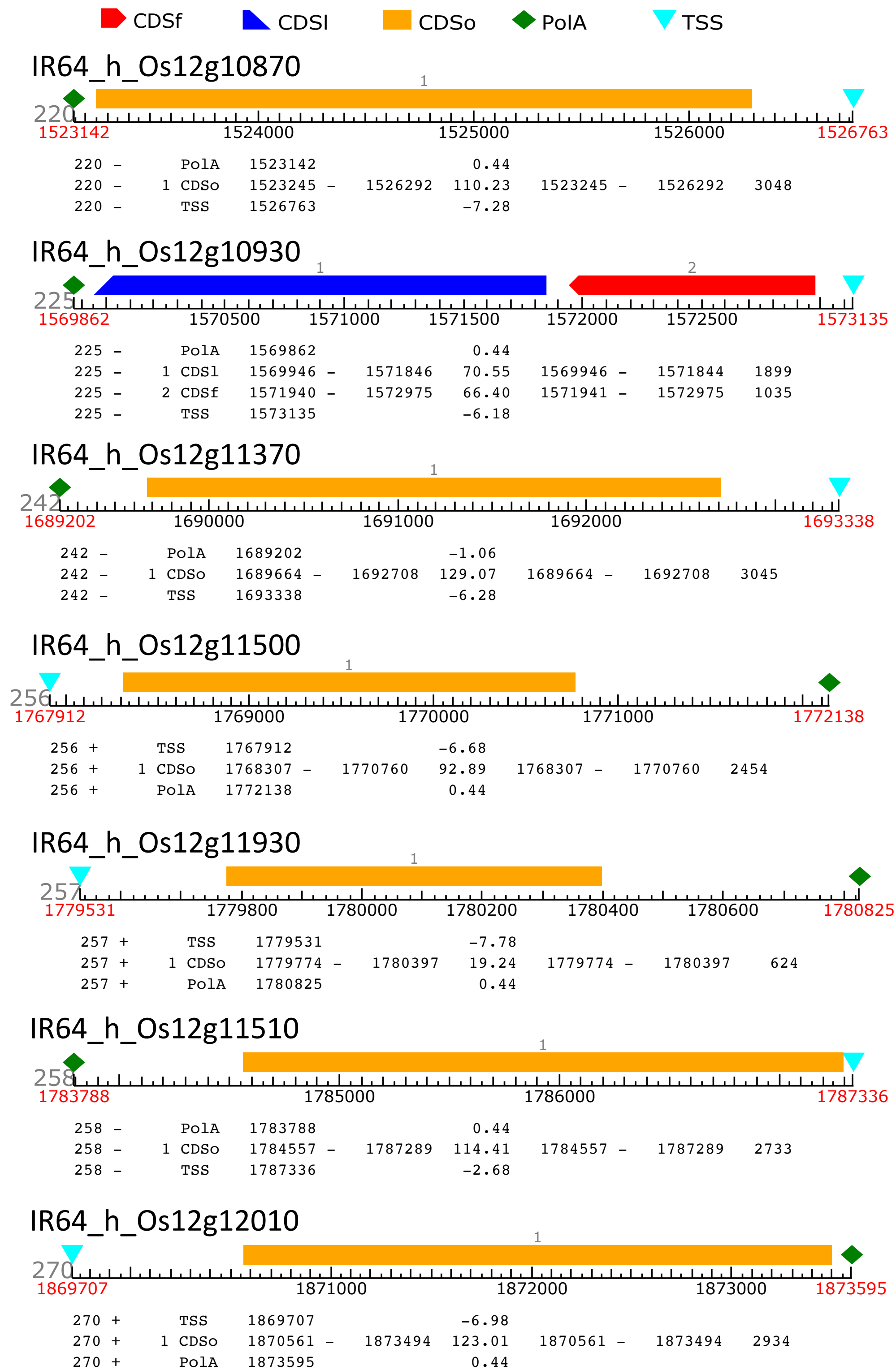


Figure 3.4. IR64 gene prediction models for genes showing homology to candidate *Striga hermonthica* resistance genes in Nipponbare, predicted by Fgenesh gene-finder (Softberry, Inc). Names refer to the closest Nipponbare homolog. Number of exons and their coordinates on the PacBio contig are shown below each prediction. TSS: transcription start site; PolA: polyadenylation signal sequence; CDSo: coding sequence, solo (single exon); CDSf: coding sequence, first; CDSl: coding sequence, last.

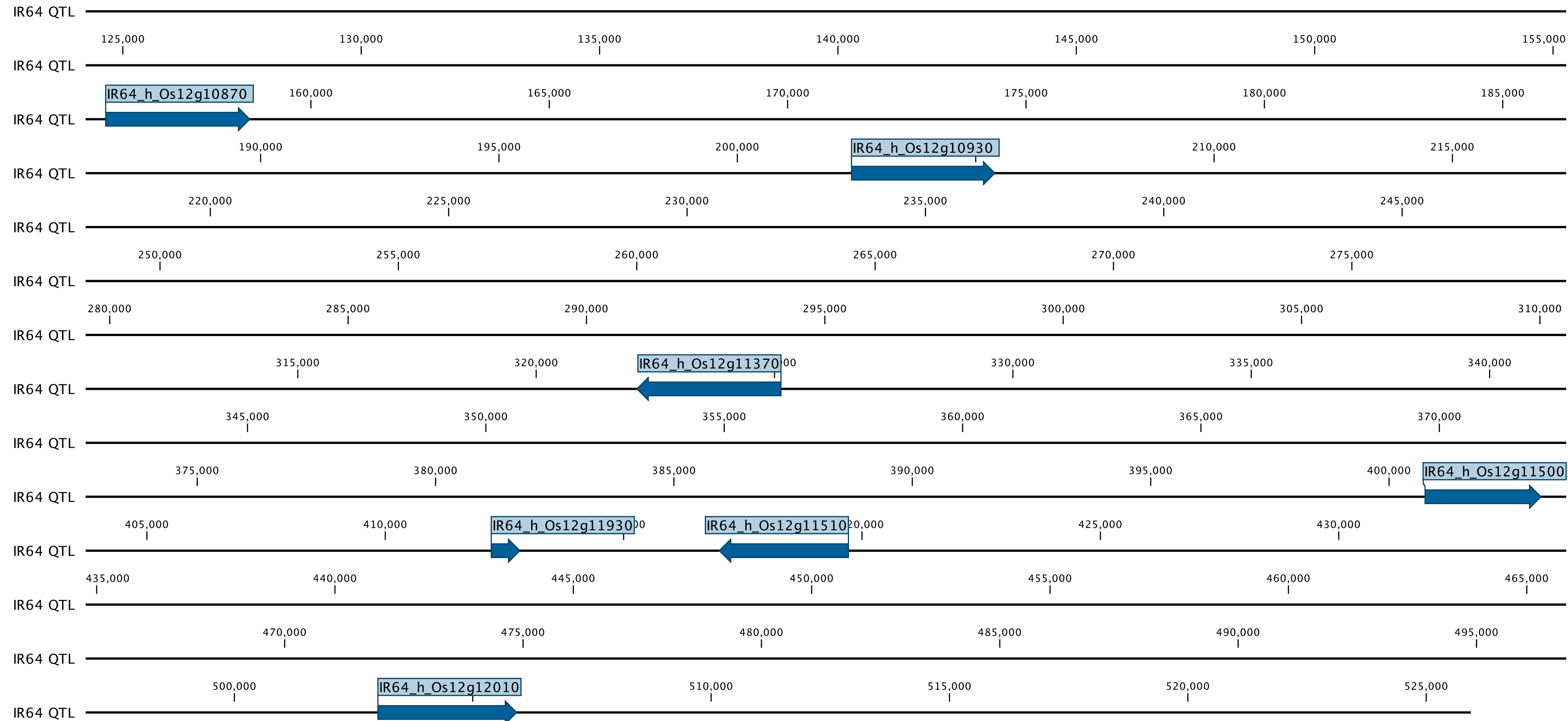


Figure 3.5 Position of IR64 gene predictions on the PacBio contig for genes showing homology to candidate *Striga hermonthica* resistance RLP genes in Nipponbare. Gene predictions were made by Fgenesh gene-finder (Softberry, Inc). Names refer to the closest Nipponbare homolog. Numbers above sequence refer to the length of the QTL in bp and does not correlate with position on the genome. Arrows indicate gene orientation. IR64 PacBio contig sequence was obtained from the Schatz lab (confidential)

Figure 3.6. Nucleic acid sequenced alignment from part of the IR64 homolog to Os12g10870. Illumina and PacBio sequnces are aligned against genes cloned from IR64. The sequencing error of the PacBio sequence is highlighted in red.

A phylogenetic tree was constructed for the 13 candidate *S. hermonthica* resistance RLP genes in the Nipponbare QTL, and the homologs identified in the IR64 *S. hermonthica* resistance QTL (Figure 3.8). This indicates that 4 of the RLP genes in Nipponbare have undergone a duplication event. One of each of the duplicated genes is present in the “Nipponbare-specific region” that was observed when the sequence of the Nipponbare QTL was aligned against the equivalent region of the Koshihikari genome (Figure 3. 2). The IR64 homologs of these RLP genes are more closely related to the genes present outside this “Nipponbare-specific” region (Figure 3.8).

```
                                      Signal peptide
                                                                  40                    60
IR64_h_Os12g10870  MSSFSKRV-- PHHVAS-LLA MMLILQLVQA TTLDDLTTTS SETTPAM-CL PDQASALLRL 56
IR64_h_Os12g11370  MSSSTKRLVR PHHLAKPLLT MLHILLQVQA IA----ALTD DATAPVIQCL PDQASALLRL 56
IR64_h_Os12g10930  M-----RVAR --HLPL-LLT VLQIVLQAQA ATILT-DRTS SSVPPPIPCL PDQASALLQL 51
IR64_h_Os12g11500  M--------- ---------- --YILLQVQA TT-----NTA RTVVPPVRCH PDQASALLRL 34
IR64_h_Os12g11930  M--------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  MRSA------ -YHLMPPLAM LLILGLADHA ---------- SSTEAPAACL PDQASALLQL 43
IR64_h_Os12g12010  MA-------- ---------- ---------- ---------- ----APIQCL PGQAAALLQL 18

                              80                    100                   120
IR64_h_Os12g10870  KRSFNATAGD YSTTFRSWIP GSDCCRWESV HCDGA-DGRV TSLDLGGHNL QAGG-LDHAL 114
IR64_h_Os12g11370  KNSFNKTAGG YSTAFRSWIT GTDCCHWDGV DCGGGEDGRV TSLVLGGHNL QAGS-ISPAL 115
IR64_h_Os12g10930  KRSFNPKAGD YTTAFRSWIT GIDCCHWDGI ACGGA-DGRV TSLDLGGHHL QASI-VDPAL 109
IR64_h_Os12g11500  KHSFNATAGD YSTAFQSWVA GTDCCRWDGV GCGGA-DGRV TSLDLGGHQL QAGS-VDPAL 92
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  KRSFNATIGD YPAAFRSWVA GADCCHWDGV RCGGAGGR-V TSLDLSHRDL QASSGLDDAL 102
IR64_h_Os12g12010  KRSFDATVGD YFAAFRSWVA GADCCHWDGV RCGGNDGRAI TFLDLRCHQL QAEV-LDAAL 77

                              140                   160                   180
IR64_h_Os12g10870  FRLTSLKHLN LSGNIFTMSQ LPA-TGFEQL TELTHLDLSD TNIAGKVPAG IGRLVSLVYL 173
IR64_h_Os12g11370  FRLTSLRYLD ISGNNFSMSQ LPV-TGFENL TELTHLDLSD TNIAGEVPAG IGSLVNLVYL 174
IR64_h_Os12g10930  FRLTSLRYLD LSGNNFSMSI L-I-NGLEQL TELTHLDLSD TNIAGEVPSA IGRLTSLVYL 167
IR64_h_Os12g11500  FRLTSLKHLN LSGNDFSMSQ LPVITGFEQL TELVYLDLSD TNIAGEVPGS IGRLTNLVYL 152
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  FSLTSLEYLD LSSNDFSKSK LPA-TGFEKL TGLTHLDLSN TNFAGLVPAG IGRLTSLNYL 161
IR64_h_Os12g12010  FSLTSLEYLD ISSNDFSASK LPA-TGFELL AELTHLDLSD DNFAGEVPAG IGHLTNLVYL 136

                              200                   220                   240
IR64_h_Os12g10870  DLSTSFVIVS YDDENSITRY ATHSIGQLSA PNMETLLTNL TNLEELHMGM V--DMSNN-- 229
IR64_h_Os12g11370  DLSTSFYIIY YDDENKMMPF ASDNFWQLSV PNMETLLANL TNLEELHMGM V--DMSGN-- 230
IR64_h_Os12g10930  DLSTSFYIVE FDNENGKMTY NSDLFRQLSA PNLETLLGNL TNLEELHMGM V--NMSGN-- 223
IR64_h_Os12g11500  DLSTSFYIVE YNDDEQVT-F DSDSVWQLSA PNMETLIENL SNLEELHMGM V--DLSGN-- 207
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  DLSTTFFVEG LDDKYSITYY YSDTMAQLSE PSLETLLANL TNLEELRLGM VMVNMSSNYG 221
IR64_h_Os12g12010  DLSTSFLDEE LDEENSVLYY TSYSLSQLSE PSLDSLLANL TNLQELRLGM V--DMSSN-- 192

                              260                   280                   300
IR64_h_Os12g10870  GELWCDHIAK YTPKLQVLSL PYCSLSGPVC ASFAAMRSLT TIELHYNLLS GSVPEFLAGF 289
IR64_h_Os12g11370  GERWCDDIAK FTPKLQVLSL PYCSLSGPIC TSLSSMNSLT RIELHYNHLS GSVPEFLAGF 290
IR64_h_Os12g10930  GDQWCDHIAK STPKLQVLSL PWCLLSGPIC TSLSAMQSLN TIELHYNHLS GSVPEFLATF 283
IR64_h_Os12g11500  GERWCDNIAK YTPKLQVLSL PYCSLSGPIC ASFSALQALT MIELHYNHLS GSVPEFLAGF 267
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  TARWCDAMAR SSPKLRVISM PYCSLSGPIC HSLSALRSLS VIELHYNHLS GPVPEFLAAL 281
IR64_h_Os12g12010  GARWCDAIAR FSPKLQIISM PYCSLSGPIC QSFSALKSLV VIELHYNYLS GPIPEFLADL 252

                              320                   340                   360
IR64_h_Os12g10870  SNLTVLQLST NNFQGWFPPI IFQHKKLRTI DLSKNPGISG NLP-NFSQDS SLENLFVSRT 348
IR64_h_Os12g11370  SNLTVLQLSK NKFEGLFPPI IFQHKKLVTI NITNNPGLSG SLP-NFSQDS KLENLLISST 349
IR64_h_Os12g10930  SNLTVLQLSR NKFEGWFPPT IFQHKKLITI NIINNPGLSG HLP-NFSQAS SLENVFISLT 342
IR64_h_Os12g11500  SNLTVLQLSK NKFQGSFPPI IFQHKKLRTI NLSKNPGISG NLP-NFSQDT SLENLFLNNT 326
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  PSLSVLQLSN NMFEGVFPPI IFQHEKLTTI NLTKNLGISG NLPTSFSGDS SLQSLSVSNT 341
IR64_h_Os12g12010  SNLSVLQLSN NNFEGWFPPI IFQHKKLRGI DLSKNFGISG NLP-NFSADS NLQSISVSNT 311

                              380                   400                   420
IR64_h_Os12g10870  NFTGMIPSSI SNLRSLKKLG IGASGFSGTL PSSLGSFLYL DLLEVSGFQI VGSMPSWISN 408
IR64_h_Os12g11370  NFTGIIPSSI SNLKSLTKLD LGASGFSGML PSSLGSLKYL DLLEVSGIQL TGSMAPWISN 409
IR64_h_Os12g10930  NFTA------ ---------- ---------- -----TLKYL DLLEVSGLQL VGSIPSWISN 371
IR64_h_Os12g11500  NFTGTIPGSI INLISVKKLD LGASGFSGSL PSSLGSLKYL DMLQLSGLQL VGTIPSWISN 386
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  NFSGTIPGSI SNLRSLKELA LGASGFSGVL PSSIGKLKSL SLLEVSGLEL VGSIPSWISN 401
IR64_h_Os12g12010  NFSGTIPSSI SNLKSLKELA LGASGFSGEL PSSIGKLKSL DLLEVSGLEL VGSMPSWISN 371

                              440                   460                   480
IR64_h_Os12g10870  LTSLTVLQFS NCGLSGHVPS SIGNLRELIK LALYNCKFSG KVPPQILNLT HLETLVLHSN 468
IR64_h_Os12g11370  LTSLTVLKFS DCGLSGEIPS SIGNLKKLSM LALYNCKFSG KVPPQIFNLT QLQSLQLHSN 469
IR64_h_Os12g10930  LTSLTVLQFS NCGLSGQVPS SIGNLRELRK LALYNCKFSG KMPPQILNLT RLQTLLLHSN 431
IR64_h_Os12g11500  LTSLTVLRIS NCGLSGPVPS SIGNLRELTT LALYNCNFSG TVHPQILNLT RLQTLLLHSN 446
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  LTSLTVLKFF SCGLSGPIPA SIGNLKKLTK LALYNCHFSG VIAPQILNLT HLQYLLLHSN 461
IR64_h_Os12g12010  LTSLTVLNFF HCGLSGRLPA SIVYLTKLTK LALYDCHFSG EVVNLILNLT QLETLLLHSN 431

                              500                   520                   540
IR64_h_Os12g10870  NFDGTIELTS FSKLKNLSVL NLSNNKLVVV DGENISSLVS FPNLEFLSLA SCSMSTFPNI 528
IR64_h_Os12g11370  NLAGTVELTS FTKLKNLSVL NLSNNKLLVL HGENSSSLVP FPKIKLLRLA SCSISTFPNI 529
IR64_h_Os12g10930  NFTGTVEITS FSKLENLSVL NLSNNELLVV DGENSTKVLS FPKIKFLRLA SCSISTFPNI 491
IR64_h_Os12g11500  NFAGTVDLTS FSKLKNLTFL NLSNNKLLVV EGKNSSSLVL FPKLQLLSLA SCSMTTFPNI 506
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  NLVGTVELSS YSKMQNLSAL NLSNNKLVVM DGENSSSVVS YPNIILLRLA SCSISSFPNI 521
IR64_h_Os12g12010  NFVGTAELTS LSKLQNLSVL NLSNNKLVVI DGENSSSEAT YPSISFLRLS SCSISSFPNI 491

                              560                   580                   600
IR64_h_Os12g10870  LKHLDKMFSL DISHNQIQGA IPQWAWKTWK GLQFLLLNMS HNNFTSLGSD PLLPLHIEFL 588
IR64_h_Os12g11370  LKHLHEITTL DLSHNKIQGA IPQWAWETWR GMYFLLLNIS HNNITSLGSD PLLPLEIDFF 589
IR64_h_Os12g10930  LKSLNEITSL DLSCNQIQGA IPQWAWGTWK GLQFYLLNIS HNNFTSLGPD SLLPLHIDYF 551
IR64_h_Os12g11500  LRDLPDITSL DLSNNQIQGA IPQWAWKTWK GLQFIVLNIS HNNFTSLGSD PFLPLYVEYF 566
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ---------- 1
IR64_h_Os12g11510  LRHLHEITFL DLSYNQIQGA IPQWAWKT-L NLGFALFNLS HNKFTSIGSH PLLPVYIEFF 580
IR64_h_Os12g12010  LRHLPEITSL DLSYNQIRGA IPQWVWKT-S GY-FSLLNLS HNKFTSTGSD PLLPLNIEFF 549
```

```
                                             620                           640                           660
IR64_h_Os12g10870  DLSFNSIEGP IPIPQEGSST LDYSSNQFSS IPLHYLTYLG ETLTFKASRN KLSGDIPPSI  648
IR64_h_Os12g11370  DLSFNSIEGP IPVPQEGSTM LDYSSNQFSS MPLHYSTYLG ETFTFKASKN KLSGNIP-SI  648
IR64_h_Os12g10930  DVSFNSIEGP IPIPRDGSST LDYSSNQFSA MPLHYSTYLG ETLVFKASKN KLSGNIPSSI  611
IR64_h_Os12g11500  DLSFNSIEGP IPIPQEGSST LDYSSNQFSS MPLRYSTYLG ETVTFKASKN KLSGNVPPLI  626
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ----------  1
IR64_h_Os12g11510  DLSFNNIEGV IPIPKEGSVT LDYSNNRFSS LPLNFSTYLT NTVLFKASNN SISRNIPPSI  640
IR64_h_Os12g12010  DLSFNKIEGV IPIPQKGSIT LDYSNNQFSS MPLNFSTYLK KTIIFKVSKN NLSGNIPPSI  609
                                             680                           700                           720
IR64_h_Os12g10870  CTAATNLQLF DLSYNNLSGS IPSCLMEDAI ELQVLSLKEN KLVGNLPDSI KEGCSLEAID  708
IR64_h_Os12g11370  CSAF-RLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI KEGCALEAID  707
IR64_h_Os12g10930  CTAVRRLQLI DLSYNNFSGP IPSCLMEDLT ALQVLSLKEN KLVGKLPDSI KEGCVLEALD  671
IR64_h_Os12g11500  CTTARKLQLI DLSYNNLSGS IPSCLLESFS ELQVLSLKAN KFVGKLPDII KEGCALEALD  686
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ----------  1
IR64_h_Os12g11510  CDGIKSLQLI DLSNNNLTGL IPSCLMEDAD ALQVLSLKDN HLTGELPDNI KEGCALSALD  700
IR64_h_Os12g12010  CDRIKSLQLI DLSNNYLTGI IPSCLMEDAS ALQVLSLKEN NLTGELPDNI KEGCALSALD  669
                                             740                           760                           780
IR64_h_Os12g10870  LSGNLIDGKI PRSLVSCRNL EILDVGNNQI SDSFPCWMSK LCKLQVLVLK SNKFTGQVMD  768
IR64_h_Os12g11370  LSGNLFEGRI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLALK SNKFTGQIMD  767
IR64_h_Os12g10930  LSGNLIEGKI PRSLVACKNL EFLDIGSNQI SDTFPCWMSE LPKLQVLVLK SNKFTGQVMD  731
IR64_h_Os12g11500  LSDNSIEGKI PRSLVSCRNL EILDIGSNQI SDSFPCWLSQ LPKLQVLVLK SNKLTGQVMD  746
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- ---------- ----------  1
IR64_h_Os12g11510  FSGNSIQGQL PRSLVACRNL EILDIGNNKI SDSFPCWMSK LPQLQVLVLK SNKFIGQILD  760
IR64_h_Os12g12010  FSGNLIQGKL PRSLVACRNL EILDIGNNQI SDSFPCWMSK LPQLQVLVLK SNRFIGQ-MD  728
                                             800                           820                           840
IR64_h_Os12g10870  PSYTVDRNSC AFTQLRIADM ASNNFNGTLP EAWFKMLKSM IAMTQNDTLV MENK--YYHG  826
IR64_h_Os12g11370  PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISDNDTLV MENQ--YYHG  825
IR64_h_Os12g10930  PSYMAGGDTC EFTELRIADM ASNNFNGTLP EAWFKMLKSM MSRSDNETLV MENQ--YYHG  789
IR64_h_Os12g11500  PSYTGRQISC EFPALRIADM ASNNLNGMLM EGWFNVTPRN LAQ------- ----------  789
IR64_h_Os12g11930  ---------- ---------- ---------- ---------- -ARSDNDTLV MENQ--YYHG  18
IR64_h_Os12g11510  PSYTGGNNC QFTKLQFADM SSNNLSGTLP EEWFKMLKSM IMDTCDNDML MREQHLYYRG  820
IR64_h_Os12g12010  ISYTGDANNC QFTKLRIADI ASNNFSGMLP EEWFKMLKSM MTSSDNGTSV MESQ--YYHG  786
                                             860                           880                           900
IR64_h_Os12g10870  --QTYQFTAS VTYKGSDMTI SKILRTLMLI DFSNNAFHGT IPETVGGLVL LHGLNMSHNA  884
IR64_h_Os12g11370  --QTYQFTAA VTYKGNYITI SKILRTLVLI DFSNNAFHGT IPETIGELVL LHGLNMSHNS  883
IR64_h_Os12g10930  --QTYQFTAT VTYKGSSMTI SKILRTLVLI DLSNNAFHGT IPGTIGELIL LHGLNMSHNA  847
IR64_h_Os12g11500  ---------- ---------I SRLFCVL--- ---------- ---------- --------NP  799
IR64_h_Os12g11930  --QTYQFTAT VTYKGNDRTI SKILRSLVLI DVSGNAFHGA IPDTIGELVL LRGLNLSHNA  76
IR64_h_Os12g11510  KMQSYQFTAG ISYKGSGLTI SKTLRTLVLI DVSNNAFHGR IPRSIGELVL LRALNMSHNA  880
IR64_h_Os12g12010  --QTYQFTAA LTYKGNDITI SKILTSLVLI DVSNNDFHGS IPSSIGELAL LHGLNMSRNM  844
                                             920                           940                           960
IR64_h_Os12g10870  LTGSIPTQFG RLNQLESLDL SSNELTGGIP KELASLNFLS TLNLSYNMLV GTIPNSYQFS  944
IR64_h_Os12g11370  LTGPIPTQFG RLNQLESLDL SSNELFGEIP KELASLNFLS ILNLSYNTLV GRIPNSYQFS  943
IR64_h_Os12g10930  LTGPVPPQLG RLNQLETLDL SSNKLSGEIP DELASLNFLS TLNLSYNVLV GRIPDSYQFS  907
IR64_h_Os12g11500  CPGPARV--- ---------- --HKTTSNI- ---------- ---------- -------QFQ  816
IR64_h_Os12g11930  LTGPIPSQFG RLDQLESLDL SFNELSGEIP KELASLNFLS TLNLSNNTLV GRIPDSYQFS  136
IR64_h_Os12g11510  LTGPIPVQFA NLKQLELLDL SSNELSGEI- ---------- ---------- ----------  909
IR64_h_Os12g12010  LTGPIPTQFG NLNNLESLDL SSNKLSNEIP EKLASLNFLA TLNLSYNMLA GRIPQSSHFS  904
                                             980                         1,000    Transmembrane region
IR64_h_Os12g10870  TFSNNSFLGN IGLCGPPLSK QCDNPKEPIV MTYTSEKS-T DVVLVLFTAL GFGVSYAMTI  1003
IR64_h_Os12g11370  TFSNNSFLGN TGLCGPPLSK QCDNPQESTV MPYVSEKS-I DVLLVLFTAL GFGVSFAITI  1002
IR64_h_Os12g10930  TFSNSSFLGN TGLCGPPLSR QCDNPKGPTE MPYTSEKS-I DVVLLLFTAL GFGLSFAMTI  966
IR64_h_Os12g11500  T*-------- ---------- ---------- ---------- ---------- ----------  818
IR64_h_Os12g11930  TFSNSSFLGN TGLCGPPLSR QCDNPEEPIT IPYTSEKS-I DAVLLLFTAL GFGIFFAMTI  195
IR64_h_Os12g11510  ---------- ---------- ---------- ---------- ---------- ----------  909
IR64_h_Os12g12010  TFSNASFEGN IGLCGAPLSK QCSYRSEPNI MPHASKKDPI DVLLFLFTGL GFGVCFGITI  964
                       Cytoplasmic tail
IR64_h_Os12g10870  LILWGRCMRK QR*-  1016
IR64_h_Os12g11370  LIVWGRHMKK QR*-  1015
IR64_h_Os12g10930  LIVWGSHMRK EH*-  979
IR64_h_Os12g11500  ---------- ----  818
IR64_h_Os12g11930  LIVWGSHMRK RH*-  208
IR64_h_Os12g11510  ---------- --L*  911
IR64_h_Os12g12010  LVIWGSNKRN QQA*  978
```

Key

- Signal peptide
- Transmembrane domain
- Cytoplasmic tail
- LRR region

Figure 3.7. Amino acid alignment for 7 predicted IR64 homologs of candidate *Striga hermonthica* resistance RLP genes in Nipponbare. Polymorphic residues are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding (Zhang *et al.*, 2014).

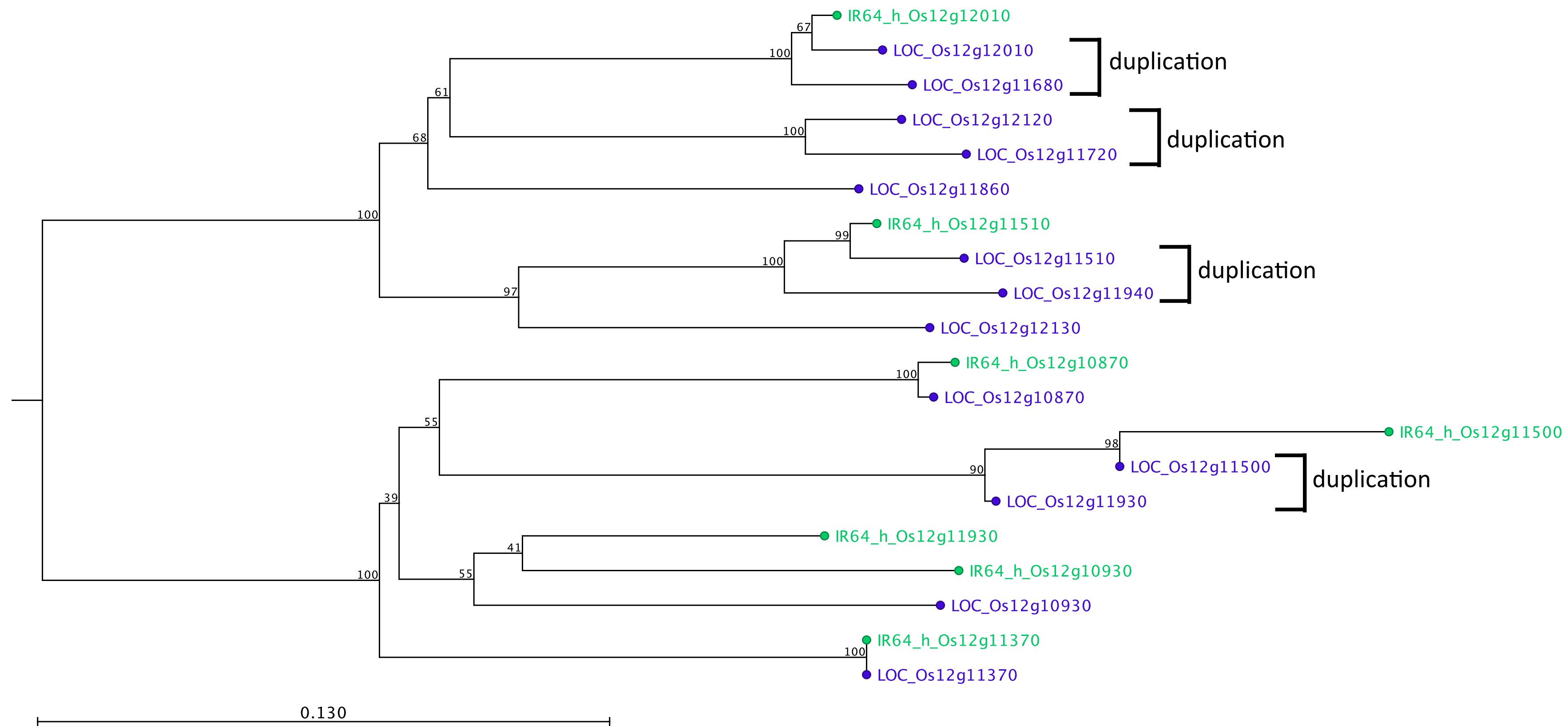


Figure 3.8. Phylogenetic relationship of 13 RLP genes in the *Striga hermonthica* resistance QTL for the rice cultivar Nipponbare, and the 7 identified homologs in the IR64 *S. hermonthica* resistance QTL. Nipponbare genes are shown in blue, IR64 homologs are shown in green. Four Nipponbare genes have undergone duplication (indicated). Constructed in CLC bio Main Workbench version 7.0.3 using the Neighbor Joining method and 1,000 bootstrap replicates. Branch lengths are proportional to phylogenetic distances. Numbers at nodes are bootstrap confidence values.

3.3.3 Koshihikari contains homologs of candidate *S. hermonthica* resistance genes in Nipponbare

Fgenesh gene-finder programme was also used to predict gene models for Koshihikari using Bac sequence for the equivalent region to the IR64 QTL. Nineteen genes were predicted from the NIAS Koshihikari Bac sequence. The amino acid sequence of all predicted genes was blasted against the rice reference sequence to obtain gene annotations (Table 3.5). Like IR64 and Nipponbare, the largest proportion of these predicted genes were annotated as transposons or retrotransposons (8), 4 were annotated as expressed proteins, and 3 were unidentified. Four genes were identified that shared closest homology to candidate *S. hermonthica* resistance RLP genes in Nipponbare, sharing closest homology to Os12g11370, Os12g11500, Os12g11940 and Os12g12130. These will be named Kosh_h_ + the name of the Nipponbare gene with the closest homology, from here on. Eight additional genes in Koshihikari were predicted by Fgenesh using sequence obtained from NCBI for the region not covered by the NIAS Bac sequence. These included 2 expressed proteins, 2 glutathione S-transferases and a leucine-rich repeat family protein, and are listed in Table 3.5. However, it should be noted that due to the poor quality of this sequence, other genes may also be present that could not be detected from the available sequence.

The Fgenesh gene model for Kosh_h_Os12g11370 predicts a protein encoded by 4 exons, which is in contrast to the single exon for this gene for both Nipponbare and IR64. The remaining 3 RLP homologs are encoded by a single exon (Figure 3.9). Kosh_h_Os12g11500 is 70 amino acids shorter than Os12g11500, meaning this protein lacks a transmembrane domain (Figure 3.10). Kosh_h_Os12g11940 and Kosh_h_Os12g12130 are 497 and 509 amino acids shorter than their Nipponbare homologs, respectively, and therefore also lack transmembrane domains (Figure 3.10).

Table 3.5 Genes in Koshihikari for the equivalent region to the IR64 *S. hermonthica* resistance QTL, as predicted by Fgenesh gene-finder (Softberry, Inc). Annotation was obtained by blasting the sequences against the Nipponbare reference genome (MSU). % identity to Nipponbare refers to the query coverage for each BLAST hit. Genes highlighted are homologs of candidate resistance RLP genes in the Nipponbare *S. hermonthica* resistance QTL. Genomic sequence for the first 8 models was obtained from NCBI (GenBank: DG000036.1). The remaining models were predicted from the Bac sequence provided by Dr. Kiyosumi Hori (National Institute of Agrobiological Sciences (NIAS), Japan). Genome sequence between the NCBI and NIAS sequences is not continuous.

Gene	Position on contig	exons	amino acids	chain	BLAST hit (MSU)	Description	Query Coverage	% Identity to Nipponbare	Source:
>FGENESH:1a	2255 - 6664	1	1469	-	LOC_Os12g10710.1	protein|NB-ARC domain containing protein, expressed	96.32	100	
>FGENESH:2a	11822 - 14693	8	196	+	LOC_Os12g10720.1	protein|glutathione S-transferase, putative, expressed	100	91.59	
>FGENESH:3a	15940 - 18731	9	212	+	LOC_Os12g10730.2	protein|glutathione S-transferase, putative, expressed	100	100	
>FGENESH:4a	21453 - 28350	18	927	+	LOC_Os12g10740.1	protein|leucine-rich repeat family protein, putative, expressed	100	89.86	
>FGENESH:5a	48049 - 49433	3	136	+	LOC_Os12g10770.1	protein|retrotransposon protein, putative, unclassified, expressed	41.18	100	
>FGENESH:6a	57039 - 62363	13	299	+	LOC_Os12g10784.1	protein|2-dehydro-3-deoxyphosphooctonate aldolase, putative, expressed	100	82.13	
>FGENESH:7a	73408 - 76160	4	299	+	LOC_Os12g10810.1	protein|expressed protein	57.53	97.09	
>FGENESH:8a	78925 - 79491	1	188	+	OC_Os12g10810.1	protein|expressed protein	96.28	84.53	NCBI
>FGENESH:1	12833 - 13141	1	102	+	No hits				
>FGENESH:2	14733 - 18059	4	979	-	LOC_Os12g11370.1	protein|verticillium wilt disease resistance protein, putative, expressed	100	85.65	
>FGENESH:3	20289 - 21195	2	88	-	LOC_Os12g11430.1	protein|retrotransposon protein, putative, unclassified	51.14	97.78	
>FGENESH:4	23235 - 23765	1	176	-	LOC_Os03g56910.1	protein|retrotransposon protein, putative, LINE subclass, expressed	100	73.3	
>FGENESH:5	30176 - 30463	1	95	+	No hits				
>FGENESH:6	44484 - 47312	1	942	+	LOC_Os12g11500.1	protein|resistance protein SlVe1 precursor, putative, expressed	99.89	93.94	
>FGENESH:7	54358 - 55908	1	516	-	LOC_Os12g12130.1	protein|verticillium wilt disease resistance protein, putative, expressed	99.81	78.34	
>FGENESH:8	56994 - 60902	3	1226	+	LOC_Os03g33640.1	protein|retrotransposon protein, putative, Ty1-copia subclass, expressed	100	95.84	
>FGENESH:9	65928 - 68224	3	666	+	LOC_Os01g68220.1	protein|transposon protein, putative, unclassified, expressed	78.53	98.28	
>FGENESH:10	71018 - 71671	1	217	-	LOC_Os09g04590.1	protein|retrotransposon protein, putative, unclassified, expressed	48.39	99.05	
>FGENESH:11	73488 - 75970	2	391	+	LOC_Os02g20080.1	protein|retrotransposon protein, putative, unclassified	100	98.98	
>FGENESH:12	78274 - 78927	1	217	-	No hits				
>FGENESH:13	80744 - 90788	13	2218	+	LOC_Os12g39800.1	protein|retrotransposon protein, putative, Ty1-copia subclass, expressed	43.51	94.33	
>FGENESH:14	91003 - 92076	1	357	-	LOC_Os12g11940.1	protein|disease resistance family protein, putative, expressed	100	91.69	
>FGENESH:15	95091 - 95564	1	157	+	LOC_Os12g11950.1	protein|expressed protein	100	97.45	
>FGENESH:16	119080 - 119289	1	69	-	LOC_Os12g11550.1	protein|hypothetical protein	100	97.1	
>FGENESH:17	126982 - 127188	1	68	-	LOC_Os12g11980.1	protein|expressed protein	100	100	
>FGENESH:18	130499 - 130705	1	68	-	LOC_Os12g11990.1	protein|expressed protein	100	100	
>FGENESH:19	131433 - 137432	5	1138	-	LOC_Os06g13430.1	protein|retrotransposon protein, putative, unclassified, expressed	36.99	95.85	NIAS

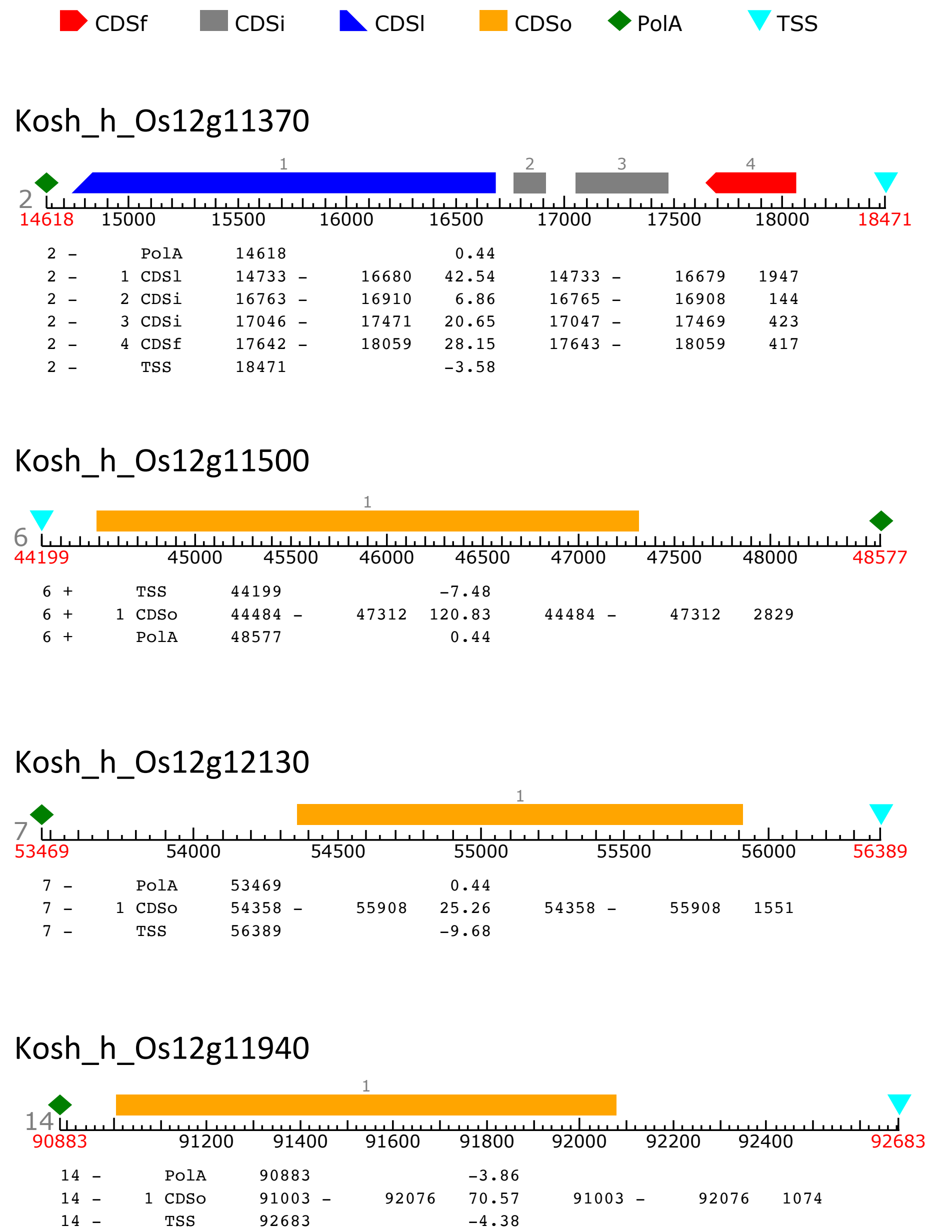


Figure 3.9 Koshihikari gene prediction models for 4 genes showing homology to candidate *Striga hermonthica* resistance genes in Nipponbare, predicted by Fgenesh gene-finder (Softberry, Inc). Names refer to the closest Nipponbare homolog. Number of exons and their coordinates on the NIAS Bac clone sequence are shown below each prediction. TSS: transcription start site; PolA: polyadenylation signal sequence; CDSo: coding sequence, solo (single exon); CDSf: coding sequence, first; CDSi: coding sequence, internal; CDSl: coding sequence, last.

```
                  Signal peptide
                                                                40                      60
Kosh_h_Os12g11370 MSSSTKRPVR PHHLAKPLLL TMLHILLQLK AITALTDDAT APVIQCLPDQ ASALLRLKHS 60
Kosh_h_Os12g11500 M--------- ---------- ---YILLQVQ ATTN-TARTV VPPVPCHPDQ ASALLRLKHS 37
Kosh_h_Os12g11940 MRSAYH---- --------LM PLLAMLLILV LADHTSSTEA VAPAACLPDQ AAALLQLKRS 48
Kosh_h_Os12g12130 M--------- ---------- ---------- ---------- ---------- ---------- 1

                             80                     100                     120
Kosh_h_Os12g11370 FNTTAGGYST TFRSWIT--G TDCCHWEGIH CSG-EDGRVT SLVLGGHNLQ TTI-VDPALF 116
Kosh_h_Os12g11500 FNATAGDYST AFRSWVA--G TDCCRWDGVG CGGGADGRVT SLDLGGHNLQ AGS-VDPALF 94
Kosh_h_Os12g11940 FNATIGDYSA AFRSWVAVAG ADCCSWDGVR CGG-AGGRVT SLDLSHRDLQ AASGLDDALF 107
Kosh_h_Os12g12130 ---------- ---------- ---------- ---------- ---------- ---------- 1

                            140                     160                     180
Kosh_h_Os12g11370 RLNSLRYLDL SGNNFSMSQL PV-------- ---------- ---------- ---------- 138
Kosh_h_Os12g11500 RLTSLKHLNL SGNDFSMSQL PVITGFEQLT ELVHLHLSDT NITGEVPGSI GRLTNLVYLD 154
Kosh_h_Os12g11940 SLTSLGYLDL SSNDFGKSQM PA-TGFEKLT GLTHLDLSNT NFAGLAPAGI GGLTSLNYLD 166
Kosh_h_Os12g12130 ---------- ---------- ---------- ---------- ---------- ---------- 1

                            200                     220                     240
Kosh_h_Os12g11370 ---------- ---------- TDYFWQLSLP SMETLLANLT NLEELHMGMV DMSGNGERWC 178
Kosh_h_Os12g11500 LSTSFYIVEY NDDEQVTFN- SDSVWQLSAP NMETLLENLT NLEKLHMGMV DLSGNGERWC 213
Kosh_h_Os12g11940 LSTTFFEEEL DNENSIPYYY SDTISQLSEP SLETLLANLT NLEELRLGMV NMSSNGARWC 226
Kosh_h_Os12g12130 ---------- ---------- ---------- ---------- ---------- ---------- 1

                            260                     280                     300
Kosh_h_Os12g11370 DDVAKFAPKL QVLSLPYCSL SGPICTSLSS MNSLTRIELH YNHLSGPVPE FLAGFSNLTV 238
Kosh_h_Os12g11500 YNIAKYTPKL QVLSLPYCSL SGPICASFSA LQALTMIELH YNHLSGSVPE FLAGFSNLTV 273
Kosh_h_Os12g11940 DAMARSSSKL RVISMPYCSL SGPICHSLSA LRLLSVIELH YNHLSGPVPE FLAALSNLSV 286
Kosh_h_Os12g12130 ---------- ---------- ---------- ---------- ---------- ---------- 1

                            320                     340                     360
Kosh_h_Os12g11370 LQLSKNKFEG LFPPIIFQHK KLVTINITNN PGLSGSLPNF SQESSLKYLD SLEVSGLQLA 298
Kosh_h_Os12g11500 LQLSKNKFQG SFPPIIFQHK KLRTINLSKN PGISGNLPNF SQDTSL---E NLFLSNTNFT 330
Kosh_h_Os12g11940 LQLSNNMFEG AFPPIIFQHE KLTTINLTKN ---------- ---------- ---------- 316
Kosh_h_Os12g12130 ---------- ---------- ---------- ---------- ---------- ---------- 1

                            380                     400                     420
Kosh_h_Os12g11370 GSMAPWISNL TSLTVLKFSD CGLSGEIPSS IGKKLDLGAS GFSGMLPSSL GSLKYLDSLE 358
Kosh_h_Os12g11500 GTIPSSIINL ISV------- ---------- --KKLDLGAS GFSGSLPSSL GSLKYLDMLQ 371
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 316
Kosh_h_Os12g12130 ---------- ---------- ---------- ---------- ---------- ---------- 1

                            440                     460                     480
Kosh_h_Os12g11370 VSGLQLAGSM APWISNLTSL TVLKFSDCGL SGEIPSSIGN LKKLSMLALY NCKFSGKVPP 418
Kosh_h_Os12g11500 LSGLQLVGTI PSWISNLSSL TVLRFSNCGL SGQVPSSIGN LRELTTLALY NCNFSGTVPP 431
Kosh_h_Os12g11940 ---------- ---------- -------LGI SGNLPSFPGD -SSLQSLSVS NTNFSGTIPS 348
Kosh_h_Os12g12130 ---------- PSWISNLTSL NVLKFFSCGL SGPIPSSIGS LTKLTKLALY NCQFSGEIPS 51

                            500                     520                     540
Kosh_h_Os12g11370 QIFNLTQLQS LQLHSNNLAG TVELTSFTKL KNLSVLNLSN NKLLVLRGEN SSSLVPFPKI 478
Kosh_h_Os12g11500 QILNLTRLQT LLLHSNNFAG TVELTSFSKL KNLTFLNLSN NKLLVVEGKN SSSLVSLHKL 491
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 348
Kosh_h_Os12g12130 LILNLTQLET LLLHSNSFVG TVELTSYTKL QNLYVLNLSN NKLVVIDGEI TSSVVSNSSM 111

                            560                     580                     600
Kosh_h_Os12g11370 KLLRLASCSI STFPNILRHL HEITTLDLSH NKIQGAIPQW AWETWRGMYF LLLNMSHNNI 538
Kosh_h_Os12g11500 QLLSLASCSM TTFPNILRHL PEITSLDLSN NQIQGAIPQW AWKTWKGLQF IVLNISHNNF 551
Kosh_h_Os12g11940 --------SI STIPSL---- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 SFLRLASCSI SSFPKILRHL PEIYSLDLSH NQIEGAIPQW AWETWT-TDF FFLNLSHNKF 170

                            620                     640                     660
Kosh_h_Os12g11370 TSLGSDPLLP LEIDFFDLSF NSIEGPIPVP QEGSTMLDYS SNQFSSMPLH YSAYLGQTFT 598
Kosh_h_Os12g11500 TSLGSDPFLP LSVEYFDLSF NSIGGPIPIP QEGSSTLDYS SNQFSSMPLR YSTYLGETLT 611
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 TNIGTNPLLP LYIEYFDLSF NNFEGDIPIP EEGSVTLDFS NNQFSSVPSN FYTYLINSLV 230

                            680                     700                     720
Kosh_h_Os12g11370 FKASKNKLSG NIPS-ICTAP R-LQLIDLSY NNLSGSIPSC LMEDVTALQI LNLKENKLVG 656
Kosh_h_Os12g11500 FKASKNKLSG NVPPLICTTA RKLQLIDLSY NNLSGSIPSC LLESFSELQV LSLKANKFVG 671
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 FKASRNSLSG NIPPMICNSI KTLQIIDLSY NNLNGSIPSC LMENLGSLQV LSLKENQLAG 290

                            740                     760                     780
Kosh_h_Os12g11370 TIPDNIKEGC ALEAIDLSGN LFEGKIPRSL VACRNLEILD IGNNEISDSF PCWMSKLPKL 716
Kosh_h_Os12g11500 KLPDIIKEGC ALEALDLSDN SIEGKIPRSL VSCRNLEILD IGSNQISDSF PCWLSQLPKL 731
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 ELPDNIKEGC ALSALDFSCN LIQGQLPRSL VACMNLEILD IGNNQISDSF PCWMSKLTEL 350

                            800                     820                     840
Kosh_h_Os12g11370 QVLVLKSNKF TGQIMDPSYT VDGNSCEFTE LRIADMASNN FNGTLPEAWF TMLKSM-NAI 775
Kosh_h_Os12g11500 QVLVLKSNKL TGQVMDPSYT GRQNSCEFPS LRIADMASNN LNGMLMEGWF KKLKSM-MAR 790
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 RVLVLKSNKF IGQILDPSYT GGGNNCQFTK LQFADISSNN LSGTLPEEWF KMLKSMIMVT 410

                            860                     880                     900
Kosh_h_Os12g11370 SENDTLVMEN Q-YYHG--QT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET 832
Kosh_h_Os12g11500 SDNDTLVMEN Q-YYHG--QT YQFTATVTYK GKSMTFSKVL RTLVLIDFSN NAFHGTIPET 847
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 SDNDMLIKEQ HLYYRGKMQS YQFTAGISYK GSHVRISKTL TTLVLIDVSN NAFHGRIPRS 470

                            920                     940                     960
Kosh_h_Os12g11370 IGELVLLHGL NMSHNELTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL 892
Kosh_h_Os12g11500 IGELILLHGL NISHNALTGP IPPQLGRLNQ LESLDLSSNK LSGKIPNELE SLNFLSTLNL 907
Kosh_h_Os12g11940 ---------- ---------- ---------- ---------- ---------- ---------- 356
Kosh_h_Os12g12130 IGELVLLRAL NMSHNALTGP IPVQFANLKQ LELLDLSSNE LSGEIL*--- ---------- 517
```

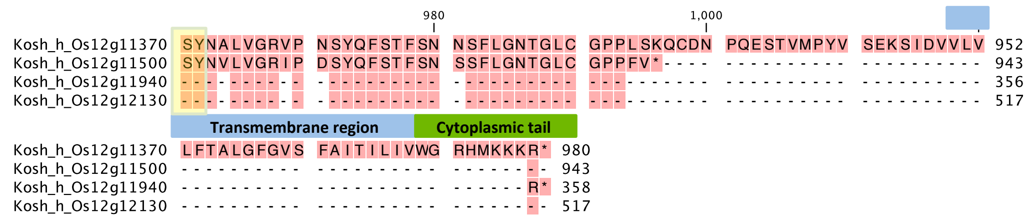


Figure 3.10 Amino acid alignment for 4 predicted Koshihikari homologs of candidate *Striga hermonthica* resistance RLP genes in Nipponbare. Polymorphic residues are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding (Zhang *et al.*, 2014).

3.3.4 Azucena shares 3 homologs of candidate *S. hermonthica* resistance genes with IR64 and Nipponbare

As the genome sequence of Azucena was not available, primers designed against candidate *S. hermonthica* resistance RLP genes in IR64 were used to try to PCR amplify full-length gene sequences from Azucena genomic DNA. PCR products were amplified from Azucena using primers designed against two genes: IR64_h_Os12g12010, and the RALF IR64_h_Os12g12000, however non-specific amplification was also observed with the IR64_h_Os12g12000 primers (Figure 3.11A). Primers designed against Nipponbare genes sequences were also used on Azucena where primer sequences differed between the IR64 and Nipponbare for a particular gene. Clear bands of the same size were observed for Os12g11500, Os12g12000 and Os12g12010 (Figure 3.11B). These products were sequenced and their sequences translated to amino acids for alignment against the IR64 and Nipponbare homologs.

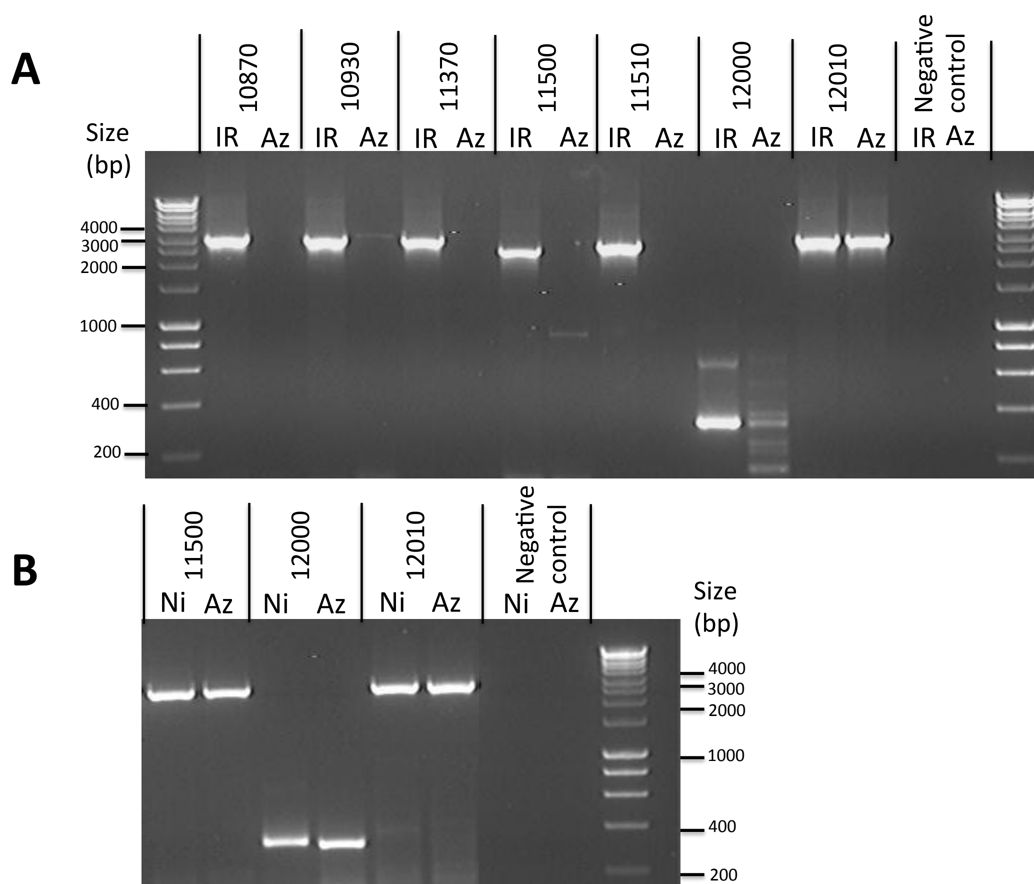


Figure 3.11 Amplification of full-length candidate *S. hermonthica* resistance genes from IR64 and Azucena genomic DNA. **A:** primers were designed against IR64 RLP homologs. **B:** where primer sequences differed between IR64 and Nipponbare for a particular gene, amplification of Azucena was also attempted with Nipponbare specific primers. IR = IR64, Az = Azucena, Ni = Nipponbare. Numbers above names refer to the Os12g gene numbers for the closest homolog in Nipponbare, according the MSU rice database.

3.3.5 Comparison of candidate *S. hermonthica* resistance RLP genes between cultivars

The amino acid sequences of the candidate *S. hermonthica* resistance genes in IR64 were aligned against those of Nipponbare, and where sequences were known, Koshihikari and Azucena. These are shown in Figures 3.12 - 3.15. In all cases, the IR64 homologs were more similar to the Nipponbare genes outside the "Nipponbare-specific region", rather than the duplicated gene within the Nipponbare specific region (see Figure 3.8 for phylogenetic tree). This was also the case for the Azucena and Koshihikari homologs, with the exception of one small gene in Koshihikari that was more similar to Os12g11940 than Os12g11510. Only two genes were identified in IR64 where homologs were not detected in either of the two susceptible cultivars. These were Os12g10870 and Os12g10930. The similarity in amino acid sequence for each of the IR64 candidate *S. hermonthica* resistance genes to Nipponbare, and also Koshihikari and Azucena where identified, is described below.

IR64_h_Os12g10870: this gene shared 98.8% identity with Nipponbare (Figure 3.12). No homologs were identified in Koshihikari or Azucena.

IR64_h_Os12g10930: When compared to Nipponbare, IR64_h_Os12g10930 differed in several ways, with nearly 400 differences in amino acids, including a region of 237 amino acids not present in the Nipponbare gene (see Appendix Figure S.1). Additionally, only one intron was predicted for IR64 for this gene, whereas there are two present in Nipponbare. No homologs for this gene were identified in Koshihikari or Azucena.

IR64_h_Os12g11370: this gene was the only gene identified in IR64 that was identical to Nipponbare at the amino acid level (Figure 3.13). Although this gene was also detected in Koshihikari, identity was only 86.6% and, in contrast to both *S. hermonthica* resistant cultivars, was predicted to contain 3 introns. No homolog was identified in Azucena.

IR64_h_Os12g11500: homologs of Os12g11500 were found in all 4 cultivars investigated. When compared to Nipponbare, the IR64 gene sequence was almost identical over the first 800 amino acids, but this was followed by an early stop codon soon after. In both Azucena and Koshihikari, Os12g11500 homologs had over 100 amino acid differences compared to IR64 and Nipponbare, in addition to an early stop codon 52 amino acids from the end (Figure 3.14). Os12g11500 in Nipponbare was therefore the only gene to posses a transmembrane domain. The smaller IR64_h_Os11930 gene aligned to the last 207 amino acids of both Os12g11500 and Os12g11930, and differed by only 5 and 4 amino acid respectively over the length covered (Figure 3.15). Hence, IR64_h_Os11930 is only 1 amino acid more similar to Os12g11930 than it is to Os12g11500. IR64_h_Os11930 was predicted to contain a transmembrane domain.

IR64_h_Os12g11510: in contrast to Nipponbare, this homolog in IR64 was not predicted to have an intron, and contained a sequence of over 100 amino acids absent from the Nipponbare sequence (Appendix Figure S.2). Ten other single amino acid differences were present between the two cultivars for this gene. Neither gene was predicted to contain a transmembrane domain. Homologs of Os12g11510 were not detected in Azucena by PCR. The two shorter Koshihikari genes (Kosh_h_Os12g11940 and Kosh_h_Os12g12130) have the closest homology to IR64_h_Os12h11510 in IR64.

IR64_h_Os12g12010: homologs of Os12g12010 were identified in both IR64 and Azucena. Over 95% identity was shared between IR64, Nipponbare and Azucena, with the Azucena sequence having greater similarity to Nipponbare (99.8% identity) (Appendix Figure S.3) than it did to the IR64 sequence (95.5 % identity). When compared to Nipponbare, several of the amino acid differences between IR64_h_12010 were the same as those of Os12g11680, the Nipponbare homolog to Os12g12010, indicating this gene in IR64 is a combination of these two duplicated genes in Nipponbare (Appendix Figure S.4).

IR64_h_Os12g12000: a homolog of the Nipponbare Rapid Alkalization factor Os12g12000 was also present in IR64, Nipponbare and Azucena. Os12g12000 also had higher identity to the Azucena homolog (94.9 %) than to the IR64 homolog (90.7 %) (Appendix Figure S.5).

Table 3.6 summarises the length and % identity of all the genes predicted in the IR64 *S. hermonthica* resistance QTL that are also present in the Nipponbare QTL. Koshihikari genes are included where known.

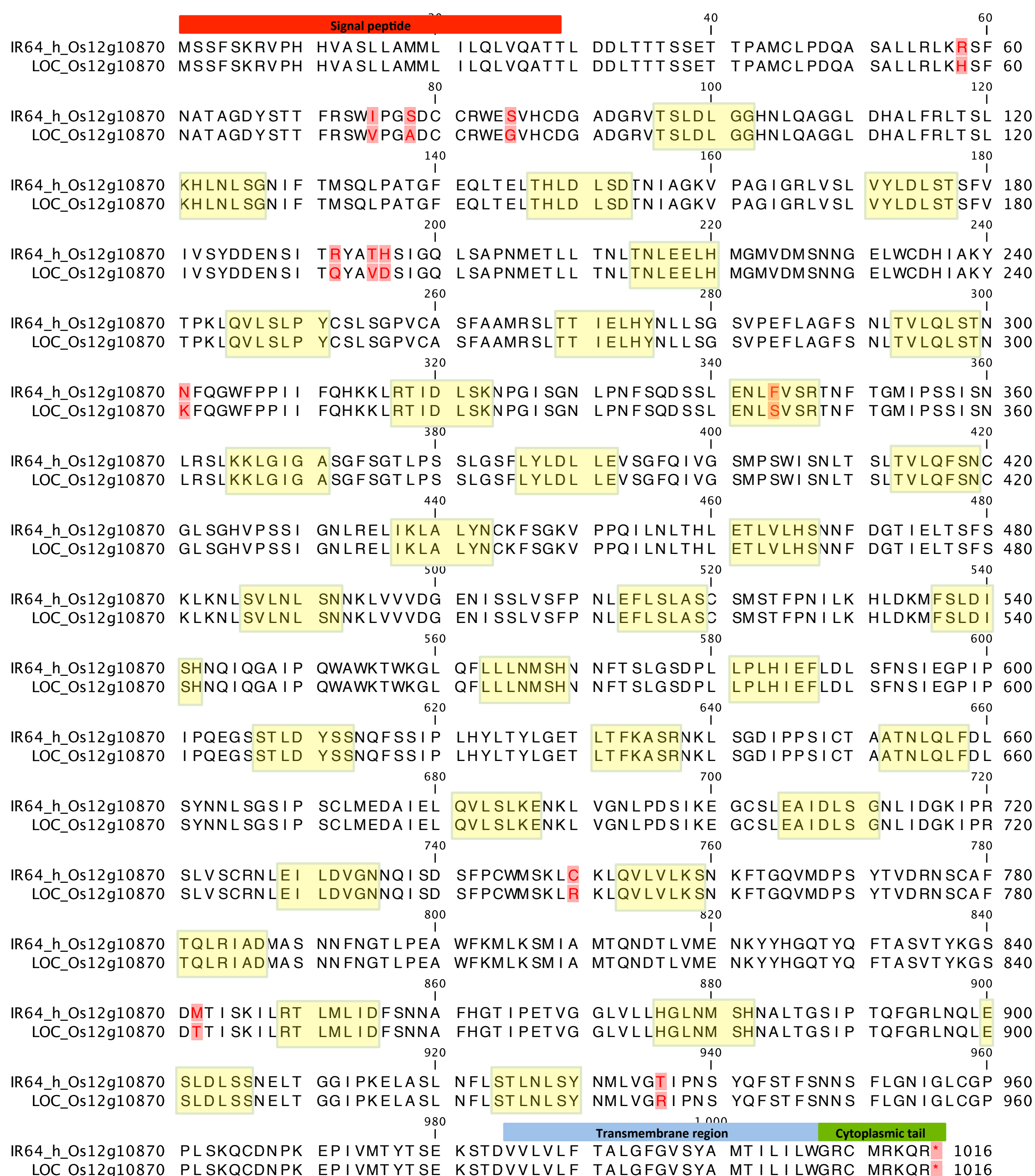


Figure 3.12 Amino acid alignment of Os12g10870 from Nipponbare and the IR64 homolog (IR64_h_Os12g10870) predicted by Fgenesh gene-finder (Softberry, Inc). Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding.

```
                   Signal peptide
IR64_h_Os12g11370  MSSSTKRLVR PHHLAKPLL- TMLHILLQVQ AIAALTDDAT APVIQCLPDQ ASALLRLKNS 59
LOC_Os12g11370     MSSSTKRLVR PHHLAKPLL- TMLHILLQVQ AIAALTDDAT APVIQCLPDQ ASALLRLKNS 59
Kosh_h_Os12g11370  MSSSTKRPVR PHHLAKPLLL TMLHILLQLK AITALTDDAT APVIQCLPDQ ASALLRLKHS 60

IR64_h_Os12g11370  FNKTAGGYST AFRSWITGTD CCHWDGVDCG GGEDGRVTSL VLGGHNLQAG SISPALFRLT 119
LOC_Os12g11370     FNKTAGGYST AFRSWITGTD CCHWDGVDCG GGEDGRVTSL VLGGHNLQAG SISPALFRLT 119
Kosh_h_Os12g11370  FNTTAGGYST TFRSWITGTD CCHWEGIHCS G-EDGRVTSL VLGGHNLQTT IVDPALFRLN 119

IR64_h_Os12g11370  SLRYLDISGN NFSMSQLPVT GFENLTELTH LDLSDTNIAG EVPAGIGSLV NLVYLDLSTS 179
LOC_Os12g11370     SLRYLDISGN NFSMSQLPVT GFENLTELTH LDLSDTNIAG EVPAGIGSLV NLVYLDLSTS 179
Kosh_h_Os12g11370  SLRYLDLSGN NFSMSQLPVT ---------- ---------- ---------- ---------- 139

IR64_h_Os12g11370  FYIIYYDDEN KMMPFASDNF WQLSVPNMET LLANLTNLEE LHMGMVDMSG NGERWCDDIA 239
LOC_Os12g11370     FYIIYYDDEN KMMPFASDNF WQLSVPNMET LLANLTNLEE LHMGMVDMSG NGERWCDDIA 239
Kosh_h_Os12g11370  ---------- -------DYF WQLSLPSMET LLANLTNLEE LHMGMVDMSG NGERWCDDVA 182

IR64_h_Os12g11370  KFTPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGSVPEFLAG FSNLTVLQLS 299
LOC_Os12g11370     KFTPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGSVPEFLAG FSNLTVLQLS 299
Kosh_h_Os12g11370  KFAPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGPVPEFLAG FSNLTVLQLS 242

IR64_h_Os12g11370  KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQDS KL---ENLLI SSTNFTGIIP 356
LOC_Os12g11370     KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQDS KL---ENLLI SSTNFTGIIP 356
Kosh_h_Os12g11370  KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQES SLKYLDSLEV SGLQLAGSMA 302

IR64_h_Os12g11370  SSISNLKSLT ---------- ---------K LDLGASGFSG MLPSSLGSLK YLDLLEVSGI 397
LOC_Os12g11370     SSISNLKSLT ---------- ---------K LDLGASGFSG MLPSSLGSLK YLDLLEVSGI 397
Kosh_h_Os12g11370  PWISNLTSLT VLKFSDCGLS GEIPSSIGKK LDLGASGFSG MLPSSLGSLK YLDSLEVSGL 362

IR64_h_Os12g11370  QLTGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 457
LOC_Os12g11370     QLTGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 457
Kosh_h_Os12g11370  QLAGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 422

IR64_h_Os12g11370  LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLHGENSSSL VPFPKIKLLR 517
LOC_Os12g11370     LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLHGENSSSL VPFPKIKLLR 517
Kosh_h_Os12g11370  LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLRGENSSSL VPFPKIKLLR 482

IR64_h_Os12g11370  LASCSISTFP NILKHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN ISHNNITSLG 577
LOC_Os12g11370     LASCSISTFP NILKHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN ISHNNITSLG 577
Kosh_h_Os12g11370  LASCSISTFP NILRHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN MSHNNITSLG 542

IR64_h_Os12g11370  SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSTY LGETFTFKAS 637
LOC_Os12g11370     SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSTY LGETFTFKAS 637
Kosh_h_Os12g11370  SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSAY LGQTFTFKAS 602

IR64_h_Os12g11370  KNKLSGNIPS ICSAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 697
LOC_Os12g11370     KNKLSGNIPS ICSAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 697
Kosh_h_Os12g11370  KNKLSGNIPS ICTAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 662

IR64_h_Os12g11370  KEGCALEAID LSGNLFEGRI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLALK 757
LOC_Os12g11370     KEGCALEAID LSGNLFEGRI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLALK 757
Kosh_h_Os12g11370  KEGCALEAID LSGNLFEGKI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLVLK 722

IR64_h_Os12g11370  SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISDNDTLV 817
LOC_Os12g11370     SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISDNDTLV 817
Kosh_h_Os12g11370  SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISENDTLV 782

IR64_h_Os12g11370  MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 877
LOC_Os12g11370     MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 877
Kosh_h_Os12g11370  MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 842

IR64_h_Os12g11370  NMSHNSLTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNTLVGRIP 937
LOC_Os12g11370     NMSHNSLTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNTLVGRIP 937
Kosh_h_Os12g11370  NMSHNELTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNALVGRVP 902

                                                                   Transmembrane region
IR64_h_Os12g11370  NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVLLV LFTALGFGVS 997
LOC_Os12g11370     NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVLLV LFTALGFGVS 997
Kosh_h_Os12g11370  NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVVLV LFTALGFGVS 962

                              Cytoplasmic tail
IR64_h_Os12g11370  FAITILIVWG RHMKKQR* 1015
LOC_Os12g11370     FAITILIVWG RHMKKQR* 1015
Kosh_h_Os12g11370  FAITILIVWG RHMKKKR* 980
```

Figure 3.13 Amino acid alignment of Os12g11370 from Nipponbare and the IR64 and Koshihikari homologs predicted by Fgenesh gene-finder (Softberry, Inc). Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding.

```
                      Signal peptide                                    40                     60
IR64_h_Os12g11500     M--------- ---------- YILLQVQATT NTARTVVPPV RCHPDQASAL LRLKHSFNAT  41
LOC_Os12g11500        MSSSTKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV RCHPDQASAL LRLKHSFNAT  60
Kosh_h_Os12g 11500    M--------- ---------- YILLQVQATT NTARTVVPPV PCHPDQASAL LRLKHSFNAT  41
Azucena_h_Os12g11500  MSSSAKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV PCHPDQASAL LRLKHSFNAT  60

                                 80                    100                    120
IR64_h_Os12g11500     AGDYSTAFQS WVAGTDCCRW DGVGCGG-AD GRVTSLDLGG HQLQAGSVDP ALFRLTSLKH 100
LOC_Os12g11500        AGDYSTAFQS WVAGTDCCRW DGVGCGG-AD GRVTSLDLGG HQLQAGSVDP ALFRLTSLKH 119
Kosh_h_Os12g 11500    AGDYSTAFRS WVAGTDCCRW DGVGCGGGAD GRVTSLDLGG HNLQAGSVDP ALFRLTSLKH 101
Azucena_h_Os12g11500  AGDYSTAFRS WVAGTDCCRW DGVGCGGGAD GRVTSLDLGG HNLQAGSVDP ALFRLTSLKH 120

                                140                    160                    180
IR64_h_Os12g11500     LNLSGNDFSM SQLPVITGFE QLTELVYLDL SDTNIAGEVP GSIGRLTNLV YLDLSTSFYI 160
LOC_Os12g11500        LNLSGNDFSM SQLPVITGFE QLTELVYLDL SDTNIAGEVP GSIGRLTNLV YLDLSTSFYI 179
Kosh_h_Os12g 11500    LNLSGNDFSM SQLPVITGFE QLTELVHLHL SDTNITGEVP GSIGRLTNLV YLDLSTSFYI 161
Azucena_h_Os12g11500  LNLSGNDFSM SQLPVITGFE QLTELVHLHL SDTNITGEVP GSIGRLTNLV YLDLSTSFYI 180

                                200                    220                    240
IR64_h_Os12g11500     VEYNDDEQVT FDSDSVWQLS APNMETLIEN LSNLEELHMG MVDLSGNGER WCDNIAKYTP 220
LOC_Os12g11500        VEYNDDEQVT FDSDSVWQLS APNMETLIEN FSNLEELHMG MVDLSGNGER WCDNIAKYTP 239
Kosh_h_Os12g 11500    VEYNDDEQVT FNSDSVWQLS APNMETLLEN LTNLEKLHMG MVDLSGNGER WCYNIAKYTP 221
Azucena_h_Os12g11500  VEYNDDEQVT FNSDSVWQLS APNMETLLEN LTNLEKLHMG MVDLSGNGER WCYNIAKYTP 240

                                260                    280                    300
IR64_h_Os12g11500     KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 280
LOC_Os12g11500        KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 299
Kosh_h_Os12g 11500    KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 281
Azucena_h_Os12g11500  KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 300

                                320                    340                    360
IR64_h_Os12g11500     QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLNNTNFTG TIPGSIINLI 340
LOC_Os12g11500        QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLNNTNFTG TIPGSIINLI 359
Kosh_h_Os12g 11500    QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLSNTNFTG TIPSSIINLI 341
Azucena_h_Os12g11500  QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLSNTNFTG TIPSSIINLI 360

                                380                    400                    420
IR64_h_Os12g11500     SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLTSL TVLRISNCGL 400
LOC_Os12g11500        SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLTSL TVLRISNCGL 419
Kosh_h_Os12g 11500    SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLSSL TVLRFSNCGL 401
Azucena_h_Os12g11500  SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLSSL TVLRFSNCGL 420

                                440                    460                    480
IR64_h_Os12g11500     SGPVPSSIGN LRELTTLALY NCNFSGTVHP QILNLTRLQT LLLHSNNFAG TVDLTSFSKL 460
LOC_Os12g11500        SGPVPSSIGN LRELTTLALY NCNFSGTVHP QILNLTRLQT LLLHSNNFAG TVDLTSFSKL 479
Kosh_h_Os12g 11500    SGQVPSSIGN LRELTTLALY NCNFSGTVPP QILNLTRLQT LLLHSNNFAG TVELTSFSKL 461
Azucena_h_Os12g11500  SGQVPSSIGN LRELTTLALY NCNFSGTVPP QILNLTRLQT LLLHSNNFAG TVELTSFSKL 480

                                500                    520                    540
IR64_h_Os12g11500     KNLTFLNLSN NKLLVVEGKN SSSLVLFPKL QLLSLASCSM TTFPNILRDL PDITSLDLSN 520
LOC_Os12g11500        KNLTFLNLSN NKLLVVEGKN SSSLVLFPKL QLLSLASCSM TTFPNILRDL PDITSLDLSN 539
Kosh_h_Os12g 11500    KNLTFLNLSN NKLLVVEGKN SSSLVSLHKL QLLSLASCSM TTFPNILRHL PEITSLDLSN 521
Azucena_h_Os12g11500  KNLTFLNLSN NKLLVVEGKN SSSLVSLHKL QLLSLASCSM TTFPNILRHL PEITSLDLSN 540

                                560                    580                    600
IR64_h_Os12g11500     NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LYVEYFDLSF NSIEGPIPIP 580
LOC_Os12g11500        NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LYVEYFDLSF NSIEGPIPIP 599
Kosh_h_Os12g 11500    NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LSVEYFDLSF NSIGGPIPIP 581
Azucena_h_Os12g11500  NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LSVEYFDLSF NSIGGPIPIP 600

                                620                    640                    660
IR64_h_Os12g11500     QEGSSTLDYS SNQFSSMPLR YSTYLGETVT FKASKNKLSG NVPPLICTTA HKLQLIDLSY 640
LOC_Os12g11500        QEGSSTLDYS SNQFSSMPLR YSTYLGETVT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 659
Kosh_h_Os12g 11500    QEGSSTLDYS SNQFSSMPLR YSTYLGETLT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 641
Azucena_h_Os12g11500  QEGSSTLDYS SNQFSSMPLR YSTYLGETLT FKASKNKTSG NVPPLICTTA RKLQLIDLSY 660

                                680                    700                    720
IR64_h_Os12g11500     NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 700
LOC_Os12g11500        NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 719
Kosh_h_Os12g 11500    NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 701
Azucena_h_Os12g11500  NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 720

                                740                    760                    780
IR64_h_Os12g11500     VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQISCEFPA 760
LOC_Os12g11500        VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQISCEFPA 779
Kosh_h_Os12g 11500    VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQNSCEFPS 761
Azucena_h_Os12g11500  VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQNSCEFPS 780

                                800                    820                    840
IR64_h_Os12g11500     LRIADMASNN LNGMLMEGWF NVTPRNLAQ- ---------- ---------- ---------- 789
LOC_Os12g11500        LRIADMASNN LNGMLMEGWF KMLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGNDR 839
Kosh_h_Os12g 11500    LRIADMASNN LNGMLMEGWF KKLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGKSM 821
Azucena_h_Os12g11500  LRIADMASNN LNGMLMEGWF KKLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGKSM 840

                                860                    880                    900
IR64_h_Os12g11500     -ISRLFCVL- ---------- ---------- ---------- NPCPGP---- -ARVHK---- 808
LOC_Os12g11500        TISKILRSLV LIDVSGNAFH GAIPDTIGEL VLLRGLNLSH NALTGPIPSQ FCRLDQLESL 899
Kosh_h_Os12g 11500    TFSKVLRTLV LIDFSNNAFH GTIPETIGEL ILLHGLNISH NALTGPIPPQ LGRLNQLESL 881
Azucena_h_Os12g11500  TFSKVLRTLV LIDFSNNAFH GTIPETIGEL ILLHGLNISH NALTGPIPPQ LGRLNQLESL 900

                                920                    940                    960
IR64_h_Os12g11500     ---------- ---------- -TTSNI---- ---------Q FQT*------ ---------- 818
LOC_Os12g11500        DLSFNELSGE IPKELASLNF LSTLNLSNNT LVGRIPDSYQ FSTFSNSSFL GNTGLCGLPL 959
Kosh_h_Os12g 11500    DLSSNKLSGK IPNELESLNF LSTLNLSYNV LVGRIPDSYQ FSTFSNSSFL GNTGLCGPPF 941
Azucena_h_Os12g11500  DLSSNKLSGK IPNELESLNF LSTLNLSYNV LVGRIPDSYQ FSTFSNSSFL GNTGLCGPPF 960
```

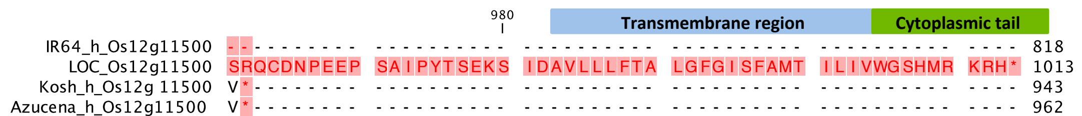


Figure 3.14 Multiple Amino acid alignment of Os12g11500 with IR64, Azucena and Koshihikari homologs. IR64 and Koshihkari sequences were predicted by Fgenesh gene-finder (Softberry, Inc), Azucena sequence was amplified by PCR and sequenced by Sanger sequencing . Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding. Only Os12g11500 was predicted to contain a transmembrane domain.

Figure 3.15 Amino acid alignment of IR64_h_Os12g11930 aligned to the end sequence of both Os12g11500 and Os12g11930 from Nipponbare. IR64 sequence was predicted by Fgenesh gene-finder (Softberry, Inc). Structural domains of the protein predicted by NCBI BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding.

3.3.6 Comparison of candidate *S. hermonthica* resistance gene in IR64 with RLP genes in different plant species

Receptor-like proteins often share conserved motifs between different plant species. These are thought to play a role in conformation, stability and possibly function of these proteins. The amino acid sequences of the RLP genes in IR64 were aligned against 5 other RLP genes identified in tomato; these were *Ve1*, which provides resistance to race 1 strains of *V. dahlia* and *V. albo-atrum*, and its close homolog *Ve2*; *Cf-4* and *Cf-9* which confer resistance to different races of *C. fulvum* through recognition of different avirulence factors (Thomas, 1997) and the recently identified *CuRe1*, essential for the perception of the stem holoparasite *C. reflexa* (Hegenauer *et al.*, 2016). Their evolutionary relationship is shown in Figure 3.16. The IR64 RLP genes were most closely related to *Ve1* and *Ve2*, and most distantly related to *CuRe1*.

Table 3.6 (overleaf). Gene comparisons for across the *S. hermonthica* resistance QTL between the annotated Nipponbare genes and those predicted within the IR64 QTL. Koshihikari gene predictions are included where known. Only genes present in both IR64 and Nipponbare QTL are shown. Transposable elements have been omitted. % identity is over the full length of the gene alignment.

Region	Query Gene model (Nipponbare)	start	stop	Annotation	Nipponbare Length (aa)	FgenesH IR64 Length (aa)	FgenesH IR64 % identity	FgenesH Koshihikari Length (aa)	FgenesH Koshihikari % identity
Top of IR64									
	LOC_Os12g10710.1	5746099	5752994	NB-ARC domain containing protein, expressed	1549	1470	89.26	1470	90.4
	LOC_Os12g10720.1	5756436	5760088	glutathione S-transferase, putative, expressed	215	197	91.63	197	91.63
	LOC_Os12g10730.2	5760466	5763724	glutathione S-transferase, putative, expressed	213	213	99.06	213	100
Top of Nipponbare mapping									
Nipponbare and IR64 QTL region	LOC_Os12g10740.1	5765954	5773383	leucine-rich repeat family protein, putative, expressed	968	855	82.77	968	90.71-100?
Nipponbare and IR64 QTL region	LOC_Os12g10750.1	5775600	5776808	ARGOS, putative, expressed	127	174	69.54	?	?
Nipponbare and IR64 QTL region	LOC_Os12g10760.1	5786022	5786569	expressed protein	173	173	100	?	?
Nipponbare and IR64 QTL region	LOC_Os12g10810.1	5817169	5819154	expressed protein	271	671	33.83	300	61.91
						408	51.87	189	55.6
Nipponbare and IR64 QTL region	LOC_Os12g10820.1	5822599	5829161	expressed protein	422	see 10810 above		?	?
Nipponbare and IR64 QTL region	LOC_Os12g10850.1	5849759	5852388	hhH-GPD superfamily base excision DNA repair protein, putative, expressed	475	466	96.63	?	?
Nipponbare and IR64 QTL region	LOC_Os12g10870.1	5856217	5859264	verticillium wilt disease resistance protein, putative, expressed	1016	1016	98.82	?	?
Nipponbare and IR64 QTL region	LOC_Os12g10880.1	5866457	5866925	expressed protein	64	69	84.06	?	?
Nipponbare and IR64 QTL region	LOC_Os12g10910.1	5901106	5902772	A/G-specific adenine DNA glycosylase, putative, expressed	216	39.1	40.38	?	?
Nipponbare and IR64 QTL region	LOC_Os12g10930.1	5904580	5907620	NL0E, putative, expressed	749	979	60.04	-	-
Nipponbare and IR64 QTL region	LOC_Os12g11370.1	6138570	6142158	verticillium wilt disease resistance protein, putative, expressed	1015	1015	100	980	86.61
Nipponbare and IR64 QTL region	LOC_Os12g11400.1	6156332	6156661	hypothetical protein	110	110	98.18	-	-
Nipponbare and IR64 QTL region	LOC_Os12g11420.1	6164435	6164653	expressed protein	73	88	52.27	-	-
Nipponbare and IR64 QTL region	LOC_Os12g11500.1 IRGSP	6224244	6227225	resistance protein SlVe1 precursor, putative, expressed	1013	818	78.18	943	87.28
Nipponbare and IR64 QTL region	LOC_Os12g11510.1	6231368	6235528	hcr2-0B, putative, expressed	829	911	87.94	-	-
Nipponbare and IR64 QTL region	LOC_Os12g11550.1	6259614	6259823	hypothetical protein	70	-	-	70	97.14
Nipponbare specific region	LOC_Os12g11660.1	6324994	6325718	RALFL45 - Rapid ALkalinization Factor RALF family protein precursor, expressed	99	-	-	-	-
Nipponbare specific region	LOC_Os12g11680.1 IRGSP	6340808	6344089	verticillium wilt disease resistance protein precursor, putative, expressed	1000	-	-	-	-
Nipponbare specific region	LOC_Os12g11720.1	6356947	6360006	verticillium wilt disease resistance protein precursor, putative, expressed	1020	-	-	-	-
Nipponbare specific region	LOC_Os12g11860.1	6447258	6450275	verticillium wilt disease resistance protein precursor, putative, expressed	1006	-	-	-	-
Nipponbare specific region	LOC_Os12g11930.1	6506509	6509781	disease resistance protein SlVe2 precursor, putative, expressed	1016	208	20.08	-	-
Nipponbare specific region	LOC_Os12g11940.1	6517869	6520549	disease resistance family protein, putative, expressed	855	-	-	358	38.83
Nipponbare and IR64 QTL region	LOC_Os12g11950.1	6525869	6526486	expressed protein	158	-	-	158	97.47
Nipponbare and IR64 QTL region	LOC_Os12g11980.1	6563886	6564092	expressed protein	69	70	88.57	69	100
Nipponbare and IR64 QTL region	LOC_Os12g11990.1	6567115	6567660	expressed protein	69	69	97.1	69	100
Nipponbare and IR64 QTL region	LOC_Os12g12000.1	6576304	6577029	RALFL46 - Rapid ALkalinization Factor RALF family protein precursor, expressed	97	95	90.72	-	-
Nipponbare and IR64 QTL region	LOC_Os12g12010.1	6583346	6586345	verticillium wilt disease resistance protein precursor, putative, expressed	1000	978	95.7	-	-
Nipponbare and IR64 QTL region	LOC_Os12g12120.1	6647316	6650612	verticillium wilt disease resistance protein precursor, putative, expressed	1006	-	-	-	-
Nipponbare and IR64 QTL region	LOC_Os12g12130.1	6650911	6654239	verticillium wilt disease resistance protein, putative, expressed	1026	-	-	517	39.48
Bottom of Nipponbare mapping									
	LOC_Os12g12170	6677657	6679930	cytochrome b5-like Heme/Steroid binding domain containing protein, expressed	151	151	98.68		
Bottom of IR64 mapping									
	LOC_Os12g12260.2	6729628	6734373	diacylglycerol kinase 1, putative, expressed	664	820	80.49		

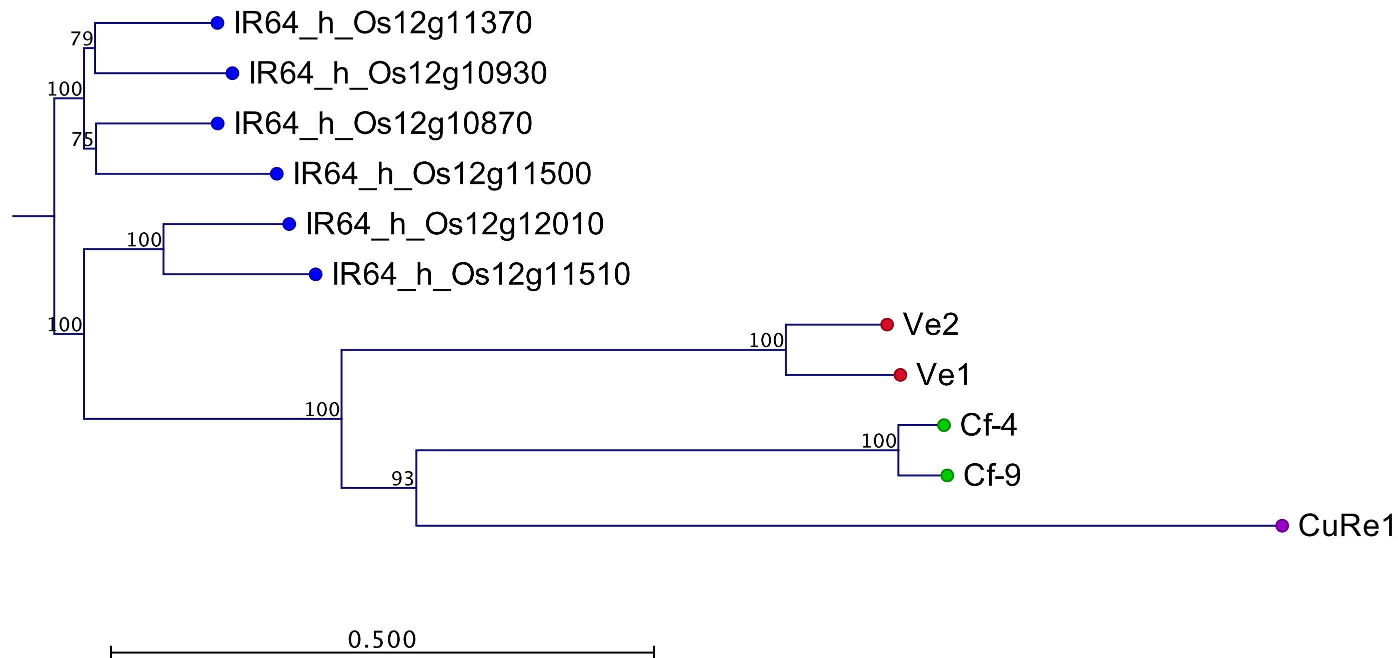


Figure 3.16 Phylogenetic relationship of RLP genes in the S. *hermonthica* resistance QTL from the rice cultivar IR64 and 5 homologous RLP genes identified in tomato, involved in resistance to *Verticillium* wilt (*Ve1* and *Ve2*) *Cladosporum fulvum* (Cf-4 and Cf-9), and the stem parasitic plant *Cuscuta reflexa* (*CuRe1*). Neighbor Joining built using 1,000 bootstrap replicates. Branch lengths are proportional to phylogenetic distances. Numbers at nodes are bootstrap confidence values.

3.3.7 The phenotype of resistance during *S. hermonthica* infection on IR64

In the resistance interaction on IR64, *S. hermonthica* successfully penetrated the root and had grown into the cortex by 2 dai (Figure 3.17, 2 dai). However, it failed to penetrate the endodermis, and by 4 dai was seen growing through the cortex and around the vascular core. In contrast, in a susceptible interaction *S. hermonthica* had breached the endodermis at 4 dai and begun to make vascular connections (Figure 3.17, 4 dai). The *S. hermonthica* radicle continued to grow through the IR64 cortex until 8 dai, but at 10 dai the radicle had started to die. On Azucena, vascular connection were fully established at 6 dai, allowing the parasite to grow and develop rapidly thereafter (Figure 3.17).

3.3.8 Gene expression analysis of candidate *S. hermonthica* resistance on infected and uninfected IR64 root tissue

The expression of all candidate *S. hermonthica* resistance RLP genes and the IR64 homolog of Os12g12000 (RALF46) was measured in IR64 *S. hermonthicaa*-infected and uninfected root tissue by qPCR at 3 time points 2, 4 and 10 days after inoculation (Figure 3.18). Transcripts were detected for all genes measured, however for IR64_h_Os12g11510 this was barely detectable and showed poor and variable amplification of the product. Transcript levels for all genes were normalised against the uninfected tissue following normalisation relative to rice presenilin (Os01g16930), to give a relative gene expression value in uninfected tissue of 1. There was very little difference in relative gene expression between infected and uninfected tissue 2 dai with *S. hermonthica* for all genes investigated. This was also the case at 4 dai; IR64_h_Os12g12000 showed a slight increase in transcript (Figure 3.18F) but large standard error was observed at this time point. At 10 dai, both IR64_h_Os12g10870 and IR64_h_Os12g11370 were upregulated 1.7 and 1.9 fold compared to uninfected tissues respectively (Figures 3.18A and C), but this was only statistically significant for IR64_h_Os12g11370 (two-sample t-test, $t = 7.35$, $df = 3.19$, $p = 0.004$). No other significant differences were observed for any gene assessed at any time point (Figure 3.18).

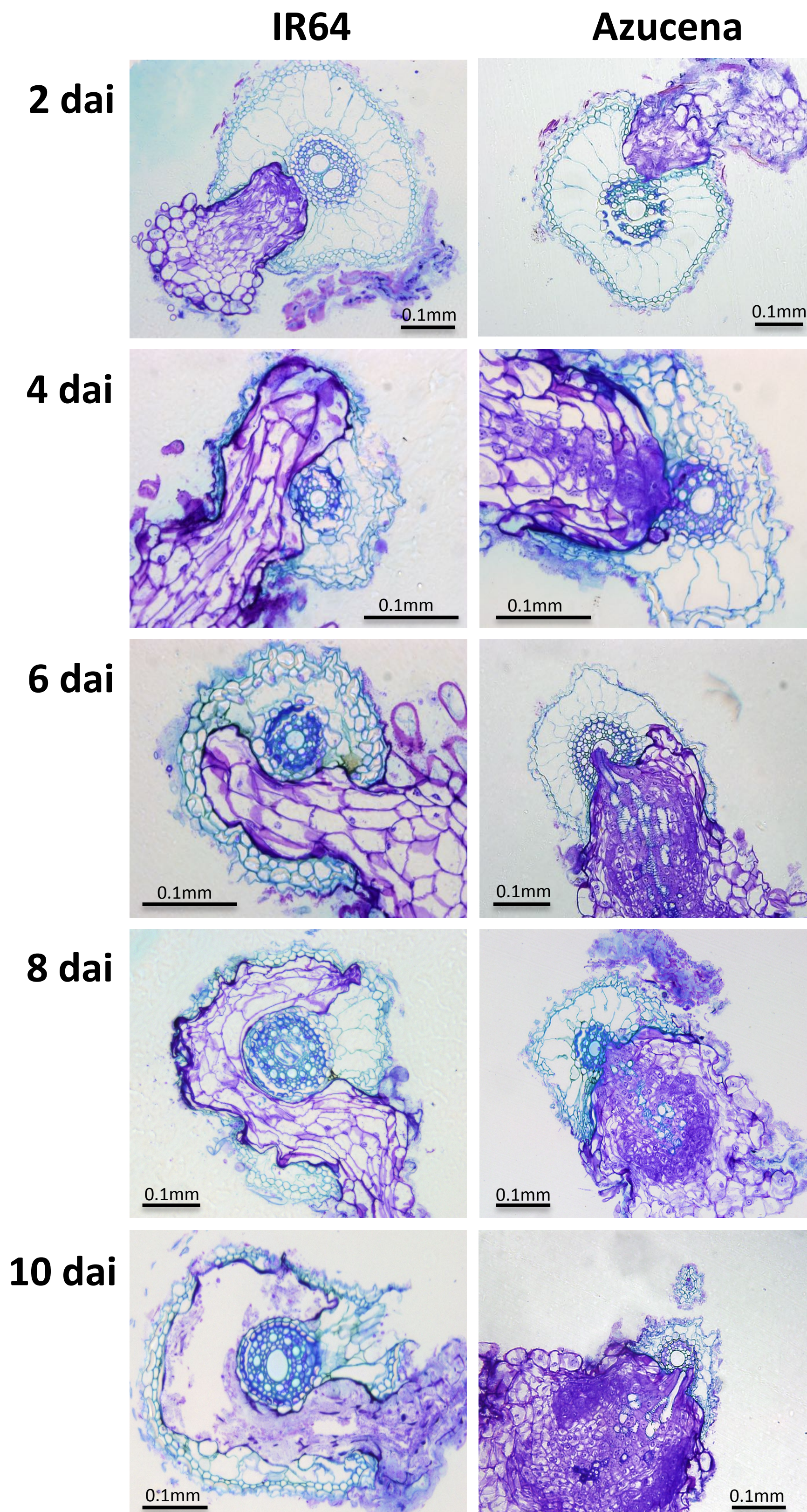


Figure 3.17 Time course of the phenotype of infection of *S. hermonthica* on the roots of resistant rice cultivar IR64 and the susceptible cultivar Azucena at 2, 4, 6, 8 and 10 dai with germinated *S. hermonthica* seed. (Kibos ecotype). IR64: resistance is shown by a failure of the parasite to breach the endodermis and attach to xylem vessels, instead growing through the cortex and around the vascular core. Azucena: by 4 dai the parasite has breached the endodermis and started to make vascular connections to the xylem. The parasite then develops rapidly.

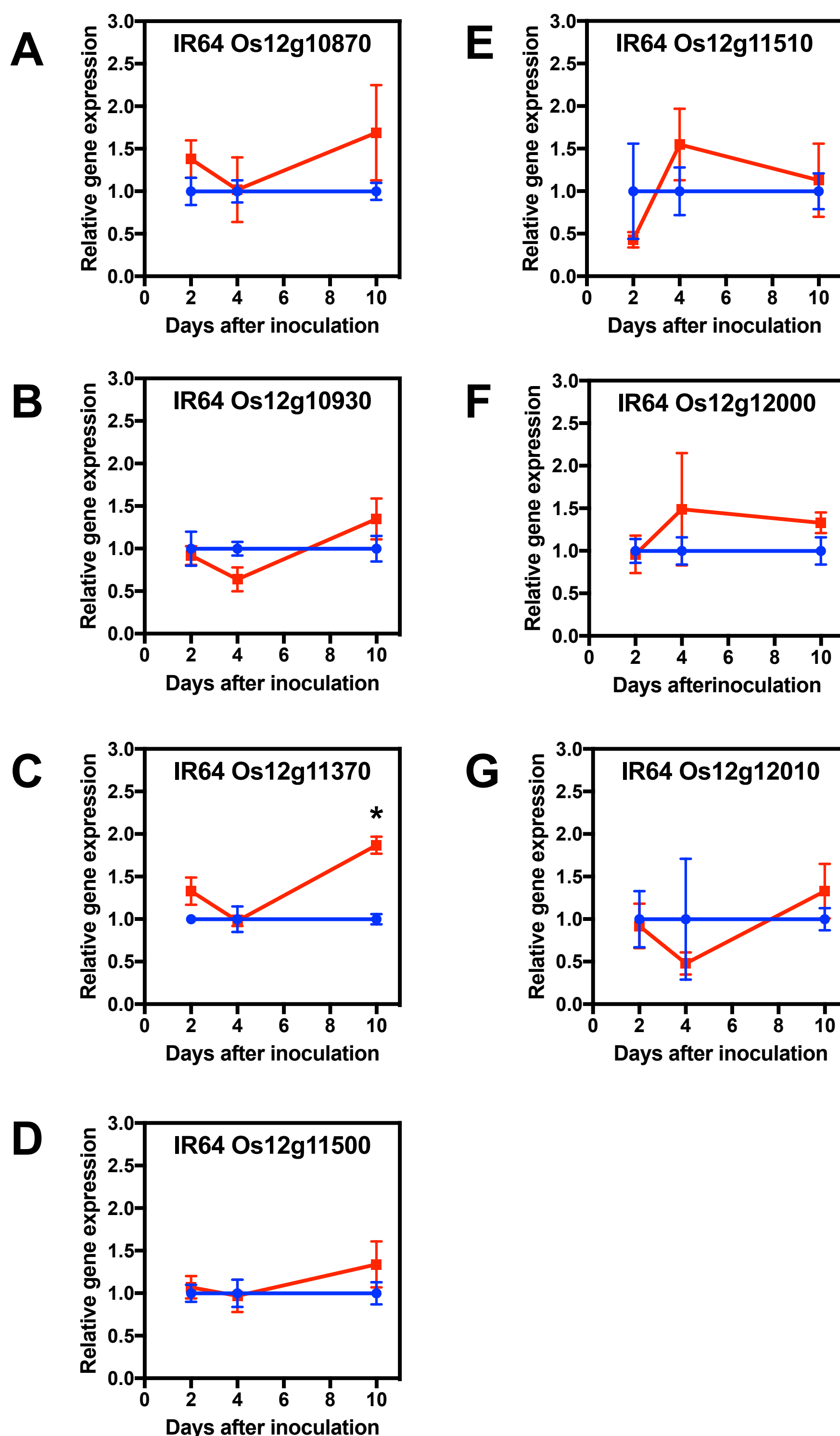


Figure 3.18 Time course of relative gene expression quantified by qPCR of candidate *S. hermonthica* resistance genes in *S. hermonthica* infected and uninfected root tissue of the resistant rice cultivar IR64. Red = infected tissue, blue = uninfected. Expression levels are normalised to rice presenilin (Os01g16930), and plotted relative to uninfected tissue at each time point. Error bars are standard error where n = 3 biological replicates, each consisting of the combined tissue from 2 individual plants. Star represents a significant difference between infected and uninfected tissue (two-sample t-test, $p < 0.05$).

3.4 Discussion

The aim of this chapter was to identify candidate resistance genes in the IR64 *S. hermonthica* resistance QTL, to test the hypothesis that the same or similar gene(s) may be present in the previously identified *S. hermonthica* resistance QTL in the rice cultivar Nipponbare. Gene prediction software and BLAST analysis identified a number of highly similar genes between the two cultivars, which were confirmed by sequencing. Only one gene (Os12g11370) was identical at the amino acid level between the two cultivars. Six genes identified within the IR64 QTL are orthologs of receptor-like proteins (RLPs) in other plant species, and BLAST analysis identified them as orthologs of *Verticillium* wilt disease resistance proteins. Given the highly similar lifestyle and infection strategy between *Verticillium* wilts and *Striga* species, it seems plausible that a *Verticillium* wilt resistance ortholog in rice may be involved in resistance to *S. hermonthica*. The recent discovery that the RLP *CuRe1* in tomato is sufficient to confer increased resistance to the holoparasite *C. reflexa* (Hegenauer *et al.*, 2016) is additional evidence that these genes are top candidates for involvement in the resistance response observed in IR64 and Nipponbare to *S. hermonthica.*

Receptor-like proteins are an important type of cell surface receptor in plants, and are known to play a role in both developmental processes and disease resistance (Kruijt *et al.*, 2005; Wang *et al.*, 2010). Genes involved in development are under evolutionary pressure to maintain a particular function and tend to be conserved (Fritz-Laylin *et al.*, 2005) while resistance genes are exposed to strong diversifying selection and often have multiple copies (Bai *et al.*, 2002; Meyers *et al.*, 2003), making their identification challenging. The protein structure of an RLP typically consist of a seven domains termed A to G, illustrated in Figure 3.19: a signal peptide (A), a cysteine rich B domain, the eLRR domain (C) that mediates protein-protein interactions, a variable domain (D), an acidic domain (E), a transmembrane helix (F) which localises the protein the plasma membrane, and a short cytoplasmic tail (G) that lacks a kinase domain. The eLLR domain can be further divided into 3 sub-domains: the C1 eLLR domain, C2 non-LRR island domain and the C3 eLLR region, although the C2 domain is not present in all RLPs (Wang *et al.*, 2010).

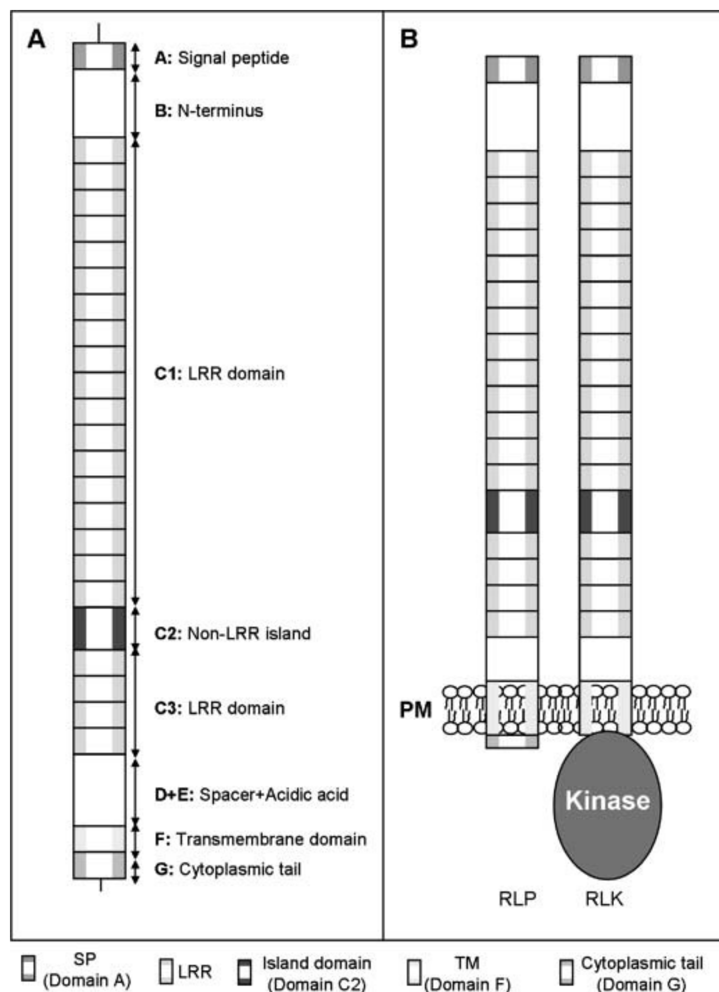


Figure 3.19 Domain structure of a typical receptor-like protein (RLP) (A) and comparison with a receptor-like kinase (B). Diagram from Wang *et al.*, 2010.

Developmental RLP genes include Too Many Mouths (TMM) which regulates stomatal distribution in *Arabidopsis* and CLAVATA2 (CLV2) which is involved in maintenance of the meristem (Wang *et al.*, 2010). RLPs involved in disease resistance include *Ve1* which provides resistance against race 1 strains of *V. dahlia* and *V. albo-atrum* in tomato (Fradin *et al.*, 2009), *HcrVf-2* which provide resistance to apple scab fungus *Venturia inaequalis* (Belfanti *et al.*, 2004), and the *Cf* genes which govern resistance to various strains of *Cladosporium fulvum* in tomato (Wang *et al.*, 2010). The ability of RLPs to perceive the presence of an attacker and activate downstream signalling is crucial for the initiation of a defence response (Yang *et al.*, 2012a). Unlike receptor-like kinases (RLKs), RLPs lack an internal kinase domain, and instead form an association with the common adaptor kinase SOBIR1 (suppressor of BAK1-interacting receptor kinase) for functionality and downstream signalling (Liebrand *et al.*, 2013). Indeed, SlSOBIR1 from *S. lycopersicum* showed constitutive interaction with the *CuRe1* receptor both

in the presence and absence of the *Cuscuta* factor stimulus (Hegenauer *et al.*, 2016). Thus, the presence of a SOBIR1 rice ortholog was also identified in rice. Unsurprisingly, this protein was highly conserved between all rice cultivars tested, showing 97.5 % identity between IR64 and Nipponbare /Koshihikari, and 100 % identity between Koshihikari and Nipponbare, indicative of its role in important biological processes. This gene would therefore be a good target for further analysis of the function of genes conferring resistance to *S. hermonthica*.

Six RLP genes were identified in the IR64 *S. hermonthica* resistance QTL, compared to 13 present in Nipponbare. Phylogenetic analysis indicates that Nipponbare has undergone a duplication event; 4 of the RLP genes are very closely related to a second RLP, with one of each of the duplicated genes occurring outside the 'Nipponbare-specific' region. This duplication event clearly took place after the divergence of Nipponbare from both IR64 and Koshihikari, as this Nipponbare specific region is absent from the genomes of both these cultivars (Figure 3.2). The IR64 RLPs are more similar to RLPs outside the Nipponbare specific region (Figure 3.8), suggesting this insertion in Nipponbare is no more likely to contain a resistance gene or genes than those outside the region, despite this region being absent from the Koshihikari genome.

Clusters of resistance genes are a well known phenomenon in plant genomes (Meyers *et al.*, 2005; McDowell & Simon, 2006), with usually one gene in the cluster providing resistance to a particular pathogen (Michelmore *et al.*, 1998; Kruijt *et al.*, 2005). For example, the *Cf-5* cluster of *Lycopersicon esculentum* var. *cerasiforme*, the wild relative of tomato, contains 4 closely related RLP genes, of which only *Hcr2-5C* provides resistance against *C. fulvum* (Dixon *et al.*, 1998). However, this is not always the case and occasionally two independent genes may be functionally equivalent; the *Cf-2* locus in tomato comprises two *Cf-2* genes, both of which provide resistance to strains of *C. fulvum* expressing the corresponding *Avr2* gene (Dixon *et al.*, 1996). Nevertheless, the presence of a highly conserved gene in both our *S. hermonthica* resistant cultivars, and its absence or lack of close identity with susceptible cultivars would be a good indicator for a possible resistance gene. Therefore to help identify the most likely resistance gene or genes in the cluster, the amino acid sequences of all RLP genes identified in IR64 were closely compared with those of Nipponbare and, the where sequences were available, with the *S. hermonthica* susceptible cultivars Koshihikari and Azucena. The two RLPs in IR64 sharing the highest amino acid identity with Nipponbare were IR64_h_Os12g10870 (98.8 %) and IR64_h_Os12g11370 (100 %). Neither gene was successfully amplified from Azucena (using primers designed against the Nipponbare or IR64 gene sequences) while only one of these genes (Kosh_h_Os12g11370) was amplified from Koshihikari. The sequence of this gene in Koshihikari shared only 86.6% identity with the IR64 and Nipponbare gene at the

amino avid level. Thus these two genes are top candidates for *S. hermonthica* resistance. Given the estimated divergence of *indica* (IR64) and *japonica* (Nipponbare) cultivars 0.4 mya (Zhu & Ge, 2005), it is surprising that Os12g11370 is identical at the amino acid level in both IR64 and Nipponbare. This suggests the gene is under evolutionary pressure not to change, which can be interpreted in two different ways: that the gene is important in a developmental process or basic cell function, or that it provides a unique and important form of resistance to the plant. The absence or lack of close identity of this gene in susceptible cultivars may suggest the latter, although it is unlikely that this gene evolved to protect against *Striga* species. Homologs of Os12g11500 were found in all cultivars tested. While the first 818 amino acids of this homolog in IR64 showed over 98% identity with Nipponbare, the last 205 amino acids were missing from the end of the IR64 version, meaning this protein in IR64 lacks a transmembrane domain and so likely localises in the cytoplasm (Bleckmann *et al.*, 2010). A smaller gene (IR64_h_Os12g11930) was also identified in IR64 over 9 Kb downstream of IR64_h_Os12g11500, and the sequence of this gene aligns to the end of Os12g11500 that is missing from IR64_h_Os12g11500. However it seems unlikely that these genes could function in the same way as Os12g11500 in Nipponbare. Interestingly, the gene model predicted using Maker2 (Schatz lab) for this gene in IR64 differed with that of Fgenesh prediction used in this study. The Maker2 annotation predicted a single gene of the same length as Os12g11500 in IR64, containing a 9 Kb intron. It is unclear which of these predictions is correct; it is also possible they could be alternatively spliced variants. Both the susceptible cultivars Koshihikari and Azucena contained a homolog of Os12g11500 that were almost identical with each other. Interestingly, where amino acids differed from IR64 or Nipponbare, these differences were often the same for Koshihikari and Azucena (see Figure 3.14). However the gene in these cultivars also has an early stop codon when compared to Nipponbare, which also means the protein lacks a transmembrane domain. It is therefore unclear how important this gene could be as a candidate for *S. hermonthica* resistance. The IR64_h_Os12g11510 amino acid sequence is clearly different from the Nipponbare homolog, and neither gene was predicted to contain a transmembrane domain (Appendix Figure S.2), suggesting they do not anchor to the membrane and thus do not function as RLPs. The closest Koshihikari homologs for this gene (Kosh_h_Os12g11940 and Kosh_h_Os12g12130) are much shorter in sequence, suggesting these genes in Koshihikari are unlikely to be functional. Interestingly, a homolog for Os12g12010 was sequenced from Azucena that differed by only 2 amino acids, while the IR64 homolog had 43 amino acid differences (Appendix Figures S.3 and S.4). Therefore unless these two mutations in Azucena underlie the basis susceptibility, it seems unlikely that this gene is involved in resistance to *S. hermonthica*. The RLP gene most different between IR64 and Nipponbare was Os12g10930. While it seems unlikely that this gene shares functionality

between the two cultivars, we cannot rule out the possibility that two genes may be acting independently to provide the same form of resistance.

Although comparisons of gene sequences between resistant and susceptible genotypes may help to identify a candidate resistance gene, interpretations must be taken cautiously, as not all amino acids are functionally important. Ligand specificity of most eLRR receptors is determined in the eLRR C1 domain (Zhang & Thomma, 2013). Zhang *et al.*, (2014) used alanine scanning on the solvent exposed residues across the eLRR domain of *Ve1*, and showed that functionality of the protein requires the presence of several eLRRs in the C1 and C3 domains, as well as the non-eLRR C2 domain. Similarly, recognitional specificity of *Cf-9* is determined by only five amino acid residues on the solvent-exposed central LRR, which occupy hypervariable positions across global *Cf* alignments. These positions also correspond to the residues determining *Cf-4* specificity (Wulff *et al.*, 2009). However, the strength of a resistance response was affected by the presence of other LRRs in the protein; amino acids in LRRs 13 to 15 were required for recognition of *Avr9* by *Cf-9*, but upstream residues in LRRs 10 to 12 contributed to the severity of the necrosis (Chakrabarti *et al.*, 2009). All this suggests that two genes, which may differ over much of their sequence, can share a common function if they also share important specificity-determining residues.

Analogous to *Striga*, resistance to *Verticillium* wilt conferred by *Ve1* is never complete, and even the most resistant plants still support low levels of proliferation of race 1 strains (Fradin *et al.*, 2011). *Ve1* was also found to provide resistance to *Fusarium oxysporum*, another fungal pathogen on tomato, which is compatible with resistance to *Striga* against a genetically diverse seed population. This form of resistance, that is broader spectrum and weaker than a race specific gene-for-gene interaction, supports the hypothesis that *Ve1* behaves more like a PRR than a typical race-specific R protein, confirming the idea of a continuum between the two. *Ve1* was shown to remain fully functional after interfamily transfer from tomato to *Arabidopsis*, suggesting direct pathogen-receptor recognition (Fradin *et al.*, 2011). If resistance to *S. hermonthica* is conferred by a similar RLP gene, this has important implications for crop breeding, as resistance may be easily transferred not only between different rice cultivars, but also between cereals.

Resistance genes are often up regulated in response to pathogen infection. However, very little change in gene expression was observed in IR64 between *S. hermonthica* infected vs. uninfected root tissue for any gene investigated here; only IR64_h_Os12g11370 was significantly up regulated at 10 dai. This is in contrast to *Ve1* and *Ve2* expression in tomato,

where both genes were up regulated in response to infection with *V. dahlia,* despite only *Ve1* conferring resistance. However up-regulation of *Ve1* was faster than that of *Ve2* (Fradin *et al.*, 2009). A resistance response to S*. hermonthica* infection in rice is seen by 4 dai, when the parasite fails to penetrate the endodermis of a resistant genotype, but is just beginning to make vascular connections on a susceptible genotype (Figure 3.17). Therefore one would expect any up regulation of resistance gene expression to occur before or around this time point. Indeed, Os12g11370 is upregulated in Nipponbare 4 dai with *S. hermonthica* (Scholes, unpublished data). There are several explanations why this was not observed in IR64. Firstly, alterations in gene expression may be local and cell-specific, occurring only in the cells around the infection site. As whole root tissue was used for analysis here, any changes in gene expression may be diluted and so not observed. Laser capture microdissection has now been used successfully to examine changes in gene expression between host-parasite interface in a cell specific manner (Honaas *et al.*, 2013), and would therefore be the method of choice for any further experiments of this kind. An alternative explanation is that the resistance gene or genes are not transcriptionally regulated, and that basal expression levels may be sufficient to confer resistance to *S. hermonthica*. However, it is also possible that although the RLP genes identified in IR64 appear to be good candidates for resistance to *S. hermonthica*, they may not actually play a role in *S. hermonthica* resistance.

Other genes were identified within the IR64 QTL region, some of which share high levels of similarity with genes in the Nipponbare QTL. The Rapid Alkalization Factor RALF46 protein is present in both cultivars. The role of RALF peptides in plants is not fully understood, but they are found throughout plant kingdom, often highly conserved, and are thought to be involved in essential processes such as regulation of cell expansion during growth and development (Pearce *et al.*, 2001; Murphy & De Smet, 2014). They are also known to act as negative regulators of pollen tube elongation (Covey *et al.*, 2010). Studies on tomato and poplar suggest RALFs are unlikely to be involved in defence, due to the lack of induction following treatment with elicitors and failure to induce a defence responses (Pearce *et al.*, 2001; Haruta, 2003). However in chickpea, attack by *Fusarium oxysporum* triggered increased RALF expression in resistant plants but not in susceptible plants. A model was proposed whereby the RALF may act as a “decoy”, mimicking the target of the pathogen effector, competing for effector-binding and thus lowering effector load. However, there is currently no evidence to support this hypothesis (Gupta *et al.*, 2010). In *Medicago truncatula* roots, the RALF *MtRALFL1* is up-regulated in response to Nod factors (Combier *et al.*, 2008), signalling molecules involved in the initial stages of nodule formation in leguminous plants (Oldroyd *et al.*, 2009). Overexpression of *MtRALF1* resulted in the formation of fewer nodules (Combier *et al.*, 2008).

Plants initially perceive rhizobia as alien invaders and mount a defence response (Zamioudis & Pieterse, 2011). It is therefore tempting to speculate a role for RALF in *S. hermonthica* resistance that may act in a similar manner to the repression of nodule formation observed here. Interestingly, RALF homologs are also known to be present in the genomes of a diverse range of plant-pathogenic fungi species (Thynne *et al.*, 2016). Synthesis of a RALF peptide from *Fusarium oxysporum* f. sp. *lycopersici*, a fungal pathogen of tomato, and treatment of this peptide on tomato seedlings, arrested growth and development of seedlings in a manner typical of endogenous RALF peptides, indicating fungal RALF peptides are perceived by plants. The RALF-encoding gene was expressed during infection on tomato, leading to the suggestion that RALFs may act as fungal effectors. However, RALF expression in *F. oxysporum* was not required for infection. The role of RALF peptides in plant-pathogenic interactions therefore remains to be determined (Thynne *et al.*, 2016). In this chapter, a RALF homolog was also sequenced from Azucena, which was more similar to Nipponbare than the IR64 homolog was. Additionally, no significant up regulation was observed in IR64 infected tissue. Taken together, this suggests that the RALF gene in IR64 is unlikely to provide resistance to *S. hermonthica*. However, it cannot be ruled out entirely given the diverse role that RALFs are known to play, and how little is currently known about their function.

Also in the *S. hermonthica* resistance QTL were ARGOS, a gene known to increase the size of organs such as leaves by increasing both cell number and cell size (Wang *et al.*, 2009), a DNA glycosylase and a DNA repair protein, which are involved in recognising damaged bases and repairing DNA (Bruner *et al.*, 2000; Jacobs & Schär, 2012), and two glutathione S-transferase genes, involved in the detoxification of xenobiotic chemicals (Edwards *et al.*, 2000). Given their annotation, it is unclear how these genes might be involved in resistance to *S. hermonthica*. Additionally, the putative glutathione S-transferases in Koshihikari each share 100 % identity with either IR64 or Nipponbare. Another potential gene of interest is the LRR containing protein Os12g10740, which shares 82 % identity between IR64 and Nipponbare. The available sequence for this gene in Koshihikari is incomplete, with 90 nucleotides missing. However, all other nucleotides are an exact match with Os12g10740 of Nipponbare, making this gene 90-100 % identical. Without knowing the exact sequence of this gene in the susceptible cultivar it cannot be ruled out as a candidate for *S. hermonthica* resistance, especially given the prevalence of LRRs found in resistance proteins (Meyers et al., 2003).

The QTL region for both IR64 and Nipponbare, and equivalent region in Koshihikari, also contains large numbers of expressed proteins of unknown function (Figure 3.3). Several of the expressed proteins are identical between Koshihikari and Nipponbare, or are absent or have

poor identity between resistant cultivars, making them unlikely resistance gene candidates. This is not always the case however, and so they cannot as yet be completely discounted.

Finally, both IR64 and Nipponbare possess large numbers of transposable elements (TEs) in their QTL regions; 31 retrotransposons (class I TEs) and 11 DNA transposons (class II TEs) are present in the IR64 QTL, while 51 retrotransposons and 21 DNA transposons are present in the Nipponbare QTL. Class I TEs use reverse transcriptase to transpose through a 'copy-and-paste' mechanism, and contribute to the expansion in genome size, while class II TEs transpose through a 'cut-and-paste' mechanism, and are often associated with genic regions, meaning they have had closer interactions with plant genes (Dooner & Weil, 2007; Song & Cao, 2017). Both classes of TEs are known to have had a substantial impact on plant genomes, and be a major driving force for gene and genome evolution. Their mobility can induce mutations in genes, as well as chromosomal rearrangements such as deletions, translocations and inversions (Dooner & Weil, 2007; Chenais et al., 2012). Indeed, the exceptional genetic variability of the QTL region in terms of structural rearrangements (Figure 3.2), duplications and SNPs, observed between IR64, Nipponbare and Koshihikari, are likely to be a result of the activity of TEs on the rice genomes. Rice is known to harbour large numbers of TEs scattered across its genome (Song & Cao, 2017). The mobilisation of TE can also disrupt genome stability, and insertion into a gene or a gene regulatory element can disrupt the function of the gene (Chenais *et al.*, 2012). This highly mutagenic nature of TEs means that their activity is usually repressed by the host genome through a variety of epigenetic mechanisms, including DNA methylation, histone modifications and regulation by small RNAs (Slotkin & Martienssen, 2007). However, TEs are also known to modulate the expression of nearby genes themselves, by inducing heritable epigenetic changes that contribute to phenotypic variation and adaptation (Lippman *et al.*, 2004; Huettel *et al.*, 2006; Kinoshita *et al.*, 2006). It is therefore possible that the large number of TEs present in the *S. hermonthica* resistance QTL could influence the expression of the nearby genes through epigenetic changes. These may differ between cultivars, adding a further layer of complexity to the resistance phenotype.

3.4.1 Conclusion

As mentioned previously, the comparison of gene sequences between genotypes to help identify candidate resistance genes must be taken cautiously as functional amino acids are unknown, and that different genes may be providing resistance in IR64 and Nipponbare. However, the identification of a cluster of highly similar RLP genes in both cultivars suggests otherwise. Their similarity with *CuRe1*, the *C. refexa* resistance gene, and *Ve1*, which provides resistance to *Verticillium* wilts, xylum-invading pathogens with a remarkably similar lifestyle to

Striga species, means these RLPs are top candidate for *S. hermonthica* resistance genes. The next step will be to functionally show whether the RLPs identified in IR64 and Nipponbare provide increased resistance to *S. hermonthica*, and which gene, or combination of genes, is necessary for this. Isolation of the resistance gene or genes will then allow further examination of the functional amino acids determining specificity, as well as a better understanding of downstream signalling and metabolic processes that prevent parasite penetration.

Chapter 4

Functional analysis of *S. hermonthica* candidate resistance genes in rice

4.1 Introduction

A major challenge facing geneticists this century is the identification of genes within QTL that underlie the phenotype of the quantitative trait (Price, 2006). The availability of plant genomes now makes the identification of genes within a mapped QTL more straightforward; however, relating the phenotype to a given gene can prove challenging. A major QTL for post-attachment resistance to *S. hermonthica* has been mapped in a rice IR64 x Azucena RIL mapping population (Chapter 2). The QTL in IR64 mapped to the same position on chromosome 12 as the *S. hermonthica* resistance QTL previously mapped in Nipponbare, a temperate *japonica* rice cultivar. Gene prediction software and BLAST analysis identified a cluster of RLP genes that are present in both the IR64 and Nipponbare QTL regions. Close comparison of gene sequences between resistant and susceptible cultivars (where available) confirmed many of these genes were either missing in susceptible cultivars or differed in amino acid sequence when compared to the resistant cultivars. These RLP genes are annotated as orthologs of *Ve1*, a RLP gene in tomato providing resistance to *Verticillium* wilt (Fradin *et al.*, 2009), a fungus with a remarkably similar lifestyle and infection strategy to *Striga*. A recent study has also shown that the RLP gene *CuRe1* in tomato conferred enhanced resistance to the parasitic plant *Cuscuta reflexa* (Hegenauer *et al.*, 2016). All this means that the cluster of RLP genes in both Nipponbare and IR64 are top candidates for *S. hermonthica* resistance genes. It is therefore important to functionally test whether these RLP genes provide increased resistance to *S. hermonthica*.

A common approach to understanding gene function is to knock out or reduce the expression of the gene of interest in order to induce a mutant phenotype and thus elucidate its function. This has an advantage over transforming an R gene into susceptible plants, because requirements for downstream defence signalling may be missing in the different genetic background of susceptible plants. Insertional mutagenesis is also a useful tool in forward genetic approaches of this kind (Page & Grossniklaus, 2002). Two different biological mutagens can be used to create random genome-wide mutagenesis in plants; transposons, and T-DNA (transferred DNA) of *Agrobacterium tumefaciens* (Page & Grossniklaus, 2002). Libraries of both types of insertion line have been created in rice. Miyao *et al.*, (2003) produced over 47,000 rice insertion mutants from the endogenous *copia*-like retrotransposon *Tos17*. Nipponbare, the cultivar selected for mutagenesis studies, carries 2 copies of *Tos17* that are activated by tissue culture but inactive in the regenerated plants under normal conditions. Over 16,800 T-DNA insertion lines have also been generated in Nipponbare (Lorieux *et al.*, 2012). High-throughput PCR-based recovery and sequencing from the known DNA insertion and flanking region allows the position of the insertions to be precisely located (Lorieux et al., 2012).

Another approach to studying gene function is RNA interference (RNAi). In contrast to insertional mutagenesis, RNAi works by suppressing the mRNA levels of a targeted gene, and has been used extensively to manipulate gene expression, becoming a powerful tool in functional genomics (Hannon, 2002; Wilson & Doudna, 2013). RNAi was first discovered in the nematode *Caenorhabditis elegans* (Fire *et al.*, 1998) and is now known to be a conserved mechanism in eukaryotic organisms (Agrawal *et al.*, 2003). RNAi is induced by the presence of non-coding double-stranded small RNAs. There are 2 groups of small RNAs in plants, small interfering RNAs (siRNAs) and microRNAs (miRNAs), which both play diverse roles in regulating endogenous gene expression in processes such as growth and development, epigenetic inheritance and stress responses (Yu & Kumar, 2003; Weiberg & Jin, 2015). They also have essential roles in plant defence, targeting transposable elements and viral infections and the activation of PTI and ETI (Weiberg & Jin, 2015). miRNAs are endogenous to the plants genome and transcribed in the nucleus, while siRNAs may derive from the plants genome or be introduced by viral infection or another exogenous source (Wilson & Doudna, 2013).

The process of RNA interference is initiated by DICER, an endoribonuclease protein that binds to double-stranded RNA molecules and cleaves them into 21 - 24 nucleotide pieces (siRNAs). The siRNAs are then incorporated into a protein complex called the RNA-induced silencing complex (RISC), which includes an Argonaute protein (Figure 4.1). RISC mediates the unwinding of the siRNA molecule; the passenger strand is released, while the guide strand remains bound to Argonaute and serves to direct RISC to a homologues single-stranded mRNA substrate. This mRNA is cleaved by Argonaute and then degraded, inhibiting its translation into protein (Figure 4.1) (Hannon, 2002; Agrawal *et al.*, 2003; Baulcombe, 2004; Wilson & Doudna, 2013).

Simple RNAi vectors have been developed for introducing siRNA trigger sequences into plants by *Agrobacterium* transformation to manipulate gene expression experimentally (Miki & Shimamoto, 2004). Trigger sequences from the target gene are cloned into a vector either side of a *gus*-linker region (920bp) in a sense and anti-sense (Inverted Repeat (IR)) direction, in this case driven by a strong maize ubiquitin promoter (Figure 4.2) optimised for monocot transformation. Transcription results in the formation of a hairpin loop and double-stranded RNA of homologous sequence to the target gene, resulting in suppression of gene expression by RNAi (Miki & Shimamoto, 2004).

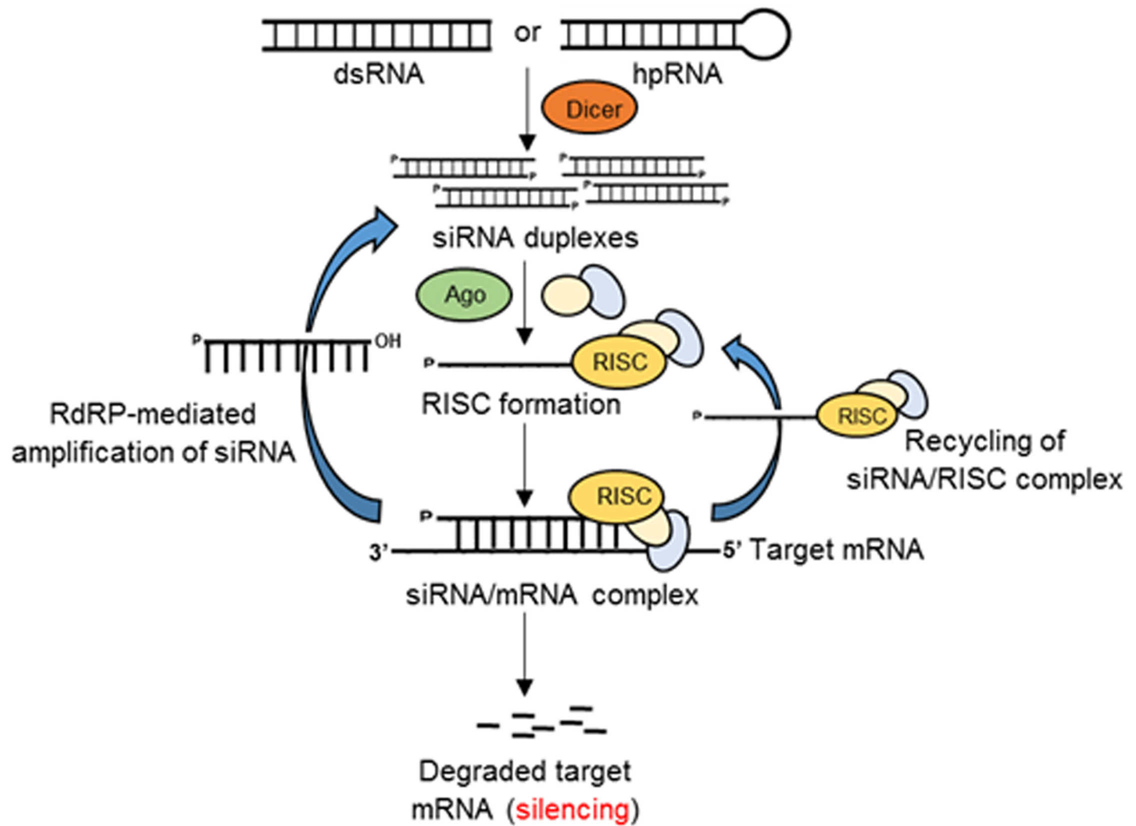


Figure 4.1 RNAi–mediated gene silencing. Double-stranded or hairpin RNA is cleaved into siRNAs by DICER. siRNAs bind to Argonaute (Ago) and are incorporated into the RNA-induced silencing complex (RISC), directing RISC to complementary mRNA substrates. This mRNA is cleaved by Argonaute and then degraded, inhibiting its translation into protein. The components of RISC can be recycled, or additional siRNA duplexes can be generated by the enzyme RNA-dependent RNA-polymerase (RdRP). Figure from Majumdar *et al.*, 2017.

Figure 4.2 Diagram of the Gateway pANDA vector developed by Miki & Shimamoto (2004) for suppression of gene function in rice by RNAi, showing position of the trigger sequence either side of the gus-linker sequence. Boxes and line downstream of the promoter are exons and an intron of the maize ubiquitin gene, respectively. Re-drawn from Miki & Shimamoto (2004).

As plants often contain families of genes sharing high similarity in their sequence, a construct designed against a single gene may result in the silencing of multiple members of the gene family. Miki *et al.*, (2005) showed that a single IR construct, using a highly conserved coding region from the *OsRac* rice family as a dsRNA trigger, successfully reduced the transcript of the target gene to less than 1 % of the wildtype, while mRNA levels of 4 other genes in the family were reduced to less than 10 %. When constructs were designed against gene-specific regions of the *OsRac* family, significant reductions in transcript were only observed for the target gene corresponding to the construct used (Miki *et al.*, 2005). RNAi is therefore a promising approach to supressing gene function in rice for both individual genes and gene clusters.

As part of a BBSRC grant in our laboratory, an RNAi approach was taken to investigate the function of the RLP genes in the Nipponbare QTL region which underlies resistance to *S. hermonthica* (Kibos isolate). Dr. Alexis Moschopoulos, a postdoctoral researcher, designed RNAi IR constructs targeting three key candidate resistance genes in the cluster, and these were stably transformed into Nipponbare using *Agrobacterium*-mediated transformation. Detailed protocols for *Agrobacterium*-mediated transformation of *japonica* rice cultivars have been published and optimised (Sallaud *et al.*, 2003; Toki *et al.*, 2004), however transformation of *indica* rice is much more difficult (Sahoo *et al.*, 2011). Transformation of either subspecies and regeneration and selection of transformants is a long process taking several months. Because of the difficulty in transforming IR64, the time required for regeneration of plants, selection of transformants and propagation by single seed decent to at least the T_2 generation, functional validation of the IR64 target genes was not possible within the timescale of this PhD. However, as the candidate *S. hermonthica* resistance genes in IR64 are close homologs of those in Nipponbare, and the T_2 generation of RNAi lines were just available at the time of my PhD, they were used to test the hypothesis that reducing the expression of one or more of the candidate RLP genes in Nipponbare will result in increased susceptibility to *S. hermonthica* (Kibos isolate). As Nipponbare is also the cultivar used for creation of the *Tos17* and some T-DNA insertion line libraries, lines with an insertion in a candidate RLP gene, where present, were also phenotyped for altered susceptibility to *S. hermonthica*.

4.1.1 Aims of Chapter 4

The aim of this chapter is to test the hypothesis that the candidate receptor-like protein genes in Nipponbare underlie resistance to *S. hermonthica* by:

1. Phenotyping the RNAi lines targeting candidate RLP genes in Nipponbare for altered susceptibility / resistance to *S. hermonthica* (Kibos ecotype).

2. Measuring the expression of the RLP genes by qPCR for a selection of RNAi lines to determine whether down regulation of gene expression is correlated with increased susceptibility to *S. hermonthica* (Kibos ecotype).
3. Phenotyping T-DNA and *Tos17* insertion lines targeting single candidate RLP genes (where available) for altered resistance / susceptibility to *S. hermonthica* (Kibos ecotype).

4.2 Materials and Methods

4.2.1 Plant materials

The RNAi constructs targeting three RLP genes (Os12g11370, Os12g11500 and Os12g11680) were produced at the University of Sheffield. Trigger sequences for the constructs were designed by Dr. Alexis Moschopoulos at the University of Sheffield, using psRNATarget, a plant small RNA target analysis server designed by Dai & Zhao (2011). Trigger sequences were cloned into the pANDA Gateway vector developed by Miki & Shimamoto (2004). The constructs, their target gene and other predicted gene targets in the RLP cluster are shown in Table 4.1. Trigger sequences are shown in Appendix Table S.1. Constructs were transformed into the rice cultivar Nipponbare at The National Institute of Agricultural Botany (NIAB), Cambridge, and regenerated plantlets returned to Sheffield. Zygosity and copy number of RNAi lines were determined by qPCR quantification of the *gus*-linker gene within the construct by iDNA Genetics Ltd, Norwich, UK (Table 4.1). All RNAi lines were phenotyped at the T_2 generation.

Tos17 insertion lines were produced in a Nipponbare background by Miyao *et al.*, (2003) and obtained from the National Agriculture and Food Research Organization (NARO), Japan. Three homozygous *Tos17* lines were available, with predicted insertions in genes Os12g10870 (line ND5288), Os12g11680 (line ND00064), or Os12g12120 (line ND2342). A single T-DNA insertion line was obtained from CIRAD, Montpellier, France, with a homozygous insertion in Os12g11370. All *Tos17* insertion lines and the T-DNA insertion lines were in a Nipponbare background.

The presence of a homozygous insertion in each line was confirmed by PCR using the Thermo Scientific Phire Plant Direct PCR kit (section 2.2.4). Primers used are listed in Table 4.2.

4.2.2 Phenotyping RNAi lines for post-attachment resistance to *S. hermonthica*

RNAi rice lines (T_2 generation) were phenotyped for post-attachment resistance to *S. hermonthica* according to the protocol described in section 2.2.2. Lines were labelled according to the construct, transformation event and generation. For example, JS6.8-7 refers to the p11370-4 construct (see Table 4.1), JS6 transformation event, with 8 referring to the T_0 individual and 7 to the T_1 individual. Before transfer to rhizotrons, all seedlings were genotyped for the presence of the transfer DNA using primers designed against the hygromycin gene (forward primer: ATGTCCTGCGGGTAAATAGC, reverse primer: CATTGTTGGAGCCGAAATCC). Genotyping was carried out 6 das using the Thermo Scientific Phire Plant Direct PCR kit according to the protocol described in section 2.2.4. Rice seedlings

Table 4.1 RNAi constructs produced for down-regulation of suites of RLP genes in Nipponbare. Construct name refers to the target gene against which the trigger sequence was designed. JS / AM number refers to independent transformation events for a particular construct. Transgene copy number and zygosity is listed for each line; Hom: homozygous line, Hemi: hemizygous line, Multi: multiple copies due to multiple insertion sites of the transgene. Hemi /Hom lines have 2+ copies but these were predicted to be at a single insertion site.

Construct name	Transformation identifier	Primary target(s)	Secondary targets	Lines	Transgene copy number	Zygosity
p11370-4	JS6	Os12g11370	Os12g10870 Os12g11860 Os12g11940	JS6.8-7	2 - 5	Multi
p11370-10	JS7	Os12g11370	Os12g10870 Os12g11500	JS7.8-12	3 (6)	Hemi *
p11680-1	JS4 JS8	Os12g11680 Os12g12010	Os12g10870 Os12g11510 Os12g11720 Os12g11940 Os12g12120 Os12g12130	JS4.8-7 JS4.12B-11 JS8.14-19 JS8.16-11	3 (6) 5 / 10 8 18	Hemi * Hemi / Hom Hom Hom
p11500-4	JS10	Os12g11500 Os12g11930	Os12g10870 Os12g11370 Os12g11940 Os12g12120 Os12g12130	JS10.10A-8 JS10A.1-14	2 - 15 8	Multi Hom
p11500-9	AM3	Os12g11500 Os12g11930	Os12g10870 Os12g11720	AM3.9-1	12 - 24	Hemi
pANDA emtpy	JS9	None	None	JS9.8-15	2	Unknown

* predominant zygosity

Table 4.2 Primer sequences for confirmation of the presence of the Tos17 / T-DNA insertions for 4 insertion lines: Os12g10870, Os11680 and Os12g12120 (*Tos17* insertion lines), and Os12g11370 (T-DNA insertion line).

Primers for testing presence of insertion			
Gene	**Primer name**	**Primer sequence (5' – 3')**	**Amplicon length (bp)**
Tos17 insertion	Tos17	TACTGAGGCTGAACTTCGGGC	~ 1000 (with any of below primers)
Os12g10870	10870 Tos17 F	TCACTGTCAGGTCCCGTATGC	
	10870 Tos17 R	GAGGTCAAGGAATTCTATATGAAGAGG	
Os12g11680	11680 Tos17 F	GATTTACTAGAGGTGTCTGGGTTAC	
	11680 Tos17 R	AAACTAATACTGGAGTAGGATGCTG	
Os12g12120	12120 Tos17 F	ATGTTGTCCAACTCCAAGAGAGTTC	
	12120 Tos17 R	GTTGGTGTTGCTCACAGATATACTC	
T-DNA insertion	hyg8	GTCTGGACCGATGGCTGTGTAGAAG	~ 900 (with below primers)
Os12g11370	11370 cDNA F	ATGTCGTCGTCCACCAAGAGG	
	11370 TDNA R	CTGAAGCTTAGGTGTAAACTTAGC	

were inoculated with germinated *S. hermonthica* seeds 15 das, and *S. hermonthica* plants harvested from rice roots 25 dai. Where possible, 5 – 8 replicates were phenotyped for each line. However, fewer replicates were phenotyped if rice seed germination was poor and / or the transgene was segregating. The minimum number of replicates used was 3 (line JS8.16-11 only). Azygous plants were used as controls where available, in addition to wildtype Nipponbare and an empty vector control. Nipponbare and Koshihikari were included in all phenotyping experiments as resistant and susceptible cultivars respectively.

In order to examine the microscopic phenotype of *S. hermonthica* attachments on a susceptible RNAi line (JS8.14-1), small sections of root plus attached parasite were harvested 11 dai together with those from Nipponbare and Koshihikari. Sections were embedded in Technovit resin, stained and observed under a microscope as described in section 2.2.5.

4.2.3 Growth and collection of RNAi root material for analysis of gene expression by qPCR

In order to measure the expression of the RLP genes in wildtype Nipponbare, the RNAi lines and empty vector control lines, 6 replicate plants were grown per line. Collection of root tissue was carried out 21 das according to the protocol described in section 3.2.6. RNA extraction, quantification, cDNA synthesis and qPCR were carried out as described in section 3.2.7. For qPCR, 4 µl cDNA (1 in 10 dilution) was used in each reaction, and 2 technical replicates carried out for each sample. In order to determine how well an RNAi construct was transcribed, the expression of *gus,* derived from the linker region of the construct, was measured for each sample (Miki & Shimamoto, 2004). Ribonuclease regulator (Os01g52460) was used as a reference gene. Primer sequences used to amplify the 12 expressed RLP genes within the Nipponbare QTL region, the reference gene and the *gus* linker are shown in Table 4.3.

4.2.4 Confirming the presence of the insertion in the target genes and phenotyping the insertion lines for post-attachment resistance to *S. hermonthica*

Before phenotying the insertion lines for resistance to *S. hermonthica* it was necessary to confirm that they contained an insertion in the genes of interest. To confirm knockout of the target gene transcript from cDNA, uninfected root tissue for each line was harvested from 6 plants 15 das according to the protocol described in section 3.2.6. RNA was extracted from root tissue and cDNA synthesised as described previously (section 3.2.7). A PCR of the target gene sequence was carried out on cDNA from all knockout lines, as well as both cDNA and gDNA from wildtype Nipponbare as positive controls. Primers used are in listed in Table 4.4. Each 20 µl consisted of 2 µl 10 x Buffer for KOD DNA polymerase, 1.2 µl of 1mM $MgSO_4$, 2 µl

Table 4.3. Primer sequences for qPCR amplification of candidate *S. hermonthica* RLP resistance genes, the reference gene ribonuclease regulator (Os01g52460) in the rice cultivar Nipponbare, and the *gus*-linker region of the construct.

Gene	Primer name	Primer sequence (5' – 3')	Amplicon length (bp)
Os12g10870	10870 F 10870 R	GATCCTTCATATACAGTTGACAGG GCCATGTCAGCAATTCTGAGC	65
Os12g10930	10930_2 F 10930_2 R	TTCTCTAACTGTGGATTGTCTGG TTATTTTCCTTCCCTGAAAAGTTGC	107
Os12g11370	11370 F 11370 R	GGATCCTTCATATACAGTTGATGG GAAGTTATTTGAGGCCATATCGG	79
Os12g11500 (JS10 construct)	11500_3 F 11500_3 F	GCACGCACCGTAGTACC GAAGGCAGTGGAGTAGTCG	108
Os12g11500 (AM3 construct)	11500_x F 11500_x R	TTAACCTCTCCGGTAACGACTTC AAGATAAACCAGTTCGGTGAGCT	83
Os12g11680	11680_2 F 11680_2 R	GTGGCTTGTCTGGACCCG TGACAATTGTACAGTGCCAAATC	76
Os12g11720	11720 F 11720 R	CGCCTGAAGAACTCATTCG GATCCATGACCGGAATGC	63
Os12g11860	11860 F 11860 R	TGCAGCATGAGAAACTAACG ACTGTCAGCAGAGAAATTAGG	83
Os12g11930	11930 F 11930 R	CGTACGAGCTCCTCCATACC CGACCCACGACCGAAAGG	124
Os12g12010	12010_2 12010_2	TACTAGAAGTGTCTGGGTTGG GGAAGCTGGTAAGCGTCCG	119
Os12g12120	12120_2 F 12120_2 R	CAGATCCAAGGTGCAATACC AAGGGGAAGTAAAGGATTGG	117
Os12g12120	12130 F 12130 R	CAGTTTTCAGGCGAGATACC TCAGGACATACAGGTTTTGC	113
LOC_Os01g52460	Os01g52460-F1 Os01g52460-R1	GGCAAACAAGAAGGGAATAGG AGTCCTCGAGATGAGAATGC	115
gus-linker	GUS pANDA F GUS pANDA R	CGATAACGTGCTGATGGTGCA CTC TTC AGC GTA AGG GTA ATG C	91

2mM dNTPs, 3 µl of each 2 µM primer, 0.4 µl KOD DNA polymerase, 2 µl cDNA / gDNA and 6.4 µl nuclease-free water. The PCR cycling conditions were: 2 min at 95 °C (initial denaturation); then 35 cycles of 20 s at 95 °C (denaturation), 10 s at 58 °C (annealing), and 1 min at 70 °C (extension); followed by a final extension of 1 min at 70 °C. Ten µl of each product was mixed with 5 x DNA Loading Buffer (Bioline) and run on a 1 % gel to check product sizes. Sanger sequencing was used to confirm the presence of the insertions in the target genes according to the protocol described in section 3.2.4**.**

Table 4.4. Primer sequences for testing whether gene transcripts were present for target genes in 4 insertion lines: Os12g10870, Os11680 and Os12g12120 (*Tos17* insertion lines), and Os12g11370 (T-DNA insertion line).

Primers for testing expression of target gene			
Gene	**Primer name**	**Primer sequence (5' – 3')**	**Amplicon length (bp)**
Os12g10870	10870 Tos17 F	TCACTGTCAGGTCCCGTATGC	1014
	10870 Tos17 R	GAGGTCAAGGAATTCTATATGAAGAGG	
Os12g11680	11680_seq1_F	GGTCCGATTTGTCGGTCCT	776
	11680 Tos17 R	AAACTAATACTGGAGTAGGATGCTG	
Os12g12120	12120 F2	GATCCGATGGCCATATCACC	
	Os12g12120-R1	AAGGGGAAGTAAAGGATTGG	1476
	11370genoR1	AAGATCTGCGGAGGCACCTT	
Os12g11370	11370genoF1	AATTCGCTCACTAGGATTGAGCTT	2247
	11370 cDNA R	TCATCGCTGTTTTTCATATGCCTTCC	

Lines were phenotyped for post-attachment resistance to *S. hermonthica* according to the protocol described in section 2.2.2. Six - ten replicates were infected for each line, in addition to Nipponbare and Koshihikari controls.

4.2.5 Statistical analysis

Two-sample t-tests were carried out to assess whether *S. hermonthica* biomass was significantly different between the RNAi / insertion lines and Nipponbare wild type, and between RNAi lines and azygous segregants where present. Tests were carried out on $\log_{10}$ transformed data to adjust for non-normal distribution where necessary.

Relative quantification of gene expression was calculated according to the comparative Ct method described in Pfaffl (2004) and in section 3.2.9. For each gene of interest, gene expression values of RNAi lines were divided by the mean expression of that gene in Nipponbare to determine any difference between them. Two-sample t-tests were carried out to identify significant differences. A Pearson's product-moment correlation analysis was used to determine the relationship between *gus*-linker expression and expression of RLP genes. The statistical package R, version 3.3.0 (http://www.r-project.org) was used for all analyses.

4.3 Results

4.3.1 Phenotyping RNAi lines for resistance to *S. hermonthica* and examination of RLP gene expression

RNAi lines exhibited varying degrees of susceptibility to *S. hermonthica*. In most cases, increased susceptibility was associated with suppression of transcript of multiple RLP genes (Figures 4.3 – 4.5). No increase in susceptibility or significant alteration in the expression of the RLP genes was observed for empty vector controls (Figure 4.3A and E).

RNAi lines targeting Os12g11500: A small but significant increase in susceptibility to *S. hermonthica* was seen for three out of four independent lines targeting Os12g11500 (Figure 4.3A). Gene expression analysis was carried out for two of the susceptible lines and the resistant line. In all cases, transcript levels of the target gene Os12g11500 were significantly reduced. No other significant reductions in transcript level for the other RLP genes was observed for the resistant line AM3.9-1. However, significant reductions in transcript of two additional genes (Os12g10870 and Os12g10930) and three addition genes (Os12g11680, Os12g11869 and Os12g12010) were seen for lines JS10A.1-14 and JS10.10A-8 respectively (Figure 4.3C and B), both of which were significantly more susceptible to *S. hermonthica* than Nipponbare.

RNAi lines targeting Os12g11680: A significant increase in susceptibility to *S. hermonthica* was observed in two out of six independent RNAi lines targeting Os12g11680 (Figure 4.4A). All azygous controls were resistant. Gene expression analysis was carried out for the two susceptible RNAi lines (JS8.14-19 and JS4.12B-11) and two lines showing greater resistance (JS8.16-11 and JS4.8-7). Transcript of all RLP genes investigated was reduced in the most susceptible line JS8.14-19, however no reduction was statistically significant (Figure 4.4B). No reduction in transcript was observed for any other gene or line investigated (Figures 4.4C –E).

RNAi lines targeting Os12g11370: A significant increase in susceptibility to *S. hermonthica* was observed for one of the two RNAi lines targeting Os12g11370 (Figure 4.5A). Increased susceptibility of the line JS6.8-7 was associated with a significant reduction in transcript abundance of the target gene Os12g11370, as well as a reduction in transcript of Os12g10870 and Os12g11860 (Figure 4.5B). In contrast, the resistant line JS7.8-12 showed no significant alterations in gene expression (Figure 4.5C).

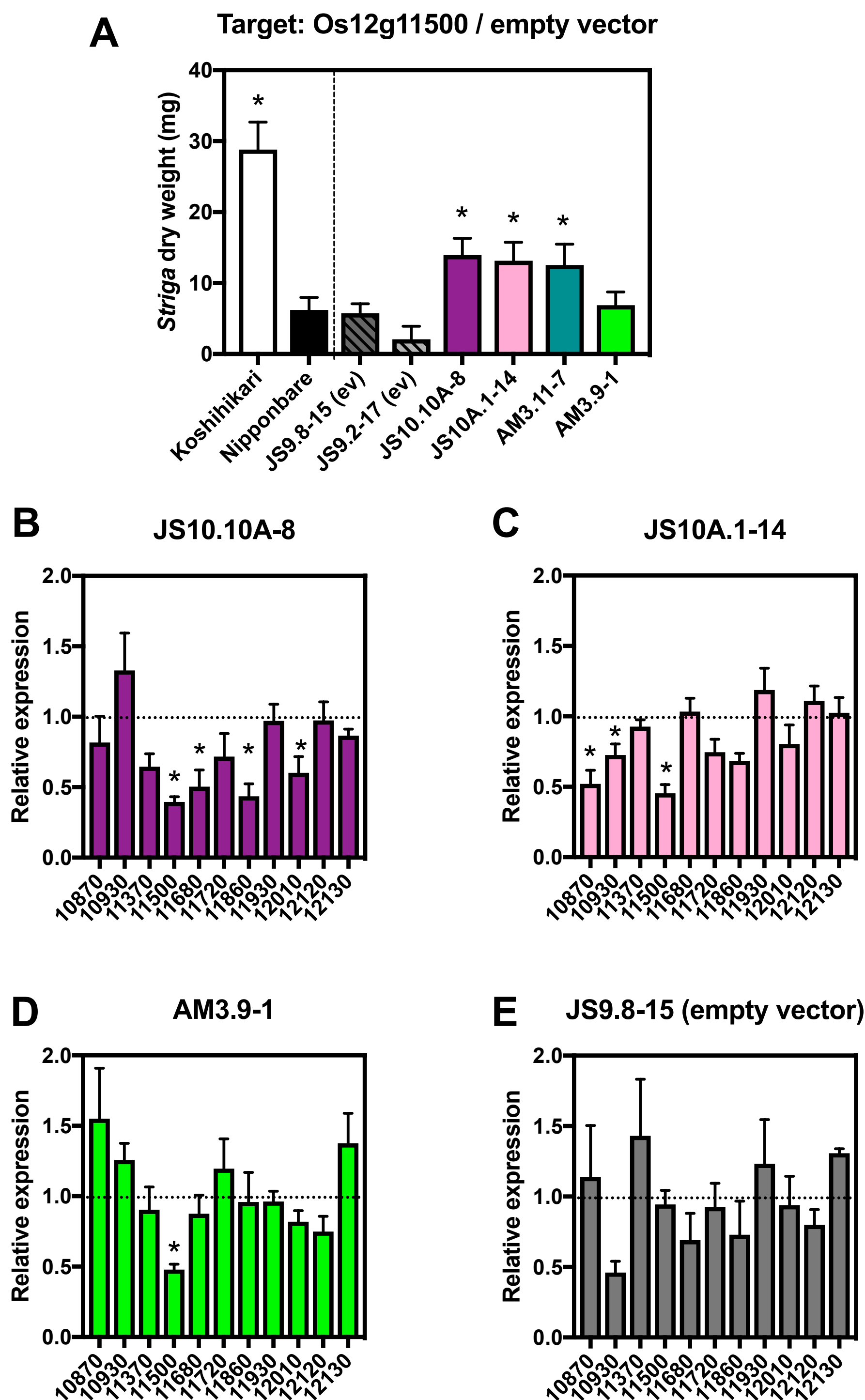


Figure 4.3 The dry biomass of *S. hermonthica* (A) and differential expression of RLP genes (B-E) for RNAi rice lines targeting the gene Os12g11500, and empty vector controls. JS/AM numbers represent independent transformations of the construct. Hashed bars are empty vector controls. **A:** *S. hermonthica* was harvested from the roots of rice plants 25 dai. Nipponbare and Koshihikari were included as a reference for resistance and susceptibility respectively. Values for dry weight are means ± SE where n = between 5 - 8. **B-E):** Differential gene expression of RLP genes measured by qPCR; numbers refer to Os12g gene numbers. Gene expression was normalised using the reference gene Os01g52460 and plotted as fold change in comparison with Nipponbare wt where Nipponbare = 1. A value of 0.5 = 50 % reduction in expression relative to Nipponbare. Data represent mean expression values ± SE from 4 – 6 biological replicates. Asterisks indicate significant differences between an RNAi line and Nipponbare (two-sample t-test $p < 0.05$).

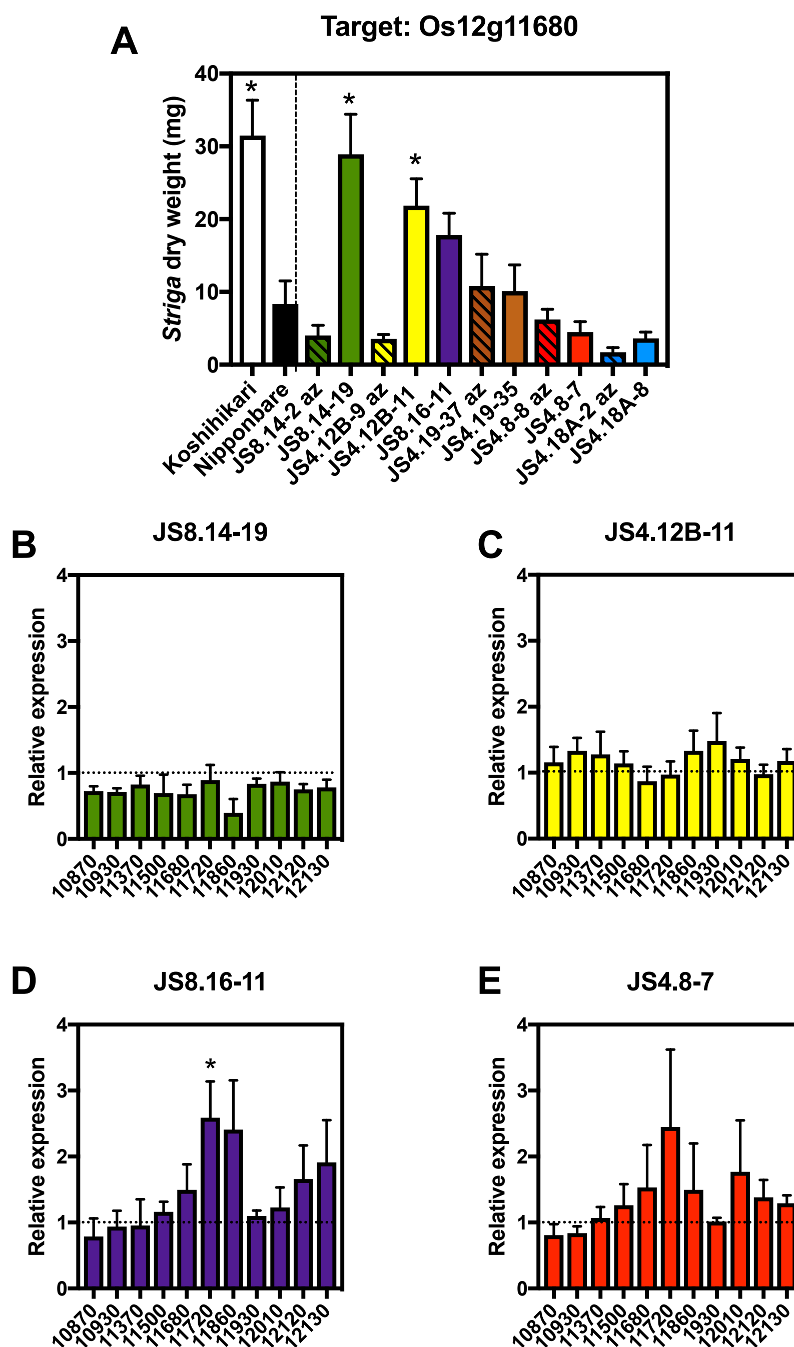


Figure 4.4 The dry biomass of *S. hermonthica* (A) and differential expression of RLP genes (B-E) for RNAi rice lines targeting the gene Os12g11680. JS numbers represent independent transformations of the construct. Hashed bars are azygous plants where the T-DNA has segregated out. **A:** *S. hermonthica* was harvested from the roots of rice plants 25 dai. Nipponbare and Koshihikari were included as a reference for resistance and susceptibility respectively. Values for dry weight are means ± SE where n = between 5 - 8. **B-E):** Differential gene expression of RLP genes measured by qPCR; numbers refer to Os12g gene numbers. Gene expression was normalised using the reference gene Os01g52460 and plotted as fold change in comparison with Nipponbare wt where Nipponbare = 1. A value of 0.5 = 50 % reduction in expression relative to Nipponbare. Data represent mean expression values ± SE from 4 – 6 biological replicates. Asterisks indicate significant differences between an RNAi line and Nipponbare (two-sample t-test $p < 0.05$).

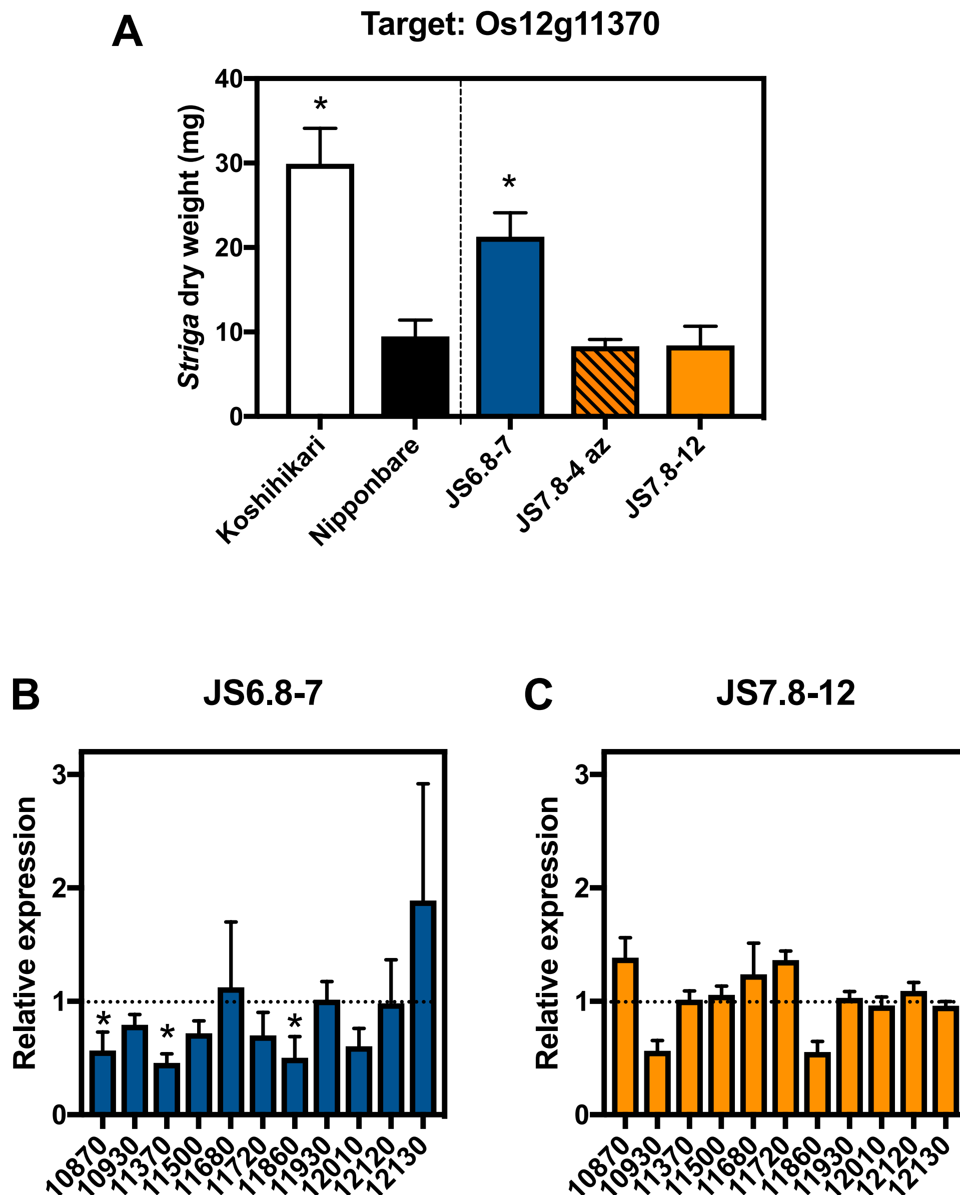


Figure 4.5 The dry biomass of *S. hermonthica* (A) and differential expression of RLP genes (B-C) for RNAi rice lines targeting the gene Os12g11370. JS numbers represent independent transformations of the construct. Hashed bars are azygous plants where the T-DNA has segregated out. **A:** *S. hermonthica* was harvested from the roots of rice plants 25 dai. Nipponbare and Koshihikari were included as a reference for resistance and susceptibility respectively. Values for dry weight are means ± SE where n = between 5 - 8. **B-E):** Differential gene expression of RLP genes measured by qPCR; numbers refer to Os12g gene numbers. Gene expression was normalised using the reference gene Os01g52460 and plotted as fold change in comparison with Nipponbare wt where Nipponbare = 1. A value of 0.5 = 50 % reduction in expression relative to Nipponbare. Data represent mean expression values ± SE from 4 – 6 biological replicates. Asterisks indicate significant differences between an RNAi line and Nipponbare (two-sample t-test $p < 0.05$).

The phenotype of resistance and susceptibility of Nipponbare, Koshihikari and a representative plant from the susceptible RNAi line JS8.14-19 is shown in Figure 4.6. Nipponbare exhibited good post-attachment resistance to *S. hermonthica*, and only a few small parasites were harvested from the root systems (Figure 4.6A). Transverse sections through the root showed a failure of the parasite to penetrate the endodermis and attach to xylem vessels, instead growing through the cortex around the vascular core. (Figure 4.6B). In contrast, many larger parasites were harvested from Koshihikari root systems (Figure 4.6C), and transverse sections revealed *S. hermonthica* had successfully penetrated the endodermis and attached to xylem vessels (Figure 4.6D), allowing continued growth of the parasite. A similar phenotype was observed for the susceptible RNAi line JS8.14-19; large parasites were harvested from rice roots (Figure 4.6E), and transverse sections confirmed fully established xylem-xylem connections between host and parasite (Figure 4.6F).

4.3.2 The efficiency of gene suppression for RNAi constructs

In order to determine how well the RNAi constructs were transcribed, transcript of the *gus*-linker region of the construct was measured by qPCR. The efficiency of gene suppression for each RNAi line was tested by plotting the expression of the *gus* transgene against expression of the target gene (Figure 4.7). Expression of the *gus* gene varied considerably between lines. There was a significant negative relationship between target gene expression and expression of the *gus*-linker (Pearson's correlation $r = -0.303$, d.f. = 45, $p = 0.038$, Figure 4.7), indicating that where RNAi constructs were not well transcribed, efficiency of their silencing was reduced.

Although the RNAi constructs were designed against a single gene, the high similarity in sequence of the RLP genes in the QTL meant that a single construct was expected to suppress the expression of more than one gene in the cluster, as many secondary targets were predicted (see Table 4.1). This makes it very difficult to relate any susceptibility to suppression of a single gene. Some lines exhibited only a small reduction in transcript levels, but this was seen for many of the RLP genes. It is also possible that more than one gene may be acting to confer resistance to *S. hermonthica*, which might obscure any clear correlation between transcript level for a single gene and susceptibility. Therefore, the sum of gene expression values for all RLP genes investigated was calculated and plotted against the dry weight of *S. hermonthica* to determine whether there was a correlation with increased susceptibility (Figure 4.8A). The SUM of RLP expression was also plotted against expression of the *gus*

transcript to test for a correlation between expression of the construct and its ability to silence the RLP genes generally. No significant correlation was observed between the SUM of RLP expression and *S. hermonthica* dry weight, or between the SUM of RLP expression and the expression of the gus-linker (Figure 4.8).

4.3.3 Insertion lines exhibited no increase in susceptibility to *S. hermonthica*

Four rice lines with homozygous insertions in the genes Os12g10870, Os12g11370, Os12g11680 and Os12g12120 were phenotyped for resistance to *S. hermonthica*. There was no significant difference in *S. hermonthica* dry weight harvested from the roots of any insertion line and that of the Nipponbare control. Unusually the Os12g11370 T-DNA insertion line showed almost full resistance; only 4 of the 16 replicates had any *S. hermonthica* infection, and < 0.3 mg *S. hermonthica* dry weight was harvested from each plant (Two-sample t-test: *Tos17* Os12g10870 line: $t = 0.44$, d.f. = 14, $p = 0.66$; *Tos17* Os12g11680 line: $t = 0.19$, d.f. = 12, $p = 0.84$; *Tos17* Os12g12120 line: $t = 1.32$, d.f. = 19, p-value = 0.20; T-DNA Os12g11370 line: $t = 2.09$, d.f. = 6.00, $p = 0.08$) (Figure 4.9A-D).

4.3.4 Verification of gene knockouts for *Tos17* and T-DNA insertion lines

A PCR was carried out on cDNA for the genes carrying the insertion to confirm knockout was successful. No expression was observed for Os12g10870, Os12g11680 or Os12g12120 for *Tos17* insertion lines, while expression of these genes in wild type Nipponbare was confirmed (Figure 4.10A). However, transcript was observed for Os12g11370 in the T-DNA insertion line for this gene; the final 2247 bp of the gene was successfully amplified from cDNA (Figure 4.10B). This was of greater abundance in the insertion line than in Nipponbare, as seen by brighter bands on the gel. Sequencing using primers inside the insertion site and at the end of the gene revealed the T-DNA insertion ended at position 695bp of Os12g11370, which is 3045 bp in length (Figure 4.11).

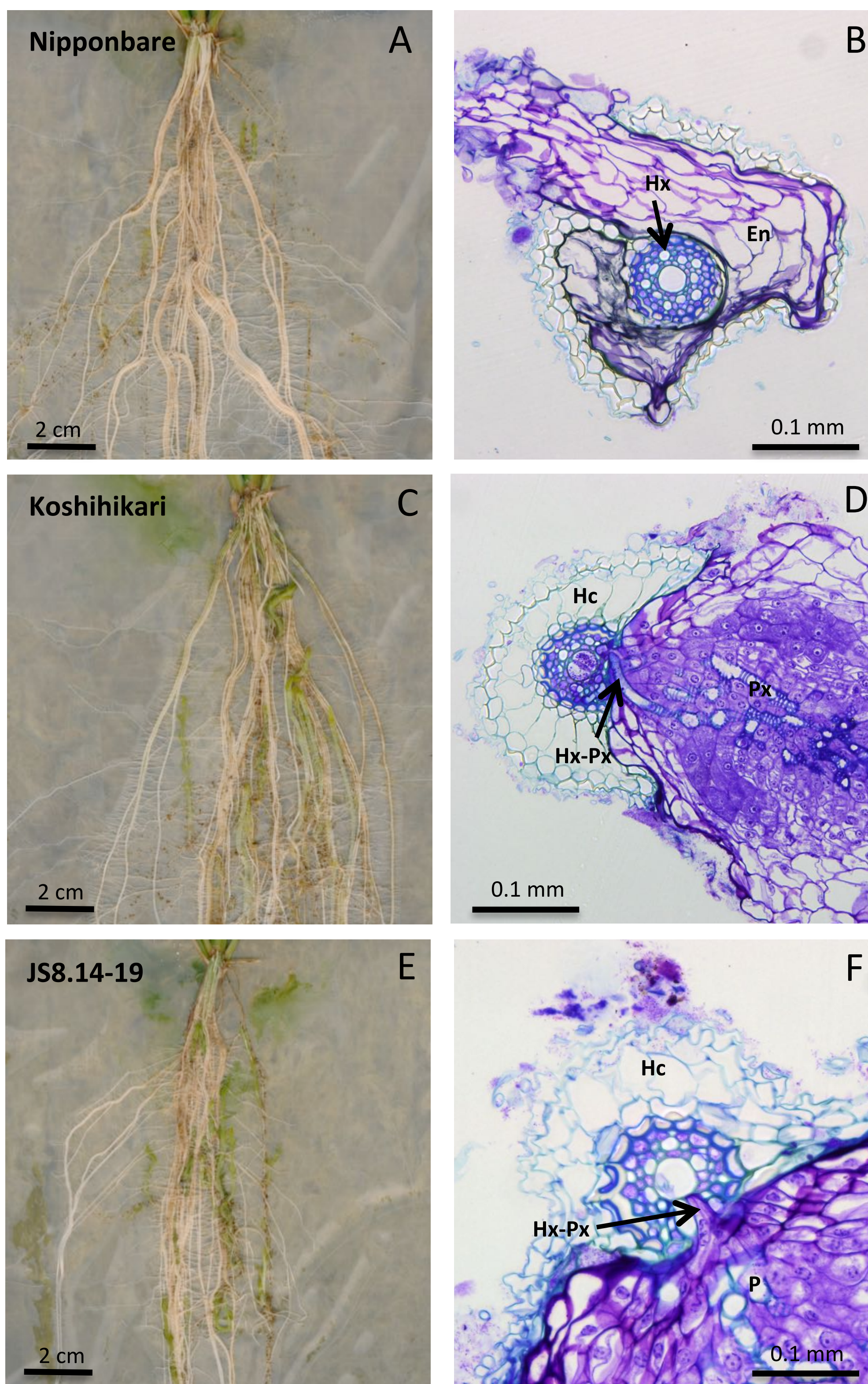


Figure 4.6 The phenotype of resistance and susceptibility of rice cultivars Nipponbare, Koshihikari and the susceptible RNAi line JS8.14-19. Left column: Images of the root systems infected with *S. hermonthica* (Kibos ecotype) 25 days after inoculation. Right column: transverse sections through the rice root and *S. hermonthica* attachment 11 days after inoculation. Resistance in Nipponbare is shown by the failure of the parasite to breach the endodermis and form vascular connections (B). Susceptible interactions on Koshihikari and JS8.14-19 revealed fully established xylem-xylem connections between host and parasite (D and F). Hx-Px, host-parasite xylem; Hc, host cortext; En, parasite endophyte; P, parasite.

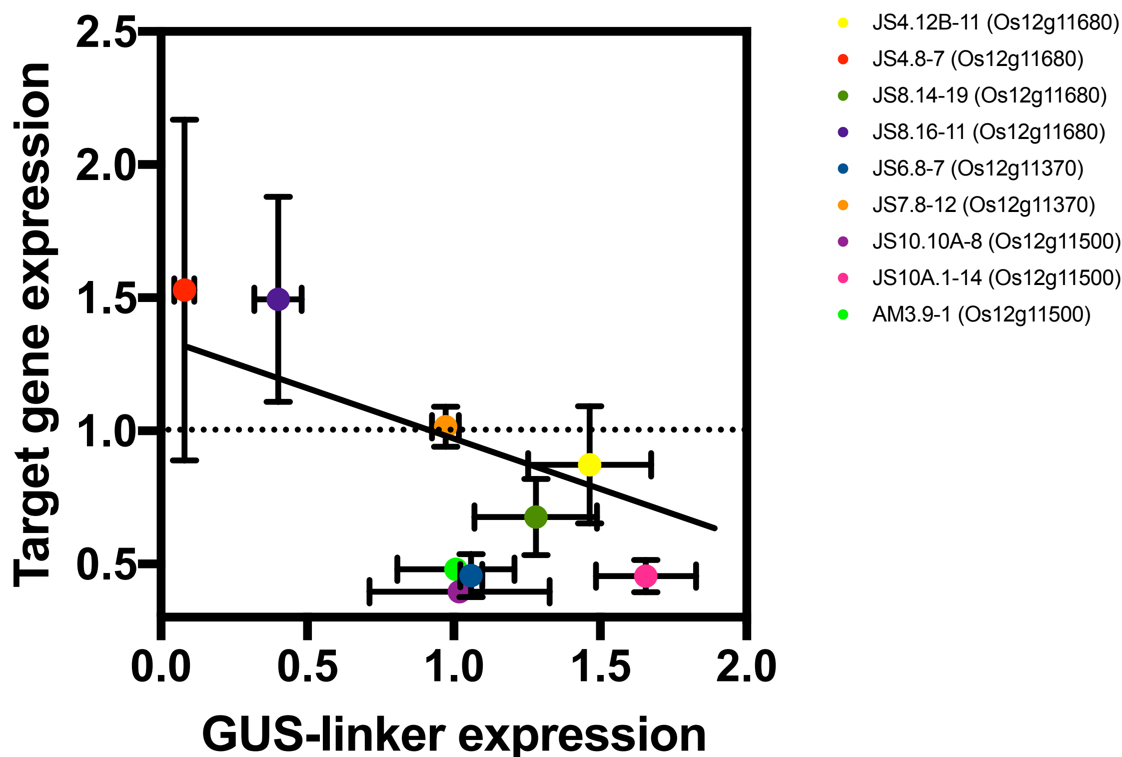


Figure 4.7 The efficiency of gene silencing for RNAi constructs, determined by the relationship between the expression of the *gus* transgene derived from the linker region of the construct, and the expression of the primary gene target. Target gene for each construct is indicated in the legend in brackets. Gene expression was measured by qPCR and normalised using the reference gene Os01g52460. Target gene expression is plotted as the fold change relative to Nipponbare wt, where Nipponbare = 0. A value of -0.5 = 50 % reduction in expression relative to Nipponbare. Dots represent mean expression values and standard error from 4 – 6 biological replicates each with 2 technical replicates. There was a significant negative relationship between target gene expression and expression of the *gus*-linker (Pearson's correlation r = -0.303, d.f. = 45, p = 0.038).

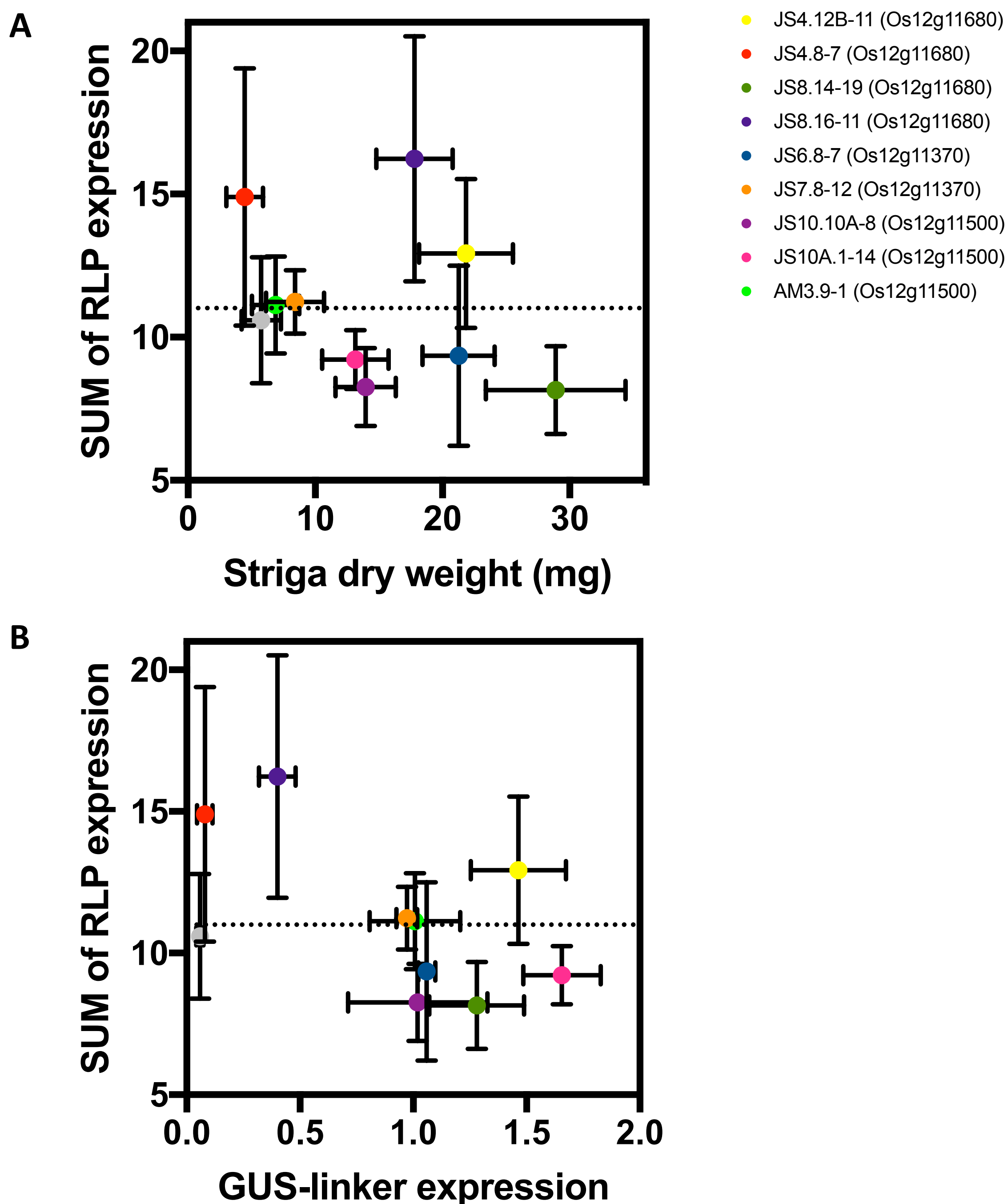


Figure 4.8 The relationship between the sum of total RLP gene expression for each RNAi line and resistance (mg dry weight) to *S. hermonthica* **(A)** and expression of the *gus*-linker transcript from the pANDA vector **(B)**. Coloured dots are different RNAi lines. Target gene for each line is indicated in the legend in brackets. Gene expression was measured by qPCR and normalised using the reference gene Os01g52460. Sum of RLP expression is plotted as the fold change levels relative to Nipponbare control where SUM of Nipponbare RLP gene expression = 11 (dotted line). Values are mean expression ± SE from 4 – 6 biological replicates each with 2 technical replicates.

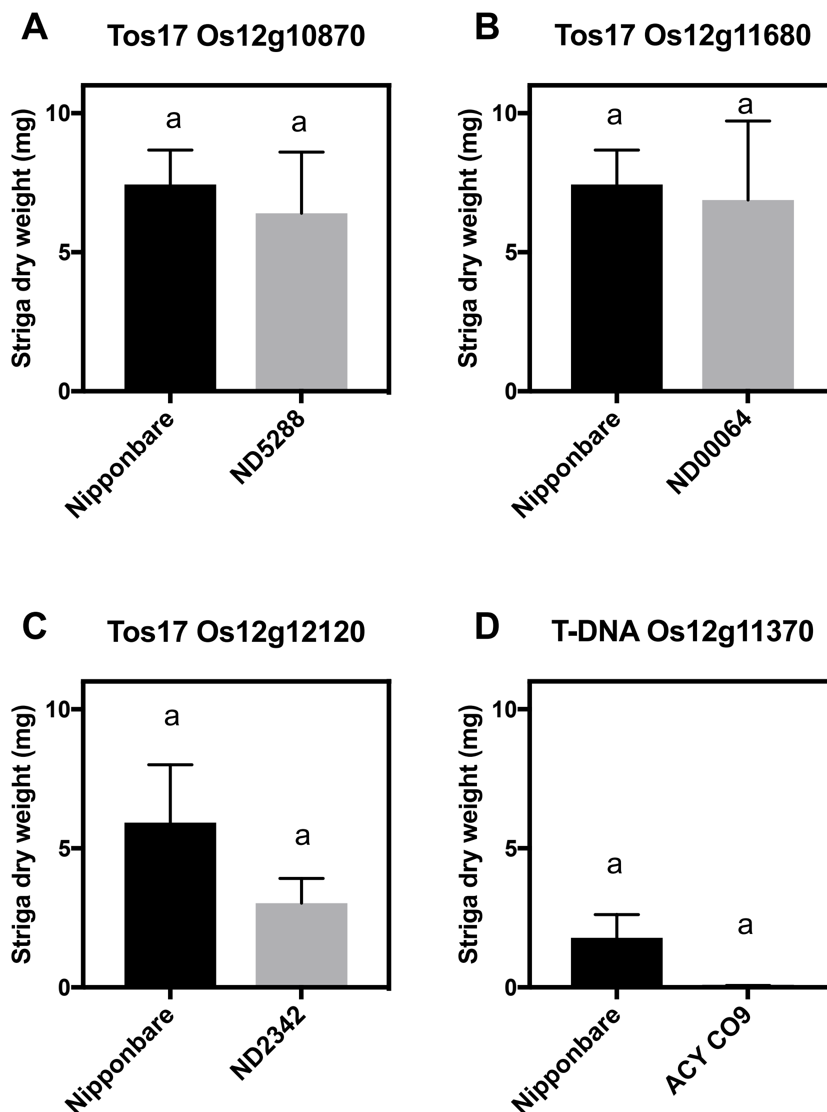


Figure 4.9 *S. hermonthica* dry weight harvested from the roots of lines carrying insertion in the genes Os12g10870, Os12g1168, Os12g12120 (all *Tos17* insertions) or Os12g11370 (T-DNA insertion) 25 dai. Values are means ± SE where n is a minimum of 7. There were no significant differences between insertion lines and Nipponbare wild type.

Figure 4.10 Verification of gene knockout for insertion lines by PCR amplification of cDNA. For each line, PCR amplification of cDNA was carried out on 4 replicates, with 2 replicates of Nipponbare cDNA and 1 of Nipponbare gDNA as positive controls. No transcript was observed for the *Tos17* insertion lines containing insertions in Os12g11680, Os12g10870 or Os12g12120 **(A)**. Transcript of Os12g11370 was observed for the T-DNA insertion line containing an insertion in this gene **(B)**.

Figure 4.11 Sequenced products of Os12g11370 T-DNA insertion line from cDNA, using primers within the insertion and at the end of the gene, aligned to Os12g11370 sequence (top), showing the position of the insertion. Blue indicates the insertion sequence position; red indicates where the sequenced products match that of Os12g11370.

4.4 Discussion

In this chapter RNAi lines and insertion lines were used to test the hypothesis that reducing the expression of one or more of the candidate RLP genes in the resistant cultivar Nipponbare will result in increased susceptibility to *S. hermonthica* (Kibos isolate). RNAi lines were phenotyped for post-attachment resistance to *S. hermonthica*, and half the lines were found to be significantly more susceptible than Nipponbare. To test the role of the RLP genes in the increase in susceptibility, the expression of the RLP genes were measured in a selection of RNAi lines showing both an increase in susceptibility or no change in susceptibility to *S. hermonthica.*

4.4.1 The role of the RLP genes in *S. hermonthica* resistance

Transcript levels of the target gene (against which the trigger sequence was designed) were significantly reduced for three out of the six susceptible RNAi lines, however in all these cases other RLP genes were also significantly reduced. This was expected given the similarity of the gene sequences, but meant that it was not possible to relate the increase in susceptibility to *S. hermonthica* to any one gene in the cluster, as no individual gene had significantly reduced expression in all susceptible lines. This strongly suggests that if the RLP genes are involved in resistance to *S. hermonthica*, more than one RLP gene is likely to be involved. The insertion line data is consistent with this hypothesis; knocking out a single gene in the cluster did not lead to any increase is susceptibility, although it is possible that none of the genes knocked out by the insertion are involved in resistance. RNAi lines that did not show a significant increase in susceptibility compared to wild type Nipponbare exhibited less (if any) down regulation of gene expression.

Although a significant increase in susceptibility to *S. hermonthica* was observed in several RNAi lines, this increase was only modest and rice plants did not support the huge numbers of parasites seen on very susceptible cultivars such as Azucena. This was unexpected; the highly significant LOD score of the QTL and the fact that no other major QTL were detected suggests the majority of resistance to *S. hermonthica* derives from this QTL region. It is therefore possible that other genes in the region may also contribute to the resistance. This is consistent with the findings of Hegenauer *et al.*, (2016) who showed that although the RLP gene *CuRe1* in tomato was sufficient to confer increased resistance to the parasitic plant *C. reflexa*, when transformed into a susceptible wild tomato, this resistance was incomplete. Thus full resistance to *C. reflexa* required more than just perception of the parasite by *CuRe1* alone (Hegenauer *et al.*, 2016). If resistance to *S. hermonthica* in Nipponbare is conferred by a similar

mechanism, it is possible that an additional gene or genes in the QTL may be acting in addition to one or multiple RLP genes to provide resistance.

Previous studies have found that clusters of both highly homologous resistance genes and clusters of dissimilar genes at a locus can provide enhanced resistance to biotic stresses. The *Bph3* locus in rice confers broad-spectrum and durable resistance to the brown planthopper (BPH) *Nilaparvata lugens*, a phloem-feeding herbivore. Liu *et al.* (2015) showed that the *Bhp3* locus contains a cluster of three plasma membrane-localised lectin receptor kinase genes (OsLecRK1 – OsLecRK3), all of which contribute to enhanced resistance. Sequence comparisons between resistant and susceptible rice cultivars revealed that all amino acid sequences were identical between resistant cultivars, but several nucleotide polymorphisms causing substitutions, frame shifts or early stop codons were identified in susceptible varieties. Transformation of each of the three genes individually into a susceptible cultivar resulted in significantly higher resistance to BPH, however transgenic plants co-expressing all three genes exhibited higher resistance than single-gene transformants. Analysis of transcripts by qPCR revealed differing degrees of correlation between OsLecRK gene expression and resistance. However, suppression of gene expression for all three OsLecRK genes in RNAi NIL lines carrying the *Bhp3* locus correlated with a reduction in BPH resistance. All this confirmed that the OsLecRK genes act together to confer resistance to BPH (Liu *et al.*, 2015). Overexpression of individual and multiple RLP genes from Nipponbare into a susceptible cultivar, or a susceptible BIL line from the mapping population lacking the QTL sequence, could therefore help identify which RLP genes are involved in resistance, and whether they act together to provide further resistance.

Different types of R genes are also known to act together. A study on resistance to the soybean cycst nematode (SCN: *Heterodera glycines*) showed a set of dissimilar genes to be involved. Interestingly, copy number of these genes, rather than their presence, was important in conferring the resistance (Cook *et al.*, 2012). Three genes within the *Rhg1* locus, encoding an amino acid transporter, an α-SNAP protein and a wound-inducible domain protein, were shown to be involved in SCN resistance. RNAi silencing of any one of three genes in the resistant soybean variety significantly reduced SCN resistance, which was dependent on reduction of the target transcript. Fluorescence *in situ* hybridisation was used to show the arrangement of the 31 kb region containing these genes in both resistant and susceptible lines. One copy of this region was detected in susceptible varieties, but 10 tandem copies were present in resistant varieties. The greater copy number of these three genes, and thus their increased level of expression, was responsible for the resistance observed (Cook *et al.*, 2012).

This suggests resistance in this system is conferred in a dose-dependent manner. Resistance in Nipponbare to *S. hermonthica* is also thought to be dose-dependent, as the F_1 plants exhibited intermediate resistance between the parents (Chapter 2). If this dose-dependent resistance to *S. hermonthica* is conferred by many RLP genes, some of which have undergone duplication, suppression of multiple RLP genes would be required in order to see a significant increase in susceptibility.

The resistance to *S. hermonthica* observed in the *Tos17* insertion lines is consistent with the hypothesis that multiple RLP genes provide resistance to *S. hermonthica* in Nipponbare, as no increase in susceptibility was observed, and is consistent with the findings of RNAi lines. If a single gene confers resistance however, none of the genes knocked out are involved. Unusually, mRNA transcript of the target gene was detected in the Os12g11370 T-DNA insertion line. The presence of the insertion was confirmed by sequencing, and the final 2247 bp of Os12g11370 (from the insertion site to the end of the gene) was transcribed. Although rare, the presence of an insertion within a gene does not always result in gene knock out. Wang (2008) investigated the effectiveness of T-DNA insertion mutagenesis in *Arabidopsis*, and found that over 90 % of insertions present within the protein coding region of a gene generated a knockout. In most cases, even if transcript was observed, the presence of the T-DNA sequence resulted in early stop codons and the termination of translation. Insertion in the 5' end of a gene can sometimes fail to knock it out, and instead lead to an overall reduction in protein levels and a less pronounced phenotype (Pruzinská *et al.*, 2007). However, this is the exception not the rule (Wang, 2008). Like the insertion line used here, most T-DNA insertions contain a promoter sequence such as the Cauliflower Mosaic Virus 35S. In some cases, this promoter has been shown to drive the expression of the downstream gene, producing a chimeric transcript, which can result in increased transcript abundance. This may or may not result in a protein synthesis, or the protein may be translated with poor efficiency or in the incorrect frame (Delatte *et al.*, 2005; Wilmoth *et al.*, 2005). Increased transcript abundance was observed for Os12g11370 in the T-DNA line tested here when compared to the expression level seen in Nipponbare (Figure 4.10). Interestingly, these rice plants also showed almost full resistance to *S. hermonthica*, which is not usually seen even on the most resistant rice cultivars. However, without further investigation, it is unclear whether this transcript is translated into a protein, or whether or not it could be functional.

4.4.2 The suppression of RLP transcript in RNAi lines

The efficiency of RLP gene silencing observed in RNAi lines varied between different lines for the same construct, but in most cases lines showing increased susceptibility to *S. hermonthica*

also showed down regulation in RLP gene expression. One exception was RNAi line JS4.12B-11, which was significantly more susceptible to *S. hermonthica* than Nipponbare, despite no significant RLP down regulation. It is possible that some other genetic alterations may have taken place during the transformation process that may have affected susceptibility. Indeed, the genetic instability of plants that have been regenerated from tissue culture is well documented; base pair mutations, deletions, chromosomal rearrangements or altered DNA-methylation patterns may all be imposed on cells undergoing tissue culture (Phillips *et al.*, 1994; Li *et al.*, 2007; Neelakandan & Wang, 2012). The activity of transposable elements may also be affected; for example the *Tos17* retrotransposon is known to be induced as a result of stress from tissue culture, with copy number increasing with the duration of tissue culture (Piffanelli *et al.*, 2007).

However, with the exception of JS4.12B-11, lines showing poor suppression in RLP transcript also failed to show an increase in susceptibility to *S. hermonthica*. There are multiple explanations that could account for the lack of significant gene suppression in these lines, as effectiveness of RNAi is determined by many factors. For example the presence of a large number of targets in a cell is known to reduce the efficacy of siRNAs. Arvey *et al.* (2010) showed that when faced with a larger number of predicted target transcripts, down-regulation of each individual target gene by siRNAs or miRNAs was achieved to a lesser extent. This was termed the *dilution effect*, as presence of many target molecules dilutes the effect of the small RNAs (Arvey *et al.*, 2010). As many RLP genes were predicted to be secondary targets for the constructs used in this study, the silencing efficiency of the constructs for individual genes may have been reduced.

The low levels of expression of the RLP genes in Nipponbare generally may also explain the lack of silencing in some lines, as the abundance of the target mRNA itself is also known to effect silencing; highly expressed genes are generally more susceptible to siRNA- mediated silencing (Hong *et al.*, 2014). In *C. elegans*, a stronger RNAi phenotype was observed for target genes of higher expression (Cutter *et al.*, 2003). This has also been observed in rice. Simultaneous suppression of multiple related genes with a single trigger sequence showed that genes of low expression levels were silenced to a lesser extent, despite greater sequence homology with the trigger compared to more abundantly expressed genes (Miki *et al.*, 2005). In *Arabidopsis*, genes expressed at moderate to high levels in wild-type plants were strongly reduced in RNAi lines, while those whose transcript was not detectably reduced in the RNAi lines were expressed at low level in the wild-type (Kerschen *et al.*, 2004). This is likely due to easier access of the more abundant mRNA target to siRNAs and RISC. However transcript

abundance is not the only factor determining the effectiveness of RNAi in *Arabidopsis*. The rate of RNA turnover, sequence composition and spatial and temporal expression patterns are other possible factors that may be important (Kerschen *et al.*, 2004).

Finally, the presence of multiple copies of the transgene can also affect gene silencing by siRNAs. Analysis of RNAi lines targeting 25 endogenous genes in *Arabidopsis* showed that independent single copy, homozygous lines targeting the same gene generally reduced expression levels to the same extent. Target mRNA suppression in multi-copy lines never exceeded the reduction seen in single copy lines, and mRNA levels were frequently reduced to a lesser extent in multi-copy lines (Kerschen *et al.*, 2004). This is consistent with the possibility that the transgenes in multi-copy lines are subject to some degree of silencing themselves, reducing their effectiveness. As single-copy homozygous lines were unavailable for the RNAi lines examined here, it is possible that poor silencing in some lines was due to differences in both copy number and zygosity.

Nevertheless, despite only half the RNAi lines showing reductions in RLP gene expression relative to Nipponbare wild type, expression of the *gus*-linker sequence indicated that where constructs were well transcribed, expression of the target gene was generally reduced (Figure 4.7). With the exception of JS4.12B-11, all lines showing an increase in *S. hermonthica* susceptibility also showed some level of RLP gene suppression, consistent with a role in *S. hermonthica* resistance.

4.4.3 Conclusions and future directions

Although this chapter has provided some evidence for the involvement of the RLP gene cluster in providing resistance to *S. hermonthica*, further functional evidence is required to relate this to individual genes or gene combinations. It is likely that the low abundance of these genes and large numbers of potential targets, as well as variations in copy number and zygosity of the RNAi lines, has obscured any clear trends. A complete gene knockout of each of the RLP genes individually and in combinations is required to test the extent to which the RLP genes may provide enhanced resistance to *S. hermonthica*. CRISPR/Cas9 technology offers a promising new approach to achieve this. As CRISPR/Cas9 edits gene sequences at the DNA level (Belhaj *et al.*, 2015; Shalem *et al.*, 2015), problems due to poor silencing of mRNA would be avoided. Work on this has already begun. RLP genes are known to associate with the LRR-receptor-like kinase SOBIR1 which is essential for the triggering of downstream defence responses (Liebrand *et al.*, 2013), and SOBIR1 orthologs are found throughout the Plant kingdom (Liebrand *et al.*, 2014). Therefore targeting the SOBIR1 rice homolog would also be

advantageous, as it should result in loss-of function of all the RLPs in the cluster. This would address the extent of the RLP genes involvement in resistance. Full susceptibility in plants lacking SOBIR1 would suggests the RLP genes are very important, while a partial increase in susceptibility could indicate other genes are involved. Overexpression of the RLP genes from IR64 and / or Nipponbare into a susceptible cultivar such as Azucena will also help elucidate the role of the RLP genes in providing *S. hermonthica* resistance. Finally, knowledge of the presence and allelic diversity of the genes within the *S. hermonthica* resistance QTL in diverse rice cultivars, and how this correlates with resistance, will not only be important in marker assisted breeding but could also help narrow down top candidate resistance genes or help identify new alleles for breeding of more durable resistance.

Chapter 5

How important is the resistance on chromosome 12 in diverse rice cultivars?

5.1 Introduction

In chapter 4 an analysis of the involvement of the RLP genes in resistance to *S. hermonthica* was carried out by down-regulating the expression of suites of these genes in Nipponbare by RNAi. Although this study provided the first evidence that these genes are likely to be involved in resistance to *S. hermonthica*, as some transformants showed an increase in susceptibility, it did not reveal which gene or combinations of genes underlie the resistance. The identification of the *S. hermonthica* resistance QTL on chromosome 12 in both IR64 and Nipponbare suggests that this region may be important in *S. hermonthica* resistance in diverse rice germplasm. The aim of this chapter is to examine the diversity of candidate resistance genes in the *S. hermonthica* resistance QTL in a range of rice genotypes with known resistance to *S. hermonthica* to test the hypothesis that their diversity can help identify genes or combinations of genes underlying resistance to *S. hermonthica*.

5.1.1 Diversity of the genus *Oryza*

Rice is the world's most important food crop, and the staple cereal for more than half the world's population. Rice is grown in a wide range of climates worldwide, in tropical, sub-tropical and temperate regions at both high and low altitudes, including irrigated, rainfed or deep water environments (Balasubramanian *et al.*, 2007; Seck *et al.*, 2012). In order to meet the growing demands of an increasing population, rice production will need to increase by at least 25 % by 2035. In Africa, where demand for rice is growing faster than any food source, an increase of 130 % will be required compared to 2010 production levels, posing a significant challenge to rice research and the development of improved genotypes (Zhang, 2007; Seck *et al.*, 2012).

The rice genus (*Oryza*) comprises 22 wild species and 2 cultivated species. These have been classified into 10 genome groups, six diploid (AA, BB, CC, EE, FF, GG) and 4 allotetraploid (BBCC, CCDD, HHJJ and KKLL) (Ge *et al.*, 2001; Vaughan *et al.*, 2003; Ammiraju *et al.*, 2010). The AA genome includes 6 wild rice species and the 2 cultivated species *O. glaberrima* and *O. sativa.* The *O. glaberrima* species is indigenous to Africa, having originated around 3,500 years ago in the swampy basins of Niger in West Africa (Sarla & Swamy, 2005). It possesses many unique traits including good resistance to pests and diseases, drought tolerance, weed competitiveness and ability to grow in low nutrient conditions, but has low grain numbers and high seed dormancy and shattering (Sarla & Swamy, 2005). In contrast, *O. sativa* has higher yields and reduced seed shattering, but more limited resistance to the stresses of the African climate. Although *O. sativa* originated in Asia, it was introduced into East Africa around 2,000

years ago, and subsequently into West Africa around 500 years ago, and is now grown worldwide (Sarla & Swamy, 2005; Sweeney & McCouch, 2007; Seck et al., 2012).

During the course of domestication, much of the genetic diversity has been lost in cultivated rice. The selection of desirable agronomic traits has resulted in lower heterozygosity and as much as a 60 % reduction in allele number (Sun *et al.*, 2001), leaving it more prone to pests, diseases and abiotic stresses. Nevertheless, considerable within-species diversity still exists, reflecting the complex course of domestication and divergence of the different subspecies. Two major subspecies exist within the *O. sativa* lineage: *indica,* which is predominantly found in the lowlands throughout tropical Asia; and *japonica,* of which the tropical *japonica* variety are mostly upland and temperate *japonica* variety are irrigated*.* The divergence of these 2 subspecies occurred around 0.44 millions years ago, which predates domestication, and suggests *indica* and *japonica* were domesticated independently from differentiated ancestral pools (Sang & Ge, 2013). Indeed, the source of the *japonica* gene pool is now thought to derive from south China and the Yangtze valley, while the *indica* gene pool is thought to have originated from Indochina and the Brahmaputra valley (Civán *et al.*, 2016). Although there is considerable debate over whether these species were derived from a single domesticated ancestor or domesticated independently in different locations, most recent evidence suggests three independent domestications of rice (Civán *et al.*, 2016). However multiple introgressions have taken place between them during their evolution which have played an important role in shaping their current genetic structure (Sang & Ge, 2013). Gene flow has largely been unidirectional from *japonica* to *indica* (Yang *et al.*, 2012b). Hybridisation of *indica* with local wild rice is also thought to have taken place (Yang *et al.*, 2012b; Gross & Zhao, 2014).

Indica and *japonica* cultivars can be further divided into five genetically distinct groups: *indica*, *aus*, tropical *japonica*, temperate *japonica* and *aromatic*. Tropical *japonica*, temperate *japonica* and *aromatic* cultivars share a closer evolutionary relationship, while *indica* and *aus* cultivars are more closely related to each other (Garris *et al.*, 2005). These varietal groups have now been confirmed using SNP data for 3000 rice genomes (The 3000 Rice Genomes Project, 2014), shown in Figure 5.1.

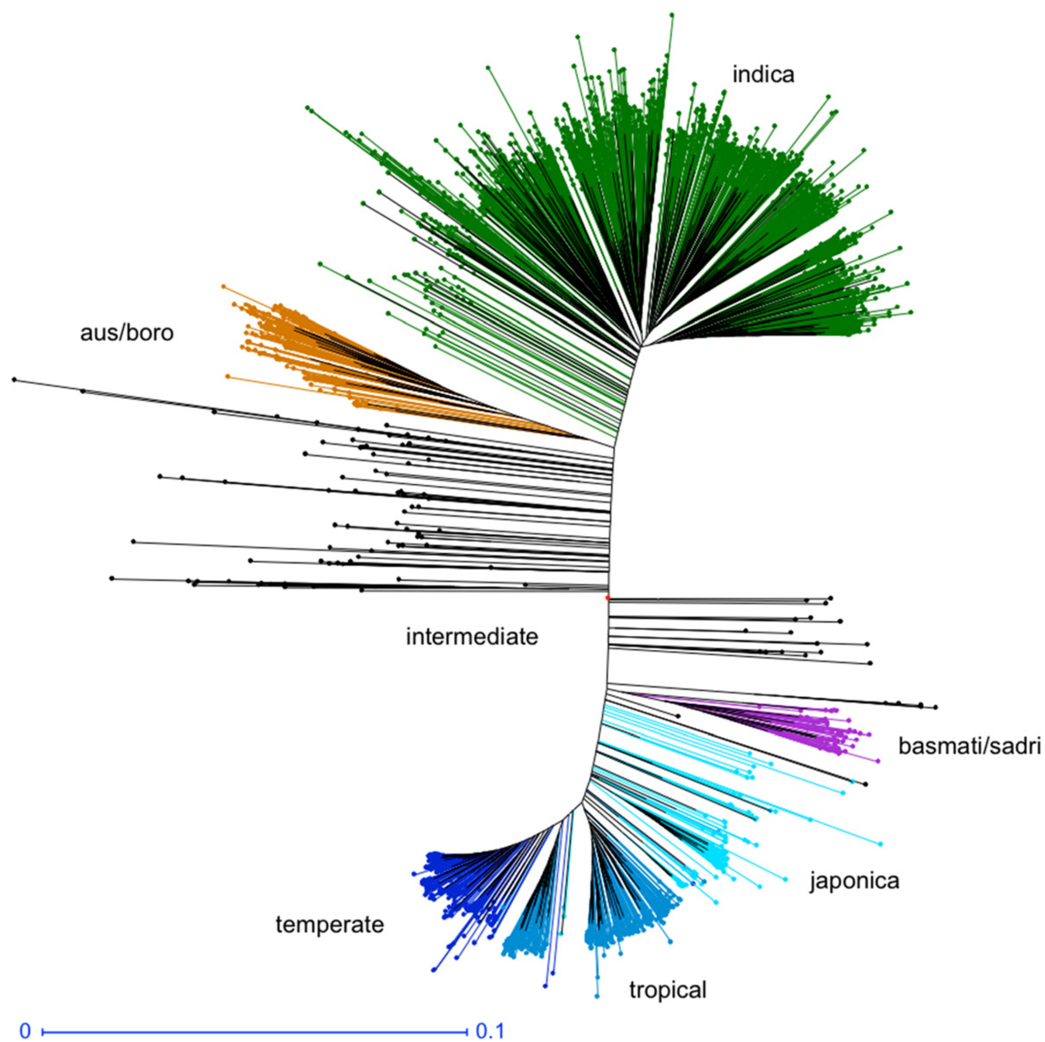


Figure 5.1 Classification of 3000 rice genomes into 5 varietal groups using 5 x 200,000 random sets of SNPs from a total of 18.9 million SNP variants. The *aromatic* group here is labelled as basmati/sadri. Figure taken from The 3000 Rice Genomes Project, 2014.

Indica cultivars are the most diverse of these groups (Garris *et al.*, 2005; Huang *et al.*, 2010), and provide the greatest rice production globally. Tropical and temperate *japonicas* are closely related to each other, and have lower genetic diversity and larger allele size, particularly the temperate *japonicas*. The lower diversity of tropical *japonicas* may reflect their cultivation on Indonesian islands, with fewer opportunities for cross pollination, leading to genetic bottlenecks (Garris *et al.*, 2005). Temperate *japonica* cultivars are thought to have derived from tropical *japonicas*, and show environmental adaptations to temperate latitudes such as day length and temperature. *Aromatic* cultivars are closely related to *japonicas* and are now believed to have arisen as a result of hybridization between *japonica* and *aus* (Figure 5.2) (Civán *et al.*, 2016). They posses a similar level of genetic diversity to the *japonica*s (Garris *et al.*, 2005) with a high proportion of monomorphic markers suggesting a recent or severe

bottleneck (Nagaraju *et al.*, 2002) and are valued for their aroma and quality, including *basmati* from Nepal, India and Pakistan, and *sadri* from Iran. In contrast, *aus* varieties from Bangladesh are more diverse and harbour important alleles for tolerance of drought (Bernier *et al.*, 2009), submergence (Xu *et al.*, 2015), phosphorous deficiency (Gamuyao *et al.*, 2012), disease resistance (Garris *et al.*, 2003) and early maturity (Yano *et al.*, 2000), meaning they are of particular interest for the development of genetically improved rice.

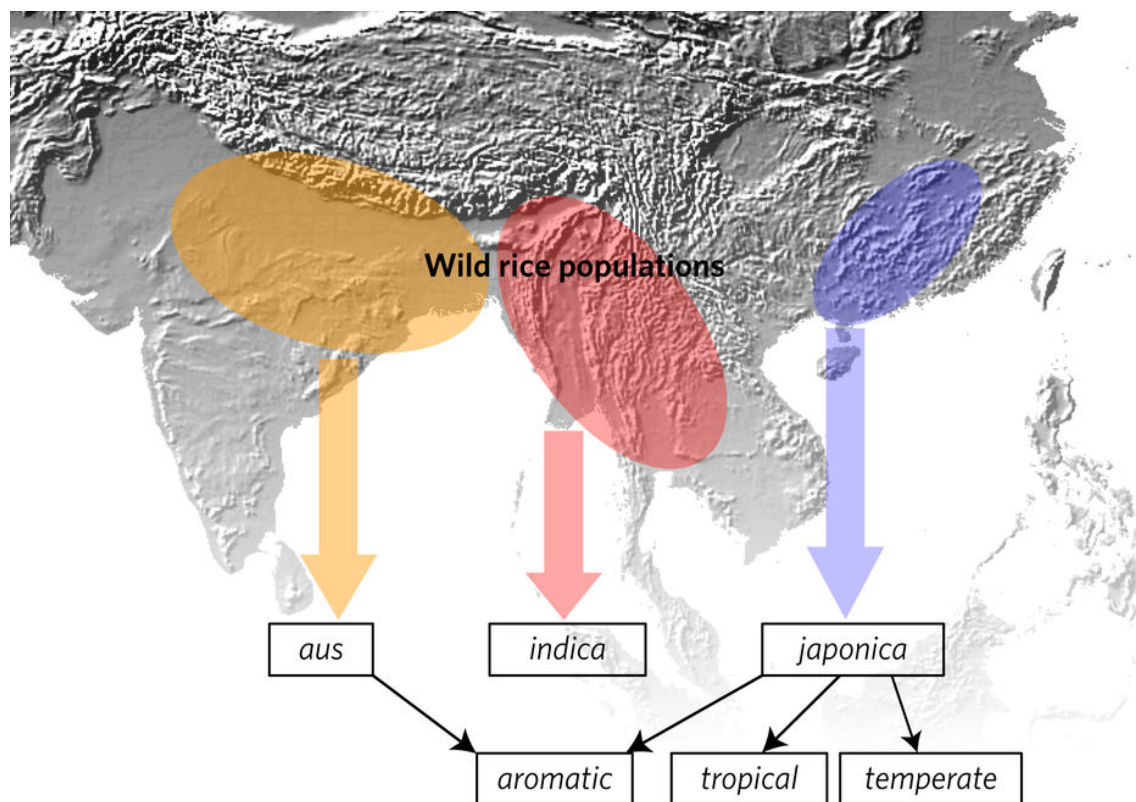


Figure 5.2 Proposed origins for domesticated rice derived from phylogenetic analysis. Three independent domestications are thought to have taken place in southern and southeast Asia, with *aus* and *japonica* hybridising to form *aromatic*, and later the divergence of *japonica* into temperate and tropical versions. Figure taken from Civán *et al.*, 2016.

Rapid advances in next-generation sequencing technologies over the last decade have been used to generate whole genome sequences of a wide variety of plant species and subspecies, including rice. Hundreds of rice varieties have now been genotyped from both wild and cultivated varieties (Huang *et al.*, 2010; Xu *et al.*, 2011; The 3000 Rice Genomes Project, 2014; Duitama *et al.*, 2015), which have lead to a much greater understanding of the diversity within this species. In a diversity panel developed by Duitama *et al.*, (2015) 104 elite *O. sativa* rice varieties were sequenced to identify high quality SNPs, indels, repeats and copy number variants. Genetic distances between accessions were estimated from SNPs, which separated *indica*, *aromatic*, *aus*, temperate *japonica* and tropical *japonica* populations.

5.1.2 Aims of Chapter 5

In this chapter SNP data of these diverse rice cultivars was used to examine the diversity of candidate resistance genes in the *S. hermonthica* resistance QTL, to test the hypothesis that their diversity can help identify genes or combinations of genes underlying resistance to *S. hermonthica*. The specific objectives of this chapter are:

1) To compare the allelic diversity of 112 rice genotypes with that of Nipponbare across the *S. hermonthica* resistance QTL region.
2) To compare SNP similarity to Nipponbare of individual genes in the QTL region for a selection of diverse rice cultivars.
3) To phenotype rice cultivars for post-attachment resistance to *S. hermonthica*.
4) For each gene, to carry out a correlation analysis between SNP similarity to Nipponbare and resistance to *S. hermonthica*.

However, there are also many assumptions and caveats that must be taken into account when performing this kind of analysis, which means that care must be taken when interpreting the data. The assumptions are: (1) resistance is due to large differences in gene structure rather than a single SNP; (2) homologs of any candidate resistance genes in susceptible cultivars share low similarity to those of Nipponbare; (3) resistant cultivars may be resistant because they contain resistance genes in other parts of the genome and (4) resistance in IR64 and Nipponbare is due to the same gene cluster. It must also be mentioned that only 26 rice accessions were available for phenotyping in this study, which is not sufficient for a robust association analysis. Therefore the work in this chapter must be regarded as additional data to support or refine our understanding of the genetic basis of resistance provided in previous chapters.

5.2 Materials and Methods

5.2.1 Plant materials

Seeds of the rice diversity panel were supplied by French Agricultural Research Centre for International Development (CIRAD), Montpellier, France. Illumina sequencing was carried out to identify SNP markers. Sequencing of 18 of these cultivars was carried out at > 8 x coverage as described in Duitama *et al*., (2015). The remaining cultivars were sequenced at > 15 x coverage as described in Xu e*t al*., (2011). *S. hermonthica* seeds were collected from plants parasitising maize in Kibos, Western Kenya, in 2013.

5.2.2 Comparison of SNPs between cultivars and predictions of resistance and susceptibility to *S. hermonthica*

SNP data covering the *S. hermonthica* resistance QTL for 112 diverse rice cultivars was provided by Dr. Mathias Lorieux (CIAT/IRD), and used to make broad predictions of resistance or susceptibility to *S. hermonthica* for a selection of the cultivars. Cultivars with greatest SNP similarity to Nipponbare across the QTL were predicted to be more resistant, while those with least similarity were predicted to be more susceptible. Twenty-six rice cultivars were selected for detailed analysis (Table 5.1) to including examples of the five rice genetic groups *indica*, *aus*, temperate *japonica*, tropical *japonica* and *aromatic.* These cultivars were selected to show similarity and differences in genotype with respect to Nipponbare across the *S. hermonthica* resistance QTL region.

SNP data for individual genes within the QTL region was provided in an Excel spreadsheet by Dr. Mathias Lorieux (CIAT/IRD) (Table 5.2). Rice cultivars were sorted and compared by SNP similarity to Nipponbare across 19 genes covering the *S. hermonthica* resistance QTL. SNP similarity to Nipponbare was calculated for each cultivar and gene; the sum of the matching SNPs was divided by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent).

Table 5.1 The 26 rice cultivars selected for phenotyping for post attachment resistance to *S. hermonthica* (Kibos isolate) and their genetic grouping. All cultivars are *Oryza sativa* species.

Subspecies	**Cultivar name**
Aromatic	Firooz [a]
Aromatic	Kitrana508 [a]
Aromatic	Darmali [a]
Aus	Kalamkati [a]
Aus	Mehr [a]
Aus	Miriti [a]
Aus	Jhona349 [a]
Indica	IR8 [b]
Indica	IR36 [a]
Indica	JC91 [a]
Indica	IR64 [b]
Indica	Gie57 [a]
Indica	Guan-Yin-Tsan [a]
Indica	Ai-Chiao-Hong [a]
Indica	Jasmine85 [b]
Temperate japonica	Nipponbare [a]
Temperate japonica	Bengal [b]
Temperate japonica	Fanny [b]
Temperate japonica	Mansaku [a]
Temperate japonica	Chodongji [a]
Tropical japonica	Dixiebelle [b]
Tropical japonica	Davao [a]
Tropical japonica	Trembese [a]
Tropical japonica	Curinga [b]
Tropical japonica	Azucena [a]
Tropical japonica	Binulawan [a]

a) Sequenced by Xu et al 2011
b) Sequenced by Duitama et al 2015

Table 5.2 The 19 genes across the Nipponbare *S. hermonthica* resistance QTL examined in this study, and their annotation.

Genes in Nipponbare	Annotation
LOC_Os12g10670	AAA-type ATPase family protein, putative, expressed
LOC_Os12g10820	Expressed protein
LOC_Os12g10850	hhH-GPD superfamily base excision DNA repair protein, putative, expressed
LOC_Os12g10870	Verticillium wilt disease resistance protein, putative, expressed
LOC_Os12g10930	NL0E, putative, expressed
LOC_Os12g11370	Verticillium wilt disease resistance protein, putative, expressed
LOC_Os12g11500	resistance protein SlVe1 precursor, putative, expressed
LOC_Os12g11510	hcr2-0B, putative, expressed
LOC_Os12g11660	RALFL45 - Rapid ALkalinization Factor RALF family protein precursor, expressed
LOC_Os12g11680	Verticillium wilt disease resistance protein precursor, putative, expressed
LOC_Os12g11720	Verticillium wilt disease resistance protein precursor, putative, expressed
LOC_Os12g11860	Verticillium wilt disease resistance protein precursor, putative, expressed
LOC_Os12g11930	Disease resistance protein SlVe2 precursor, putative, expressed
LOC_Os12g11940	Disease resistance family protein, putative, expressed
LOC_Os12g12000	RALFL46 - Rapid ALkalinization Factor RALF family protein precursor, expressed
LOC_Os12g12010	Verticillium wilt disease resistance protein precursor, putative, expressed
LOC_Os12g12120	Verticillium wilt disease resistance protein precursor, putative, expressed
LOC_Os12g12130	Verticillium wilt disease resistance protein, putative, expressed
LOC_Os12g12514	NADP-dependent oxidoreductase, putative, expressed

5.2.3 Phenotyping rice cultivars for post-attachment resistance to *S. hermonthica*

Rice cultivars (Table 5.1) were phenotyped for resistance to *S. hermonthica* according to the protocol described in section 2.2.2. Five rice plants of each cultivar were infected with *S. hermonthica* (Kibos ecotype), making a total of 145 plants. Plants were split into 3 batches of approximately 49 plants each. The 5 replicates were distributed randomly between the 3 batches to help calibrate for any variation that may occur between them. Batches 1 and 2 were separated by 2 days, and batches 2 and 3 were separated by 5 days. Dry biomass, number and length of parasites harvested from the roots of each plant were recorded 21 days after inoculation.

5.2.4 DNA extraction and quantification, and genotyping of cultivars by PCR amplification

In order to verify the SNP sequencing for our own rice material, rice cultivars (listed in Table 5.1) were genotyped using PCR markers for 5 genes (Os12g11370, Os12g11500, Os12g11680, Os12g11860 and Os12g12010) across the *S. hermonthica* resistance QTL. The susceptible cultivar Koshihikari is known to contain a homolog of both Os12g11370 and Os12g11500, therefore primers for these genes were designed against both the Nipponbare and Koshihikari gene sequences. For Os12g11370 it was possible to design primers that amplified PCR products of different lengths for these cultivars. However this was not possible for Os12g11500, therefore two different sets of primers were used, one specific to the Nipponbare sequence, and one specific to the Koshihikari sequence. All other primers were designed against the Nipponbare gene sequences. Primers and product sizes are shown in Table 5.3. For DNA extraction, approximately 80 – 100 mg leaf material was collected from each rice cultivar for genotyping. DNA extraction was carried out according to the Qiagen DNeasy Plant Mini Kit as described in section 3.2.4.1. The quality and quantity of DNA in the samples was checked using a Nanodrop spectrophotometer (ND-8000, Thermo Scientific). DNA samples were diluted to a working concentration of 10 ng / µl, and 5 µl of this was used in each 50 µl PCR reaction. This consisted of 25 µl Taq Master Mix, 5 µl DNA, 1 µl of each primer (10 µM each, to give a final concentration of 0.2 µM) and 18 µl nuclease-free water. The PCR program was: 3 min at 94 °C (initial denaturation); then 32 cycles of 30 s at 94 °C (denaturation), 15 s at 58 °C (annealing), and 1 min 30 s at 72 °C (extension); followed by a final extension of 10 min at 72 °C. Where product sizes were less than 500bp, extension was reduced to 1 min. Eight µl PCR product and 2 µl DNA Loading Buffer (Bioline) was loaded onto a 1.5 % agarose gel.

5.2.5 PCR amplification and sequencing of full-length genes from rice cultivars

Following examination of the genotyping results for the 5 genes, the genes Os12g11370 and Os12g11500 were selected for amplification of the full-length gene by PCR from 4 cultivars (Jhona349, Jasmine85, Fanny and Darmali) as described in section 3.2.5. Primers were designed against the Nipponbare sequence a short distance outside the start / stop sites to ensure amplification of the entire gene sequence. Where successful amplification was achieved, the PCR purified genes were sent for Sanger sequencing (described in section 3.2.5) and their sequences compared to the Nipponbare reference sequences. The primers used for gene amplification are shown in Tables 5.4. Internal sequencing primers are in Table 3.2.

Table 5.3 Primer sequences used for genotyping rice cultivars for five candidate resistance genes in the *S. hermonthica* resistance QTL region. Primer sequences were designed against Nipponbare gene sequences, with the exception of Os12g11500 where a second set of primers was designed against the Koshihikari sequence.

Gene	Primer name	Primer sequence (5' – 3')	Amplicon length in Nipponbare (bp)
Os12g11370	11370_genotype_F1	AATTCGCTCACTAGGATTGAGCTT	569
	11370_genotype_R1	AAGATCTGCGGAGGCACCTT	(852 in Koshihikari)
Os12g11500	11500 Nipp-IR64 F2	CTACAGTTACATACAAAGGGAATGA	240
	11500 Nipp-IR64 R1	TCT CTC CAG AAA GCT CAT TGA AA	
Os12g11500	11500 Kosh F1	TTCATCACTGGTATCACTCCAC	1229
	11500 Kosh R1	TCT TCC CAG AAA GCT TAT TTG AG	
Os12g11680	11680 specific F1	TGTCTGGACCCGTACCATCT	267
	11680 specific R1	TGGAGTAGGATGCTGCCGAA	
Os12g11860	11860 specific F1	ATCCCACAGGAGCTACCAT	207
	11860 specific R1	CTTCTCGGAAGCATGTGTCA	
Os12g12010	12010 specific F2	GGAGTTGGCACTTGGAGCTA	446
	12010 specific R2	AACTAATACTGGGGTAGGTCGC	

Table 5.4 Primer sequences used for amplification of full-length genes across the *S. hermonthica* resistance QTL for a selection of rice cultivars. Primer sequences were designed against Nipponbare sequences. Gene refers to the Nipponbare homolog.

Gene	Primer name	Primer sequence (5' – 3')	Amplicon length in Nipponbare (bp)
Os12g11370	11370 locus SEQ F2	GAACTGGGAAGAACTCAATGTGC	3624
	11370 locus SEQ R2	GTAGAGATCACATTTAGATGTGGG	
Os12g11500	11500 locus seq F1	ATTCATAAGAGCACACGCTTTCC	3421
	11500 locus seq R2	GCAATTCGTCCATATATCAGCTATG	

5.2.6 Statistical analysis

The statistical package R, version 3.3.0 (http://www.r-project.org) was used for all analyses. A one-way ANOVA was carried out to assess significant differences in *S. hermonthica* dry weight, and the number and length of *S. hermonthica* plants harvested from the roots of rice cultivars. A Tukey Multiple Comparison test was used to identify significant differences between cultivars. A Linear Mixed Model was also carried out to determine how much variation was due to differences between experimental batches compared to differences between cultivars, with batch and cultivar fitted as random effects. Tests were carried out on log_{10} transformed data, to adjust for non-normal distribution.

For each gene in the *S. hermonthica* resistance QTL, a Pearson's product-moment correlation analysis was carried out between S*. hermonthica* dry biomass and the SNP similarity to the Nipponbare allele, using SNP data for the 26 rice cultivars. The % variance explained by the linear model was determined by calculating the R^2 value.

5.3 Results

5.3.1 The genomic structure of the *S. hermonthica* resistance QTL in diverse rice cultivars, and predictions of resistance and susceptibility

Comparison of the SNP data for 112 rice cultivars revealed broad trends in genomic structure across the *S. hermonthica* resistance QTL (Figure 5.3). Of the thirteen temperate *japonica* cultivars, only four appeared to be similar to Nipponbare across the QTL region. All tropical *japonica* cultivars were very different to Nipponbare and appeared to be missing much of the region (Figure 5.3). In contrast, *aromatic* cultivars appeared to contain the region from 6.2 – 6.45 Mb, although differences with Nipponbare were observed, and poor similarity to Nipponbare was seen between 6.45 – 6.56 Mb (Figure 5.3). *Indica* cultivars varied in their similarity to Nipponbare; for most *indica* cultivars SNP data was alternate to Nipponbare and missing in many places, however four cultivars showed a greater similarity to Nipponbare across most of the region. Only five *aus* cultivars were included in the analysis and these varied in similarity to Nipponbare; in most cases SNP data was missing or very different to Nipponbare, although one cultivar did show greater similarity over much of the region (Figure 5.3). The two *Oryza glaberrima* cultivars showed poor similarity to Nipponbare. *Oryza rufipogon and Oryza nivara* accessions varied in their similarity to Nipponbare; where the region was present in these genotypes, differences to Nipponbare were often observed (Figure 5.3). Thus, extensive differences were seen not only between but also within subspecies for this region. For example some *indica* cultivars were more similar to some temperate *japonica* cultivars than other *indica* cultivars (Figure 5.3), which is in contrast to the diversity seen at the genome-wide level (Figure 5.2).

Twenty-six cultivars (Table 5.1) were selected for detailed analysis of the candidate *S. hermonthica* resistance genes from the five *O. sativa* subgroups (*indica*, *aus*, temperate *japonica*, tropical *japonica* and *aromatic)* to include the broad range of the diversity observed. Ideally more genotypes would have been examined but access to the seed was not available. For each cultivar, the SNP similarity to Nipponbare was calculated for 19 individual genes spanning the *S. hermonthica* resistance QTL (Table 5.5). This revealed different trends in gene similarity between the 26 rice cultivars and Nipponbare. For example, the temperate *japonica* cultivars Fanny and Bengal and the *aus* cultivar Mehr shared greatest similarity with Nipponbare alleles for most genes across the *S. hermonthica* resistance QTL (Table 5.5). With the exception of Os12g10930 where SNP data was missing or alternate to Nipponbare for most cultivars, cultivars fell broadly into 2 groups for the region between genes Os12g10820 and Os12g11510 (Table 5.5). All *indica* cultivars shared good similarity (75 – 98 %) with Nipponbare for these genes, along with Firooz and Kitrana508 (*aromatic*), and Kalamkati and Mehr (*aus*).

However, SNP data was missing or alternate for these genes in tropical *japonica* cultivars, as well as Darmali (*aromatic*), Miriti and Jhona349 (*aus*), and Mansaku and Chodongji (temperate *japonica*) (Table 5.5).

Homologs of the Nipponbare RLP genes Os12g11680, Os12g11720 and Os12g11860 and the RALF gene, Os12g11660, were clearly present in 7 cultivars Firooz, Kitrana508, Mehr, Guan-Yin-Tsan, Ai-Chiao-Hong, Bengal and Fanny, and SNP similarity varied from 63 - 100 %. Only 3 of these cultivars (Mehr, Bengal and Fanny) had any close similarity with Nipponbare for the RLP genes *Os12g11930* and *Os12g11940* (Table 5.5).

All cultivars, with the exception of Firooz, Kitrana508, Guan-Yin-Tsan and Ai-Chiao-Hong, possessed a homolog of the RALF gene Os12g12000. The cultivars missing Os12g12000 also appeared to be missing much of the neighbouring gene Os12g12010 that was present in all other cultivars (Table 5.5), although many SNPs were observed for some *indica* cultivars.

The 7 cultivars that contained a homolog of RLP gene Os12g11720 also contained a homolog of Os12g12120 (Table 5.5). SNP data was missing or alternate to Nipponbare in all other cultivars for both these genes. Os12g12130 appeared to be absent from *indica* cultivars along with Firooz (*aromatic*), Kitrana508 (*aromatic*) and Kalamkati (*aus*), however it was present in all tropical and temperate japonica cultivars and the 3 aus cultivars.

Two genes outside the QTL were included in the analysis (Os12g10670 and Os12g12514) to determine whether any correlation between *S. hermonthica* resistance and gene similarity to Nipponbare extended beyond the QTL. Homologs of these genes were present in all cultivars. *Indica* cultivars possessed several SNPs in Os12g10670 compared to Nipponbare, while many SNPs and gaps were seen in all cultivars for Os12g12514 (Table 5.5).

Predictions of resistance or susceptibility to *S. hermonthica* were made for the 26 selected cultivars based on broad trends of gene similarity to Nipponbare across the QTL. These predictions are shown in Table 5.6.

6.212 Mb ↑: predicted resistance genes 6.561 Mb

MSU 7 annotation

6300k 6400k 6500k

Accessions

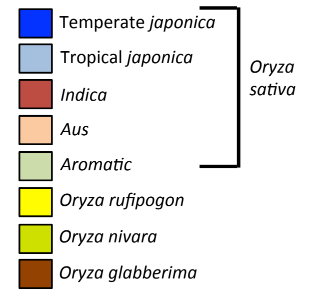


Figure 5. 3 Genomic structure of 112 rice accessions in the *S. hermonthica* resistance QTL (6.212 – 6.561 Mb) obtained from SNP sequencing data. Allele colour scheme is set to 'Similarity of SNPs to Nipponbare'. red: Nipponbare SNP; green: alternate SNP; white: missing / no data. Species and subspecies of accessions are indicated by colour on the right. Figure provided by Mathias Lorieux, CIAT.

Table 5.5 Similarity of SNPs to the Nipponbare allele for 19 genes across the *S. hermonthica* resistance QTL, in 26 diverse rice cultivars. SNP similarity to Nipponbare for each gene is the sum of the matching SNPs divided by the total number of SNPs, to give a value of similarity between 1 (identical allele) to 0 (no SNP similarity / gene absent). Colours are on a blue – yellow –white scale; blue: 100 % similarity, white: 0 % similarity. The QTL region is indicated above in green. Cultivars are grouped according to subspecies (all *Oryza sativa*).

Similarity to Nipponbare

			QTL region																	
Subspecies	**Cultivar**	**Os12g10670**	**Os12g10820**	**Os12g10850**	**Os12g10870**	**Os12g10930**	**Os12g11370**	**Os12g11500**	**Os12g11510**	**Os12g11660**	**Os12g11680**	**Os12g11720**	**Os12g11860**	**Os12g11930**	**Os12g11940**	**Os12g12000**	**Os12g12010**	**Os12g12120**	**Os12g12130**	**Os12g12514**
Aromatic	Firooz	1.000	0.778	0.724	0.856	0.229	0.942	0.039	0.762	0.636	0.880	0.784	0.791	0.143	0.247	0.000	0.329	0.825	0.000	0.800
Aromatic	Kitrana508	1.000	0.775	0.687	0.896	0.219	0.946	0.039	0.829	0.636	0.880	0.836	0.785	0.154	0.242	0.000	0.368	0.772	0.000	0.825
Aromatic	Darmali	0.985	0.451	0.233	0.072	0.267	0.507	0.432	0.410	0.000	0.152	0.172	0.029	0.367	0.209	0.727	0.961	0.105	0.654	0.759
Aus	Kalamkati	0.697	0.806	0.782	0.852	0.251	0.968	0.952	0.883	0.000	0.261	0.000	0.035	0.432	0.115	0.803	0.855	0.000	0.065	0.749
Aus	Mehr	1.000	0.768	0.695	0.728	0.749	0.604	0.541	0.816	0.818	0.935	0.733	0.890	0.672	0.753	0.848	0.961	0.868	0.774	0.803
Aus	Miriti	1.000	0.425	0.218	0.092	0.200	0.432	0.397	0.454	0.000	0.109	0.147	0.058	0.332	0.220	0.606	0.961	0.114	0.668	0.800
Aus	Jhona349	1.000	0.444	0.215	0.096	0.184	0.540	0.380	0.435	0.000	0.163	0.129	0.006	0.328	0.187	0.697	0.961	0.114	0.687	0.765
Indica	IR8	0.758	0.778	0.575	0.884	0.241	0.964	0.904	0.857	0.318	0.391	0.000	0.267	0.560	0.440	0.530	0.566	0.026	0.000	0.635
Indica	IR36	0.803	0.803	0.924	0.964	0.203	0.993	0.974	0.883	0.000	0.217	0.000	0.058	0.452	0.121	0.803	0.855	0.000	0.000	0.794
Indica	JC91	0.742	0.765	0.727	0.928	0.210	0.960	0.825	0.784	0.000	0.348	0.000	0.227	0.417	0.269	0.500	0.566	0.000	0.000	0.603
Indica	IR64	0.742	0.778	0.582	0.880	0.184	0.942	0.878	0.813	0.000	0.391	0.000	0.203	0.517	0.379	0.530	0.566	0.000	0.028	0.594
Indica	Gie57	0.788	0.806	0.891	0.988	0.225	0.993	0.974	0.886	0.273	0.217	0.138	0.041	0.498	0.165	0.788	0.921	0.009	0.346	0.829
Indica	Guan-Yin-Tsan	0.742	0.749	0.782	0.848	0.213	0.993	0.961	0.810	0.682	0.902	0.793	0.791	0.367	0.341	0.000	0.276	0.789	0.000	0.778
Indica	Ai-Chiao-Hong	0.727	0.781	0.913	0.952	0.000	0.964	0.965	0.816	0.636	0.891	0.802	0.773	0.405	0.313	0.000	0.368	0.746	0.000	0.581
Indica	Jasmine85	0.727	0.679	0.709	0.896	0.102	0.946	0.751	0.768	0.000	0.337	0.000	0.209	0.394	0.253	0.470	0.539	0.000	0.041	0.422
Temperate japonica	Nipponbare	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Temperate japonica	Bengal	0.879	0.727	0.829	0.960	0.889	0.838	0.908	0.844	0.955	0.935	0.940	0.901	0.907	0.802	0.788	0.895	0.877	0.912	0.495
Temperate japonica	Fanny	1.000	0.997	0.960	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000	0.991	1.000	0.914
Temperate japonica	Mansaku	1.000	0.416	0.298	0.072	0.254	0.392	0.371	0.305	0.000	0.196	0.164	0.064	0.317	0.225	0.636	0.961	0.088	0.834	0.762
Temperate japonica	Chodongji	1.000	0.359	0.236	0.152	0.263	0.529	0.389	0.406	0.000	0.141	0.155	0.064	0.382	0.187	0.697	0.961	0.105	0.608	0.806
Tropical japonica	Dixiebelle	1.000	0.470	0.309	0.172	0.308	0.482	0.406	0.495	0.000	0.337	0.129	0.221	0.398	0.269	0.682	0.947	0.096	0.724	0.822
Tropical japonica	Davao	1.000	0.425	0.247	0.096	0.200	0.468	0.410	0.416	0.000	0.152	0.129	0.006	0.336	0.225	0.652	0.947	0.088	0.687	0.810
Tropical japonica	Trembese	1.000	0.473	0.276	0.084	0.219	0.435	0.428	0.422	0.000	0.065	0.164	0.041	0.351	0.275	0.545	0.961	0.114	0.714	0.810
Tropical japonica	Curinga	0.985	0.546	0.269	0.092	0.241	0.468	0.476	0.438	0.000	0.141	0.147	0.064	0.390	0.236	0.712	0.961	0.105	0.811	0.806
Tropical japonica	AZUCENA	0.985	0.486	0.236	0.120	0.222	0.417	0.410	0.419	0.000	0.098	0.172	0.105	0.359	0.203	0.667	0.961	0.088	0.719	0.797
Tropical japonica	Binulawan	1.000	0.505	0.240	0.056	0.248	0.432	0.345	0.410	0.000	0.163	0.138	0.029	0.324	0.242	0.652	0.934	0.079	0.691	0.756

Table 5.6 Predictions of resistance or susceptibility to *S. hermonthica* for 26 genetically diverse rice cultivars, based on SNP similarity to Nipponbare for 17 genes across the S. *hermonthica* resistance QTL on chromosome 12 assuming that there are no other major genes contributing to resistance in the cultivars. Cultivars are grouped by subspecies (all *Oryza sativa*).

Subspecies	Cultivar	Prediction
Aromatic	Firooz	?
Aromatic	Kitrana508	?
Aromatic	Darmali	Susceptible
Aus	Kalamkati	Resistant
Aus	Mehr	?
Aus	Miriti	Susceptible
Aus	Jhona349	Susceptible
Indica	IR8	Resistant
Indica	IR36	Resistant
Indica	JC91	Resistant
Indica	IR64	Resistant
Indica	Gie57	Resistant
Indica	Guan-Yin-Tsan	Resistant
Indica	Ai-Chiao-Hong	Resistant
Indica	Jasmine85	Resistant
Temperate japonica	Nipponbare	Resistant
Temperate japonica	Bengal	Resistant
Temperate japonica	Fanny	Resistant
Temperate japonica	Mansaku	Susceptible
Temperate japonica	Chodongji	Susceptible
Tropical japonica	Dixiebelle	Susceptible
Tropical japonica	Davao	Susceptible
Tropical japonica	Trembese	Susceptible
Tropical japonica	Curinga	Susceptible
Tropical japonica	Azucena	Susceptible
Tropical japonica	Binulawan	Susceptible

5.3.2 Evaluation of post-attachment resistance of rice cultivars to *Striga hermonthica*

Figure 5.4 shows the resistance of 26 cultivars to *S. hermonthica* (Kibos ecotype). There was a highly significant difference in post-attachment resistance between rice cultivars for all traits measured (*S. hermonthica* dry weight: ANOVA F = 13.9, d.f. = 28,122, p < 0.001; number of *S. hermonthica* seedlings: ANOVA: F = 11.95, d.f. = 28, 122, p < 0.001; cumulative length of *S. hermonthica* seedlings: ANOVA: F = 10.2, d.f. = 28, 122, p < 0.001). Genotype explained 69.2 % variation in *S. hermonthica* dry weight, 66.2 % variation in number and 62.3 % variation in cumulative length of *S. hermonthica* seedlings. There was no significant variation due to batch (< 0.66 %) for all traits measured (Linear Mixed Model).

There was no significant difference in *S. hermonthica* dry weight, number, or cumulative length of *S. hermonthica* seedlings for the 5 most susceptible cultivars Binulawan, Azucena, Curinga, Trembese and Davao (all *tropical japonica* cultivars that were predicted to be susceptible). The 5 most resistant cultivars were IR8, JC91, IR36, IR64 (all *indica*) and Nipponbare (*temperate japonica*) and all were predicted to be resistant (Figure 5.4 A, B and C and Figure 5.5D). However, there was no significant difference in the dry weight of *S. hermonthica* between the 16 most resistant cultivars (from Fanny to IR8) (Figure 5.4A), Tukey Multiple Comparison test p > 0.05). Twelve of the 16 resistant cultivars were predicted to be resistant. For 3 genotypes, Mehr, Firooz and Kitrana it was not possible to make a prediction and for 1, Damali, the prediction was wrong. *S. hermonthica* dry weight did not differ significantly between Davao and Darmali, or any cultivar with *S. hermonthica* dry weight intermediate between these 2 cultivars (Tukey Multiple Comparison test p < 0.05) (Figure 5.4 and Figure 5.5C). In general, *indica* and *aromatic* cultivars showed good resistance to *S. hermonthica*. Temperate *japonica* and *aus* cultivars were either intermediate or resistant (Figure 5.4). The root systems of 4 rice cultivars, Curinga, (susceptible), Jhona349 (less susceptible) and Firooz (resistant) and Nipponbare (resistant) infected with *S. hermonthica* are shown in Figure 5.5.

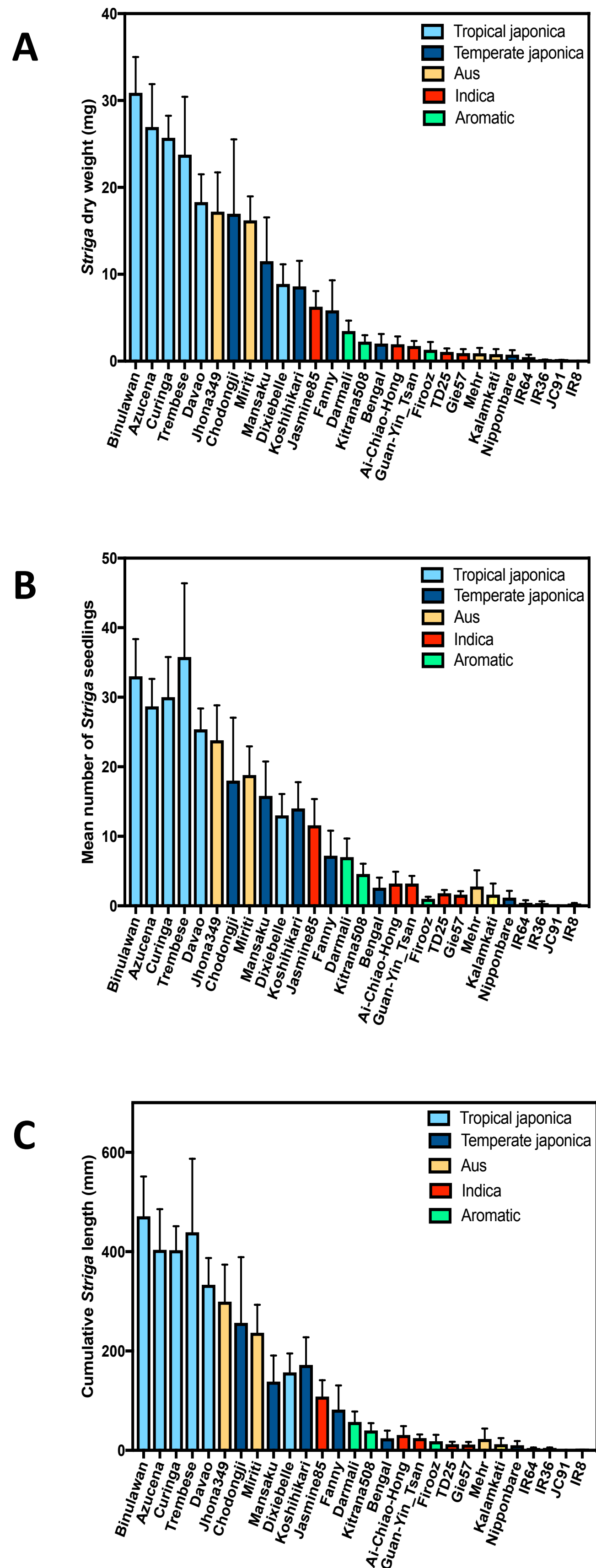


Figure 5.4 The phenotype of resistance to *S. hermonthica* for 26 genetically diverse rice cultivars, plus Koshihikari (not included in SNP analysis). A) Mean dry biomass, B) mean number, and C) mean cumulative length of *S. hermonthica* seedlings harvested from the roots of rice plants 21 dai. Data are means ± SE where n = 5.

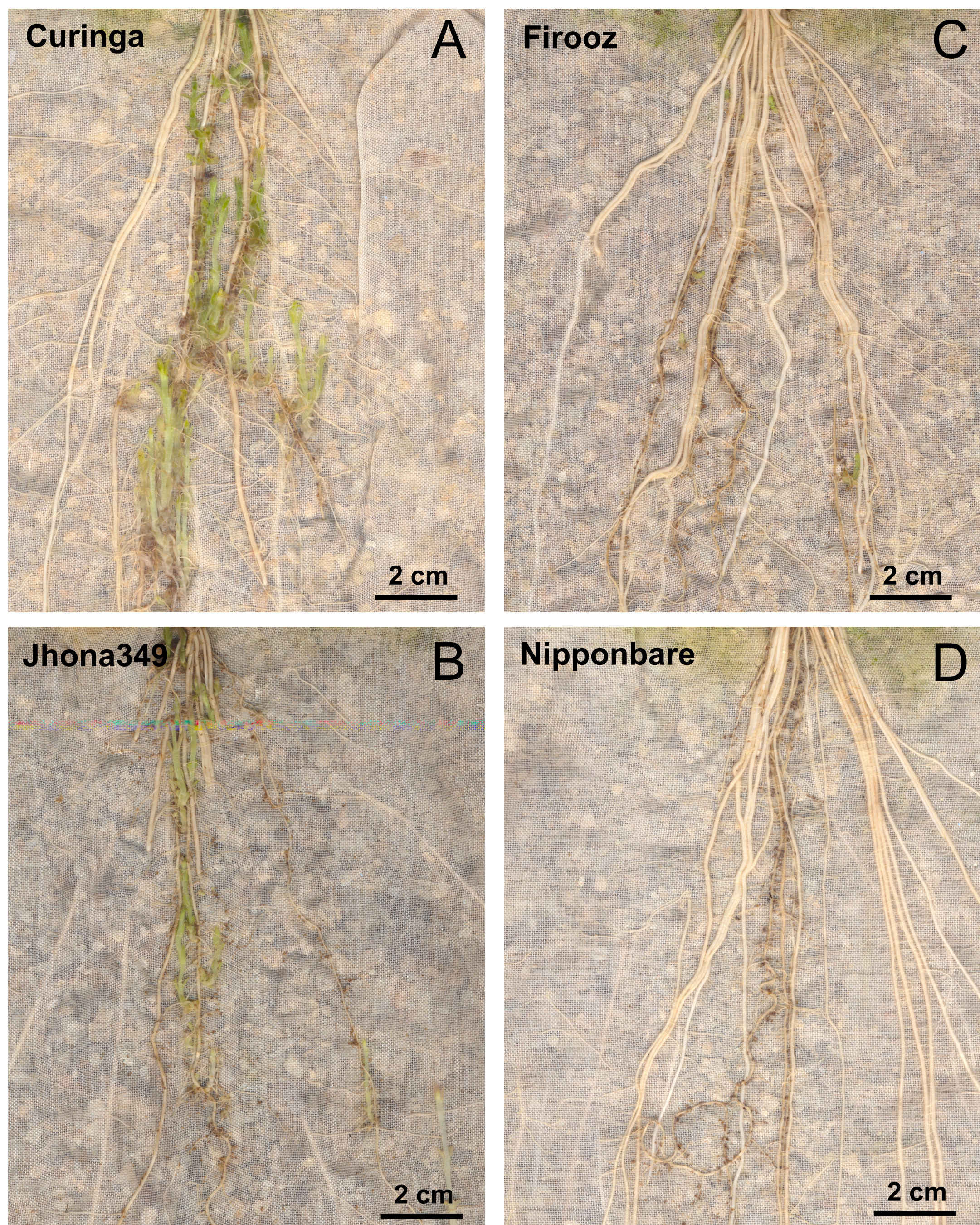


Figure 5.5 The root systems of 4 rice cultivars, Curinga (tropical japonica, susceptible), Jhona349 (*aus*, intermediate), Firooz (*aromatic*, resistant) and Nipponbare (temperate *japonica*, resistant) 21 dai with *S. hermonthica*.

5.3.3 The relationship between *S. hermonthica* resistance and SNP similarity to Nipponbare

To determine if resistance to *S. hermonthica* was correlated with the similarity of any of the candidate resistance genes in Nipponbare, a Pearson's correlation analysis was carried out between the dry biomass of *S. hermonthica* harvested from the roots of each cultivar and the SNP similarity of each gene to the Nipponbare allele. Genes were grouped according to the relationship observed; Group 1) genes showing a significant correlation between *S. hermonthica* resistance and similarity to the Nipponbare allele (Figure 5.6), Group 2) genes showing no significant correlation (Figure 5.7), or Group 3) genes where a few resistant cultivars shared high similarity to Nipponbare but where many outliers were also observed (Figure 5.8). The summary statistics are shown in Table 5.7.

Group 1: A significant correlation was observed between resistance to *S. hermonthica* and similarity to the Nipponbare allele for six genes. The gene with the highest R^2 value, and so explaining the greatest variation, was Os12g10870 (R^2 = 70.7 %), followed by Os12g11370 (R^2 = 67.2 %), Os12g11510 (R^2 = 66.3 %), Os12g10850 (R^2 = 63.3 %), Os12g10820 (R^2 = 55.4 %), and Os12g11500 (R^2 = 28.5 %) (Figure 5.6). The allelic diversity of three of these genes (Os12g10870, Os12g11370 and Os12g11500) is shown in Figure 5. 9 A-C. All six of these genes separated cultivars broadly into two groups, those with the Nipponbare-like allele and those where large differences in SNPs were observed (Figures 5.6 and 5.9 A-C). In most cases those cultivars with good similarity to Nipponbare were resistant, although there were exceptions. For example, the resistant cultivar Darmali showed very little SNP similarity to Nipponbare for any of these genes. The resistant cultivar Mehr also showed poor SNP similarity to Nipponbare for some regions of these genes (Figure 5.9 A-C), and Os12g11500 appeared to be missing entirely from two *aromatic* cultivars Firooz and Kitrana508 (Figure 5.6E and 5.9C). Resistant cultivars varied in their similarity to Nipponbare for the different genes. For example a cluster of very resistant cultivars that were > 80 % similar to Nipponbare was observed for Os12g10870, while similarity to Os12g10850 was as little as 57 %. In order to distinguish between these different distributions, cultivars were sorted by resistance, and genes highlighted that had > 80 % similarity to Nipponbare (Table 5.8). The four RLP genes Os12g10870, Os12g11370, Os12g11500 and Os12g11510 were highlighted as having the greatest similarity to Nipponbare in resistant cultivars, despite the lower R^2 value obtained from Os12g11500 correlation analysis, indicating the similarity of resistant cultivars to Nipponbare, not simply the correlation, is important in determining the importance of a gene in *S. hermonthica* resistance.

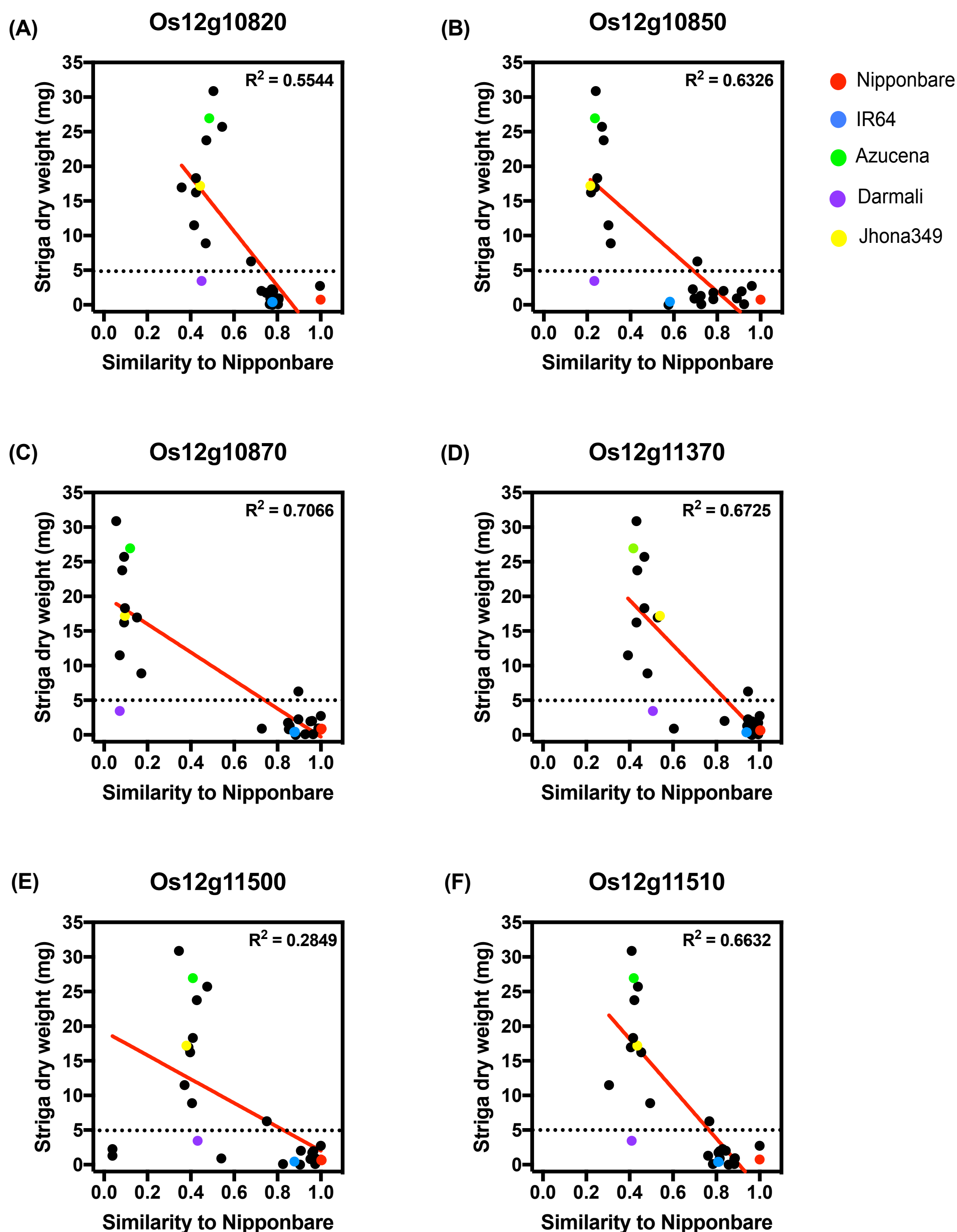
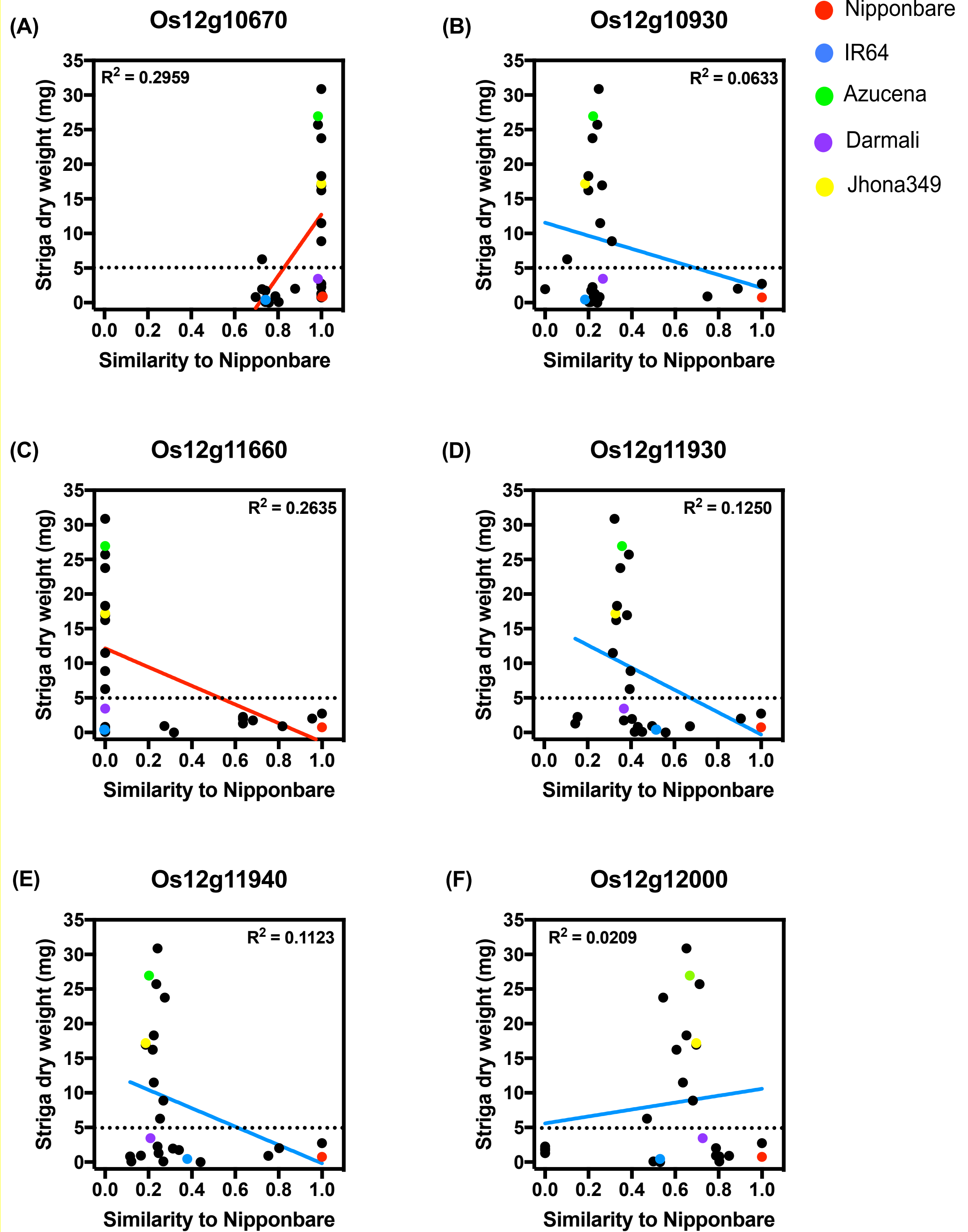


Figure 5.6 The relationship between *S. hermonthica* dry biomass of 26 diverse rice cultivars and SNP similarity to the Nipponbare allele for 6 genes in the *S. hermonthica* resistance QTL, showing a significant negative correlation between SNP similarity and resistance. Red lines indicate a significant relationship between variables ($p < 0.05$). R^2 values indicate the variance explained by the linear model. SNP similarity to Nipponbare was calculated for each cultivar and gene; the sum of the matching SNPs was divided by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). Cultivars with < 5 mg *Striga* dry weight were considered very resistant.

Group 2: No significant correlation between *S. hermonthica* resistance and similarity to Nipponbare was observed for seven genes within the QTL (Os12g10930, Os12g11660, Os12g11930, Os12g11940, Os12g12000, Os12g12010 and Os12g12130), or for the two genes investigated outside the QTL (Os12g10670 and Os12g12514) (Figure 5.7). IR64 had poor similarity to Nipponbare in all cases. SNP similarity to Nipponbare for Os12g12010 was > 93 % for all susceptible cultivars, while resistant cultivars showed both good and poor similarity for this gene (Figures 5.7G and 5.9E, Table 5.8). Only three resistant cultivars had homologs of Os12g10930, Os12g11930 and Os12g11940 with > 80 % similarity to Nipponbare (Table 5.8).

Group 3: The four genes Os12g11680, Os12g11720, Os12g11860 and Os12g12120 had > 80 % SNP similarity to Nipponbare in a small number of resistant cultivars but very poor similarity in all susceptible cultivars (Figure 5.8 and Table 5.8). The allelic diversity of Os12g11860 is shown in Figure 5.9D as an example. As seen here, many of the most resistant cultivars, including IR64, are missing or have poor similarity to Nipponbare for these genes (Table 5.8).



(A)
Os12g10670
R^2 = 0.2959
(B)
Os12g10930
R^2 = 0.0633
(C)
Os12g11660
R^2 = 0.2635
(D)
Os12g11930
R^2 = 0.1250
(E)
Os12g11940
R^2 = 0.1123
(F)
Os12g12000
R^2 = 0.0209
Striga dry weight (mg)
Similarity to Nipponbare
Nipponbare
IR64
Azucena
Darmali
Jhona349

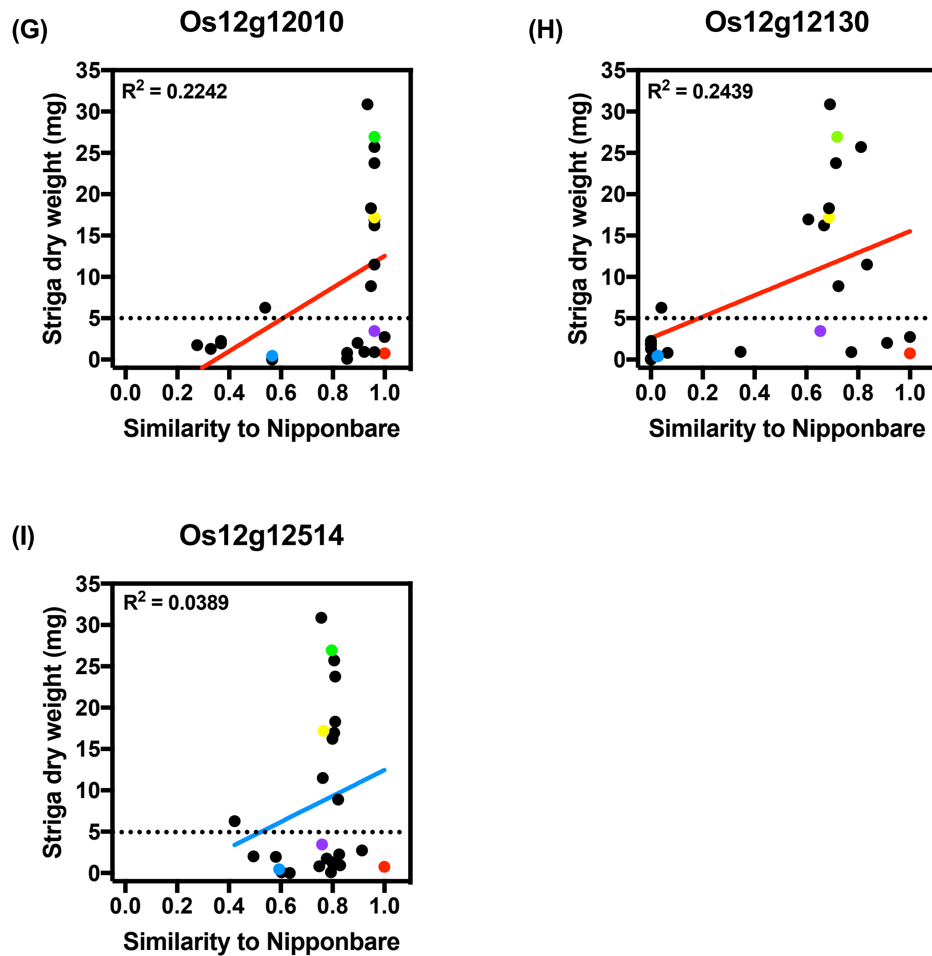


Figure 5.7 The relationship between *S. hermonthica* dry biomass of 26 diverse rice cultivars and SNP similarity to the Nipponbare allele for 9 genes across the *S. hermonthica* resistance QTL, showing no significant correlation between *S. hermonthica* resistance and similarity to Nipponbare. Red lines indicate a significant relationship between variables ($p < 0.05$). R^2 *values* indicate the variance explained by the linear model. SNP similarity to Nipponbare was calculated for each cultivar and gene; the sum of the matching SNPs was divided by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). Cultivars with < 5 mg *Striga* dry weight were considered very resistant.

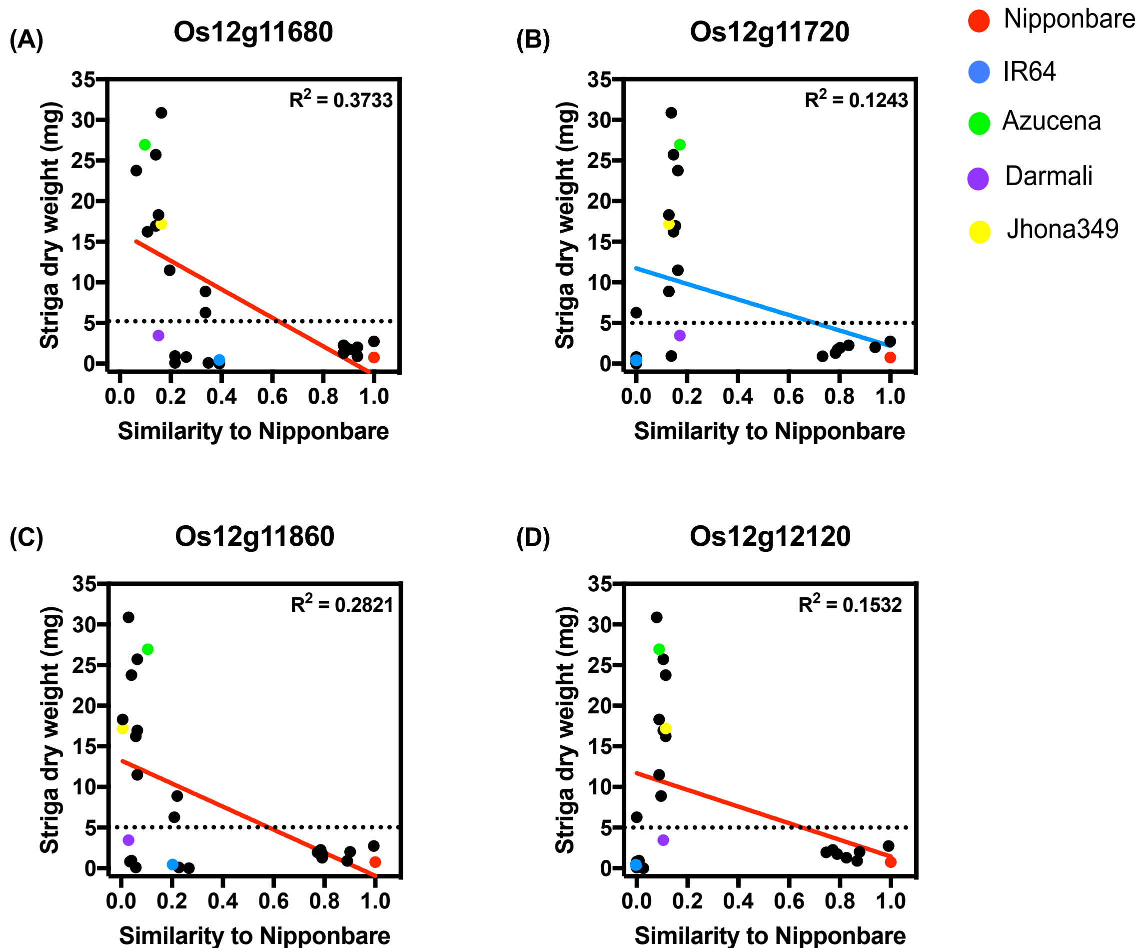


Figure 5.8 The relationship between *S. hermonthica* dry biomass of 26 diverse rice cultivars and SNP similarity to the Nipponbare allele for 4 genes in the *S. hermonthica* resistance QTL where IR64, and many other resistant cultivars, show poor SNP similarity to Nipponbare. Red lines indicate a significant relationship between variables ($p < 0.05$). R^2 values indicate the variance explained by the linear model. SNP similarity to Nipponbare was calculated for each cultivar and gene; the sum of the matching SNPs was divided by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent).

Table 5.7 Summary statistics for Pearson's product-moment correlation analysis carried out between *S. hermonthica* dry weight and SNP similarity to Nipponbare for 19 genes across the *S. hermonthica* resistance QTL gene in 26 diverse rice cultivars.

Gene	t	d.f.	p	R	R^2
Os12g10670	3.177	24	0.004065	0.544	0.296
Os12g10820	-5.464	24	1.29E-05	-0.745	0.554
Os12g10850	-6.429	24	1.20E-06	-0.795	0.633
Os12g10870	-7.603	24	7.68E-08	-0.841	0.707
Os12g10930	-1.274	24	0.215	-0.252	0.063
Os12g11370	-7.020	24	2.94E-07	-0.820	0.672
Os12g11500	-3.092	24	0.004984	-0.534	0.285
Os12g11510	-6.875	24	4.14E-07	-0.814	0.663
Os12g11660	-2.930	24	0.007323	-0.513	0.263
Os12g11680	-3.781	24	0.0009146	-0.611	0.373
Os12g11720	-1.846	24	0.07728	-0.353	0.124
Os12g11860	-3.071	24	0.005237	-0.531	0.282
Os12g11930	-1.852	24	0.07636	-0.354	0.125
Os12g11940	-1.742	24	0.09424	-0.335	0.112
Os12g12000	0.715	24	0.4816	0.144	0.021
Os12g12010	2.633	24	0.01456	0.473	0.224
Os12g12120	-2.084	24	0.04798	-0.391	0.153
Os12g12130	2.783	24	0.01034	0.494	0.244
Os12g12514	0.987	24	0.3336	0.197	0.039

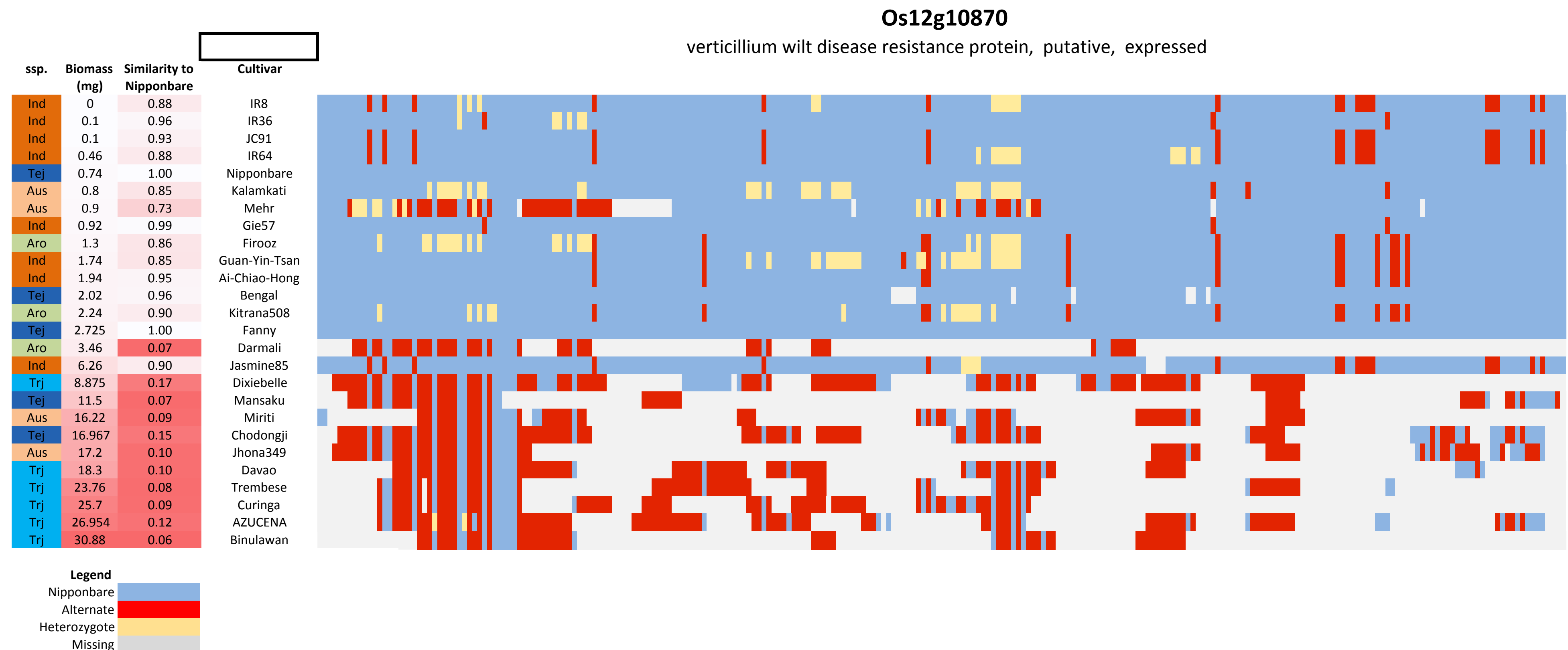


Figure 5.9 A Allelic diversity of Os12g10870 in 26 diverse *O. sativa* cultivars, based on SNP similarity to Nipponbare. SNP similarity to Nipponbare was calculated for each cultivar by dividing the sum of the matching SNPs by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). Cultivars are ordered by *S. hermonthica* dry biomass harvested from roots, from most resistant (top) to most susceptible (bottom). Blue indicates Nipponbare genotype, red indicates alternate SNPs, yellow indicates heterozygote and white indicates missing data. Total number of SNPs across this gene is 250.

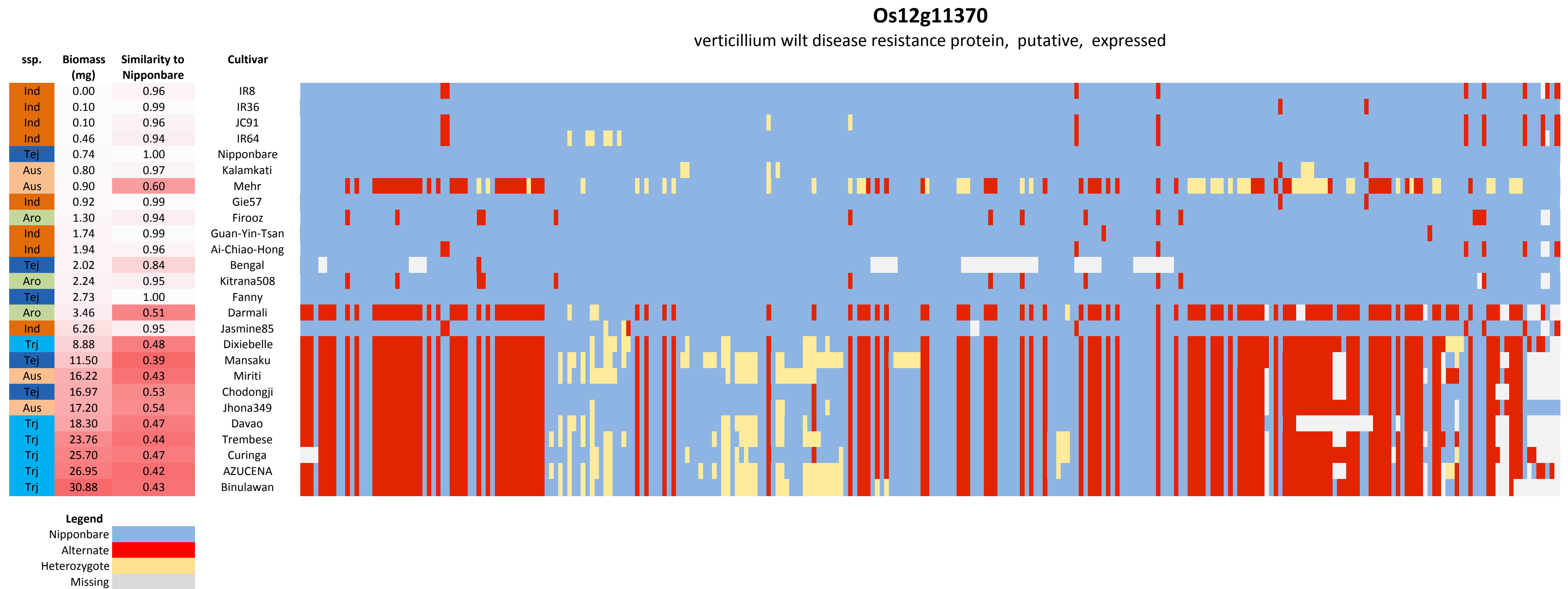


Figure 5.9 B Allelic diversity of Os12g11370 in 26 diverse *O. sativa* cultivars, based on SNP similarity to Nipponbare. SNP similarity to Nipponbare was calculated for each cultivar by dividing the sum of the matching SNPs by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). Cultivars are ordered by *S. hermonthica* dry biomass harvested from roots, from most resistant (top) to most susceptible (bottom). Blue indicates Nipponbare genotype, red indicates alternate SNPs, yellow indicates heterozygote and white indicates missing data. Total number of SNPs across this gene is 250.

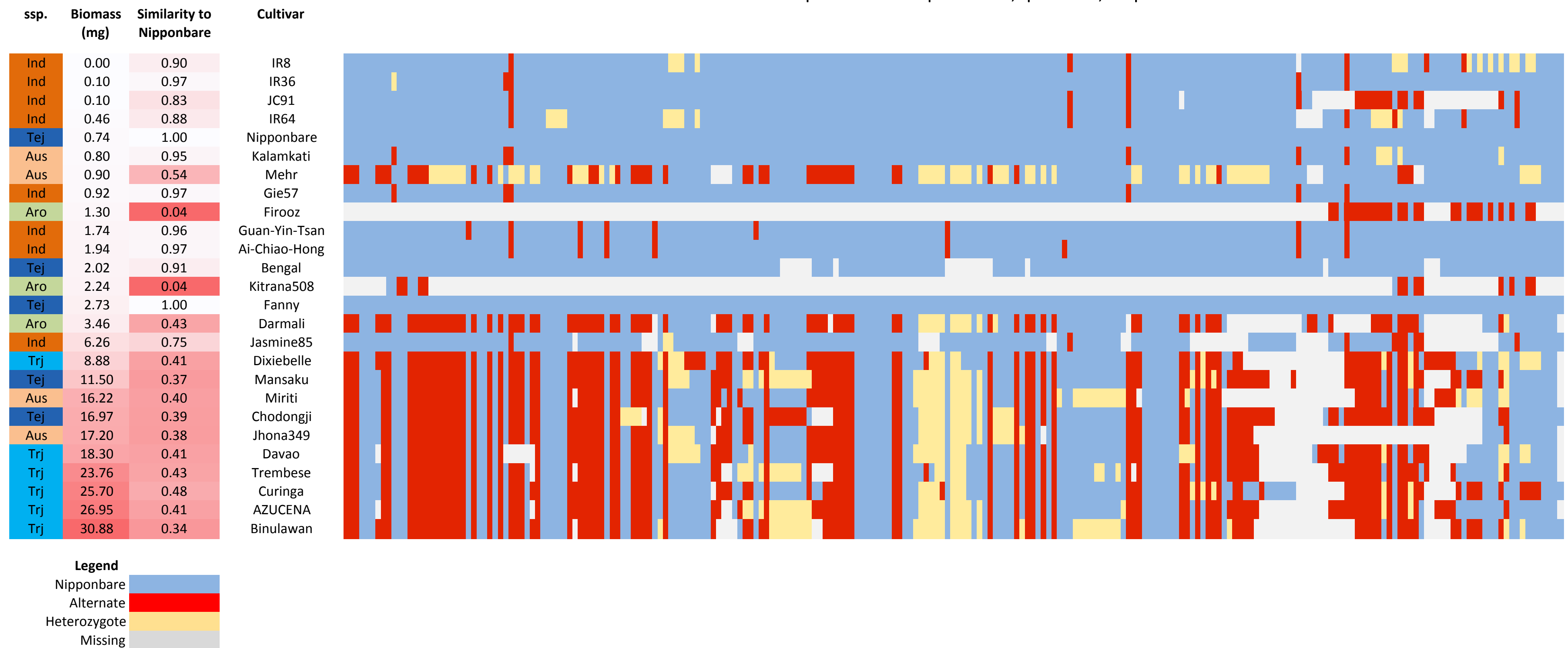


Figure 5.9 C Allelic diversity of Os12g11500 in 26 diverse *O. sativa* cultivars, based on SNP similarity to Nipponbare. SNP similarity to Nipponbare was calculated for each cultivar by dividing the sum of the matching SNPs by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). Cultivars are ordered by *S.* hermonthica dry biomass harvested from roots, from most resistant (top) to most susceptible (bottom). Blue indicates Nipponbare genotype, red indicates alternate SNPs, yellow indicates heterozygote and white indicates missing data. Total number of SNPs across this gene is 229.

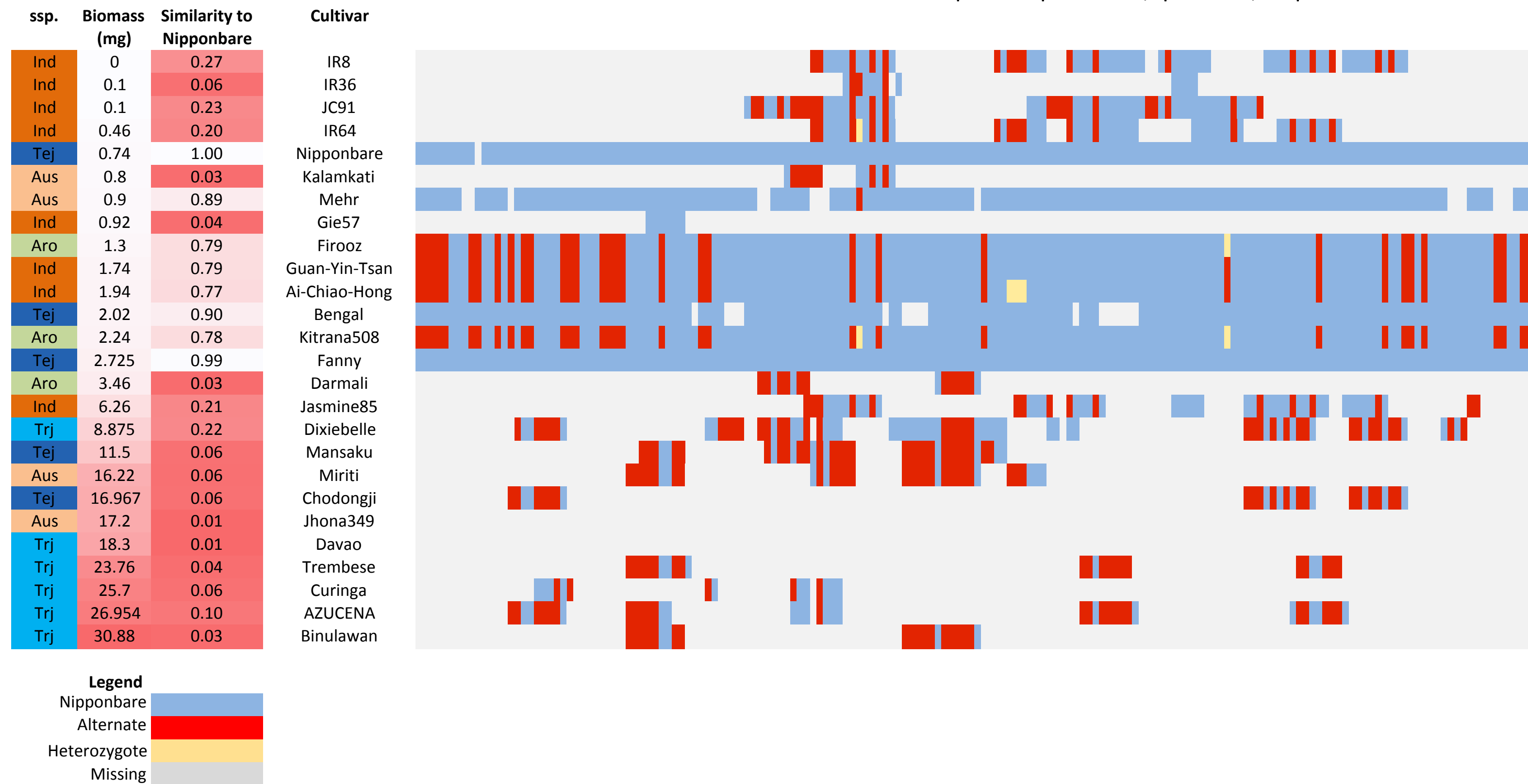


Figure 5.9 D Allelic diversity of Os12g11860 in 26 diverse *O. sativa* cultivars, based on SNP similarity to Nipponbare. SNP similarity to Nipponbare was calculated for each cultivar by dividing the sum of the matching SNPs by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). *Cultivars are ordered by S. hermonthica* dry biomass harvested from roots, from most resistant (top) to most susceptible (bottom). Blue indicates Nipponbare genotype, red indicates alternate SNPs, yellow indicates heterozygote and white indicates missing data. Total number of SNPs across this gene is 172.

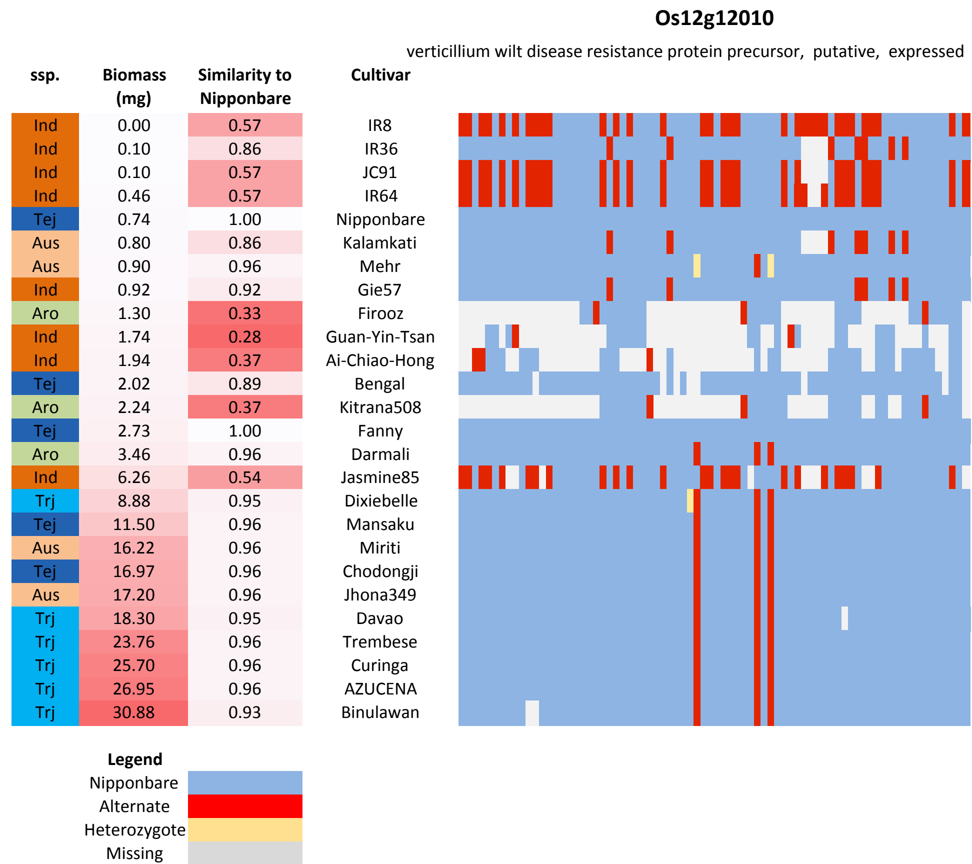


Figure 5.9 E Allelic diversity of Os12g12010 in 26 diverse *O. sativa* cultivars, based on SNP similarity to Nipponbare. SNP similarity to Nipponbare was calculated for each cultivar by dividing the sum of the matching SNPs by the total number of SNPs, to give a value of similarity ranging from 1 (identical) to 0 (no SNP similarity / gene absent). Cultivars are ordered by *S. hermonthica* dry biomass harvested from roots, from most resistant (top) to most susceptible (bottom). Blue indicates Nipponbare genotype, red indicates alternate SNPs, yellow indicates heterozygote and white indicates missing data. Total number of SNPs across this gene is 76.

Table 5.8 Similarity of SNPs to the Nipponbare allele for 19 genes across the *S. hermonthica* resistance QTL, in 26 diverse rice cultivars. Genes highlighted in red have > 80 % SNP similarity to Nipponbare are sorted from most resistant to most susceptible. Cultivars annotated as resistant have < 5 mg *S. hermonthica* dry weight. SNP similarity to Nipponbare for each gene is the sum of the matching SNPs divided by the total number of SNPs, to give a value of similarity between 1 (identical allele) to 0 (no SNP similarity / gene absent). Where two gene identifiers have the same colour the genes have arisen by gene duplication. The QTL region is indicated above in green. All cultivars are *Oryza sativa*.

				QTL region																		
				expressed protein	DNA repair protein	RLP	RLP	RLP	RLP	RLP	RALF	RLP	RLP	RLP	RLP	RLP	RALF	RLP	RLP	RLP	
	Subspecies	Cultivar	Os12g10670	Os12g10820	Os12g10850	Os12g10870	Os12g10930	Os12g11370	Os12g11500	Os12g11510	Os12g11660	Os12g11680	Os12g11720	Os12g11860	Os12g11930	Os12g11940	Os12g12000	Os12g12010	Os12g12120	Os12g12130	Os12g12514
Resistant	Indica	IR8	0.758	0.778	0.575	0.884	0.241	0.964	0.904	0.857	0.318	0.391	0.000	0.267	0.560	0.440	0.530	0.566	0.026	0.000	0.635
	Indica	IR36	0.803	0.803	0.924	0.964	0.203	0.993	0.974	0.883	0.000	0.217	0.000	0.058	0.452	0.121	0.803	0.855	0.000	0.000	0.794
	Indica	JC91	0.742	0.765	0.727	0.928	0.210	0.960	0.825	0.784	0.000	0.348	0.000	0.227	0.417	0.269	0.500	0.566	0.000	0.000	0.603
	Indica	IR64	0.742	0.778	0.582	0.880	0.184	0.942	0.878	0.813	0.000	0.391	0.000	0.203	0.517	0.379	0.530	0.566	0.000	0.028	0.594
	Temperate japonica	Nipponbare	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Aus	Kalamkati	0.697	0.806	0.782	0.852	0.251	0.968	0.952	0.883	0.000	0.261	0.000	0.035	0.432	0.115	0.803	0.855	0.000	0.065	0.749
	Aus	Mehr	1.000	0.768	0.695	0.728	0.749	0.604	0.541	0.816	0.818	0.935	0.733	0.890	0.672	0.753	0.848	0.961	0.868	0.774	0.803
	Indica	Gie57	0.788	0.806	0.891	0.988	0.225	0.993	0.974	0.886	0.273	0.217	0.138	0.041	0.498	0.165	0.788	0.921	0.009	0.346	0.829
	Aromatic	Firooz	1.000	0.778	0.724	0.856	0.229	0.942	0.039	0.762	0.636	0.880	0.784	0.791	0.143	0.247	0.000	0.329	0.825	0.000	0.800
	Indica	Guan-Yin-Tsan	0.742	0.749	0.782	0.848	0.213	0.993	0.961	0.810	0.682	0.902	0.793	0.791	0.367	0.341	0.000	0.276	0.789	0.000	0.778
	Indica	Ai-Chiao-Hong	0.727	0.781	0.913	0.952	0.000	0.964	0.965	0.816	0.636	0.891	0.802	0.773	0.405	0.313	0.000	0.368	0.746	0.000	0.581
	Temperate japonica	Bengal	0.879	0.727	0.829	0.960	0.889	0.838	0.908	0.844	0.955	0.935	0.940	0.901	0.907	0.802	0.788	0.895	0.877	0.912	0.495
	Aromatic	Kitrana508	1.000	0.775	0.687	0.896	0.219	0.946	0.039	0.829	0.636	0.880	0.836	0.785	0.154	0.242	0.000	0.368	0.772	0.000	0.825
	Temperate japonica	Fanny	1.000	0.997	0.960	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000	0.991	1.000	0.914
	Aromatic	Darmali	0.985	0.451	0.233	0.072	0.267	0.507	0.432	0.410	0.000	0.152	0.172	0.029	0.367	0.209	0.727	0.961	0.105	0.654	0.759
	Indica	Jasmine85	0.727	0.679	0.709	0.896	0.102	0.946	0.751	0.768	0.000	0.337	0.000	0.209	0.394	0.253	0.470	0.539	0.000	0.041	0.422
	Tropical japonica	Dixiebelle	1.000	0.470	0.309	0.172	0.308	0.482	0.406	0.495	0.000	0.337	0.129	0.221	0.398	0.269	0.682	0.947	0.096	0.724	0.822
	Temperate japonica	Mansaku	1.000	0.416	0.298	0.072	0.254	0.392	0.371	0.305	0.000	0.196	0.164	0.064	0.317	0.225	0.636	0.961	0.088	0.834	0.762
	Aus	Miriti	1.000	0.425	0.218	0.092	0.200	0.432	0.397	0.454	0.000	0.109	0.147	0.058	0.332	0.220	0.606	0.961	0.114	0.668	0.800
	Temperate japonica	Chodongji	1.000	0.359	0.236	0.152	0.263	0.529	0.389	0.406	0.000	0.141	0.155	0.064	0.382	0.187	0.697	0.961	0.105	0.608	0.806
	Aus	Jhona349	1.000	0.444	0.215	0.096	0.184	0.540	0.380	0.435	0.000	0.163	0.129	0.006	0.328	0.187	0.697	0.961	0.114	0.687	0.765
	Tropical japonica	Davao	1.000	0.425	0.247	0.096	0.200	0.468	0.410	0.416	0.000	0.152	0.129	0.006	0.336	0.225	0.652	0.947	0.088	0.687	0.810
	Tropical japonica	Trembese	1.000	0.473	0.276	0.084	0.219	0.435	0.428	0.422	0.000	0.065	0.164	0.041	0.351	0.275	0.545	0.961	0.114	0.714	0.810
	Tropical japonica	Curinga	0.985	0.546	0.269	0.092	0.241	0.468	0.476	0.438	0.000	0.141	0.147	0.064	0.390	0.236	0.712	0.961	0.105	0.811	0.806
	Tropical japonica	AZUCENA	0.985	0.486	0.236	0.120	0.222	0.417	0.410	0.419	0.000	0.098	0.172	0.105	0.359	0.203	0.667	0.961	0.088	0.719	0.797
	Tropical japonica	Binulawan	1.000	0.505	0.240	0.056	0.248	0.432	0.345	0.410	0.000	0.163	0.138	0.029	0.324	0.242	0.652	0.934	0.079	0.691	0.756

5.3.4 Verification of SNP data: genotyping by PCR amplification of marker genes in different rice cultivars

Genotyping using PCR markers was carried out for five genes in the *S. hermonthica* resistance QTL in order to verify the SNP similarity data for the rice material phenotyped in this study (Figure 5.10 A-F). The results are summarised in Table 5.9. Koshihikari was the only susceptible cultivar where genome sequence was available. Therefore where genes were known to be present in Koshihikari, primers were designed to distinguish between the Nipponbare and Koshihikari alleles.

Os12g11370

PCR amplification of Os12g11370 from Nipponbare and IR64 produced an amplicon of 569 bp while a product of 852bp was amplified from Koshihikari, and a larger product of approximately 1500 bp was amplified from Azucena. All other cultivars displayed either the Nipponbare genotype (560 bp amplicon) or the Koshihikari genotype (1500 bp amplicon) (Figure 5.10 A). With the exception of Darmali, Jasmine85 and all cultivars more resistant than Jasmine85 had the Nipponbare genotype for this gene. The more susceptible Jhona349 also displayed the Nipponbare genotype. All other susceptible cultivars displayed the Koshihikari genotype. This separation of cultivars into two genotypes matched the SNP data, with the exception of Jhona349 (susceptible) that had only 54 % SNP similarity with Nipponbare but appeared to have the Nipponbare genotype here. However, the fainter band on the gel may suggest poor primer binding and sequence differences between these cultivars.

Os12g11500

Two different primer combinations were used for genotyping of Os2g11500 to distinguish the known differences between the Nipponbare and Koshihikari alleles for this gene (Figure 5.10 B and C). A very similar pattern was observed to that seen for Os12g11370. A PCR product of 1229bp (Koshihikari genotype) was amplified from Koshihikari and all cultivars more susceptible than Koshihikari, with the exception of Jhona349 (Figure 5.10 B). The Koshihikari-like allele was also amplified from Darmali. A PCR product of 240bp was amplified with primers specific to the Nipponbare allele for Jhona349 and all other cultivars more resistant than Koshihikari, with the exception of Firooz and Mehr where there was no product (Figure 5.10C). This matched the SNP data, with the exception of Jhona349, which had the Nipponbare genotype here, but shared only 38 % SNP similarity with Nipponbare. This could reflect a problem in SNP data due to poor alignment, possibly due to gene duplication.

Os12g11680 and *Os12g11860*

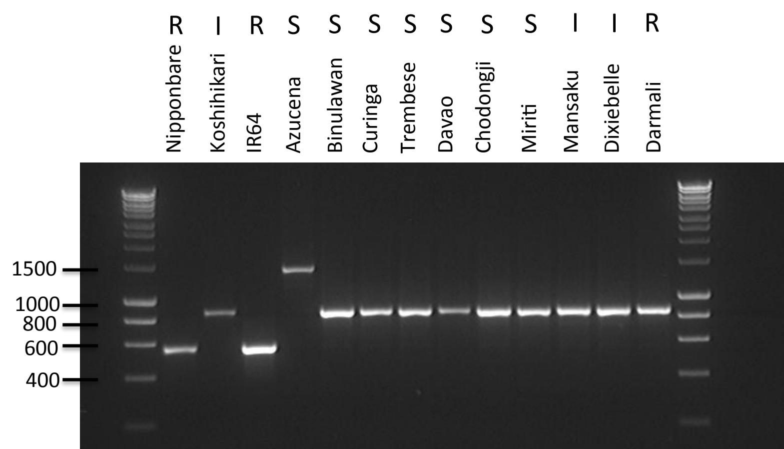
The Nipponbare genotype was observed for Nipponbare, Fanny, Bengal, Ai-Chiao-Hong, Guan-Yin-Tsan, Firooz and Mehr. Fainter bands and non-specific amplification was observed for all other cultivars, suggesting these genes were absent or very different in sequence (Figure 5.10 D and E). These results agreed with the SNP data; Nipponbare, Fanny, Bengal, Ai-Chiao-Hong, Guan-Yin-Tsan, Firooz and Mehr all possessed a homolog of each of Os12g11680 and Os12g11860 that were missing in the other cultivars.

Os12g12010

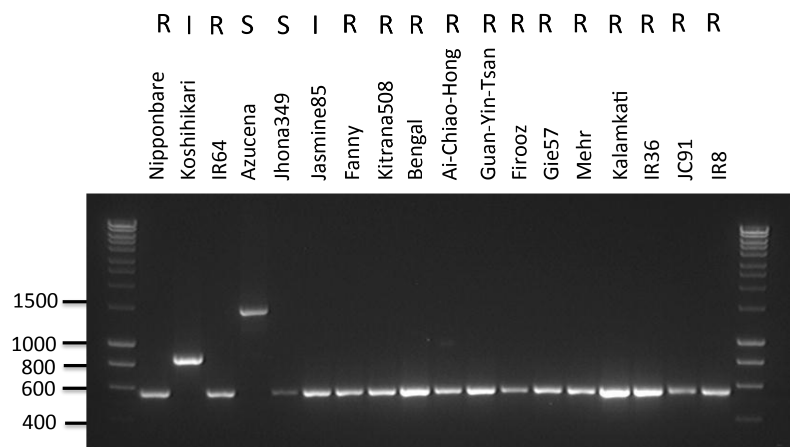
A single PCR product of 446bp was amplified from all cultivars (Figure 5.10 F). This agrees with the high SNP similarity observed for most cultivars for this gene.

Os12g11370

A1

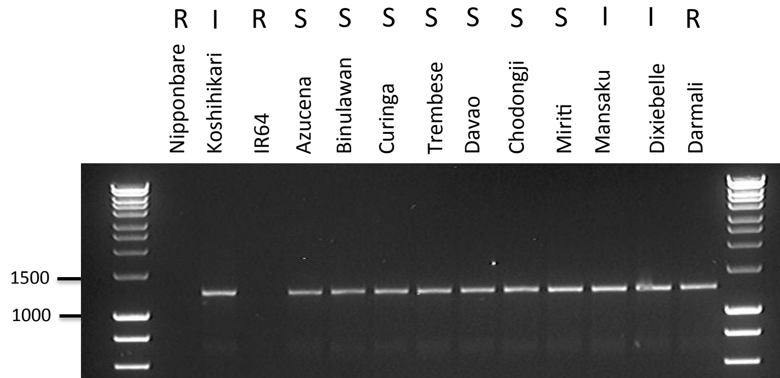


A2



Os12g11500 Koshihikari specific primers

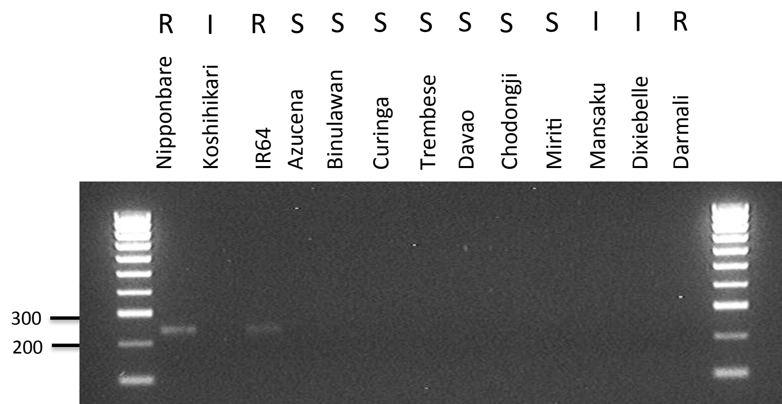
B1



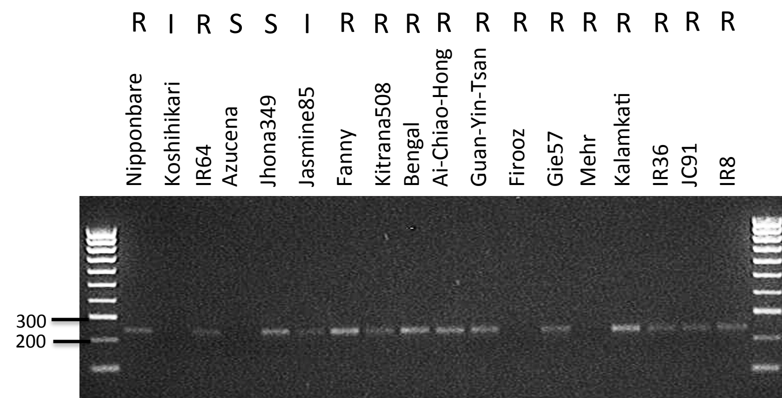
B2

Os12g11500 Nipponbare specific primers

C1

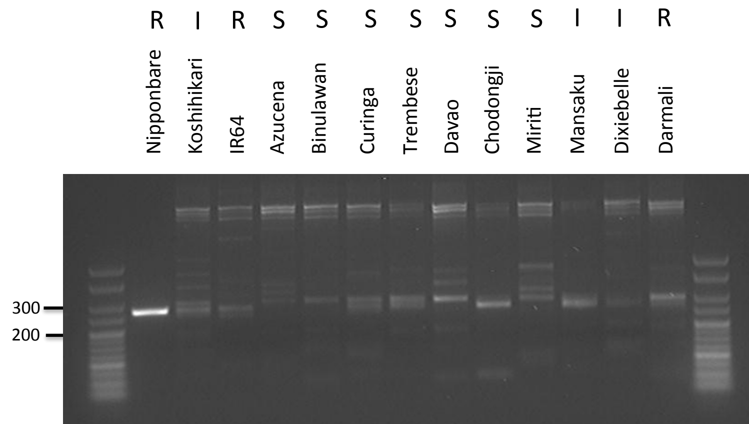


C2

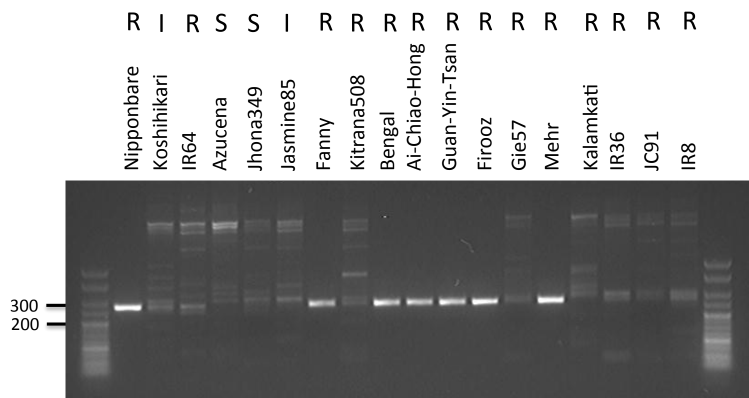


Os12g11680 Nipponbare specific primers

D1

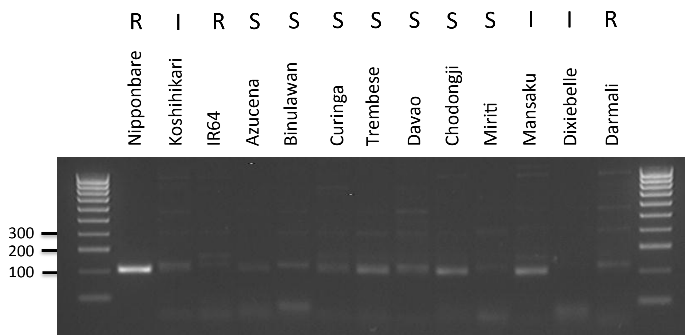


D2

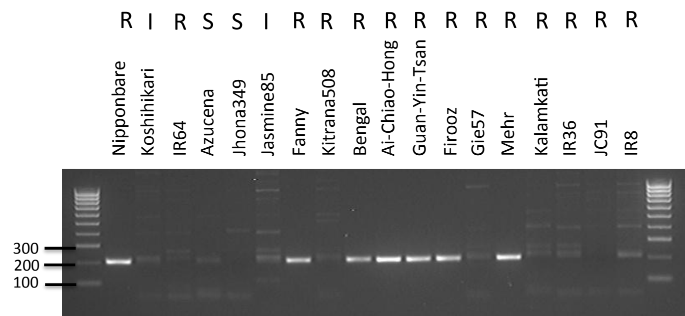


Os12g11860 Nipponbare specific primers

E1



E2



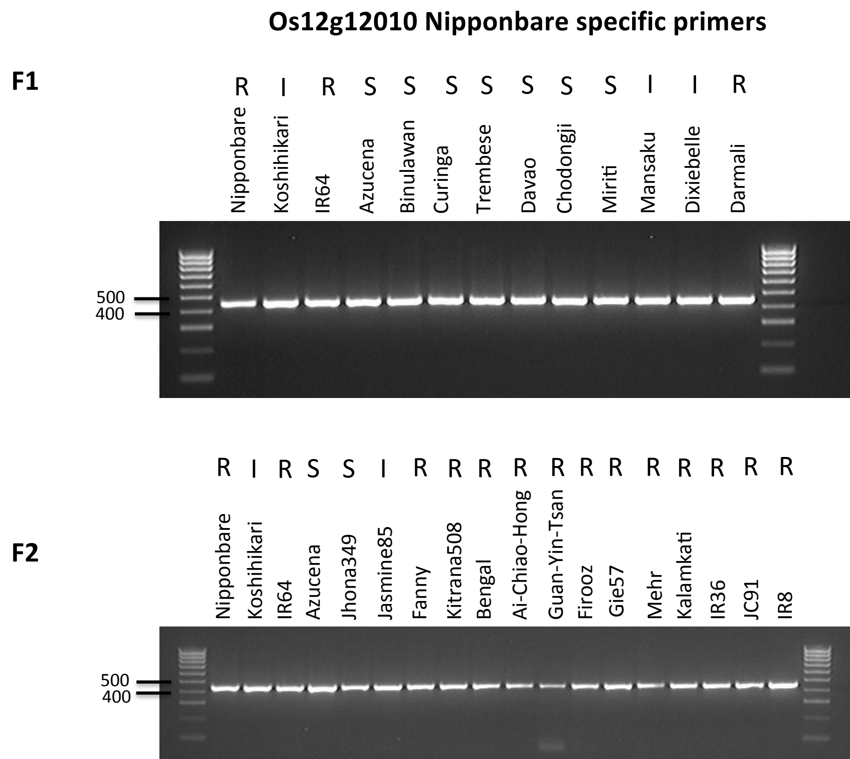


Figure 5.10 A-F Genotyping 27 rice cultivars using PCR markers for 5 candidate resistance genes across the *S. hermonthica* resistance QTL. Letters above cultivar names relate to the phenotype of resistance to *S. hermonthica*; R: resistant; I: intermediate resistance; S: susceptible. **A**: Primers used to amplify Os12g11370 were designed to amplify different size products for Nipponbare and Koshihikari alleles. **B and C**: Primers used were either specific to the Koshihikari allele (B) or the Nipponbare allele (C). **D – F**: Primers were designed against the Nipponbare gene sequences.

Table 5.9 Summary table of genotyping results for five genes in the Nipponbare *S. hermonthica* resistance QTL for the 26 rice cultivars genotyped by SNP markers, plus Koshihikari. N: Nipponbare allele, K: Koshihikari allele, A: Azucena allele, '-': no PCR amplification. Primers used for Os12g11370 and Os12g11500 were designed to distinguish between Nipponbare and Koshihikari alleles. All other primers were designed against the Nipponbare genes sequences. R: resistant; I: intermediate resistance; S: susceptible to *S. hermonthica*. Cultivars are listed from most resistant to most susceptible.

Cultivar	ssp.	R/S	Os12g11370	Os12g11500	Os12g11680	Os12g11860	Os12g12010
IR8	Ind	R	N	N	-	-	N
IR36	Ind	R	N	N	-	-	N
JC91	Ind	R	N	N	-	-	N
IR64	Ind	R	N	N	-	-	N
Nipponbare	TeJ	R	N	N	N	N	N
Kalamkati	Aus	R	N	N	-	-	N
Mehr	Aus	R	N	-	N	N	N
Gie57	Ind	R	N	N	-	-	N
Firooz	Aro	R	N	-	N	N	N
Guan-Yin-Tsan	Ind	R	N	N	N	N	N
Ai-Chiao-Hong	Ind	R	N	N	N	N	N
Bengal	TeJ	R	N	N	N	N	N
Kitrana508	Aro	R	N	N	-	-	N
Darmali	Aro	R	K	K	-	-	N
Fanny	TeJ	I	N	N	N	N	N
Jasmine85	Ind	I	N	N	-	-	N
Dixiebelle	TrJ	I	K	K	-	-	N
Koshihikari	TeJ	I	K	K	-	-	N
Mansaku	TeJ	I	K	K	-	-	N
Miriti	Aus	S	K	K	-	-	N
Chodongji	TeJ	S	K	K	-	-	N
Jhona349	Aus	S	(N)	N	-	-	N
Davao	TrJ	S	K	K	-	-	N
Trembese	TrJ	S	K	K	-	-	N
Curinga	TrJ	S	K	K	-	-	N
Azucena	TrJ	S	A	K	-	-	N
Binulawan	TrJ	S	K	K	-	-	N

5.3.5 Amplification and sequencing of candidate *S. hermonthica* resistance genes from rice cultivars

Genotyping rice cultivars using PCR markers with primers designed against Os12g11370 or Os21g11500 separated cultivars into two distinct groups: those with the Nipponbare allele and those with the Koshihikari allele (Figure 5.10 A-C). In most cases this correlated well with *S. hermonthica* resistance or susceptibility; most cultivars with the Nipponbare alleles were resistant while those with the Koshihikari alleles were susceptible (Table 5.9). However there were exceptions to this. Jhona349 (susceptible) and Jasmine85 (intermediate) both possessed the Nipponbare allele for these genes. Replicates of Fanny varied considerably in *S. hermonthica* resistance, but all possessed the Nipponbare allele. In contrast, Darmali possessed the Koshihikari allele but was very resistant (Table 5.9 and Figure 5.4). It was therefore important to know if the Os12g11370 and Os12g11500 gene sequences for these

cultivars differed to those of Nipponbare. PCR was used to amplify the full-length gene sequences of Os12g11370 and Os12g11500 from Darmali, Fanny, Jasmine85 and Jhona349, as well as Nipponbare, Koshihikari, IR64 and Azucena (Figure 5.11). For Os12g11370, a clear band was observed for Nipponbare, IR64, Fanny and Jasmine85. These products were successfully sequenced and assembled for Fanny and Jasmine85. A faint band was observed for Koshihikari and Darmali, and a very faint and larger band was observed for Azucena. Sequencing of these products was not possible. No product was seen for Jhona349 for this gene (Figure 5.11A). For Os12g11500, a bright band was observed for Nipponbare, Koshihikari, Azucena, Darmali, Fanny and Jhona349 (Figure 5.11B). These products were sequenced. No product was observed for IR64 or Jasmine85 using these primers.

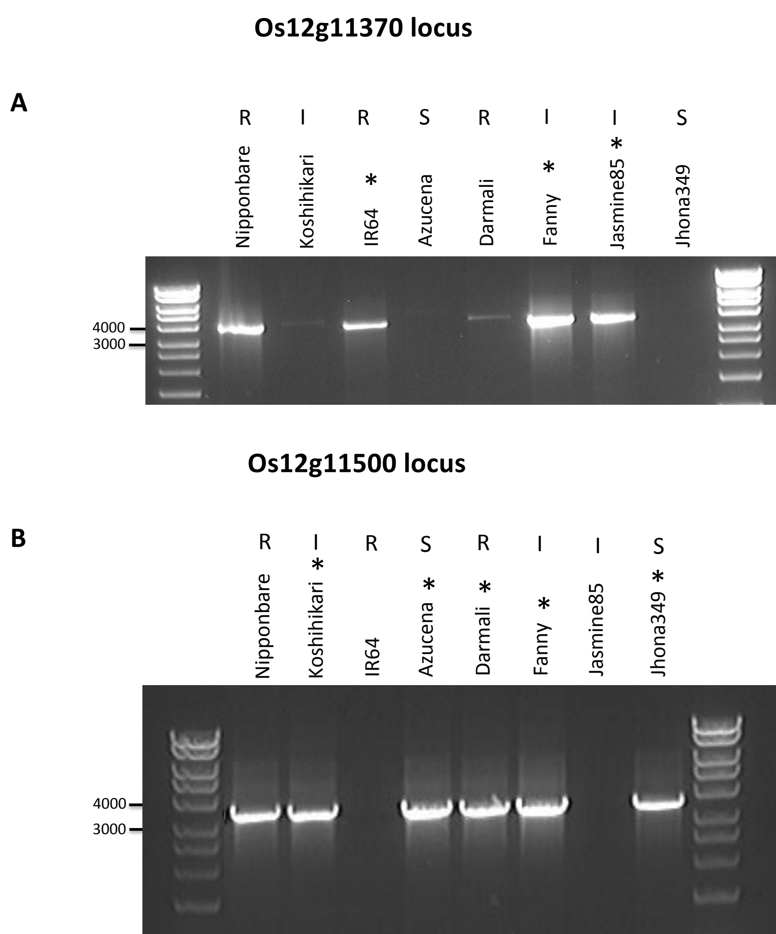


Figure 5.11 Amplification of full-length candidate *S. hermonthica* resistance genes Os12g11370 (**A**) and Os12g11500 (**B**) from 8 rice cultivars. Primers were designed against the Nipponbare sequence a short distance outside the start / stop sites to ensure amplification of the entire gene sequence. Where successful amplification was achieved, the PCR purified genes were sent for Sanger sequencing. Asterisks indicate gene products that were successfully sequenced. Letters above cultivar names relate to the phenotype of resistance to *S. hermonthica*; R: resistant; I: intermediate resistance; S: susceptible.

All sequences obtained were translated into amino acids for comparison between cultivars. At the nucleotide level, the Fanny Os12g11370 allele was identical to Nipponbare while the Jasmine85 Os12g11370 allele was identical to the IR64 sequence. However all four cultivars were identical at the amino acid level (Figure 5.12). In comparison, the Koshihikari allele shared only 86.6 % identity to Nipponbare (Chapter 3).

For Os12g11500, the cultivars appeared to have 1 of 2 alleles, falling into 2 distinct groups: Nipponbare, Fanny and Jhona349 containing one version, and Koshihikari, Darmali and Azucena containing the other. Single differences in amino acids between these two groups were observed at multiple places across the sequence (Figure 5.13). The Fanny allele was identical to that of Nipponbare, while Jhone349 had only 6 amino acid differences to Nipponbare. Koshihikari, Darmali and Azucena all had over 100 amino acid differences to Nipponbare, and lacked a transmembrane domain (Figure 5.13). The high similarity of Jhona349 with Nipponbare is in contrast to the SNP data where Jhona349 had only 38 % identity with Nipponbare (Figure 5.9C), suggesting the material sequenced to identify SNPs was not the same as used in this study, and should thus be excluded when interpreting the SNP data.

```
                         Signal peptide                     40         60
LOC_Os12g11370       MSSSTKRLVR PHHLAKPLL- TMLHILLQVQ AIAALTDDAT APVIQCLPDQ ASALLRLKNS 59
Fanny_h_Os12g11370   MSSSTKRLVR PHHLAKPLL- TMLHILLQVQ AIAALTDDAT APVIQCLPDQ ASALLRLKNS 59
Jasmine85_h_Os12g11370 MSSSTKRLVR PHHLAKPLL- TMLHILLQVQ AIAALTDDAT APVIQCLPDQ ASALLRLKNS 59
Kosh_h_Os12g11370    MSSSTKRPVR PHHLAKPLLL TMLHILLQLK AITALTDDAT APVIQCLPDQ ASALLRLKHS 60
                                           80                    100                   120
LOC_Os12g11370       FNKTAGGYST AFRSWITGTD CCHWDGVDCG GGEDGRVTSL VLGGHNLQAG SISPALFRLT 119
Fanny_h_Os12g11370   FNKTAGGYST AFRSWITGTD CCHWDGVDCG GGEDGRVTSL VLGGHNLQAG SISPALFRLT 119
Jasmine85_h_Os12g11370 FNKTAGGYST AFRSWITGTD CCHWDGVDCG GGEDGRVTSL VLGGHNLQAG SISPALFRLT 119
Kosh_h_Os12g11370    FNTTAGGYST TFRSWITGTD CCHWEGIHCS G-EDGRVTSL VLGGHNLQTT IVDPALFRLN 119
                                           140                   160                   180
LOC_Os12g11370       SLRYLDISGN NFSMSQLPVT GFENLTELTH LDLSDTNIAG EVPAGIGSLV NLVYLDLSTS 179
Fanny_h_Os12g11370   SLRYLDISGN NFSMSQLPVT GFENLTELTH LDLSDTNIAG EVPAGIGSLV NLVYLDLSTS 179
Jasmine85_h_Os12g11370 SLRYLDISGN NFSMSQLPVT GFENLTELTH LDLSDTNIAG EVPAGIGSLV NLVYLDLSTS 179
Kosh_h_Os12g11370    SLRYLDLSGN NFSMSQLPVT ---------- ---------- ---------- ---------- 139
                                           200                   220                   240
LOC_Os12g11370       FYIIYYDDEN KMMPFASDNF WQLSVPNMET LLANLTNLEE LHMGMVDMSG NGERWCDDIA 239
Fanny_h_Os12g11370   FYIIYYDDEN KMMPFASDNF WQLSVPNMET LLANLTNLEE LHMGMVDMSG NGERWCDDIA 239
Jasmine85_h_Os12g11370 FYIIYYDDEN KMMPFASDNF WQLSVPNMET LLANLTNLEE LHMGMVDMSG NGERWCDDIA 239
Kosh_h_Os12g11370    ---------- -------DYF WQLSLPSMET LLANLTNLEE LHMGMVDMSG NGERWCDDVA 182
                                           260                   280                   300
LOC_Os12g11370       KFTPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGSVPEFLAG FSNLTVLQLS 299
Fanny_h_Os12g11370   KFTPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGSVPEFLAG FSNLTVLQLS 299
Jasmine85_h_Os12g11370 KFTPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGSVPEFLAG FSNLTVLQLS 299
Kosh_h_Os12g11370    KFAPKLQVLS LPYCSLSGPI CTSLSSMNSL TRIELHYNHL SGPVPEFLAG FSNLTVLQLS 242
                                           320                   340                   360
LOC_Os12g11370       KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQDS KL---ENLLI SSTNFTGIIP 356
Fanny_h_Os12g11370   KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQDS KL---ENLLI SSTNFTGIIP 356
Jasmine85_h_Os12g11370 KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQDS KL---ENLLI SSTNFTGIIP 356
Kosh_h_Os12g11370    KNKFEGLFPP IIFQHKKLVT INITNNPGLS GSLPNFSQES SLKYLDSLEV SGLQLAGSMA 302
                                           380                   400                   420
LOC_Os12g11370       SSISNLKSLT ---------- ---------K LDLGASGFSG MLPSSLGSLK YLDLLEVSGI 397
Fanny_h_Os12g11370   SSISNLKSLT ---------- ---------K LDLGASGFSG MLPSSLGSLK YLDLLEVSGI 397
Jasmine85_h_Os12g11370 SSISNLKSLT ---------- ---------K LDLGASGFSG MLPSSLGSLK YLDLLEVSGI 397
Kosh_h_Os12g11370    PWISNLTSLT VLKFSDCGLS GEIPSSIGKK LDLGASGFSG MLPSSLGSLK YLDSLEVSGL 362
                                           440                   460                   480
LOC_Os12g11370       QLTGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 457
Fanny_h_Os12g11370   QLTGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 457
Jasmine85_h_Os12g11370 QLTGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 457
Kosh_h_Os12g11370    QLAGSMAPWI SNLTSLTVLK FSDCGLSGEI PSSIGNLKKL SMLALYNCKF SGKVPPQIFN 422
                                           500                   520                   540
LOC_Os12g11370       LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLHGENSSSL VPFPKIKLLR 517
Fanny_h_Os12g11370   LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLHGENSSSL VPFPKIKLLR 517
Jasmine85_h_Os12g11370 LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLHGENSSSL VPFPKIKLLR 517
Kosh_h_Os12g11370    LTQLQSLQLH SNNLAGTVEL TSFTKLKNLS VLNLSNNKLL VLRGENSSSL VPFPKIKLLR 482
                                           560                   580                   600
LOC_Os12g11370       LASCSISTFP NILKHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN ISHNNITSLG 577
Fanny_h_Os12g11370   LASCSISTFP NILKHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN ISHNNITSLG 577
Jasmine85_h_Os12g11370 LASCSISTFP NILKHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN ISHNNITSLG 577
Kosh_h_Os12g11370    LASCSISTFP NILRHLHEIT TLDLSHNKIQ GAIPQWAWET WRGMYFLLLN MSHNNITSLG 542
                                           620                   640                   660
LOC_Os12g11370       SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSTY LGETFTFKAS 637
Fanny_h_Os12g11370   SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSTY LGETFTFKAS 637
Jasmine85_h_Os12g11370 SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSTY LGETFTFKAS 637
Kosh_h_Os12g11370    SDPLLPLEID FFDLSFNSIE GPIPVPQEGS TMLDYSSNQF SSMPLHYSAY LGQTFTFKAS 602
                                           680                   700                   720
LOC_Os12g11370       KNKLSGNIPS ICSAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 697
Fanny_h_Os12g11370   KNKLSGNIPS ICSAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 697
Jasmine85_h_Os12g11370 KNKLSGNIPS ICSAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 697
Kosh_h_Os12g11370    KNKLSGNIPS ICTAPRLQLI DLSYNNLSGS IPSCLMEDVT ALQILNLKEN KLVGTIPDNI 662
                                           740                   760                   780
LOC_Os12g11370       KEGCALEAID LSGNLFEGRI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLALK 757
Fanny_h_Os12g11370   KEGCALEAID LSGNLFEGRI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLALK 757
Jasmine85_h_Os12g11370 KEGCALEAID LSGNLFEGRI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLALK 757
Kosh_h_Os12g11370    KEGCALEAID LSGNLFEGKI PRSLVACRNL EILDIGNNEI SDSFPCWMSK LPKLQVLVLK 722
                                           800                   820                   840
LOC_Os12g11370       SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISDNDTLV 817
Fanny_h_Os12g11370   SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISDNDTLV 817
Jasmine85_h_Os12g11370 SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISDNDTLV 817
Kosh_h_Os12g11370    SNKFTGQIMD PSYTVDGNSC EFTELRIADM ASNNFNGTLP EAWFTMLKSM NAISENDTLV 782
                                           860                   880                   900
LOC_Os12g11370       MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 877
Fanny_h_Os12g11370   MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 877
Jasmine85_h_Os12g11370 MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 877
Kosh_h_Os12g11370    MENQYYHGQT YQFTAAVTYK GNYITISKIL RTLVLIDFSN NAFHGTIPET IGELVLLHGL 842
                                           920                   940                   960
LOC_Os12g11370       NMSHNSLTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNTLVGRIP 937
Fanny_h_Os12g11370   NMSHNSLTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNTLVGRIP 937
Jasmine85_h_Os12g11370 NMSHNSLTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNTLVGRIP 937
Kosh_h_Os12g11370    NMSHNELTGP IPTQFGRLNQ LESLDLSSNE LFGEIPKELA SLNFLSILNL SYNALVGRVP 902
```

```
                                          980                              1,000          Transmembrane region
LOC_Os12g11370          NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVLLV LFTALGFGVS 997
Fanny_h_Os12g11370      NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVLLV LFTALGFGVS 997
Jasmine85_h_Os12g11370  NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVLLV LFTALGFGVS 997
Kosh_h_Os12g11370       NSYQFSTFSN NSFLGNTGLC GPPLSKQCDN PQESTVMPYV SEKSIDVVLV LFTALGFGVS 962

                                   Cytoplasmic tail
LOC_Os12g11370          FAITILIVWG RHMKKQR* 1015
Fanny_h_Os12g11370      FAITILIVWG RHMKKQR* 1015
Jasmine85_h_Os12g11370  FAITILIVWG RHMKKQR* 1015
Kosh_h_Os12g11370       FAITILIVWG RHMKKKR* 980
```

Figure 5.12 Amino acid alignment of the candidate *S. hermonthica* resistance RLP gene Os12g11370 from 4 rice cultivars, Nipponbare (top), Fanny, Jasmine85 and Koshihikari. Polymorphic residues are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding (Zhang *et al.*, 2014). Koshihikari sequence used is the FgenesH predicted sequence from Chapter 3.

```
                     Signal peptide                       40                        60
LOC_Os12g11500      MSSSTKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV RCHPDQASAL LRLKHSFNAT 60
Fanny_h_Os12g11500  MSSSTKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV RCHPDQASAL LRLKHSFNAT 60
Jhona349_h_Os12g11500 MSSSTKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV RCHPDQASAL LRLKHSFNAT 60
Darmali_h_Os12g11500 MSSSAKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV PCHPDQASAL LRLKHSFNAT 60
Kosh_h_Os12g11500   M--------- ---------- YILLQVQATT NTARTVVPPV PCHPDQASAL LRLKHSFNAT 41
Azucena_h_Os12g11500 MSSSAKRVAH HLPSLLLTAM YILLQVQATT NTARTVVPPV PCHPDQASAL LRLKHSFNAT 60
                                80                  100                   120
LOC_Os12g11500      AGDYSTAFQS WVAGTDCCRW DGVGCGG-AD GRVTSLDLGG HQLQAGSVDP ALFRLTSLKH 119
Fanny_h_Os12g11500  AGDYSTAFQS WVAGTDCCRW DGVGCGG-AD GRVTSLDLGG HQLQAGSVDP ALFRLTSLKH 119
Jhona349_h_Os12g11500 AGDYSTAFQS WVAGTDCCRW DGVGCGG-AD GRVTSLDLGG HQLQAGSVDP ALFRLTSLKH 119
Darmali_h_Os12g11500 AGDYSTAFRS WVAGTDCCRW DGVGCGGGAD GRVTSLDLGG HNLQAGSVDP ALFRLTSLKH 120
Kosh_h_Os12g11500   AGDYSTAFRS WVAGTDCCRW DGVGCGGGAD GRVTSLDLGG HNLQAGSVDP ALFRLTSLKH 101
Azucena_h_Os12g11500 AGDYSTAFRS WVAGTDCCRW DGVGCGGGAD GRVTSLDLGG HNLQAGSVDP ALFRLTSLKH 120
                                140                 160                   180
LOC_Os12g11500      LNLSGNDFSM SQLPVITGFE QLTELVYLDL SDTNIAGEVP GSIGRLTNLV YLDLSTSFYI 179
Fanny_h_Os12g11500  LNLSGNDFSM SQLPVITGFE QLTELVYLDL SDTNIAGEVP GSIGRLTNLV YLDLSTSFYI 179
Jhona349_h_Os12g11500 LNLSGNDFSM SQLPVITGFE QLTELVYLDL SDTNIAGEVP GSIGRLTNLV YLDLSTSFYI 179
Darmali_h_Os12g11500 LNLSGNDFSM SQLPVITGFE QLTELVHLHL SDTNITGEVP GSIGRLTNLV YLDLSTSFYI 180
Kosh_h_Os12g11500   LNLSGNDFSM SQLPVITGFE QLTELVHLHL SDTNITGEVP GSIGRLTNLV YLDLSTSFYI 161
Azucena_h_Os12g11500 LNLSGNDFSM SQLPVITGFE QLTELVHLHL SDTNITGEVP GSIGRLTNLV YLDLSTSFYI 180
                                200                 220                   240
LOC_Os12g11500      VEYNDDEQVT FDSDSVWQLS APNMETLIEN HSNLEELHMG MVDLSGNGER WCDNIAKYTP 239
Fanny_h_Os12g11500  VEYNDDEQVT FDSDSVWQLS APNMETLIEN HSNLEELHMG MVDLSGNGER WCDNIAKYTP 239
Jhona349_h_Os12g11500 VEYNDDEQVT FDSDSGWQLS APNMETLIEN LSNLEELHMG MVDLSGNGER WCDNIAKYTP 239
Darmali_h_Os12g11500 VEYNDDEQVT FNSDSVWQLS APNMETLLEN LTNLEKLHMG MVDLSGNGER WCYNIAKYTP 240
Kosh_h_Os12g11500   VEYNDDEQVT FNSDSVWQLS APNMETLLEN LTNLEKLHMG MVDLSGNGER WCYNIAKYTP 221
Azucena_h_Os12g11500 VEYNDDEQVT FNSDSVWQLS APNMETLLEN LTNLEKLHMG MVDLSGNGER WCYNIAKYTP 240
                                260                 280                   300
LOC_Os12g11500      KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 299
Fanny_h_Os12g11500  KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 299
Jhona349_h_Os12g11500 KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNI TVLQLSKNKF 299
Darmali_h_Os12g11500 KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 300
Kosh_h_Os12g11500   KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 281
Azucena_h_Os12g11500 KLQVLSLPYC SLSGPICASF SALQALTMIE LHYNHLSGSV PEFLAGFSNL TVLQLSKNKF 300
                                320                 340                   360
LOC_Os12g11500      QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLNNTNFTG TIPGSIINLI 359
Fanny_h_Os12g11500  QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLNNTNFTG TIPGSIINLI 359
Jhona349_h_Os12g11500 QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLSNTNFTG TIPGSIINLI 359
Darmali_h_Os12g11500 QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLSNTNFTG TIPSSIINLI 360
Kosh_h_Os12g11500   QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLSNTNFTG TIPSSIINLI 341
Azucena_h_Os12g11500 QGSFPPIIFQ HKKLRTINLS KNPGISGNLP NFSQDTSLEN LFLSNTNFTG TIPSSIINLI 360
                                380                 400                   420
LOC_Os12g11500      SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLTSL TVLRISNCGL 419
Fanny_h_Os12g11500  SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLTSL TVLRISNCGL 419
Jhona349_h_Os12g11500 SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLTSL TVLRISNCGL 419
Darmali_h_Os12g11500 SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLSSL TVLRFSNCGL 420
Kosh_h_Os12g11500   SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLSSL TVLRFSNCGL 401
Azucena_h_Os12g11500 SVKKLDLGAS GFSGSLPSSL GSLKYLDMLQ LSGLQLVGTI PSWISNLSSL TVLRFSNCGL 420
                                440                 460                   480
LOC_Os12g11500      SGPVPSSIGN LRELTTLALY NCNFSGTVHP QILNLTRLQT LLLHSNNFAG TVDLTSFSKL 479
Fanny_h_Os12g11500  SGPVPSSIGN LRELTTLALY NCNFSGTVHP QILNLTRLQT LLLHSNNFAG TVDLTSFSKL 479
Jhona349_h_Os12g11500 SGPVPSSIGN LRELTTLALY NCNFSGTVHP QILNLTRLQT LLLHSNNFAG TVDLTSFSKL 479
Darmali_h_Os12g11500 SGQVPSSIGN LRELTTLALY NCNFSGTVPP QILNLTRLQT LLLHSNNFAG TVELTSFSKL 480
Kosh_h_Os12g11500   SGQVPSSIGN LRELTTLALY NCNFSGTVPP QILNLTRLQT LLLHSNNFAG TVELTSFSKL 461
Azucena_h_Os12g11500 SGQVPSSIGN LRELTTLALY NCNFSGTVPP QILNLTRLQT LLLHSNNFAG TVELTSFSKL 480
                                500                 520                   540
LOC_Os12g11500      KNLTFLNLSN NKLLVVEGKN SSSLVLFPKL QLLSLASCSM TTFPNILRDL PDITSLDLSN 539
Fanny_h_Os12g11500  KNLTFLNLSN NKLLVVEGKN SSSLVLFPKL QLLSLASCSM TTFPNILRDL PDITSLDLSN 539
Jhona349_h_Os12g11500 KNLTFLNLSN NKLLVVEGKN SSSLVLFPKL QLLSLASCSM TTFPNILRDL PDITSLDLSN 539
Darmali_h_Os12g11500 KNLTFLNLSN NKLLVVEGKN SSSLVSLHKL QLLSLASCSM TTFPNILRHL PEITSLDLSN 540
Kosh_h_Os12g11500   KNLTFLNLSN NKLLVVEGKN SSSLVSLHKL QLLSLASCSM TTFPNILRHL PEITSLDLSN 521
Azucena_h_Os12g11500 KNLTFLNLSN NKLLVVEGKN SSSLVSLHKL QLLSLASCSM TTFPNILRHL PEITSLDLSN 540
                                560                 580                   600
LOC_Os12g11500      NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LYVEYFDLSF NSIEGPIPIP 599
Fanny_h_Os12g11500  NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LYVEYFDLSF NSIEGPIPIP 599
Jhona349_h_Os12g11500 NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LYVEYFDLSF NSIEGPIPIP 599
Darmali_h_Os12g11500 NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LSVEYFDLSF NSIGGPIPIP 600
Kosh_h_Os12g11500   NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LSVEYFDLSF NSIGGPIPIP 581
Azucena_h_Os12g11500 NQIQGAIPQW AWKTWKGLQF IVLNISHNNF TSLGSDPFLP LSVEYFDLSF NSIGGPIPIP 600
                                620                 640                   660
LOC_Os12g11500      QEGSSTLDYS SNQFSSMPLR YSTYLGETVT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 659
Fanny_h_Os12g11500  QEGSSTLDYS SNQFSSMPLR YSTYLGETVT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 659
Jhona349_h_Os12g11500 QEGSSTLDYS SNQFSSMPLR YSTYLGETVT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 659
Darmali_h_Os12g11500 QEGSSTLDYS SNQFSSMPLR YSTYLGETLT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 660
Kosh_h_Os12g11500   QEGSSTLDYS SNQFSSMPLR YSTYLGETLT FKASKNKLSG NVPPLICTTA RKLQLIDLSY 641
Azucena_h_Os12g11500 QEGSSTLDYS SNQFSSMPLR YSTYLGETLT FKASKNKTSG NVPPLICTTA RKLQLIDLSY 660
                                680                 700                   720
LOC_Os12g11500      NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 719
Fanny_h_Os12g11500  NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 719
Jhona349_h_Os12g11500 NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 719
Darmali_h_Os12g11500 NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 720
Kosh_h_Os12g11500   NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 701
Azucena_h_Os12g11500 NNLSGSIPSC LLESFSELQV LSLKANKFVG KLPDIIKEGC ALEALDLSDN SIEGKIPRSL 720
```

```
                                                  740                 760                 780
LOC_Os12g11500       VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQISCEFPA 779
Fanny_h_Os12g11500   VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQISCEFPA 779
Jhona349_h_Os12g11500 VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQISCEFPA 779
Darmali_h_Os12g11500 VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQNSCEFPS 780
Kosh_h_Os12g11500    VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQNSCEFPS 761
Azucena_h_Os12g11500 VSCRNLEILD IGSNQISDSF PCWLSQLPKL QVLVLKSNKL TGQVMDPSYT GRQNSCEFPS 780
                                                  800                 820                 840
LOC_Os12g11500       LRIADMASNN LNGMLMEGWF KMLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGNDR 839
Fanny_h_Os12g11500   LRIADMASNN LNGMLMEGWF KMLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGNDR 839
Jhona349_h_Os12g11500 LRIADMASNN LNGMLMEGWF KMLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGNDR 839
Darmali_h_Os12g11500 LRIADMASNN LNGMLMEGWF KKLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGKSM 840
Kosh_h_Os12g11500    LRIADMASNN LNGMLMEGWF KKLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGKSM 821
Azucena_h_Os12g11500 LRIADMASNN LNGMLMEGWF KKLKSMMARS DNDTLVMENQ YYHGQTYQFT ATVTYKGKSM 840
                                                  860                 880                 900
LOC_Os12g11500       TISKILRSLV LIDVSGNAFH GAIPDTIGEL VLLRGLNLSH NALTGPIPSQ FCRLDQLESL 899
Fanny_h_Os12g11500   TISKILRSLV LIDVSGNAFH GAIPDTIGEL VLLRGLNLSH NALTGPIPSQ FCRLDQLESL 899
Jhona349_h_Os12g11500 TISKILRSLV LIDVSGNAFH GAIPDTIGEL VLLRGLNLSH NALTGPIPSQ FGRLDQLESL 899
Darmali_h_Os12g11500 TFSKVLRTLV LIDFSNNAFH GTIPETIGEL ILLHGLNISH NALTGPIPPQ LGRLNQLESL 900
Kosh_h_Os12g11500    TFSKVLRTLV LIDFSNNAFH GTIPETIGEL ILLHGLNISH NALTGPIPPQ LGRLNQLESL 881
Azucena_h_Os12g11500 TFSKVLRTLV LIDFSNNAFH GTIPETIGEL ILLHGLNISH NALTGPIPPQ LGRLNQLESL 900
                                                  920                 940                 960
LOC_Os12g11500       DLSFNELSGE IPKELASLNF LSTLNLSNNT LVGRIPDSYQ FSTFSNSSFL GNTGLCGLPL 959
Fanny_h_Os12g11500   DLSFNELSGE IPKELASLNF LSTLNLSNNT LVGRIPDSYQ FSTFSNSSFL GNTGLCGLPL 959
Jhona349_h_Os12g11500 DLSFNELSGE ISKELASLNF LSTLNLSNNT LVGRIPDSYQ FSTFSNSSFL GNTGLCGLPL 959
Darmali_h_Os12g11500 DLSSNKLSGK IPNELESLNF LSTLNLSYNV LVGRIPDSYQ FSTFSNSSFL GNTGLCGPPF 960
Kosh_h_Os12g11500    DLSSNKLSGK IPNELESLNF LSTLNLSYNV LVGRIPDSYQ FSTFSNSSFL GNTGLCGPPF 941
Azucena_h_Os12g11500 DLSSNKLSGK IPNELESLNF LSTLNLSYNV LVGRIPDSYQ FSTFSNSSFL GNTGLCGPPF 960
                                                  980
                                            Transmembrane region        Cytoplasmic tail
LOC_Os12g11500       SRQCDNPEEP SAIPYTSEKS IDAVLLLFTA LGFGISFAMT ILIVWGSHMR KRH* 1013
Fanny_h_Os12g11500   SRQCDNPEEP SAIPYTSEKS IDAVLLLFTA LGFGISFAMT ILIVWGSHMR KRH* 1013
Jhona349_h_Os12g11500 SRQCDNPEEP SAIPYTSEKS IDAVLLLFTA LGFGISFAMT ILIVWGSHMR KRH* 1013
Darmali_h_Os12g11500 V*-------- ---------- ---------- ---------- ---------- ---- 962
Kosh_h_Os12g11500    V*-------- ---------- ---------- ---------- ---------- ---- 943
Azucena_h_Os12g11500 V*-------- ---------- ---------- ---------- ---------- ---- 962
```

Figure 5.13 Amino acid alignment of the candidate *S. hermonthica* resistance RLP gene Os12g11500 from 6 rice cultivars, Nipponbare (top), Darmali, Fanny, Koshihikari, Jhona349 and Azucena. Polymorphic residues are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding (Zhang *et al.*, 2014). Koshihikari sequence used is the FgenesH predicted sequence from Chapter 3.

5.4 Discussion

This chapter investigated the hypothesis that the diversity of the candidate *S. hermonthica* resistance genes in rice can help identify genes or combinations of genes underlying the resistance to *S. hermonthica*. The genomic structure of the QTL was first examined at a broad level, which revealed considerable differences both within and also between rice subspecies for this region. This is commonly observed in resistance gene clusters, where extensive duplications and reshuffling of genes often occurs. The duplicated genes then diverge through accumulating mutations in their sequence, thus increasing the complexity of the R gene sequence (Meyers *et al.*, 2005). Predictions of resistance and susceptibility of cultivars were made based on the similarity of the RLP genes to Nipponbare and cultivars were phenotyped for resistance to *S. hermonthica*. Interestingly, a good correlation was observed between *S. hermonthica* resistance and the similarity of a number of candidate resistance genes.

5.4.1 Can any of the candidate resistance genes be discounted as being involved in resistance to *S. hermonthica*?

Examination of SNP data does not reveal the biological function of a gene. In addition, a number of assumptions were made that must be taken into account. These were: (1) resistance is due to large differences in gene structure rather than a single SNP; (2) homologs of any candidate resistance genes in susceptible cultivars share low similarity to those of Nipponbare; (3) resistant cultivars may be resistant because they contain resistance genes in other parts of the genome and (4) resistance in IR64 and Nipponbare is due to the same gene cluster. The importance of the QTL region in providing resistance can also be also considered under different scenarios: 1) that resistance is provided by a single gene; or 2) that resistance is provided by two or more genes, possibly in an additive manner. Nevertheless, the presence of a highly conserved gene in several resistant cultivars, and its absence or lack of close similarity in susceptible cultivars, could indicate a possible role for the gene in *S. hermonthica* resistance.

A correlation analysis was performed for individual candidate resistance genes to search for associations between SNP similarity to the Nipponbare allele and resistance to *S. hermonthica*. The genes that explained the most variation, and are therefore top candidates for resistance genes, occurred in the proximal region of the QTL between the genes Os12g10820 – Os12g11510 (5.8 – 6.2 Mb) on the chromosome. The high similarity to Nipponbare for several genes in this region in most resistant cultivars, and their poor similarity in susceptible cultivars, is consistent with the hypothesis that more than one gene provides resistance to *S. hermonthica*, either showing functionally redundancy, or possibly acting independently to

detect different genetic variants of the highly diverse parasite. However, as all these genes are present in the same region of the QTL, the high R^2 values could be a result of genetic linkage rather than all these genes conferring resistance. This meant it was difficult to distinguish between each of these genes and their importance in *S. hermonthica* resistance. It must be noted that the correlation between SNP similarity to Nipponbare and resistance is not always informative however, and the distribution of the cultivars can be more useful in determining their importance as candidate resistance genes. By imposing a very stringent measure of resistance (< 5 mg *Striga* dry weight) and SNP similarity (> 80 %), four RLP genes were identified as the best candidates for resistance: Os12g10870, Os12g11370, Os12g11500 and Os12g11510. Previous work in Chapter 4 found homozygous *Tos17* insertion lines in Os12g10870 to be resistant to *S. hermonthica*. In addition, Os12g11510 is not expressed in Nipponbare. Therefore if a single gene acts on its own to provide resistance, the best candidate genes can be further limited to Os12g11370 and Os12g11500. It is possible that other genes may also be involved, however.

Although these four genes appeared to be good candidates for *S. hermonthica* resistance in many of the resistant cultivars investigated, this was not the case for all cultivars. This was not unexpected given that resistant cultivars were not assumed to contain the resistance gene(s) in the QTL, as alternative forms of resistance may exist elsewhere in their genome. This meant that susceptible cultivars were more useful in determining the importance of each gene in conferring resistance. The resistant *aromatic* cultivar Darmali was a clear example of this. Darmali lacked close similarity to Nipponbare or IR64 over much of the QTL region. The resistant *aus* cultivar Mehr also differed to Nipponbare over much of the region. These cultivars could therefore provide valuable genetic resources for the discovery of new sources of *S. hermonthica* resistance.

Under the assumption that the genes providing resistance in Nipponbare and IR64 are the same, the sub-region between the genes Os12g11660 and Os12g11940, as well as the genes Os12g12120 and Os12g12130, are less likely to be involved in resistance to *S. hermonthica,* as these genes are missing from the IR64 genome (Chapter 3). All these genes showed a much poorer correlation between SNP similarity to Nipponbare and resistance generally, and appeared to be missing in most cultivars investigated, including most *indica* and tropical *japonica* cultivars. This suggests the sub-region is not essential for *S. hermonthica* resistance in IR64. However, SNP data indicates that eight cultivars do contain this region, all of which exhibited strong resistance to *S. hermonthica*. Many of the genes in this region have undergone duplication (Chapter 3) and the similarity of these genes with their duplicated pair

could result in possible functional redundancy. Therefore although these duplicated genes may not be required, they may contribute to *S. hermonthica* resistance in some cultivars, as it is not known whether one, both or neither of the duplicated genes is important. This is similar to the *Cf-2* locus in tomato, containing two very similar, functionally redundant genes (Dixon *et al.*, 1996). If both copies of the duplicated genes provide resistance to *S. hermonthica*, they may both need to be deactivated to induce susceptibility.

Although no significant correlation was observed between SNP similarity for the RALF gene Os12g12000 and resistance / susceptibility to *S. hermonthica*, it is interesting to note that every cultivar possessed a copy of this gene or its duplicated pair Os12g11660. This is perhaps not surprising given the widespread nature of the RALF family in the plant kingdom (Murphy & De Smet, 2014), and may argue against its role in *S. hermonthica* resistance. In addition, the most resistant *indica* cultivars showed the greatest SNP difference to Nipponbare for this gene, which does not agree with the assumption that resistance in IR64 and Nipponbare is the same if this gene is involved in resistance.

All cultivars investigated possessed a homolog of Os12g12010, with the exception of Firooz, Kitrana508, Guan-Yin-Tsan and Ai-Chiao-Hong, which contained a homolog more similar to Os12g11680. SNP similarity to Nipponbare for Os12g12010 was over 96% for most of the susceptible cultivars (Figure 5.9E), which strongly argues against a role for this gene in resistance. This agrees with the findings in Chapter 3, where Os12010 in the susceptible cultivar Azucena differed from Nipponbare by only 2 amino acids, compared to 43 in IR64, suggesting this gene is very unlikely to be involved in resistance if conferred by a single gene.

Thus the work carried out in this chapter has allowed regions of the QTL to be identified that are more likely to be involved in resistance. However, the possible interaction between multiple genes and the genetic linkage observed between 5.8 – 6.2 Mb means that several genes remain good candidates for *S. hermonthica* resistance.

A very similar study was carried out by Garris *et al.*, (2003) on the *xa5* locus in rice conferring race-specific resistance to bacterial blight. They used SSR markers to characterise linkage disequilibrium in the 70-kb candidate region for 114 rice accessions to analyse haplotype diversity and determine if it was possible to reduce the number of candidate genes. Like the present study, extensive linkage disequilibrium (LD) was observed. A 45-kb region in resistant *aus-boro* accessions showed significant LD, a pattern not observed in susceptible accessions, potentially reducing the size of the candidate region. Different haplotypes that associated with

resistance were also observed between *aus-boro* and *indica* accessions, suggesting genetic heterogeneity of resistance exists for this pathogen in rice. Despite the much larger number of accessions investigated, the location of the candidate gene within the region remained unresolved due to significant LD (Garris *et al.*, 2003). This kind of association study is likely to prove more successful in outcrossing species such as maize where linkage disequilibrium declines more rapidly (Flint-Garcia *et al.*, 2003).

5.4.2 Interpretation of the data in the light of the assumptions

As stated in section 5.1 (Introduction), many assumptions were made that must be taken into account when performing this kind of investigation, meaning interpretation of results must be taken with great caution. The first assumption is that resistance is due to large differences in gene structure rather than a single SNP; a significant negative correlation between SNP similarity to Nipponbare and *S. hermonthica* resistance would not be observed if susceptible cultivars possessed near identical alleles to resistant cultivars. The similarity of genes between cultivars can be considered at different levels of hierarchy: presence/absence; large differences in sequence or insertions / deletions; or single differences in nucleotides. While the absence of a gene is easier to interpret, it is very difficult to determine how small differences in sequence affect the biological function of the protein simply by comparing SNPs between cultivars, as very few changes in amino acid sequence may affect the function of a protein. For example, the strength of a resistance response can be determined by differences in amino acids in other parts of a protein; the severity of necrosis conferred by the RLP gene *Cf-9* was found to be enhanced by the presence of LRRs 10 to 12 (Chakrabarti *et al.*, 2009). However, small changes in amino acid sequence do not always affect the function of a protein. Dixon *et al.*, (1996) identified two *Cf-2* genes at the *Cf-2* locus in tomato whose sequences differed by three amino acids. These two genes were functionally equivalent, each conferring resistance to *C. fulvum* strains expressing the *Avr2* avirulence gene (Dixon *et al.*, 1996). It should also be noted that while resistance in both Nipponbare and IR64 is assumed to be the same, only one gene (Os12g11370) is 100% identical between these two cultivars at the amino acid level. If other genes in the QTL are providing resistance, it is assumed that they are functionally equivalent between these two cultivars.

Wulff *et al.*, (2009) showed that recognition specificity of the resistance gene *Cf-9* to *C. fulvum* was determined by only five amino acids in the central LRR region. Therefore the second assumption, that homologs of resistance gene(s) in susceptible cultivars are absent or very different, may not be the case, as resistance may be determined by a single nucleotide polymorphism. However, only one gene (Os12g12010) investigated in this study showed high

similarity to Nipponbare in susceptible cultivars; Os12g12010 had over 93 % similarity to Nipponbare in all susceptible cultivars. In contrast, the similarity of this gene in resistant cultivars was much less (0 – 84 %). This suggests that small differences in gene sequence here are unlikely to have affected the interpretation of results in this study.

In order to see a correlation between SNP similarity for a given gene and *S. hermonthica* resistance it not only requires that resistance is due to large differences in gene structure, but also that resistant cultivars do not posses resistance in other parts of their genome, and thus obscure the association. Significant associations were observed for a number of genes, indicating other sources of resistance were not a serious problem in this study. This suggests very few novel sources of resistance are present in the cultivars investigated for this ecotype of *S. hermonthica*, reinforcing the importance of the QTL on chromosome 12 in providing resistance to this ecotype of *S. hermonthica.*

In addition to the assumptions above, interpretations of the SNP data itself must also be considered due to potential errors in sequencing and assembly. Like the vast majority of rice genomes, sequenced reads of the cultivars analysed were aligned onto the Nipponbare reference genome to identify the generated SNPs (Duitama *et al.*, 2015), rather than being assembled *de novo*. This can cause potential misalignments or elimination of critical sequences that could not be aligned with confidence, a particular problem with *indica*, *aus* and other divergent genomes (Schatz *et al.*, 2014). These cultivars may contain regions or insertions not present in the Nipponbare reference genome, and would therefore be absent in any subsequent analysis (Sakai *et al.*, 2014). This was observed in a number of cultivars when SNP data was verified by PCR amplification. PCR products of Os12g11370 were larger in Azucena, Koshihikari and nine other cultivars when compared to products amplified from IR64 or Nipponbare, which was not detectable from SNP data. Duplications in the genomes of cultivars being sequenced can also introduce errors such as false heterozygotes. For example, both copies of a duplicated gene in a diverged cultivar would map the same position on the Nipponbare reference, and appear as heterozygous. Although copy number variant (CVN) analysis can detect duplicated genes, this analysis was not performed here. Conversely, mapping a single gene onto a reference genome that has undergone recent duplication will result in it mapping to two positions. (Dr. Mathias Lorieux, personal communication). IR64 is a good example of this. Previous work in Chapter 3 showed that IR64 contains only 6 RLP homologs in the QTL, compared to 13 in Nipponbare, which is not clear from the SNP analysis in this study. For example, one gene in IR64 (IR64_h_Os12g12010) shares high similarity to two duplicated genes Os12g12010 and Os12g11680 in Nipponbare, being a hybrid of the two

Nipponbare sequences. SNP data is this study showed that IR64 contained sequence that was 65 % and 39 % similar in SNPs to Os12g12010 and Os12g11680 respectively, but does not show that these sequences are from the same gene.

Finally, the lack of SNPs in a gene does not mean it is absent, as the detection of SNPs is reliant on there being differences between those genes. Therefore if most of a gene is identical to the reference genome but the rest is not, only SNPs in a small part of that gene will be detectable if all accessions investigated have the same sequence. The use of an additional reference sequence during genome assembly could greatly improve the detection of sequence polymorphisms that cannot be detected using the Nipponbare sequence alone (Schatz *et al.*, 2014; Sakai *et al.*, 2014).

5.4.3 Other levels of complexity: post transcriptional and post-translational modifications

While the presence and sequence of a gene is clearly very important for specificity, post transcriptional and post-translational modifications add a further layer of complexity with the potential to alter the functioning of a protein, and would go undetected in direct comparisons of gene sequences. For example, alternative splicing of R genes has been shown to be involved in defence against some pathogens. The TIR-NBS-LRR-type resistance gene *RCT1* undergoes alternative splicing in *Medicargo truncatula*. The alternative transcript encodes a truncated protein lacking the C-terminal domain. Both the regular and the alternatively spliced transcripts are required for resistance to the fungal pathogen *Colletotrichum trifolii* (Tang *et al.*, 2013). Similarly, both full length and truncated forms of the *RPS4* gene transcript in *Arabidopsis* were required for resistance to *Pseudomonas syringae* pv *tomato* strain DC3000, as plants which could not produce the alternatively spliced variant were susceptible (Zhang & Grassmann, 2003). Epigenetic changes, such as DNA methylation and histone modifications, can also influence gene expression, and thus the phenotype of a species or population, in the absence of sequence variation, and are known to be heritable across generations (Latzel *et al.*, 2013; Zhang *et al.*, 2013). Latzel *et al.*, (2012) used epiRILSs of *Arabidopsis*, lines nearly identical in DNA sequence but highly variable in levels of DNA methylation, to show that variation in response to plant defence hormones may be due to epigenetic, rather than purely genetic, variation (Latzel *et al.*, 2012).

Transposable elements known to be a major player in the epigenetic control of endogenous gene expression through alterations in DNA methylation, histone modifications and production of non-coding RNAs (Song & Cao, 2017). For example, siRNAs derived from TEs have been shown to regulate endogenous gene expression through RNA-directed DNA methylation

(RdDM), whereby siRNAs direct the methylation of cytosines for DNA sequences that are complementary to the siRNA sequence (Chinnusamy & Zhu, 2009; Wei *et al.*, 2014). Kinoshita *et al.*, (2006) showed that small RNA thought to originate from a short interspersed nuclear element (SINE) retrotransposon was responsible for directing DNA methylation of cytosines around the transcriptional start site of FWA, the gene responsible for flowering time in *Arabidopsis thaliana*. Variation in expression of FWA between isolates with identical gene sequences were associated with changes in DNA methylation (Fujimoto *et al.*, 2011). It is therefore possible that some of the variation in resistance to *S. hermonthica* observed between different rice cultivars with apparently very similar gene sequences across much of the QTL might be due to different levels of expression as a result of differences in RNA-directed DNA methylation. Interestingly, the methylation of cytosines can in turn be a source of higher rates of genetic mutation. In *A. thaliana*, G:C sites showing at least partial methylation had a higher probability of mutation to A:T when compared to non-methylated C:G sites (Ossowski *et al.*, 2010). Thus, it is tempting to speculate that TEs may be responsible for some of the variation in SNPs in addition to larger genome rearrangements.

Finally, altered levels of glycosylation of cell surface pattern recognition receptors (PRRs) are known to affect downstream signalling responses in plant immunity. The *Arabidopsis* receptor kinase EFR lacking normal levels of glycosylation in the extracellular domain showed reduced levels of expression and lost the ability to bind its ligand and thus transmit a defence response (Häweker *et al.*, 2010). The RLP *Cf-9* in tomato contains 22 putative N-linked glycosylation sites in the extracellular domain, 19 of which occur in the LRR domain and nearly all of which are glycosylated (van der Hoorn *et al.*, 2005). Mutant experiments revealed that most of the glycosylation sites contributed to *Cf-9* activity, and those on the outer helices of the LRR regions were essential (van der Hoorn *et al.*, 2005). As many of the candidate *S. hermonthica* resistance genes are annotated as cell surface RLP genes, it is possible that different glycosylation patterns between cultivars could result in different levels of resistance to the parasite.

5.4.4 Conclusions and future directions

In this study diverse rice cultivars were used to further dissect the importance of the *S. hermonthica* candidate resistance genes on chromosome 12. The region of the QTL between positions 5.8 – 6.2 Mb was identified as the most likely to be important in providing resistance to *S. hermonthica*. Within this region, four RLP genes were identified as being the best candidates, showing high similarity to Nipponbare in resistant cultivars and being very different in susceptible cultivars. Of these, Os12g11370 and Os12g11500 were identified as top

candidates for single gene resistance. These genes were highly similar in genetically diverse cultivars, suggesting resistance from this region may be very important in providing resistance to this ecotype of *S. hermonthica* in rice generally, and be easily transferable into diverse genetic backgrounds. However, genetic linkage in this region meant that it was not possible to determine whether more than one gene is involved, and it is possible that two or more genes may be conferring resistance, possibly in an additive manner. A much larger association analysis using cultivars selected from The 3000 Rice Genomes Project, (2014) may provide further insight if cultivars could be selected showing historical recombination events that break apart the linked genes in the region. Once the resistance gene or genes have been functionally validated, comparing small differences in SNPs between a wide range of cultivars may help determine the functional amino acids and identify novel alleles that may provide new and alternative sources of *S. hermonthica* resistance.

Chapter 6

General Discussion

6.1 Introduction

Striga species are among the most destructive parasites on crop plants (Parker, 2009). Their impacts are particularly devastating in sub-Saharan Africa, where they are estimated to infest over 40 % of cereal producing areas (Scholes & Press, 2008). Rice is one of the major cereals affected by *Striga* species in Africa, and crop losses of rice as a result of *Striga* infection are expected to increase in the future due to rapid expansion of its cultivation, population increases and changing consumer preferences (Rodenburg *et al.*, 2010; Seck *et al.*, 2012). Control strategies must not only be affordable for resource poor farmers, but must also act early on, thus preventing the damaging effects on the host that occur within a few days of attachment. The use of resistant cultivars is thought to be a cost effective control strategy, especially when used as part of an integrated approach aimed at improving soil fertility and reducing the *Striga* seed bank (Rodenburg *et al.*, 2010; Yoder & Scholes, 2010). However, the enormous genetic diversity of *S. hermonthica* and large production of seeds with high longevity means that a single resistance gene is unlikely to be effective in the field. Broad spectrum and durable resistance will require pyramiding multiple resistance genes with different modes of action into a single cultivar (Scholes & Press, 2008). In addition, a better understanding of host resistance with respect to parasite virulence is required to allow predictive breeding to be targeted appropriately for different regions. Despite many years of research, no resistance genes to *S. hermonthica* have yet been identified.

The aims of this thesis were therefore to identify novel QTL and candidate resistance genes in rice, and to functionally validate their role in providing resistance to *S. hermonthica*. Figure 6.1 provides an overview of the work carried out, the findings, and the future research directions following on from this PhD.

6.2 Were novel QTL identified in rice for *S. hermonthica* resistance?

This PhD utilised a rice RIL mapping population derived from a cross between IR64, an *indica* cultivar, and Azucena, a tropical *japonica* cultivar, to search for novel QTL and genes for *S. hermonthica* resistance. Phenotyping and QTL analysis of this population led to the discovery of a major QTL for post-attachment resistance to *S. hermonthica* on chromosome 12, with a LOD score of 14 (Chapter 2). The very high LOD score indicated the presence of a major resistance gene, or a few genes of major effect. Previous work in our laboratory had used the same S. *hermonthica* Kibos isolate that was used in this study to map a major QTL in a BIL population derived from the resistant Nipponbare and the more susceptible Koshihikari

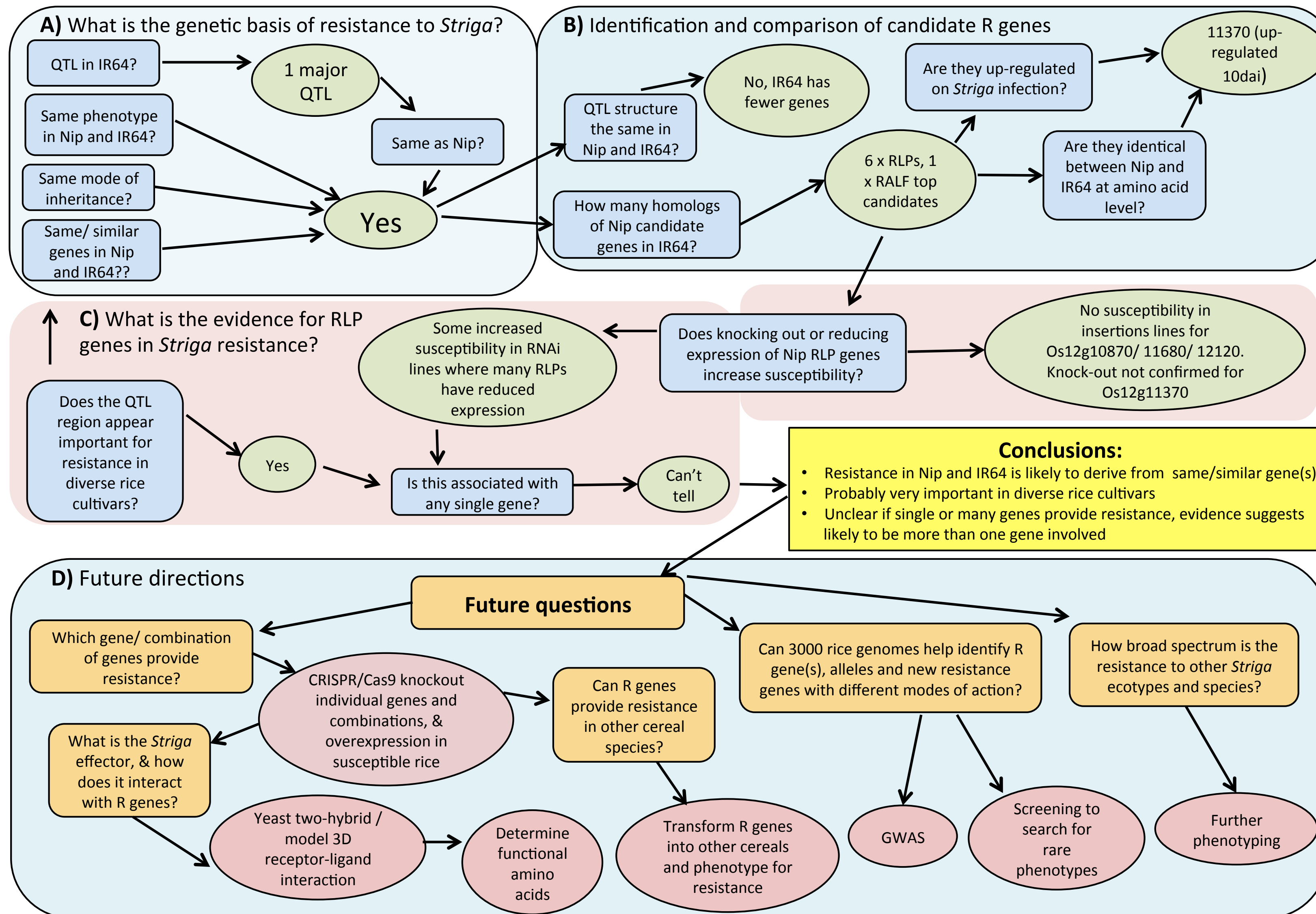


Figure 6.1. An overview of the work carried out in this PhD, the findings, and the future directions for further research into resistance to *Striga* in rice. Blue rectangles indicate questions addressed, green ovals are results determined by the experimental work carried out. Conclusions are highlighted in the yellow box. Future questions are shown in orange rectangles, and future experimental work to address these are shown pink circles.

cultivars, both of which belong to the temperate *japonica* subspecies. Therefore, a population derived from a resistant *indica* cultivar was predicted to identify novel QTL for *S. hermonthica* resistance. *Indica* and *japonica* cultivars are estimated to have diverged 0.4 mya (Zhu & Ge, 2005) which far predates domestication 10,000 years ago (Sang & Ge, 2013), therefore the discovery of a major QTL on chromosome 12 in IR64 that mapped to the same position as the QTL previously identified in Nipponbare was very surprising, as resistance was expected to be different between the different rice subspecies. This led to a number of questions regarding the genetic basis of resistance in IR64, and how this compared to that of Nipponbare (Figure 6.1 A). For example, were the phenotype and mode of inheritance the same between these two cultivars, and were the same or similar genes present in the two QTL regions? Cross sections of the rice root at the site of parasite attachment revealed the phenotype of resistance was similar to that seen in Nipponbare. The resistance response was characterised by necrosis of the rice root tissue at the site of attachment after a few days, and by an inability of the parasite to penetrate the endodermis and fuse its xylem vessels to those of the host. This phenotype is characteristic of a number of different rice cultivars (Yoshida & Shirasu, 2009; Cissoko *et al.*, 2011). The mode of inheritance of this resistance was tested by crossing both IR64 and Nipponbare to two different susceptible cultivars (Azucena and Koshihikari) to produce F_1 plants, and phenotyping them for resistance to *S. hermonthica*. In all cases, F_1 plants showed intermediate resistance between the parents, indicating co-dominance. In addition, gene prediction software was used to identify a cluster of RLP genes in the IR64 QTL that showed very high similarity to a gene cluster in the Nipponbare QTL. All this provided further evidence to support the hypothesis that resistance in Nipponbare and IR64 was likely to be derived from the same or similar RLP genes. Differences do exist between them however; whereas Nipponbare possesses thirteen tightly linked RLP genes in the QTL, the IR64 cluster contains only six, and only Os12g11370 is 100 % identical at the amino acid level (Figure 6.1 B). Phylogenetic analysis revealed this region in Nipponbare underwent a duplication event, which clearly took place after its divergence from *indica* subspecies.

The discovery of the same *S. hermonthica* resistance QTL in both *indica* and temperate *japonica* cultivars poses many intriguing questions as to the evolution of the region. Interestingly, this region in Nipponbare is more similar to IR64 than it is to Koshihikari, a temperate *japonica* cultivar and so more closely related to Nipponbare than IR64 (for phylogenetic tree of rice see Figure 5.2). The most likely explanation for this is that an ancient introgression of this region occurred between them at some point during their evolution. As Koshihikari is missing much of the region, it is possible an introgression occurred from *indica* to Nipponbare after its divergence from Koshihikari. Multiple introgressions are known to have

taken place between *indica* and *japonica* during the course of their evolution, as well as hybridisation events with wild rice species (Yang *et al.*, 2012b; Gross & Zhao, 2014). It is also possible that Koshihikari lost most of the region, or that IR64 and Nipponbare each obtained the region from an alternative source independently, followed by a duplication event in Nipponbare. Although unlikely, the final explanation is the independent evolution of similar genes by convergent evolution under similar environmental pressures.

It should be noted that while good resistance to *S. hermonthica* was observed for both IR64 and Nipponbare, neither of these cultivars evolved in the presence of the parasite. IR64 originates from the Phillipines, while Nipponbare is a lowland variety originating from Japan where *Striga* is absent. This is appears to be common in rice, as resistance to *S. hermonthica* has been observed in many cultivars that have not co-evolved with the parasite. For example, Harahap *et al.*, (1993) found *S. hermonthica* resistant genotypes from both *indica* and *japonica* groups that were indigenous to the humid tropics and sub-tropics of Asia where *S. hermonthica* is not found (Harahap *et al.*, 1993). This begs the question as to what other biotic or abiotic factors drove the evolution in this region, and what other roles any *S. hermonthica* resistance genes may have. Interestingly, the work carried out in Chapter 5 suggests that genes present within the QTL region could be very important in providing resistance to *S. hermonthica* in diverse rice cultivars. Further work is required to identify individual genes that underlie resistance in this region, however the presence of highly similar genes not only in *indica* and temperate *japonica* cultivars but also in several *aromatic* and *aus* cultivars could mean that this form of resistance may be easily transferable into diverse genetic backgrounds for crop improvement. Importantly, a number of resistant cultivars were identified that lacked close similarity to the Nipponbare alleles across the entire QTL region. This confirmed the presence of additional sources of *S. hermonthica* resistance in rice germplasm that remain unidentified, but which, if identified, could contribute to providing the much-needed multi-gene resistance required for more durable and broader spectrum resistance to this parasite.

6.3 What is the evidence that the candidate RLP genes underlie resistance to *S. hermonthica*?

The role of the candidate RLP genes in providing resistance to *S. hermonthica* was tested in Nipponbare using RNAi lines targeting suits of the RLP genes, and insertion lines targeting four RLP genes individually (Figure 6.1 C). Increased susceptibility was observed in six RNAi lines. For three of these lines, susceptibility was associated with significant down-regulation of two or three RLP genes, however the genes suppressed were not always the same between these

lines. One susceptible line showed no down-regulation of any RLP gene (JS4.12B-11), while all RLP genes were down-regulated to some extent in another susceptible line (JS8.14-19). No significant down-regulation of gene expression was observed for any of the resistant lines, with the exception of AM3.9-1 where only the expression of the target gene Os12g11500 was reduced. Thus, there is some evidence to suggest that suppression of more than one RLP gene is associated with increased susceptibility to *S. hermonthica*, rather than a single gene acting alone or that other genes in addition to the RLP genes may be required for full susceptibility, as has been suggested for the resistance to *C. reflexa* in tomato provided by the RLP gene *CuRe1* (Hegenauer *et al.*, 2016). The data from the insertion lines agrees with this; knocking out four of the RLP genes individually did not result in increased resistance. However it is also possible that none of the genes knocked out by the insertion are involved in providing *S. hermonthica* resistance.

Further evidence to support the role of the RLP genes in resistance was their presence in diverse but very resistant cultivars (Chapter 5). A comparison of the similarity of genes across the QTL region to Nipponbare for diverse cultivars allowed a region of the QTL between positions 5.8 - 6.2 Mb to be identified that is more likely to contain a resistance gene or genes. Four RLP genes Os12g10870, Os12g11370, Os12g11500 and Os12g11510 were identified as top candidates for resistance genes, with Os12g11370 sharing the greatest similarity between resistance cultivars (> 94 % similar in most cases). A *Tos17* insertion line for Os12g10870 was infected with *S. hermonthica* and found to be resistant (Chapter 4), suggesting this gene is not involved in resistance. Interestingly, a transcript of a large part of the Os12g11370 was present in the T-DNA insertion line for this gene, probably driven by the promoter in the T-DNA insertion itself. The abundance of the transcript was higher than in the wildtype Nipponbare, consistent with the idea that transcription is due to the promoter in the insertion region. It is not clear whether this transcript will produce a functional protein but it is interesting to note that the insertion line was even more resistant than Nipponbare. The lack of expression of Os12g11510 in Nipponbare indicates this gene is also unlikely to be important. Two other genes in this region, the expressed protein Os12g10820 and the DNA repair protein Os12g10850, were also identified in resistant cultivars that had a poor similarity to Nipponbare in susceptible cultivars. However, the similarity of the DNA repair protein in many cultivars was lower than that observed for the RLP genes.

The co-dominance of resistance in both IR64 and Nipponbare, demonstrated by the intermediate resistance of the F_1 plants when crossed with a susceptible cultivar, suggests resistance to *S. hermonthica* acts in a dose-dependent manner. This implies that lower

concentrations of the resistance protein limit the plants ability to activate a defence response, and is consistent with their naturally low levels of expression. This additive form of resistance agrees with the hypothesis that more than one gene may be conferring resistance, especially if they are acting in a similar manner. It is also consistent with their possible role as a receptor, as increased perception of the parasite could be achieved by greater concentrations of the receptor (Hammond-Kosack, 1994). Similar co-dominance was observed for the *Cf* RLP genes in tomato that provide resistance to the fungal pathogen *C. fulvum*. *Cf-9* and *Cf-4* were found to confer only moderate to weak resistance to the pathogen compared to *Cf-2* and *Cf-5*. The apparent differences in the relative strengths of the resistance were thought to be due to different concentrations of the gene products rather than differences in their modes of action (Hammond-Kosack, 1994). If more than one RLP gene in rice is acting in the same way to confer *S. hermonthica* resistance, it may be necessary to inactivate all copies of the genes to bring about loss of function. This is the case for the *Cf-2* locus in tomato which contains two functionally redundant copies of the *Cf-2* gene that differ from each other by only three amino acids (Dixon *et al.*, 1996). As the level of host resistance is also determined by the virulence of the parasite, it is possible that if more than one resistance gene is involved, the different genes could be detecting genetically different parasites from the very diverse *S. hermonthica* seed bank.

Other evidence in support the role of the candidate RLP genes in *S. hermonthica* resistance is their structural similarity to other resistance genes. Their annotation as orthologs of *Ve1*, the gene conferring resistance to *Verticillium* wilt in tomato, cannot be ignored, as the lifestyle and infection process of this pathogen is strikingly similar to that of *Striga* species. Like *Striga* they also have a very broad host range, infecting over two hundred dicotyledonous plant species, with different isolates exhibiting some host specificity (Fradin & Thomma, 2006). Microsclerotia of *Verticillium* wilts are triggered to germinate in the soil in response root exudates from host plants, and hyphae penetrate host roots, crossing the endodermis to reach the xylem (Fradin & Thomma, 2006). It therefore seems highly likely that a resistance mechanism that prevents infection by *Verticillium* wilt could also provide resistance to *Striga* species, making the RLP orthologs in rice excellent candidates for *S. hermonthica* resistance genes. Recently, the RLP gene *CuRe1* was identified in tomato which provides increased resistance to the parasitic plant *C. reflexa* (Hegenauer *et al.*, 2016), demonstrating that RLP genes can act against parasitic plants, and providing further support for the potential role of the RLP cluster in rice in resistance to *S. hermonthica*.

The resistance conferred by *Ve1* and *CuRe1* is not complete and low levels of proliferation of the fungus or parasite is still observed in plants expressing these genes respectively. Indeed, full immunity to *C. reflexa* in tomato is thought to involve additional layers to *CuRe1* (Hegenauer *et al.*, 2016), much like the idea of PTI and ETI acting as different layers to perceive microbial pathogens. The lack of complete resistance conferred by *Ve1* and *CuRe1* is analogous to resistance to *Striga*; even highly resistant plants can support one or two parasites. This suggests these RLP genes act like PRRs, detecting conserved fungal or parasite molecular patterns. Indeed, *Ve1* not only confers resistance against *Verticillium* wilts but also to the fungal pathogen *Fusarium oxysporum* f. sp. *lycopersici* (de Jonge *et al.*, 2012). PRR-mediated resistance is generally thought to be a broader spectrum, more durable but weaker form of resistance than R protein-mediated resistance (Tsuda & Katagiri, 2010), and is consistent with resistance against the genetically diverse *S. hermonthica*. The broad spectrum and durability of PRR-mediated resistance would be valuable source of resistance for plant breeding and crop protection against this parasite.

On-going work in our laboratory has investigated if the QTL on chromosome 12 provides broad-spectrum resistance against other isolates and species of *Striga*. A selection of lines from the BIL mapping population derived from crosses between Nipponbare and Koshihikari have been phenotyped for post-attachment resistance to a selection of both *S. hermonthica* and *S. asiatica* ecotypes. Lines possessing the Nipponbare QTL region exhibited good resistance against a range of different *S. hermonthica* ecotypes originating from both East and West Africa, and also against the USA isolate of *S. asiatica*, although they were susceptible against other *S. asiatica* ecotypes (Prof. Scholes, personal communication). Thus, this QTL region on the chromosome 12 and the resistance genes will be important in crop breeding by providing resistance to several ecotypes of *Striga* across Africa. A similar experiment using RILs from the IR64 x Azucena mapping population would confirm if this is the same for IR64.

6.4 The chromosome 12 QTL: a complex locus

It is clear from the work carried out for this PhD that the *S. hermonthica* resistance QTL on chromosome 12 in rice is a complex locus, and that the resistance governed by this region is unlikely to be a simple story of a single, dominant resistance gene. Although a cluster of RLP genes are excellent candidates for resistance genes, it is unclear how many, or which, of these is the most important. Extensive genome rearrangements and genetic mutations are also observed between different rice cultivars, both resistant and susceptible to the parasite. The

high numbers of TEs present in the QTL region may explain some of this variation both in terms of structural variants, gene duplications and SNPs. For example, a second cluster of 9 relatively similar genes, also annotated as encoding homologs *of Verticillium* wilt disease resistance proteins, is present on Nipponbare chromosome 1 between positions 1.76 – 3.27 Mbp (http://rice.plantbiology.msu.edu/). Five of these also appear to be present in the IR64 genome, although the chromosome number is not known for IR64 due to the current lack of genome assembly. Given that TEs are known to cause chromosomal breakage and large chromosomal rearrangements (Dooner & Weil, 2007; Slotkin & Martienssen, 2007; Bennetzen & Wang, 2014), it is possible that the activity of transposable elements resulted in a translocation of these genes between chromosome 1 and 12. However, no QTL was detected on chromosome 1 in this study, indicating no involvement of these genes in resistance to *S. hermonthica*.

In addition to genetic alterations, it is possible that TEs may be influencing the expression of genes by inducing epigenetic polymorphisms that affect the transcriptional responsiveness of neighbouring genes. For example TE-derived siRNAs are known to be involved in the regulation of endogenous plant genes through the process of RNA-directed DNA methylation (Matzke *et al.*, 2007; Chinnusamy & Zhu, 2009; Law & Jacobsen, 2010; Wei *et al.*, 2014) which may contribute to natural variation in plant disease resistance (Zhang *et al.*, 2016). Additionally, TE may influence flanking genes by acting as promoters or enhancers (Girard & Freeling, 1999). These regulatory and epigenetic polymorphisms, that go undetected in direct comparisons of gene sequences, could affect the transcriptional responsiveness and thus relative importance of a resistance gene between different rice cultivars, leading to variation in phenotype. In Chapter 2, resistance to *S. hermonthica* in the RIL population showed a continuum between resistance and susceptibility, rather than separating into two distinct phenotypes (Figure 2.6). The most likely explanation for this is the presence of many minor genes which contribute the phenotype, but which went undetected in the QTL analysis. However, an alternative hypothesis is that the presence of many TEs in the QTL could also affect the transcriptional responsiveness of neighbouring resistance genes. Mutuku *et al.*, (2015), showed that levels of JA are induced to a much higher extent in resistant Nipponbare compared to susceptible Koshihikari following *S. hermonthica* infection. JA is essential for S*. hermonthica* resistance in rice (Mutuku *et al.*, 2015). Therefore any variation in transcriptional responsiveness of these RILs could be investigated by determining whether enhanced resistance is associated with increased sensitivity to exogenously applied JA, by measuring levels of JA-inducible marker-gene expression.

6.5 Conclusions and future perspectives

There is good evidence to support the role of the RLP genes in conferring resistance to *S. hermonthica* in rice, but there are many questions that remain to be answered (Figure 6.1 D). Further functional evidence is required to relate resistance to an individual gene or gene combination. Work has already begun to knock out the RLP genes in IR64, both individually and in combinations, using CRIPSR/Cas9 technology. This has the advantage over an RNAi approach, as the gene knockout is targeted at the DNA level (Belhaj *et al.*, 2015; Shalem *et al.*, 2015), thus avoiding problems with poor silencing of mRNA. Further work will also be carried out to over-express these genes in the susceptible cultivar Azucena, or a RIL from the mapping population containing the Azucena haplotype for the QTL region. Isolation of a resistance gene or genes will pave the way for further examination of the mode of action, downstream signalling processes and dissection of the functional amino acids determining specificity. It may also help identify a parasite-derived effector using a yeast two-hybrid system or 3D modelling of receptor-ligand interactions. Work to identify a *Striga* effector is ongoing in our laboratory. Once identified, the role of resistance genes can then be tested in other cereals such as maize, where sources of resistance to *Striga* are scarce. Orthologs of the RLP genes identified in this PhD do not exist in the maize genome, therefore if shown to be functional in maize, these RLP genes could have enormous benefit for crop improvement in this species. Future work should make use of the enormous genetic resources now available for rice by using the 3000 sequenced rice genomes to search for new sources of resistance with different modes of action that will contribute to the long-term durability of resistance and control of this devastating parasite.

References

Aflakpui GKS, Gregory PJ, Froud-Williams RJ. **2002**. Growth and Biomass Partitioning of Maize During Vegetative Growth in Response To Striga hermonthica Infection and Nitrogen Supply. *Experimental Agriculture* **38**: 265–276.

Agrawal N, Dasaradhi PVN, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK. **2003**. RNA interference: biology, mechanism, and applications. *Microbiology and molecular biology reviews : MMBR* **67**: 657–85.

Akiyama K, Matsuzaki K, Hayashi H. **2005**. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**: 824–7.

Ali ML, Sanchez PL, Yu S, Lorieux M, Eizenga GC. **2010**. Chromosome Segment Substitution Lines: A Powerful Tool for the Introgression of Valuable Genes from Oryza Wild Species into Cultivated Rice (O. sativa). *Rice* **3**: 218–234.

Ammiraju JSS, Song X, Luo M, Sisneros N, Angelova A, Kudrna D, Kim H, Yu Y, Goicoechea JL, Lorieux M, *et al.* 2010. The Oryza BAC resource: A genus-wide and genome scale tool for exploring rice genome evolution and leveraging useful genetic diversity from wild relatives. *Breeding Science* **60**: 536–543.

Arnaud M-C, Veronesi C, Thalouarn P. **1999**. Physiology and histology of resistance to Striga hermonthica in Sorghum bicolor var. Framida. *Australian Journal of Physiology* **26**: 63–70.

Arvey A, Larsson E, Sander C, Leslie CS, Marks DS. **2010**. Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular systems biology* **6**: 363.

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**: 752–760.

Atera EA, Itoh K, Azuma T, Ishii T. **2012**. Farmers' perspectives on the biotic constraint of Striga hermonthica and its control in western Kenya. *Weed Biology and Management* **12**: 53–62.

Atera EA, Onyango JC, Thanh PT, Ishii T, Itoh K, Basin L, Authority D, Basin L, Authority D. **2015**. Identification of QTL for Striga hermonthica Resistance Using Backcross Population Derived from a Cross between Oryza sativa (cv . Nipponbare) and O . rufipogon. *Journal of Agricultural Science* **7**: 99–105.

Atokple IDK, Singh BB, Emechebe AM. **1995**. Genetics of Resistance to Striga and Alectra in Cowpea. **86**: 45–49.

Bai J, Pennill L a, Ning J, Lee SW, Ramalingam J, Webb C a, Zhao B, Sun Q, Nelson JC, Leach JE, *et al.* 2002. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome research* **12**: 1871–84.

Balasubramanian V, Sie M, Hijmans RJ, Otsuka K. **2007**. Increasing Rice Production in Sub-Saharan Africa: Challenges and Opportunities. *Advances in Agronomy* **94**: 55–133.

Baulcombe D. **2004**. RNA silencing in plants. *Nature* **43**: 356–363.

Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer B, Gianfranceschi L, Gessler C, Sansavini S. **2004**. The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 886–890.

Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V. **2015**. Editing plant genomes

with CRISPR/Cas9. *Current Opinion in Biotechnology* **32**: 76–84.

Bennett JR, Mathews S. **2006**. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *American journal of botany* **93**: 1039–51.

Bennetzen JL, Wang H. **2014**. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annual Review of Plant Biology* **65**: 505–530.

Berner DK, Cardwell KF, Faturoti BO, Ikie FO, Williams OA. **1994**. Relative roles of wind, crop seeds, and cattle in dispersal of Striga spp. *Plant Disease* **78**: 402–406.

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

Bernier J, Kumar A, Venuprasad R, Spaner D, Verulkar S, Mandal NP, Sinha PK, Peeraju P, Dongre PR, Mahto RN, *et al.* 2009. Characterization of the effect of a QTL for drought resistance in rice, qtl12.1, over a range of environments in the Philippines and eastern India. *Euphytica* **166**: 207–217.

Bleckmann A, Weidtkamp-Peters S, Seidel C a M, Simon R. **2010**. Stem cell signaling in Arabidopsis requires CRN to localize CLV2 to the plasma membrane. *Plant physiology* **152**: 166–76.

Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J. **2009**. The 'inner circle' of the cereal genomes. *Current Opinion in Plant Biology* **12**: 119–125.

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

Bouwmeester HJ, Matusova R, Zhongkui S, Beale MH. **2003**. Secondary metabolite signalling in host–parasitic plant interactions. *Current Opinion in Plant Biology* **6**: 358–364.

Bruner SD, Norman DP, Verdine GL. **2000**. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**: 859–66.

Cardoso C, Zhang Y, Jamil M, Hepworth J, Charnikhova T, Dimkpa SON, Meharg C, Wright M, Liu J, Meng X, *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* **111**: 3279–2384.

Chakrabarti A, Panter SN, Harrison K, Jones JDG, Jones D a. **2009**. Regions of the Cf-9B disease resistance protein able to cause spontaneous necrosis in Nicotiana benthamiana lie within the region controlling pathogen recognition in tomato. *Molecular plant-microbe interactions : MPMI* **22**: 1214–26.

Chenais B, Caruso A, Hiard S, Casse N. **2012**. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* **509**: 7–15.

Chinnusamy V, Zhu JK. **2009**. Rna-directed DNA methylation and demethylation in plants. *Science in China, Series C: Life Sciences* **52**: 331–343.

Cissoko M. **2012**. Understanding resistance in inter-specific rice cultivars to the parasitic witchweed Striga.

Cissoko M, Boisnard A, Rodenburg J, Press MC, Scholes JD. **2011**. New Rice for Africa (NERICA) cultivars exhibit different levels of post-attachment resistance against the parasitic weeds Striga hermonthica and Striga asiatica. *The New phytologist* **192**: 952–63.

Civán P, Craig H, Cox CJ, Brown TA. **2016**. Multiple domestications of Asian rice. *Nature Plants* **2**: 1–5.

Combier J-P, Küster H, Journet E-P, Hohnjec N, Gamas P, Niebel A. **2008**. Evidence for the Involvement in Nodulation of the Two Small Putative Regulatory Peptide-Encoding Genes *MtRALFL1* and *MtDVL1*. *Molecular Plant-Microbe Interactions* **21**: 1118–1127.

Conn CE, Bythell-douglas R, Neumann D, Yoshida S, Whittington B, Westwood JH, Shirasu K, Bond CS, Dyer KA, Nelson DC. **2015**. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **2**: 7–11.

Cook DE, Lee TG, Guo XL, Melito S, Wang K, Bayless AM, Wang JP, Hughes TJ, Willis DK, Clemente TE, *et al.* 2012. Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. *Science* **338**: 1206–1209.

Cook DE, Mesarich CH, Thomma BPHJ. **2015**. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annual Review of Phytopathology* **53**: 541–63.

Covey PA, Subbaiah CC, Parsons RL, Pearce G, Lay FT, Anderson MA, Ryan CA, Bedinger PA. **2010**. A Pollen-Specific RALF from Tomato That Regulates Pollen Tube Elongation. *Plant Physiology* **153**: 703–715.

Cutter AD, Payseur BA, Salcedo T, Estes AM, Good JM, Wood E, Hartl T, Maughan H, Strempel J, Wang B, *et al.* 2003. Molecular correlates of genes exhibiting RNAi phenotypes in Caenorhabditis elegans. *Genome Research* **13**: 2651–2657.

Dai X, Zhao PX. **2011**. PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Research* **39**: 155–159.

Dangl JL, Jones JDG. **2001**. Plant pathogens and integrated defence responses to infection. *Nature* **411**: 826–833.

Delatte T, Trevisan M, Parker ML, Zeeman SC. **2005**. Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. *Plant Journal* **41**: 815–830.

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JD. **1998**. The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* **10**: 1915–1925.

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JDG. **1996**. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**: 451–459.

Dodds PN, Rathjen JP. **2010**. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* **11**: 539–548.

Dooner HK, Weil CF. **2007**. Give-and-take: interactions between DNA transposons and their host plant genomes. *Current Opinion in Genetics and Development* **17**: 486–492.

Dorr I. **1997**. How Striga Parasitizes its Host : a TEM and SEM Study. *Annals of Botany* **79**: 463–472.

Duitama J, Silva A, Sanabria Y, Cruz DF, Quintero C, Ballen C, Lorieux M, Scheffler B. **2015**. Whole Genome Sequencing of Elite Rice Cultivars as a Comprehensive Information Resource for Marker Assisted Selection. *PloS one* **10**: 1–20.

Echevarría-Zomeño S, Pérez-de-Luque A, Jorrín J, Maldonado AM. **2006**. Pre-haustorial

resistance to broomrape (Orobanche cumana) in sunflower (Helianthus annuus): cytochemical studies. *Journal of experimental botany* **57**: 4189–200.

Edwards R, Dixon DP, Walbot V. **2000**. Plant glutathione S -transferases: enzymes with multiple functions in sickness and in health. *Trends in biotechnology* **5**: 193–198.

Ejeta G. **2007**. The Striga scourge in Africa: A growing pandemic. In: Ejeta G,, In: Gressel J, eds. Integrating New Technologies for Striga Control: Towards Ending the Witch-hunt. World Scientific, 3–15.

Ejeta G, Gressel J. **2007**. Preface. In: Ejeta G,, In: Gressel J, eds. Integrating New Technologies for Striga Control: Towards Ending the Witch-hunt. World Scientific, v–vi.

Emechebe a. M, Ellis-Jones J, Schulz S, Chikoye D, Douthwaite B, Kureh I, Tarawali G, Hussaini M a., Kormawa P, Sanni a. 2004. Farmers Perception of the Striga Problem and Its Control in Northern Nigeria. *Experimental Agriculture* **40**: 215–232.

Fernie AR, Klee HJ. **2011**. The use of natural genetic diversity in the understanding of metabolic organization and regulation. *Frontiers in plant science* **2**: 59.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. **1998**. Potent and specific genetic interference by double- stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806–811.

Flint-Garcia SA, Thornsberry JM, Buckler ES. **2003**. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* **54**: 357–374.

Flor HH. **1971**. Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **9**: 275–296.

Fradin EF, Abd-El-Haliem A, Masini L, van den Berg GCM, Joosten MH a J, Thomma BPHJ. **2011**. Interfamily transfer of tomato Ve1 mediates Verticillium resistance in Arabidopsis. *Plant physiology* **156**: 2255–2265.

Fradin EF, Thomma BPHJ. **2006**. Physiology and molecular aspects of Verticillium wilt diseases caused by V. dahliae and V. albo-atrum. *Molecular Plant Pathology* **7**: 71–86.

Fradin EF, Zhang Z, Juarez Ayala JC, Castroverde CDM, Nazar RN, Robb J, Liu C-M, Thomma BPHJ. **2009**. Genetic dissection of Verticillium wilt resistance mediated by tomato Ve1. *Plant physiology* **150**: 320–32.

Fragoso CA, Heffelfinger C, Zhao H, Dellaporta SL. **2016**. Imputing genotypes in biallelic populations from low-coverage sequence data. *Genetics* **202**: 487–495.

Fragoso CA, Moreno M, Wang Z, Heffelfinger C, Arbelaez LJ, Aguirre JA, Franco N, Romero LE, Labadie K, Zhao H, *et al.* 2017. Genetic Architecture of a Rice Nested Association Mapping Population. *G3: Genes|Genomes|Genetics*: g3.117.041608.

Fritz-Laylin LK, Krishnamurthy N, Tor M, Sjolander V, Jones JDG. **2005**. Phylogenomic Analysis of the Receptor-Like Proteins of Rice and Arabidopsis. *Plant Physiology* **138**: 611–623.

Frost DL, Gurney AL, Press MC, Scholes JD. **1997**. Striga hermonthica reduces photosynthesis in sorghum : the importance of stomatai limitations and a potentiai role for ABA ? *Plant, Cell and Environment* **20**: 483–492.

Fujimoto R, Sasaki T, Kudoh H, Taylor JM, Kakutani T, Dennis ES. **2011**. Epigenetic variation in the FWA gene within the genus Arabidopsis. *Plant Journal* **66**: 831–843.

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S. **2012**. The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* **488**: 535–9.

Garris AJ, McCouch SR, Kresovich S. **2003**. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (Oryza sativa L.). *Genetics* **165**: 759–69.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. **2005**. Genetic structure and diversity in Oryza sativa L. *Genetics* **169**: 1631–8.

Gbehounou G, Adango E. **2003**. Trap crops of Striga hermonthica : in vitro identification and effectiveness in situ. *Crop Protection* **22**: 395–404.

Ge SS, Lu BR, Hong DY. **2001**. Phylogeny of the genus Oryza as revealed by molecular approaches. In: Khush G,, In: Brar D,, In: Hardy B, eds. Rice Genetics IV. Proceedings of the fourth international rice genetics symposium. The Philippines, IRRI, 89–105.

Girard L, Freeling M. **1999**. Regulatory changes as a consequence of transposon insertion. *Developmental Genetics* **25**: 291–296.

Glazebrook J. **2005**. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual review of phytopathology* **43**: 205–27.

Gobena D, Shimels M, Rich PJ, Ruyter-Spira C, Bouwmeester H, Kanuganti S, Mengiste T, Ejeta G. **2017**. Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance. *Proceedings of the National Academy of Sciences of the United States of America*: 201618965.

Goldwasser Y, Plakhine D, Kleifeld Y, Zamski E, Rubin B. **2000**. The differential susceptibility of vetch (Vicia spp.) to Orobanche aegyptiaca: Anatomical studies. *Ann Bot (Lond)* **85**: 257–262.

Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pagès V, Dun EA, Pillot J-P, Letisse F, Matusova R, Danoun S, Portais J-C, *et al.* **2008**. Strigolactone inhibition of shoot branching. *Nature* **455**: 189–194.

Grenier C, Ibrahim Y, Haussmann BIG, Kiambi D, Ejeta G. **2007**. Marker-assisted selection for striga resistance in sorghum. In: Ejeta G,, In: Gressel J, eds. Integrating New Technologies for Striga Control: Towards Ending the Witch-hunt. World Scientific, 159–171.

Gross BL, Zhao Z. **2014**. Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 6190–7.

Gupta S, Chakraborti D, Basu D, Das S. **2010**. In search of Decoy/Guardee to R Genes. *Plant Signaling & Behavior* **5**: 1081–1087.

Gurney AL, Grimanelli D, Kanampiu F, Hoisington D, Scholes JD, Press MC. **2003**. Novel sources of resistance to Striga hermonthica in Tripsacum dactyloides, a wild relative of maize. *New Phytologist* **160**: 557–568.

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

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

Hammond-Kosack KE. **1994**. Incomplete Dominance of Tomato Cf Genes for Resistance to

Cladosporium fulvum. *Molecular Plant-Microbe Interactions* **7**: 58.

Han B, Huang X. **2013**. Sequencing-based genome-wide association study in rice. *Current opinion in plant biology* **16**: 133–8.

Hannon GJ. **2002**. RNA interference. *Nature* **418**: 244–251.

Harahap Z, Ampong-Nyarko K, Olela JC. **1993**. Striga hermonthica resistance in upland rice. *Crop Protection* **12**: 229–231.

Haruta M. **2003**. Rapid Alkalinization Factors in Poplar Cell Cultures. Peptide Isolation, cDNA Cloning, and Differential Expression in Leaves and Methyl Jasmonate-Treated Cells. *Plant Physiology* **131**: 814–823.

Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Reddy BVS, Kayentao M, Welz HG, Geiger HH. **2004**. Genomic regions influencing resistance to the parasitic weed Striga hermonthica in two recombinant inbred populations of sorghum. *Theoretical and Applied Genetics* **109**: 1005–16.

Haussmann BIG, Hess DE, Omanya GO, Reddy BVS, Welz HG, Geiger HH. **2001**. Major and minor genes for stimulation of Striga hermonthica seed germination in Sorghum, and interaction with different Striga populations. *Crop Science* **41**: 1507–1512.

Häweker H, Rips S, Koiwa H, Salomon S, Saijo Y, Chinchilla D, Robatzek S, Von Schaewen A. **2010**. Pattern recognition receptors require N-glycosylation to mediate plant immunity. *Journal of Biological Chemistry* **285**: 4629–4636.

Hayward A, Stirnberg P, Beveridge C, Leyser O. **2009**. Interactions between auxin and strigolactone in shoot branching control. *Plant physiology* **151**: 400–12.

Hearne SJ. **2009**. Control--the Striga conundrum. *Pest management science* **65**: 603–14.

Hegenauer AV, Fürst U, Kaiser B, Smoker M, 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**: 478–81.

Hewitt E. **1966**. *Sand and Water Culture Methods Used in the Study of Plant Nutrition*. London, UK: Common Agricultural Bureau.

Hibberd JM, Jeschke WD. **2001**. Solute flux into parasitic plants. *Journal of experimental botany* **52**: 2043–9.

Hibberd JM, Quick WP, Press MC, Scholes JD. **1996**. The influence of the parasitic angiosperm Striga gesnerioides on the growth and photosynthesis of its host , Vigna unguiculata. *Journal of Experimental Botany* **47**: 507–512.

Hiraoka Y, Sugimoto Y. **2008**. Molecular Responses of Sorghum to Purple Witchweed (Striga hermonthica) Parasitism. *Weed Science* **56**: 356–363.

Hirochika H. **2001**. Contribution of the Tos17 retrotransposon to rice functional genomics. *Current Opinion in Plant Biology* **4**: 118–122.

Honaas LA, Wafula EK, Yang Z, Der JP, Wickett NJ, Altman NS. **2013**. Functional genomics of a generalist parasitic plant : Laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression Functional genomics of a generalist parasitic plant : Laser microdissection of host-parasite. *BMC Plant Biology* **13**: 1471–2229.

Hong SW, Jiang Y, Kim S, Li CJ, Lee D. **2014**. Target gene abundance contributes to the

efficiency of siRNA-mediated gene silencing. *Nucleic acid therapeutics* **24**: 192–8.

Hooper AM, Tsanuo MK, Chamberlain K, Tittcomb K, Scholes J, Hassanali A, Khan ZR, Pickett JA. **2010**. Isoschaftoside, a C-glycosylflavonoid from Desmodium uncinatum root exudate, is an allelochemical against the development of Striga. *Phytochemistry* **71**: 904–908.

van der Hoorn R a L, Wulff BBH, Rivas S, Durrant MC, van der Ploeg A, de Wit PJGM, Jones JDG. **2005**. Structure-function analysis of cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. *The Plant cell* **17**: 1000–1015.

Huang K, Mellor KE, Paul SN, Lawson MJ, Mackey AJ, Timko MP. **2012a**. 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**: 1471–2164.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, *et al.* 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetics* **42**: 961–7.

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

Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJM, Matzke M. **2006**. Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO Journal* **25**: 2828–2836.

Jacobs AL, Schär P. **2012**. DNA glycosylases: In DNA repair and beyond. *Chromosoma* **121**: 1–20.

Jamil M, Charnikhova T, Cardoso C, Jamil T, Ueno K, Verstappen F, Asami T, Bouwmeester HJ. **2011a**. Quantification of the relationship between strigolactones and striga hermonthica infection in rice under varying levels of nitrogen and phosphorus. *Weed Research* **51**: 373–385.

Jamil M, Charnikhova T, Houshyani B, van Ast A, Bouwmeester HJ. **2012**. Genetic variation in strigolactone production and tillering in rice and its effect on Striga hermonthica infection. *Planta* **235**: 473–84.

Jamil M, Rodenburg J, Charnikhova T, Bouwmeester HJ. **2011b**. Pre-attachment Striga hermonthica resistance of New Rice for Africa (NERICA) cultivars based on low strigolactone production. *The New phytologist* **192**: 964–75.

Johnson DE, Riches CR, Diallot R, Jones MJ. **1997**. Striga on rice in West Africa ; crop host range and the potential of host resistance. *Crop Protection* **16**: 153–157.

Jones JDG, Dangl L. **2006**. The plant immune system. *Nature* **444**: 323–329.

Jones MP, Dingkuhn M, Aluko GK, Semon M. **1997**. Interspecific Oryza Sativa L. X O. Glaberrima Steud. progenies in upland rice improvement. *Euphytica* **94**: 237–246.

de Jonge R, Peter van Esse H, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao K V., *et al.* 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proceedings of the National Academy of Sciences* **109**: 5110–5115.

Kaewchumnong K, Price AH. **2008**. A study on the susceptibility of rice cultivars to Striga hermonthica and mapping of Striga tolerance quantitative trait loci in rice. *New Phytologist* **180**: 206–16.

Kanampiu FK, Kabambe V, Massawe C, Jasi L, Friesen D, Ransom JK, Gressel J. **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**: 697–706.

Keen NT. **1990**. Gene-for-gene complementarity in plant-pathogen interactions. *Annual Review of Genetics* **24**: 447–463.

Kerschen A, Napoli CA, Jorgensen RA, Müller AE. **2004**. Effectiveness of RNA interference in transgenic plants. *FEBS Letters* **566**: 223–228.

Keyes WJ, Palmer AG, Erbil WK, Taylor J V, Apkarian RP, Weeks ER, Lynn DG. **2007**. Semagenesis and the parasitic angiosperm Striga asiatica. *The Plant Journal* **51**: 707–16.

Keyes WJ, Taylor J V, Apkarian RP, Lynn DG. **2001**. Dancing Together . Social Controls in Parasitic Plant Development. *Plant Physiology* **127**: 1508–1512.

Khan ZR, Pickett JA, Hassanali A, Hooper AM, Midega CAO. **2008**. Desmodium species and associated biochemical traits for controlling Striga species: present and future prospects. *Weed Research* **48**: 302–306.

Kifuko-Koech M, Pypers P, Okalebo JR, Othieno CO, Pickett JA, Kipkoech AK, Vanlauwe B. **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**: 97–207.

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M. **2006**. Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. : 38–45.

Kruijt M, MJ DEK, de Wit PJ. **2005**. Receptor-like proteins involved in plant disease resistance. *Molecular plant pathology* **6**: 85–97.

Kunkel BN, Brooks DM. **2002**. Cross talk between signaling pathways in pathogen defense. *Current Opinion in Plant Biology* **5**: 325–331.

Kusumoto D, Goldwasser Y, Xie X, Yoneyama K, Takeuchi Y, Yoneyama K. **2007**. Resistance of red clover (Trifolium pratense) to the root parasitic plant Orobanche minor is activated by salicylate but not by jasmonate. *Ann Bot* **100**: 537–544.

Lander ES, Botstein D. **1989**. Mapping Mendelian factors underlying quantitative traits using RFLP limkage maps. *Genetics* **121**: 185–199.

Latzel V, Allan E, Bortolini Silveira A, Colot V, Fischer M, Bossdorf O. **2013**. Epigenetic diversity increases the productivity and stability of plant populations. *Nature communications* **4**: 2875.

Latzel V, Zhang Y, Karlsson Moritz K, Fischer M, Bossdorf O. **2012**. Epigenetic variation in plant responses to defence hormones. *Annals of Botany* **110**: 1423–1428.

Law JA, Jacobsen SE. **2010**. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204–220.

Leyser O. **2003**. Regulation of shoot branching by auxin. *Trends in plant science* **8**: 541–5.

Li J, Lis KE, Timko MP. **2009**. Molecular genetics of race-specific resistance of cowpea to Striga gesnerioides (Willd.). *Pest Management Science* **65**: 520–527.

Li J, Timko MP. **2009**. Gene-for-gene resistance in Striga-cowpea associations. *Science* **325**:

1094.

Li X, Yu X, Wang N, Feng Q, Dong Z, Liu L, Shen J, Liu B. **2007**. Genetic and epigenetic instabilities induced by tissue culture in wild barley (Hordeum brevisubulatum (Trin.) Link). *Plant Cell, Tissue and Organ Culture* **90**: 153–168.

Liebrand TWH, van den Berg GCM, Zhang Z, Smit P, Cordewener JHG, America AHP, America AHP, Sklenar J, Jones AME, Tameling WIL, *et al.* 2013. Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 10010–5.

Liebrand TWH, van den Burg HA, Joosten MHAJ. **2014**. Two for all: Receptor-associated kinases SOBIR1 and BAK1. *Trends in Plant Science* **19**: 123–132.

Lippman ZL, GENDREL A-V, BLACK M, VAUGHN MW, DEDHIA N, MCCOMBIE WR, LAVINE K, MITTAL V, MAY B, KASSCHAU KD, *et al.* 2004. Role of transposable elements in heterochromatic and epigenetic control. *Nature* **430**: 471–476.

Liu Y, Wu H, Chen H, Liu Y, He J, Kang H, Sun Z, Pan G, Wang Q, Hu J, *et al.* 2015. A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice. *Nature Biotechnology* **33**: 301–305.

Lorieux M. **2012**. MapDisto: fast and efficient computation of genetic linkage maps. *Molecular Breeding* **30**: 1231–1235.

Lorieux M, Blein M, Lozano J, Bouniol M, Droc G, Diévart A, Périn C, Mieulet D, Lanau N, Bès M, *et al.* 2012. In-depth molecular and phenotypic characterization in a rice insertion line library facilitates gene identification through reverse and forward genetics approaches. *Plant Biotechnology Journal* **10**: 555–568.

Lozano-Torres JL, Wilbers RHP, Gawronski P, Boshoven JC, Finkers-Tomczak A, Cordewener JHG, America AHP, Overmars HA, Van 't Klooster JW, Baranowski L, *et al.* 2012. Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* **109**: 10119–24.

Luderer R, Rivas S, Nürnberger T, Mattei B, Van den Hooven HW, Van der Hoorn R a, Romeis T, Wehrfritz JM, Blume B, Nennstiel D, *et al.* 2001. No evidence for binding between resistance gene product Cf-9 of tomato and avirulence gene product AVR9 of Cladosporium fulvum. *Molecular plant-microbe interactions : MPMI* **14**: 867–876.

Majumdar R, Rajasekaran K, Cary JW. **2017**. RNA Interference (RNAi) as a Potential Tool for Control of Mycotoxin Contamination in Crop Plants: Concepts and Considerations. *Frontiers in Plant Science* **8**: 1664–462.

Matsumoto T, Wu J, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio B, Baba T, *et al.* 2005. The map-based sequence of the rice genome. *Nature* **436**: 793–800.

Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJM. **2007**. Targets of RNA-directed DNA methylation. *Current Opinion in Plant Biology* **10**: 512–519.

Mauricio R. **2001**. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature reviews. Genetics* **2**: 370–81.

McDowell JM, Simon SA. **2006**. Recent insights into R gene evolution. *Molecular Plant Pathology* **7**: 437–448.

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

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. **2003**. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *The Plant cell* **15**: 809–34.

Michelmore RW, Meyers BC, Young ND. **1998**. Cluster of resistance genes in plants evolveby divergent selection and a birth-and-death process. *Current Opinion in Plant Biology* **3**: 285–90.

Midega CAO, Pittchar J, Salifu D, Pickett JA, Khan ZR. **2013**. Effects of mulching, N-fertilization and intercropping with Desmodium uncinatum on Striga hermonthica infestation in maize. *Crop Protection* **44**: 44–49.

Miki D, Itoh R, Shimamoto K. **2005**. RNA silencing of single and multiple members in a gene family of rice. *Plant physiology* **138**: 1903–13.

Miki D, Shimamoto K. **2004**. Simple RNAi vectors for stable and transient suppression of gene function in rice. *Plant & cell physiology* **45**: 490–5.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H. **2003**. Target Site Specificity of the Tos17 Retrotransposon Shows a Preference for Insertion within Genes and against Insertion in Retrotransposon-Rich Regions of the Genome. *The Plant Cell* **15**: 1771–1780.

Mohamed KI, Bolin JF, Musselman LJ, Peterson AT. **2007**. Genetic diversity of Striga and implications for control and modeling future distributions. In Integrating New Tchnologies for Striga Control: Towards Ending the Witch-Hunt. *World scientific*: 71–84.

Mohamed A, Ellicott A, Housley TL, Ejeta G. **2003**. Hypersensive response to Striga infection in Sorghum. *Crop Science* **43**: 1320–1324.

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

Molinero-Ruiz ML, Melero-Vara JM, Garc??a-Ruiz R, Dom??nguez J. **2006**. Pathogenic diversity within field populations of Orobanche cumana and different reactions on sunflower genotypes. *Weed Research* **46**: 462–469.

Morrell PL, Buckler ES, Ross-Ibarra J. **2012**. Crop genomics: advances and applications. *Nature reviews. Genetics* **13**: 85–96.

Mur LAJ, Kenton P, Atzorn R, Miersch O, Wasternack C. **2006**. The Outcomes of Concentration-Specific Interactions between Salicylate and Jasmonate Signaling Include Synergy , Antagonism , and Oxidative Stress Leading to Cell Death. *Plant Physiology* **140**: 249–262.

Murphy E, De Smet I. **2014**. Understanding the RALF family: A tale of many species. *Trends in Plant Science* **19**: 664–671.

Mutuku JM, Yoshida S, Shimizu T, Ichihashi Y, Wakatake T, Takahashi A, Seo M, Shirasu K. **2015**. The WRKY45-Dependent Signaling Pathway Is Required For Resistance against Striga hermonthica Parasitism. *Plant physiology* **168**: 1152–63.

Nagaraju J, Kathirvel M, Kumar RR, Siddiq EA, Hasnain SE, Published F, Sattar AKMA, Jeong E, Satchi M, Ah J, *et al.* **2002**. Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *PNAS* **9**: 12–15.

Neelakandan AK, Wang K. **2012**. Recent progress in the understanding of tissue culture-

induced genome level changes in plants and potential applications. *Plant Cell Reports* **31**: 597–620.

Nwanze KF, 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**: 675–677.

Oldroyd GED, Harrison MJ, Paszkowski U. **2009**. for Endosymbiosis. : 23–25.

Ossowski S, Schneeberger K, Lucas-lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M, Ossowski S, Schneeberger K, *et al.* 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science (New York, N.Y.)* **327**: 92–94.

Oswald A, Ransom JK, Kroschel J, Sauerborn J. **2002**. Intercropping controls striga in maize based farming systems. *Crop Protection* **21**: 367–374.

Ouedraogo JT, Maheshwari V, Berner K, St-Pieere A, Belzile F, Timko P. **2001**. Identification of AFLP markers linked to resistance of cowpea (Vigna unguiculata L .) to parasitism by Striga gesnerioides. *Theoretical and Applied Genetics* **102**: 1029–1036.

Page DR, Grossniklaus U. **2002**. the Art and Design of Genetic Screens : Arabidopsis Thaliana. **3**: 124–136.

Parker C. **2009**. Observations on the current status of Orobanche and Striga problems worldwide. *Pest management science* **65**: 453–9.

Parker C. **2013**. The Parasitic Weeds of the Orobanchaceae. Parasitic Orobanchaceae. Parasitic Mechanisms and Control Strategies.313–344.

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

Pearce G, Daniel S M, Stratmann J, Ryan A Jr. C. **2001**. RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. *PNAS* **98**: 12843–12847.

Pérez-de-Luque A, González-Verdejo CI, Lozano MD, Dita M a, Cubero JI, González-Melendi P, Risueño MC, Rubiales D. **2006a**. Protein cross-linking, peroxidase and beta-1,3-endoglucanase involved in resistance of pea against Orobanche crenata. *Journal of experimental botany* **57**: 1461–9.

Pérez-de-Luque A, Lozano MD, Cubero JI, González-Melendi P, Risueño MC, Rubiales D. **2006b**. Mucilage production during the incompatible interaction between Orobanche crenata and Vicia sativa. *Journal of experimental botany* **57**: 931–42.

Pfaffl MW. **2004**. Relative quantification. In: Dorak T, ed. Real-time PCR. International University Line, 63–82.

Phillips RL, Kaeppler SM, Olhoft P. **1994**. Genetic instability of plant tissue cultures: breakdown of normal controls. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 5222–6.

Piedras P, Rivas S, Dröge S, Hillmer S, Jones JDG. **2000**. Functional, c-myc-tagged Cf-9 resistance gene products are plasma-membrane localized and glycosylated. *Plant Journal* **21**: 529–536.

Pieterse CMJ, Leon-Reyes A, Van der Ent S, Van Wees SCM. **2009**. Networking by small-molecule hormones in plant immunity. *Nature chemical biology* **5**: 308–16.

Piffanelli P, Droc G, Mieulet D, Lanau N, Bès M, Bourgeois E, Rouvière C, Gavory F, Cruaud C,

Ghesquière A, *et al.* 2007. Large-scale characterization of Tos17 insertion sites in a rice T-DNA mutant library. *Plant Molecular Biology* **65**: 587–601.

Press MC, Stewart GR. **1987**. Growth and Photosynthesis in Sorghum bicolor Infected with Striga hermonthica. *Annals of Botany* **60**: 657–662.

Price AH. **2006**. Believe it or not, QTLs are accurate! *Trends in plant science* **11**: 213–6.

Pruzinská A, Anders I, Aubry S, Schenk N, Tapernoux-Lüthi E, Müller T, Kräutler B, Hörtensteiner S. **2007**. In vivo participation of red chlorophyll catabolite reductase in chlorophyll breakdown. *The Plant cell* **19**: 369–387.

Rank C, Rasmussen LS, Jensen SR, Pierce S, Press MC, Scholes JD. **2004**. Cytotoxic constituents of Alectra and Striga species. *Weed Research* **44**: 265–270.

Rich PJ, Ejeta G. **2007**. Biology of host-parasite interactions in Striga species. In: Ejeta G,, In: Gressel J, eds. Intergrating New Technologies for Striga control. Towards ending the witch-hunt. World Scientific, 19–32.

Rich PJ, Ejeta G. **2008**. Towards effective resistance to Striga in African maize. *Plant signalling and behavior* **3**: 618–621.

Rispail N, Dita M, González-Verdejo C, Pérez-de-Luque a, Castillejo M, Prats E, Román B, Jorrín J, Rubiales D. **2007**. Plant resistance to parasitic plants: molecular approaches to an old foe. *The New phytologist* **173**: 703–12.

Rodenburg J, Bastiaans L. **2011**. Host-plant defence against Striga spp.: reconsidering the role of tolerance. *Weed Research* **51**: 438–441.

Rodenburg J, Bastiaans L, Kropff MJ, Ast A van. **2006**. Effects of host plant genotype and seedbank density on Striga reproduction. *Weed Research* **46**: 251–263.

Rodenburg J, Cissoko M, Kayongo N, Dieng I, Bisikwa J, Irakiza R, Masoka I, Midega CAO, Scholes JD. **2017**. Genetic variation and host–parasite specificity of Striga resistance and tolerance in rice: the need for predictive breeding. *New Phytologist*.

Rodenburg J, Johnson DE. **2009**. *Weed Management in Rice-Based Cropping Systems in Africa*. Elsevier Inc.Academic Press.

Rodenburg J, Riches CR, Kayeke JM. **2010**. Addressing current and future problems of parasitic weeds in rice. *Crop Protection* **29**: 210–221.

Rubiales D, Pérez-de-luque A, Joel DM, Alcántara C, Sillero JC, Perez-de-luque A. **2003**. Characterization of resistance in chickpea to crenate broomrape (Orobanche crenata). *Weed Science* **51**: 702–707.

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

Sahoo KK, Tripathi AK, Pareek A, Sopory SK, Singla-Pareek SL. **2011**. An improved protocol for efficient transformation and regeneration of diverse indica rice cultivars. *Plant Methods* **7**: 49.

Sakai H, Kanamori H, Arai-Kichise Y, Shibata-Hatta M, Ebana K, Oono Y, Kurita K, Fujisawa H, Katagiri S, Mukai Y, *et al.* 2014. Construction of Pseudomolecule Sequences of the aus Rice Cultivar Kasalath for Comparative Genomics of Asian Cultivated Rice. *DNA research : an international journal for rapid publication of reports on genes and genomes*: 1–9.

Sallaud C, Gay C, Larmande P, Bès M, Piffanelli P, Piégu B, Droc G, Regad F, Bourgeois E,

Meynard D, *et al.* 2004. High throughput T-DNA insertion mutagenesis in rice: A first step towards in silico reverse genetics. *Plant Journal* **39**: 450–464.

Sallaud C, Meynard D, van Boxtel J, Gay C, Bes M, Brizard P, Larmande P, Ortega D, Raynal M, Portefaix M, *et al.* 2003. Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L .) functional genomics. *Theoretical and Applied Genetics* **106**: 1396–1408.

Samejima H, Babiker AG, Mustafa A, Sugimoto Y, Parker C. **2016**. Identification of Striga hermonthica -Resistant Upland Rice Varieties in Sudan and Their Resistance Phenotypes. *Frontiers in Plant Science* **7**: 1–12.

Sang T, Ge S. **2013**. Understanding rice domestication and implications for cultivar improvement. *Current Opinion in Plant Biology* **16**: 139–146.

Sarla N, Swamy BPM. **2005**. Oryza glaberrima : A source for the improvement of Oryza sativa. *Current Science* **89**: 955–963.

Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E, *et al.* 2014. Whole genome de novo assemblies of three divergent strains of rice , Oryza sativa , document novel gene space of aus and indica. *Genome Biology* **15**: 1–16.

Scholes JD, Press MC. **2008**. Striga infestation of cereal crops - an unsolved problem in resource limited agriculture. *Current opinion in plant biology* **11**: 180–6.

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

Seck PA, Diagne A, Mohanty S, Woperels MCS. **2012**. Crops that feed the world 7 : Rice. *Food Security* **4**: 7–24.

Selote D, Kachroo A. **2010**. RPG1-B-derived resistance to AvrB-expressing pseudomonas syringae requires RIN4-like proteins in soybean. *Plant Physiol* **153**: 1199–1211.

Semagn K, Bjørnstad Å, Ndjiondjop MN. **2006**. Principles , requirements and prospects of genetic mapping in plants. *African Journal of Biotechnology* **5**: 2569–2587.

Semagn K, Ndjiondjop MN, Lorieux M, Cissoko M, Jones M, Mccouch S. **2007**. Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (Oryza sativa) and CG 14 (Oryza glaberrima). **6**: 2014–2022.

Shalem O, Sanjana NE, Zhang F. **2015**. High-throughput functional genomics using CRISPR-Cas9. *Nature reviews. Genetics* **16**: 299–311.

Slotkin RK, Martienssen R. **2007**. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**: 272–285.

Solovyev V, Kosarev P, Seledsov I, Vorobyev D. **2006**. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome biology* **7 Suppl 1**: S10.1-12.

Song X, Cao X. **2017**. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Current Opinion in Plant Biology* **36**: 111–118.

Spallek T, Mutuku M, Shirasu K. **2013**. The genus Striga: a witch profile. *Molecular plant pathology* **14**: 861–9.

Sun CQ, Wang XK, Li ZC, Yoshimura A, Iwata N. **2001**. Comparison of the genetic diversity of common wild rice (Oryza rufipogon Griff.) and cultivated rice (O. sativa L.) using RFLP markers. *Theoretical and Applied Genetics* **102**: 157–162.

Swarbrick PJ, Huang K, Liu G, Slate J, Press MC, Scholes JD. **2008**. Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant Striga hermonthica. *New phytologist* **179**: 515–29.

Swarbrick PJ, Scholes JD, Press MC, Slate J. **2009**. A major QTL for resistance of rice to the parasitic plant Striga hermonthica is not dependent on genetic background. *Pest management science* **65**: 528–32.

Sweeney M, McCouch S. **2007**. The complex history of the domestication of rice. *Annals of botany* **100**: 951–7.

Tang F, Yang S, Gao M, Zhu H. **2013**. Alternative splicing is required for RCT1-mediated disease resistance in Medicago truncatula. *Plant Molecular Biology* **82**: 367–374.

The 3000 Rice Genomes Project. **2014**. The 3,000 rice genomes project. *GigaScience* **3**: 7–12.

Thomas CM. **1997**. Characterization of the Tomato Cf-4 Gene for Resistance to Cladosporium fulvum Identifies Sequences That Determine Recognitional Specificity in Cf-4 and Cf-9. *the Plant Cell Online* **9**: 2209–2224.

Thomma BPHJ, Cammue BPA, Thevissen K. **2002**. Plant defensins. *Planta* **216**: 193–202.

Thomma BPHJ, Nürnberger T, Joosten MHAJ. **2011**. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *The Plant cell* **23**: 4–15.

Thynne E, Saur IML, Simbaqueba J, Ogilvie HA, Gonzalez-Cendales Y, Mead O, Taranto A, Catanzariti AM, Mcdonald MC, Schwessinger B, *et al.* 2016. Fungal phytopathogens encode functional homologues of plant rapid alkalinization factor (RALF) peptides. *Molecular Plant Pathology*: 1–14.

Timko M, Scholes JD. **2013**. Host reaction to Attack by Root Parasitic Plants. In: Joel D,, In: Gressel J, eds. Parasitic Orobanchaceae. Parasitic Mechanisms and Control Strategies. Springer (Heidelberg Berlin), 115–141.

Toh S, Holbrook-Smith D, Stogios PJ, Onopriyenko O, Lumba S, Tsuchiya Y, Savchenko A, McCourt P. **2015**. Structure-function analysis identifies highly sensitive strigolactone receptors in Striga. *Science (New York, N.Y.)* **350**: 203–7.

Toki S, Hara N, Ono K, Onodera H, Tagiri A, Oka S, Tanaka H. **2004**. Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. *The Plant Journal* **47**: 969–976.

Tomilov A a, Tomilova NB, Wroblewski T, Michelmore R, Yoder JI. **2008**. Trans-specific gene silencing between host and parasitic plants. *The Plant journal : for cell and molecular biology* **56**: 389–97.

Tsuchiya Y, Yoshimura M, Sato Y, Kuwata K, Toh S, Holbrook-Smith D, Zhang H, McCourt P, Itami K, Kinoshita T, *et al.* 2015. Probing strigolactone receptors in Striga hermonthica with fluorescence. *Science* **349**: 864–868.

Tsuda K, Katagiri F. **2010**. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology* **13**: 459–465.

Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya

Y, Shirasu K, Yoneyama K, *et al.* 2008. Inhibition of shoot branching by new terpenoid plant hormones. *Nature* **455**: 195–200.

Vallance KB. **1950**. Studies on the Germination of the Seeds of Striga Hermonthica. *Annals of Botany* **XIV**: 347–363.

Vaughan DA, Morishima H, Kadowaki K. **2003**. Diversity in the Oryza genus. *Current Opinion in Plant Biology* **6**: 139–146.

Vogler RK, Ejeta G, Butler LG. **1996**. Vogler et al 1996. Inheritance of low production of striga germination stimulant.pdf. *Crop Protection* **36**: 1185–1191.

Wang YH. **2008**. How effective is T-DNA insertional mutagenesis in Arabidopsis? *J Biochem Tech* **1**: 11–20.

Wang G, Fiers M, Ellendorff U, Wang Z, de Wit PJGM, Angenent GC, Thomma BPHJ. **2010**. The Diverse Roles of Extracellular Leucine-rich Repeat-containing Receptor-like Proteins in Plants. *Critical Reviews in Plant Sciences* **29**: 285–299.

Wang B, Sang Y, Song J, Gao XQ, Zhang X. **2009**. Expression of a rice OsARGOS gene in Arabidopsis promotes cell division and expansion and increases organ size. *Journal of Genetics and Genomics* **36**: 31–40.

Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, Wang L, Hu F, Zhai J, Meyers BC, *et al.* 2014. Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proceedings of the National Academy of Sciences* **111**: 3877–3882.

Weiberg A, Jin H. **2015**. Small RNAs-the secret agents in the plant-pathogen interactions. *Current Opinion in Plant Biology* **26**: 87–94.

Westwood JH, Yoder JI, Timko MP, dePamphilis CW. **2010**. The evolution of parasitism in plants. *Trends in plant science* **15**: 227–35.

Wilmoth JC, Wang S, Tiwari SB, Joshi AD, Hagen G, Guilfoyle TJ, Alonso JM, Ecker JR, Reed JW. **2005**. NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. *Plant Journal* **43**: 118–130.

Wilson RC, Doudna JA. **2013**. Molecular Mechanisms of RNA Interference. *Annual Review of Biophysics* **42**: 217–239.

Wulff BBH, Heese A, Tomlinson-Buhot L, Jones D a, de la Peña M, Jones JDG. **2009**. The major specificity-determining amino acids of the tomato Cf-9 disease resistance protein are at hypervariable solvent-exposed positions in the central leucine-rich repeats. *Molecular plant-microbe interactions : MPMI* **22**: 1203–1213.

Xu R-F, Li H, Qin R-Y, Li J, Qiu C-H, Yang Y-C, Ma H, Li L, Wei P-C, Yang J-B. **2015**. Generation of inheritable and 'transgene clean' targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Scientific reports* **5**: 11491.

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, *et al.* 2011. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology* **30**: 105–111.

Yang X, Deng F, Ramonell KM. **2012a**. Receptor-like kinases and receptor-like proteins: Keys to pathogen recognition and defense signaling in plant innate immunity. *Frontiers in Biology* **7**: 155–166.

Yang C, Kawahara Y, Mizuno H, Wu J, Matsumoto T, Itoh T. **2012b**. Independent

domestication of Asian rice followed by gene flow from japonica to indica. *Molecular biology and evolution* **29**: 1471–9.

Yano M, Katoyuki Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehaha Y, Nagamura Y, *et al.* 2000. Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. *The Plant Cell* **12**: 2473–2483.

Yoder JI. **2001**. Host-plant recognition by parasitic Scrophulariaceae. *Current opinion in plant biology* **4**: 359–65.

Yoder JI, Scholes JD. **2010**. Host plant resistance to parasitic weeds; recent progress and bottlenecks. *Current opinion in plant biology* **13**: 478–84.

Yoneyama K, Awad A a, Xie X, Yoneyama K, Takeuchi Y. **2010**. Strigolactones as germination stimulants for root parasitic plants. *Plant & cell physiology* **51**: 1095–103.

Yoneyama K, Xie X, Kusumoto D, Sekimoto H, Sugimoto Y, Takeuchi Y, Yoneyama K. **2007a**. Nitrogen deficiency as well as phosphorus deficiency in sorghum promotes the production and exudation of 5-deoxystrigol, the host recognition signal for arbuscular mycorrhizal fungi and root parasites. *Planta* **227**: 125–32.

Yoneyama K, Yoneyama K, Takeuchi Y, Sekimoto H. **2007b**. Phosphorus deficiency in red clover promotes exudation of orobanchol, the signal for mycorrhizal symbionts and germination stimulant for root parasites. *Planta* **225**: 1031–8.

Yoshida S, Cui S, Ichihashi Y, Shirasu K. **2016**. The Haustorium, a Specialized Invasive Organ in Parasitic Plants. *Annual review of plant biology* **67**: 643–667.

Yoshida S, Shirasu K. **2009**. Multiple layers of incompatibility to the parasitic witchweed, Striga hermonthica. *The New phytologist* **183**: 180–9.

Yu J, Holland JB, McMullen MD, Buckler ES. **2008**. Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**: 539–51.

Yu H, Kumar PP. **2003**. Post-transcriptional gene silencing in plants by RNA. *Plant cell reports* **22**: 167–74.

Zamioudis C, Pieterse CMJ. **2011**. Modulation of Host Immunity by Beneficial Microbes. *Molecular Plant-Microbe Interactions* **25**: 139–150.

de Zélicourt A, Letousey P, Thoiron S, Campion C, Simoneau P, Elmorjani K, Marion D, Simier P, Delavault P. **2007**. Ha-DEF1, a sunflower defensin, induces cell death in Orobanche parasitic plants. *Planta* **226**: 591–600.

Zhang Q. **2007**. Strategies for developing Green Super Rice. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 16402–16409.

Zhang YY, Fischer M, Colot V, Bossdorf O. **2013**. Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytologist* **197**: 314–322.

Zhang X-C, Grassmann W. **2003**. RPS4-Mediated Disease Resistance Requires the Combined Presence of RPS4 Transcripts with Full-Length and Truncated Open Reading Frames. *the Plant Cell Online* **15**: 2333–2342.

Zhang Z, Song Y, Liu CM, Thomma BPHJ. **2014**. Mutational analysis of the Ve1 immune receptor that mediates Verticillium resistance in tomato. *PLoS ONE* **9**: 1–10.

Zhang H, Tao Z, Hong H, Chen Z, Wu C, Li X, Xiao J, Wang S. **2016**. Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nature Plants* **2**.

Zhang Z, Thomma BPHJ. **2013**. Structure-function aspects of extracellular leucine-rich repeat-containing cell surface receptors in plants. *Journal of Integrative Plant Biology* **55**: 1212–1223.

Zhu Q, Ge S. **2005**. Phylogenetic relationships among A-genome species of the genus Oryza revealed by intron sequences of four nuclear genes. *New Phytologist* **167**: 249–65.

Zhu C, Gore M, Buckler ES, Yu J. **2008**. Status and Prospects of Association Mapping in Plants. *The Plant Genome Journal* **1**: 5.

Appendix

```
                   Signal peptide
IR64_h_Os12g10930  M-----RVAR HLPLLLTVLQ IVLQAQATI  LTDRTSSS-V PPPIPCLPDQ ASALLQLKRS  54
LOC_Os12g10930     MSSFSKRVAH HLTSLLTVMH ILLQVQATPT LADRTTTSIV TTPVLCLPEQ ASALLQLKGS  60

IR64_h_Os12g10930  FNPKAGDYTT AFRSWITGID CCHWDGIACG GADGRVTSLD LGGHHLQASI VDPALFRLTS 114
LOC_Os12g10930     FNVTAGDYST VFRSWVAGAD CCHWEGVHCD GADGRVTSLD LGGHHLQADS VHPALFRLTS 120

IR64_h_Os12g10930  LRYLDLSGNN FSMSIL-ING LEQLTELTHL DLSDTNIAGE VPSAIGRLTS LVYLDLSTSF 173
LOC_Os12g10930     LKHLDLSGNN FSMSKLPFTG FQELTELMHL DLSNTNIAGE VPAGIGSIMN LVYLDLSTKF 180

IR64_h_Os12g10930  YIVEFDNENG KMTYNSDLFR QLSAPNLETL LGNLTNLEEL HMGMVNMSGN GDQWCDHIAK 233
LOC_Os12g10930     YALVYDDENN IMKFTLDSFW QLKAPNMETF LTNLTNLEQL HMGMMDMSRE GERWCDHIAK 240

IR64_h_Os12g10930  STPKLQVLSL PWCLLSGPIC TSLSAMQSLN TIELHYNHLS GSVPEFLATF SNLTVLQLSR 293
LOC_Os12g10930     STPKLQVLSL PWCSLSGPIC ASLSAMQSLN TIELHFNHLS GSIPEFFASF SNLSVLQLSK 300

IR64_h_Os12g10930  NKFEGWFPPT IFQHKKLITI NIINNPGLSG HLPNFSQASS LENVFISLTN FTATLKYLDL 353
LOC_Os12g10930     NDFQGWFPPI IFQHKKLRMI DLSKNPGISG NLPNFSQESS LENLFVSSTN FTGSLKYLDL 360

IR64_h_Os12g10930  LEVSGLQLVG SIPSWISNLT SLTVLQFSNC GLSGQVPSSI GNLRELRKLA LYNCKFSGKM 413
LOC_Os12g10930     LEVSGLQLVG SIPSWISNLT SLTALQFSNC GLSGQVPSSI GNLRKLTKLA LYNCNFSGK- 419

IR64_h_Os12g10930  PPQILNLTRL QTLLLHSNNF TGTVEITSFS KLENLSVLNL SNNELLVVDG ENSTKVLSFP 473
LOC_Os12g10930     ---------- ---------- ---------- ---------- ---------- ---------- 419

IR64_h_Os12g10930  KIKFLRLASC SISTFPNILK SLNEITSLDL SCNQIQGAIP QWAWGTWKGL QFYLLNISHN 533
LOC_Os12g10930     ---------- ---------- ---------- ---------- ---------- ---------- 419

IR64_h_Os12g10930  NFTSLGPDSL LPLHIDYFDV SFNSIEGPIP IPRDGSSTLD YSSNQFSAMP LHYSTYLGET 593
LOC_Os12g10930     ---------- ---------- ---------- ---------- ---------- ---------- 419

IR64_h_Os12g10930  LVFKASKNKL SGNIPSSICT AVRRLQLIDL SYNNFSGPIP SCLMEDLTAL QVLSLKENKL 653
LOC_Os12g10930     ---------- ---------- ---------- ---------- ---------- ------ENKL 423

IR64_h_Os12g10930  VGKLPDSIKE GCVLEALDLS GNLIEGKIPR SLVACKNLEF LDIGSNQISD TFPCWMSELP 713
LOC_Os12g10930     IGTLPDNIKE GCALEAIDIS GNLFEGKIPR SLIACRNLEI LDIGGNHFSD SFPCWMSQLP 483

IR64_h_Os12g10930  KLQVLVLKSN KFTGQVMDPS YMAGGDTCEF TELRIADMAS NNFNGTLPEA WFKMLKSMMS 773
LOC_Os12g10930     KLQVLVLKSN KFTGQLMDPS YMVGGNTCEF TELRIADMAS NDFNGTLPEA WFKMLKSMMT 543

IR64_h_Os12g10930  RSDNETLVME NQYYHGQTYQ FTATVTYKGS SMTISKILRT LVLIDLSNNA FHGTIPGTIG 833
LOC_Os12g10930     RSDNETLVME NQYYHGQTYQ FTATVTYKGN YMTISKILRT LVLIDFSNNA FHGAIPETIG 603

IR64_h_Os12g10930  ELILLHGLNM SHNALTGPVP PQLGRLNQLE TLDLSSNKLS GEIPDELASL NFLSTLNLSY 893
LOC_Os12g10930     ELILLHGLNM SHNALTGSIP TQFGRLNQLE SLDLSSNEFS GEIPEELASL NFLSTLNLSY 663

IR64_h_Os12g10930  NVLVGRIPDS YQFSTFSNSS FLGNTGLCGP PLSRQCDNPK GPTEMPYTSE KSIDVVLLLF 953
LOC_Os12g10930     NMLVGRIPNS YQFSTFSNNS FLGNTGLCGP PLSRQCNNPK EPIAMPYTLE KSIDVVLLLF 723

                   Transmembrane region   Cytoplasmic tail
IR64_h_Os12g10930  TALGFGLSFA MTILIVWGSH MRKEH* 979
LOC_Os12g10930     TASGFFISFA MMILIVWGSQ MRKQH* 749
```

Figure S.1 Amino acid alignment of Os12g10930 from Nipponbare and the IR64 homolog predicted by Fgenesh gene-finder (Softberry, Inc). Sequence differences are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding.

```
                   Signal peptide
IR64_h_Os12g11510  MRSAYHLMPP LAMLLILGLA DHASSTEAPA ACLPDQASAL LQLKRSFNAT IGDYPAAFRS  60
LOC_Os12g11510     MRSAYHLMPP LAMLLIHGLA DHASSTEAPA ACLPDQASAL LQLKRSFNAT IGDYPAAFRS  60

IR64_h_Os12g11510  WVAGADCCHW DGVRCGGAGG RVTSLDLSHR DLQASSGLDD ALFSLTSLEY LDLSSNDFSK 120
LOC_Os12g11510     WVAGADCCHW DGVRCGGAGG RVTSLDLSHR DLQASSGLDD ALFSLTSLEY LDLSSNDFSK 120

IR64_h_Os12g11510  SKLPATGFEK LTGLTHLDLS NTNFAGLVPA GIGRLTSLNY LDLSTTFFVE GLDDKYSITY 180
LOC_Os12g11510     SKLPATGFEM LTGLTHLDLS NTNFAGLVPA GIGRLTSLNY LDLSTTFFVE ELDDEYSITY 180

IR64_h_Os12g11510  YYSDTMAQLS EPSLETLLAN LTNLEELRLG MVMVNMSSNY GTARWCDAMA RSSPKLRVIS 240
LOC_Os12g11510     YYSDTMAQLS EPSLETLLAN LTNLEELRLG MVMVNMSSNY GTARWCDAMA RSSPKLRVIS 240

IR64_h_Os12g11510  MPYCSLSGPI CHSLSALRSL SVIELHYNHL SGPVPEFLAA LPSLSVLQLS NNMFEGVFPP 300
LOC_Os12g11510     MPYCSLSGPI CHSLSALRSL SVIELHYNHL SGPVPEFLAA LPSLSVLQLS NNMFEGVFPP 300

IR64_h_Os12g11510  IIFQHEKLTT INLTKNLGIS GNLPTSFSGD SSLQSLSVSN TNFSGTIPGS ISNLRSLKEL 360
LOC_Os12g11510     IIFQHEKLTT INLTKNLGIS GNLPTSFSGD SSLQSLSVSN TNFSGTIPGS ISNLRSLKEL 360

IR64_h_Os12g11510  ALGASGFSGV LPSSIGKLKS LSLLEVSGLE LVGSIPSWIS NLTSLTVLKF FSCGLSGPIP 420
LOC_Os12g11510     ALGASGFSGV LPSSIGQLKS LSLLEVSGLE LVGSIPSWIS NLTSLTVLKF FSCGLSGPI- 419

IR64_h_Os12g11510  ASIGNLKKLT KLALYNCHFS GVIAPQILNL THLQYLLLHS NNLVGTVELS SYSKMQNLSA 480
LOC_Os12g11510     ---------- ---------- ---------- ---------- ---------- ---------- 419

IR64_h_Os12g11510  LNLSNNKLVV MDGENSSSVV SYPNIILLRL ASCSISSFPN ILRHLHEITF LDLSYNQIQG 540
LOC_Os12g11510     ---TTPDQVI SDGPKPSPLT G--------- ---------- LVLHLHEITF LDLSYNQIQG 457

IR64_h_Os12g11510  AIPQWAWKTL NLGFALFNLS HNKFTSIGS- HPLLPVYIEF FDLSFNNIEG VIPIPKEGSV 599
LOC_Os12g11510     AIPLWAWKTL NLGFALFNLS HNKFTSIGSD HPLLPVYIEF FDLSFNNIEG VIPIPKEGSV 517

IR64_h_Os12g11510  TLDYSNNRFS SLPLNFSTYL TNTVLFKASN NSISRNIPPS ICDGIKSLQL IDLSNNNLTG 659
LOC_Os12g11510     TLDYSNNRFS SLPLNFSTYL SNTVLFKASN NSISGNIPPS ICDRIKSLQL IDLSNNNLTG 577

IR64_h_Os12g11510  LIPSCLMEDA DALQVLSLKD NHLTGELPDN IKEGCALSAL DFSGNSIQGQ LPRSLVACRN 719
LOC_Os12g11510     LIPSCLMEDA DALQVLSLKD NHLTGELPDN IKEGCALSAL DFSGNSIQGQ LPRSLVACRN 637

IR64_h_Os12g11510  LEILDIGNNK ISDSFPCWMS KLPQLQVLVL KSNKFIGQIL DPSYTGGGNN CQFTKLQFAD 779
LOC_Os12g11510     LEILDIGNNK ISDSFPCWMS KLPQLQVLVL KSNKFIGQIL DPSYTGGGNN CQFTKLQFAD 697

IR64_h_Os12g11510  MSSNNLSGTL PEEWFKMLKS MIMDTCDNDM LMREQHLYYR GKMQSYQFTA GISYKGSGLT 839
LOC_Os12g11510     MSSNNLSGTL PEEWFKMLKS MIMDTCDNDM LMREQHLYYR GKMQSYQFTA GISYKGSGLT 757

IR64_h_Os12g11510  ISKTLRTLVL IDVSNNAFHG RIPRSIGELV LLRALNMSHN ALTGPIPVQF ANLKQLELLD 899
LOC_Os12g11510     ISKTLRTLVL IDVSNNAFHG RIPRSIGELV LLRALNMSHN ALTGPIPVQF ANLKQLELLD 817

IR64_h_Os12g11510  LSSNELSGEI L* 911
LOC_Os12g11510     LSSNELSGEI L* 829
```

Figure S.2 Amino acid alignment of Os12g11510 from Nipponbare and the IR64 homolog predicted by Fgenesh gene-finder (Softberry, Inc). Sequence differences are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding. No transmembrane domain was predicted for either of the sequences.

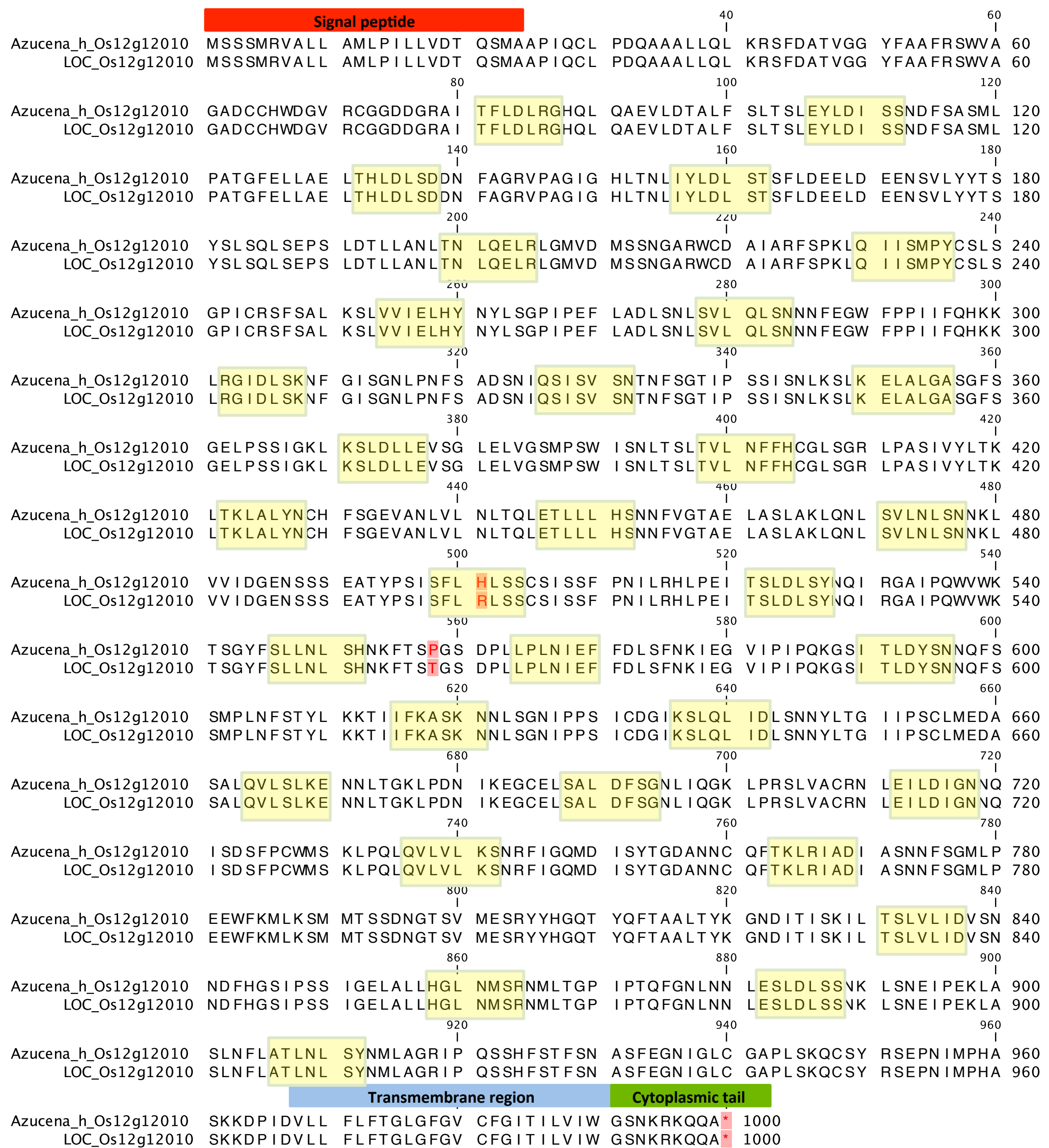


Figure S.3 Amino acid alignment of Os12g12010 from Nipponbare and the Azucena homolog. Sequence differences are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding. Only 2 differences in amino acid sequence were observed.

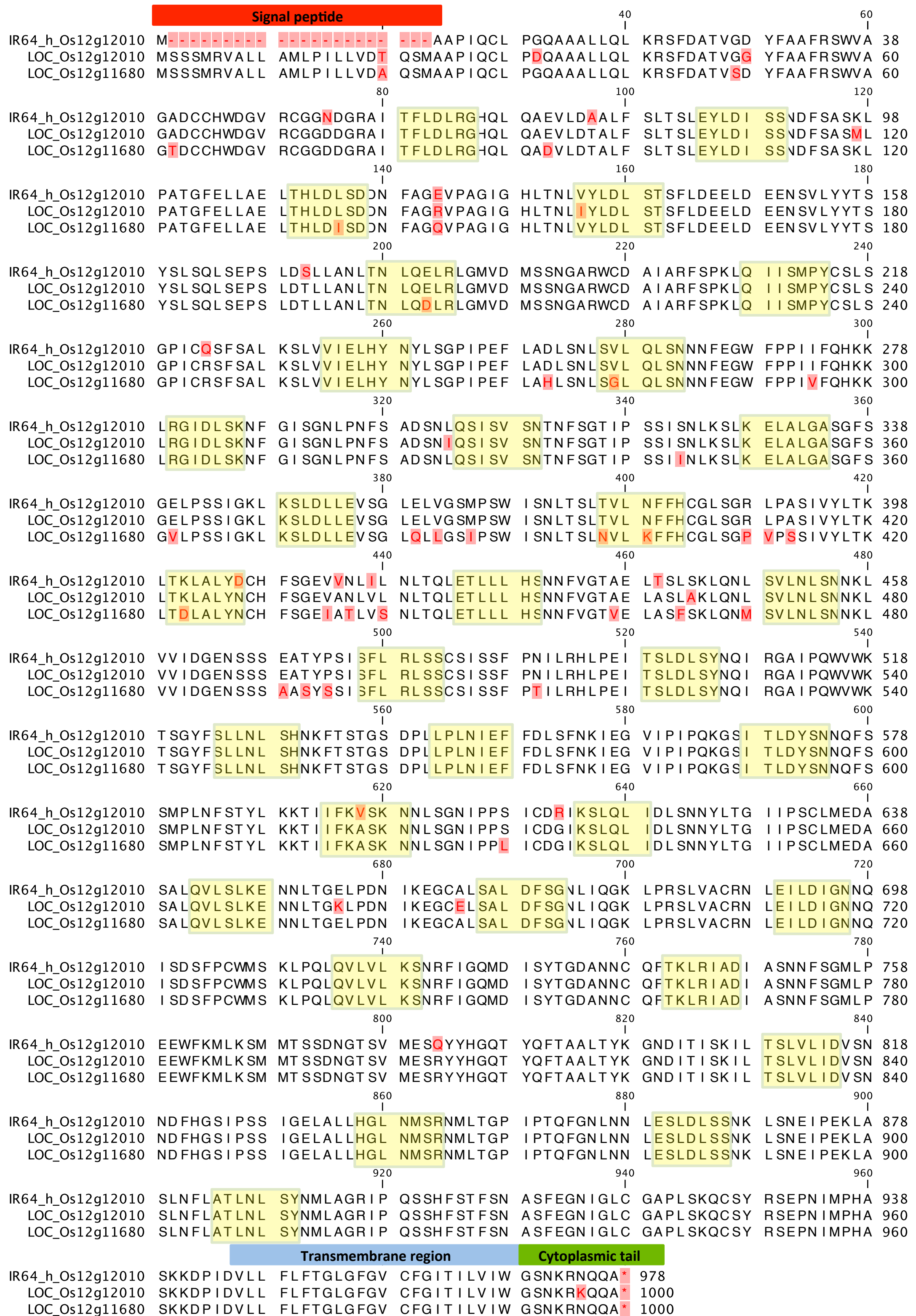


Figure S.4 Amino acid alignment of Os12g12010 and Os12g11680 from Nipponbare and the IR64 homolog (IR64_h_Os12g12010). Sequence differences are highlighted in red. Structural domains of the protein predicted by NCBI protein BLAST and SMART protein (http://smart.embl-heidelberg.de) are shown above alignments. Yellow boxes indicate amino acids that correspond to putative solvent-exposed residues xxLxLxx of the concave (inner) surface of the extracellular leucine-rich repeat (eLRR) domain thought to be involved in ligand binding.

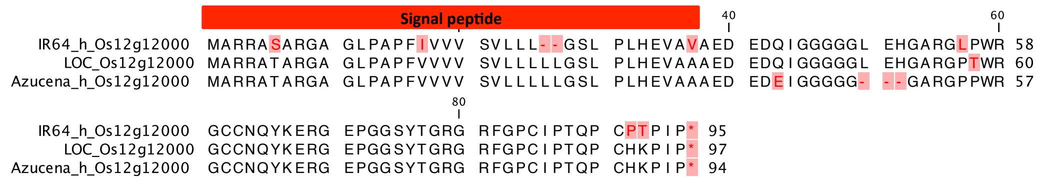


Figure S.5 Amino acid alignment of the Os12g12000 from Nipponbare and the IR64 and Azucena homologs. Sequence differences are highlighted in red. IR64 sequence was predicted by Fgenesh gene-finder (Softberry, Inc), Azucena sequence was amplified by PCR and sequenced by Sanger sequencing. Signal peptide predicted SMART protein (http://smart.embl-heidelberg.de) is shown in red above the sequence.

Chapter 4. Supplementary Figure

Table S.1 Trigger sequences for the 5 constructs used for RNAi. Identifier refers to the transformation event for a particular construct. The construct p number refers to the target gene in the *S. hermonthica* resistance QTL.

Construct	Identifier	Trigger sequence
p11370-4	JS6	AGAATCTTCTTATCAGTAGCACGAACTTCACAGGTATAATACCGAGTTCCATAAG CAATCTCAAATCTCTTACGAAGCTGGACCTTGGTGCCAGTGGCTTCTCTGGAATG CTGCCCTCTTCACTAGGTAGCCTCAAATACCTGGATTTGCTAGAAGTGTCTGGGA TACAGCTAACAGGATCCATGGCACCGTGGATATCAAACCTAACTTCTCTTACTGTT CTCAAGTTCTCTGACTGTGGATTGTCTGGAGAGAT
p11370-10	JS7	GATATGTCCGGCAACGGCGAACGGTGGTGTGACGACATAGCTAAGTTTACACCT AAGCTTCAGGTTCTAAGTTTACCTTACTGCTCATTGTCAGGTCCCATCTGCACATC CTTGTCTTCCATGAATTCGCTCACTAGGATTGAGCTTCATTACAACCACTTGTCAG GTTCAGTGCCAGAGTTCTTGGCTGGCTTTTCCAACCTCACTGTTCTTCAACTGTCC AAAAACAAGTTTGAAGGATTGTTTCCTCCCATCATCTTCC
p11680-1	JS4 JS8	CCAACGATTTCAGTGCATCCAAGCTCCCAGCCACCGGCTTCGAGCTGCTCGCCGA GCTCACCCACCTTGACATCTCCGACGACAATTTCGCCGGCCAGGTACCCGCCGGT ATCGGCCACCTCACGAATCTGGTTTACCTTGATCTTTCTACCAGCTTCCTTGATGA AGAGCTAGATGAAGAGAACAGTGTATTGTACTACACCTCATACTCACTTTCGCAG CTCTCAGAGCCAAG
p11500-4	JS10	CCACTGCCTTCCAGTCATGGGTCGCCGGCACAGACTGCTGCCGCTGGGATGGCG TCGGCTGCGGTGGCGCAGATGGCCGTGTCACCTCACTCGACCTGGGCGGCCACC AATTGCAAGCCGGCAGCGTCGACCCTGCATTGTTCAGGTTAACCTCACTCAAGCA CCTTAACCTCTCCGGTAACGACTTCAGCATGTCCCAGCTCCCGGTGATCACCGGAT TCGAGCAGCTCACCGAACTGGTTTATCTTGATCTCTCCGACACCAACATAGCGGG CGAGGTGCCAGGTAGCATCGGCCGCCTTACGAACCTGGTCTACCTCGACCTCTCC ACCAGTTTCTATATC
p11500-9	AM3	ATGTCATCGTCCACCAAGAGAGTTGCTCACCATCTTCCGTCATTGCTGCTAACCGC GATGTACATTCTCCTCCAAGTCCAGGCCACCACCAATACGGCACGCACCGTAGTA CCACCGGTTCGGTGTCATCCGGATCAAGCCTCAGCGCTGCTCCGGCTGAAGCACT CCTTCAACGCGACCGCCGGCGACTACTCCACTGCCTTCCAGTCATGGGTCGCCGG CACAGACTGCTGCCGCTGGGATGGCGTCGGCTGCGGTGGCGCAGATGGCCGTGT CACCTCACTCGACCTGGGCGGCCACCAATTGCAAGCCGGCAGCGTCGACCCTGCA TTGTTCAGGTTAACC

Chapter 5. Supplementary Figures

Table S.2 Similarity of SNPs to the IR64 allele for 11 genes across the *S. hermonthica* resistance QTL, in 26 diverse rice cultivars. SNP similarity to IR64 for each gene is the sum of the matching SNPs divided by the total number of SNPs, to give a value of similarity between 1 (identical allele) to 0 (no SNP similarity / gene absent). Colours are on a blue – yellow –white scale; blue: 100 % similarity, white: 0 % similarity. The QTL region is indicated above in green. Genes not present in IR64 are coloured grey. The 'Nipponbare specific' region (absent in IR64) is indicated in purple. Cultivars are grouped according to subspecies (all *Oryza sativa*).

Similarity to IR64

			QTL region																	
Subspecies	**Cultivar**	**Os12g10670**	**Os12g10820**	**Os12g10850**	**Os12g10870**	**Os12g10930**	**Os12g11370**	**Os12g11500**	**Os12g11510**	**Os12g11660**	**Os12g11680**	**Os12g11720**	**Os12g11860**	**Os12g11930**	**Os12g11940**	**Os12g12000**	**Os12g12010**	**Os12g12120**	**Os12g12130**	**Os12g12514**
Aromatic	Firooz	0.742	0.883	0.665	0.832	0.825	0.910	0.061	0.730	0.000	0.402	0.181	0.186	0.436	0.665	0.121	0.263	0.132	0.972	0.616
Aromatic	Kitrana508	0.742	0.879	0.709	0.832	0.860	0.914	0.061	0.797	0.000	0.402	0.129	0.192	0.432	0.698	0.121	0.276	0.202	0.972	0.575
Aromatic	Darmali	0.727	0.260	0.153	0.056	0.511	0.482	0.393	0.381	1.000	0.457	0.664	0.622	0.371	0.341	0.318	0.526	0.614	0.101	0.641
Aus	Kalamkati	0.864	0.908	0.622	0.788	0.737	0.910	0.895	0.756	1.000	0.402	1.000	0.727	0.699	0.544	0.455	0.526	0.991	0.908	0.638
Aus	Mehr	0.742	0.568	0.444	0.640	0.203	0.554	0.459	0.651	0.136	0.424	0.017	0.273	0.324	0.352	0.409	0.526	0.026	0.028	0.689
Aus	Miriti	0.742	0.238	0.135	0.088	0.530	0.424	0.380	0.460	1.000	0.446	0.647	0.640	0.409	0.302	0.364	0.526	0.482	0.060	0.698
Aus	Jhona349	0.742	0.257	0.153	0.076	0.648	0.496	0.402	0.403	1.000	0.457	0.698	0.709	0.355	0.269	0.348	0.526	0.614	0.069	0.663
Indica	IR8	0.985	0.987	0.873	0.956	0.756	0.971	0.891	0.898	0.682	0.793	1.000	0.855	0.807	0.786	0.909	0.974	0.939	0.972	0.816
Indica	IR36	0.818	0.898	0.589	0.844	0.841	0.935	0.882	0.784	1.000	0.391	1.000	0.721	0.714	0.588	0.455	0.526	1.000	0.972	0.651
Indica	JC91	0.985	0.921	0.709	0.952	0.867	0.968	0.808	0.787	1.000	0.565	1.000	0.767	0.695	0.670	0.955	0.974	1.000	0.972	0.854
Indica	IR64	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Indica	Gie57	0.803	0.914	0.596	0.868	0.819	0.935	0.882	0.768	0.727	0.380	0.767	0.674	0.699	0.533	0.439	0.513	0.982	0.571	0.648
Indica	Guan-Yin-Tsan	0.985	0.857	0.651	0.824	0.746	0.935	0.860	0.819	0.000	0.413	0.172	0.186	0.699	0.824	0.121	0.211	0.184	0.972	0.610
Indica	Ai-Chiao-Hong	0.939	0.854	0.633	0.872	0.537	0.975	0.865	0.829	0.000	0.413	0.155	0.186	0.726	0.841	0.121	0.263	0.237	0.972	0.790
Indica	Jasmine85	0.970	0.771	0.655	0.928	0.651	0.964	0.764	0.794	1.000	0.783	1.000	0.779	0.676	0.747	0.818	0.895	1.000	0.912	0.543
Temperate japonica	Nipponbare	0.742	0.778	0.582	0.880	0.184	0.942	0.878	0.813	0.000	0.391	0.000	0.203	0.517	0.379	0.530	0.566	0.000	0.028	0.594
Temperate japonica	Bengal	0.652	0.537	0.505	0.848	0.187	0.788	0.803	0.686	0.045	0.413	0.060	0.203	0.494	0.390	0.409	0.513	0.123	0.115	0.365
Temperate japonica	Fanny	0.742	0.775	0.596	0.880	0.184	0.942	0.878	0.813	0.000	0.391	0.000	0.198	0.517	0.379	0.530	0.566	0.009	0.028	0.651
Temperate japonica	Mansaku	0.742	0.238	0.185	0.048	0.476	0.385	0.384	0.327	1.000	0.380	0.638	0.570	0.351	0.280	0.364	0.526	0.500	0.060	0.632
Temperate japonica	Chodongji	0.742	0.178	0.124	0.112	0.549	0.493	0.376	0.432	1.000	0.467	0.586	0.581	0.355	0.220	0.333	0.526	0.754	0.115	0.635
Tropical japonica	Dixiebelle	0.742	0.295	0.196	0.152	0.502	0.464	0.424	0.438	1.000	0.413	0.457	0.424	0.282	0.247	0.333	0.513	0.535	0.065	0.670
Tropical japonica	Davao	0.742	0.238	0.131	0.084	0.562	0.446	0.424	0.390	1.000	0.326	0.716	0.709	0.402	0.264	0.318	0.526	0.561	0.124	0.673
Tropical japonica	Trembese	0.742	0.286	0.156	0.072	0.521	0.428	0.410	0.406	1.000	0.370	0.638	0.663	0.413	0.286	0.303	0.526	0.395	0.065	0.657
Tropical japonica	Curinga	0.727	0.359	0.164	0.080	0.556	0.460	0.472	0.416	1.000	0.402	0.655	0.669	0.359	0.231	0.348	0.526	0.605	0.074	0.676
Tropical japonica	AZUCENA	0.727	0.298	0.131	0.096	0.486	0.406	0.406	0.397	1.000	0.435	0.509	0.634	0.375	0.297	0.303	0.526	0.404	0.055	0.660
Tropical japonica	Binulawan	0.742	0.314	0.175	0.048	0.476	0.428	0.328	0.387	1.000	0.424	0.724	0.587	0.390	0.264	0.364	0.526	0.518	0.074	0.657

Table S.3 Similarity of SNPs to the IR64 allele for 11 genes across the *S. hermonthica* resistance QTL, in 26 diverse rice cultivars. Genes highlighted in red have > 80 % SNP similarity to IR64. Genes blocked out in grey are not present in IR64 (Chapter 3). Cultivars are sorted from most resistant to most susceptible. SNP similarity to IR64 for each gene is the sum of the matching SNPs divided by the total number of SNPs, to give a value of similarity between 1 (identical allele) to 0 (no SNP similarity / gene absent). The QTL region is indicated above in green. All cultivars are *Oryza sativa*.

				QTL region																	
	Subspecies	**Cultivar**	**Os12g10670**	**Os12g10820**	**Os12g10850**	**Os12g10870**	**Os12g10930**	**Os12g11370**	**Os12g11500**	**Os12g11510**	**Os12g11660**	**Os12g11680**	**Os12g11720**	**Os12g11860**	**Os12g11930**	**Os12g11940**	**Os12g12000**	**Os12g12010**	**Os12g12120**	**Os12g12130**	**Os12g12514**
Resistant	Indica	IR8	0.985	0.987	0.873	0.956	0.756	0.971	0.891	0.898							0.909	0.974			0.816
	Indica	IR36	0.818	0.898	0.589	0.844	0.841	0.935	0.882	0.784							0.455	0.526			0.651
	Indica	JC91	0.985	0.921	0.709	0.952	0.867	0.968	0.808	0.787							0.955	0.974			0.854
	Indica	IR64	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000							1.000	1.000			1.000
	Temperate japonica	Nipponbare	0.742	0.778	0.582	0.880	0.184	0.942	0.878	0.813							0.530	0.566			0.594
	Aus	Kalamkati	0.864	0.908	0.622	0.788	0.737	0.910	0.895	0.756							0.455	0.526			0.638
	Aus	Mehr	0.742	0.568	0.444	0.640	0.203	0.554	0.459	0.651							0.409	0.526			0.689
	Indica	Gie57	0.803	0.914	0.596	0.868	0.819	0.935	0.882	0.768							0.439	0.513			0.648
	Aromatic	Firooz	0.742	0.883	0.665	0.832	0.825	0.910	0.061	0.730							0.121	0.263			0.616
	Indica	Guan-Yin-Tsan	0.985	0.857	0.651	0.824	0.746	0.935	0.860	0.819							0.121	0.211			0.610
	Indica	Ai-Chiao-Hong	0.939	0.854	0.633	0.872	0.537	0.975	0.865	0.829							0.121	0.263			0.790
	Temperate japonica	Bengal	0.652	0.537	0.505	0.848	0.187	0.788	0.803	0.686							0.409	0.513			0.365
	Aromatic	Kitrana508	0.742	0.879	0.709	0.832	0.860	0.914	0.061	0.797							0.121	0.276			0.575
	Temperate japonica	Fanny	0.742	0.775	0.596	0.880	0.184	0.942	0.878	0.813							0.530	0.566			0.651
	Aromatic	Darmali	0.727	0.260	0.153	0.056	0.511	0.482	0.393	0.381							0.318	0.526			0.641
	Indica	Jasmine85	0.970	0.771	0.655	0.928	0.651	0.964	0.764	0.794							0.818	0.895			0.543
	Tropical japonica	Dixiebelle	0.742	0.295	0.196	0.152	0.502	0.464	0.424	0.438							0.333	0.513			0.670
	Temperate japonica	Mansaku	0.742	0.238	0.185	0.048	0.476	0.385	0.384	0.327							0.364	0.526			0.632
	Aus	Miriti	0.742	0.238	0.135	0.088	0.530	0.424	0.380	0.460							0.364	0.526			0.698
	Temperate japonica	Chodongji	0.742	0.178	0.124	0.112	0.549	0.493	0.376	0.432							0.333	0.526			0.635
	Aus	Jhona349	0.742	0.257	0.153	0.076	0.648	0.496	0.402	0.403							0.348	0.526			0.663
	Tropical japonica	Davao	0.742	0.238	0.131	0.084	0.562	0.446	0.424	0.390							0.318	0.526			0.673
	Tropical japonica	Trembese	0.742	0.286	0.156	0.072	0.521	0.428	0.410	0.406							0.303	0.526			0.657
	Tropical japonica	Curinga	0.727	0.359	0.164	0.080	0.556	0.460	0.472	0.416							0.348	0.526			0.676
	Tropical japonica	AZUCENA	0.727	0.298	0.131	0.096	0.486	0.406	0.406	0.397							0.303	0.526			0.660
	Tropical japonica	Binulawan	0.742	0.314	0.175	0.048	0.476	0.428	0.328	0.387							0.364	0.526			0.657