

**A Pathogenomic Approach towards Characterising the
South African Population of *Puccinia striiformis* f. sp. *tritici*,
the Causal Agent of Wheat Stripe Rust**

Hester Josina van Schalkwyk

**Thesis submitted in fulfilment of the requirements for the degree
Doctor of Philosophy**

University of the Free State
Bloemfontein
South Africa

Department of Plant Sciences (Plant Pathology and Plant Breeding)
Faculty of Natural and Agricultural Sciences

January, 2018

Promoter:

Dr R Prins

Department of Plant Sciences, University of the Free State and CenGen (Pty) Ltd

Co-promoters:

Dr DGO Saunders

John Innes Centre, Norwich, United Kingdom

Dr LA Boyd

National Institute of Agricultural Botany, Cambridge, United Kingdom

Prof. ZA Pretorius

Department of Plant Sciences, University of the Free State

Declaration

I, Hester Josina van Schalkwyk, declare this thesis hereby submitted by me for the degree Doctor of Philosophy at the University of the Free State is my own independent work and has not previously been submitted by me to another university for any degree.

I cede copyright of this thesis in favour of the University of the Free State.

Hester Josina van Schalkwyk

Date

*Dedicated to Mrs Marlize Huisamen (née Vivier),
my high school biology teacher who first taught me about DNA
and nurtured my curiosity about living things.*

Acknowledgements

I would like to express my sincere gratitude to my mentors and the funding bodies that supported me during my PhD.

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), the Department for International Development and (through a grant to BBSRC) the Bill & Melinda Gates Foundation, under the Sustainable Crop Production Research for International Development (SCPRID) programme, a joint initiative with the Department of Biotechnology of the Government of India's Ministry of Science and Technology. Two SCPRID grants supported this study: (BB/J011525/1) to Dr L Boyd, Dr R Prins and Prof. ZA Pretorius, and (BB/J012017/1) to Dr Cristobal Uauy. Additional support were received from The Monsanto Beachell-Borlaug International Scholars Program (MBBISP) and the Winter Cereal Trust (WCT), South Africa, through PhD scholarships.

The contributions of my supervisors go far beyond what I can summarise in a paragraph, nonetheless, a special thank you for the unique role they each played during my PhD. I thank Dr Diane Saunders and Dr Renée Prins for creating environments with nearly unlimited resources where I could work. I thank Dr Lesley Boyd for intense supervision while I was preparing my thesis, and Prof. Zakkie Pretorius for mentoring me in the art and science of rust pathology. I thank Dr Prins for her vision for the project and allowing me to change to this project that I so enjoyed working on. I would also like to thank Dr Cristobal Uauy

for being instrumental in arranging my placement in the Saunders lab.

I thank the following people for their involvement in obtaining the sequencing datasets: Historical South African isolates were obtained from Zakkie Pretorius. Historical East African isolates were obtained from Mogens Hovmøller. The Pakistan isolates were obtained from Sajid Ali. Samples of the recent South African *Pst* population were obtained from Driecus Lesch, Tarekegn Terefe, Zakkie Pretorius and Willem Boshoff (lost in transit). Renée Prins was instrumental in the preparatory work and shipment of the South African isolates for sequencing. Recent East African isolates were obtained from David Hodson (Ethiopia, 2014) and Ruth Wanyera (Kenya, 2014). Existing datasets of *Pst* isolates were obtained from Diane Saunders.

What fantastic opportunities to work at CenGen (Pty) Ltd, Earlham Institute, John Innes Centre, and the University of the Free State, during my PhD! A special mention to the following people for support in and out of the lab. Debbie Snyman performed qPCR assays and gel electrophoresis towards this project. Zakkie Pretorius multiplied the historical South African *Pst* urediniospores for sequencing and mentored me in inoculation and scoring of the infection assays on the differential wheat set seedlings. Sarah Holdgate for providing the United Kingdom (UK) differential wheat lines and informative discussions regarding *Pst* in the UK and UK wheat cultivars. Elsabet Wessels, Debbie Snyman, Jens Mains and Clare Lewis mentored me in specific molecular genetic procedures. I thank Philippa Borril and Oluwaseyi Shorinola for advice on RT-qPCR data analysis, and Albor Dobón for help with the planning of the time course experiment. I also thank Antoine Persoons for valuable discussions in population genetics and advice on sections of this thesis and my fellow PhD students in the Saunders lab, especially Pilar Corredor-Moreno and Vanessa Bueno-Sancho, for always being ready to advise me on the newest updates in data handling or Norwich BioScience Institutes (NBI) cluster computing. Also, thank you to the Computing NBI

Helpdesk staff, especially Tom Betteridge and Mohamed Imram, for computer support. I thank Sadie Geldenhuys for administrative support at UFS, Lizaan Rademeyr for great practical advice on best practices in laboratory record keeping, Carel van Heerden for input in the early days of the project and Anelda van der Walt for initial bioinformatics training. I thank Cari van Schalkwyk for advice on statistical analysis. I thank Prof. Ed Runge and the MBBISP panel for the very special ongoing experience of being an MBBISP scholar.

Thank you to every friend that ran, walked, climbed mountains, or performed some strange hobby with me. That helped to keep me going through the hard times. George, for your rock-solid support and your immense contribution to tailoring my skill set, thank you. I thank my family for all their love and support along the way. Thank you, dad, for reminding me that I am a finisher, and mum, for your consistent positivity, enthusiasm, and encouragement that runs through my life like a golden thread.

Abstract

Stripe (yellow) rust caused by the fungus *Puccinia striiformis* Westend. f. sp. *tritici* (*Pst*) is a major disease of wheat prevalent in most areas where wheat is cultivated across the globe. It can completely destroy a crop if left untreated. The *Pst* fungus develops feeding structures that form a close relationship with the host tissue where it facilitates extraction of water and nutrients from the plant, while manipulating the host for its own benefit using effector proteins. This parasitic behaviour reduces yield and grain quality, leading to the propagation of numerous *Pst* spores, spreading infection. In South Africa stripe rust was first detected in 1996 with the initial pathotype being designated 6E16A-. Thereafter, three more *Pst* pathotypes were detected in subsequent years (6E22A- in 1998, 7E22A- in 2001 and 6E22A+ in 2005), gaining virulence in a stepwise manner by overcoming additional resistance genes one by one. However, the source of the original pathotype and the current genetic diversity of the *Pst* population within South Africa remain open questions.

To get a better understanding of the South African *Pst* pathotypes and how they relate to *Pst* pathotypes globally, the historical population was described using a recently developed “field pathogenomics” approach. High-resolution, next-generation sequencing data utilised in this method aided in determining the genomic relationships between the four historical pathotypes and investigating their potential origin. Historic South African isolates representing the four identi-

fied pathotypes were re-sequenced, and their comparison with isolates from the United Kingdom, France, Pakistan, Ethiopia, Eritrea and Kenya revealed that the closest relatives of the historical South African isolates were a group of isolates from East Africa.

We further described polymorphisms in the South African *Pst* population that supported the existing hypothesis of stepwise evolution. Through applying pairwise comparisons between polymorphic sites across isolates, 27 potential effector proteins that could be instrumental in the stepwise virulence gain, were identified. To study the role these candidates may play during the infection processes in different pathotypes, gene expression profiling was conducted using RT-qPCR. Preliminary patterns of up- or down-regulation of these effectors between time points, over a time course of compatible interactions, were described. Furthermore, infected wheat tissues collected from locations across South Africa during the 2013, 2014 and 2015 cropping seasons, were sequenced. The “field pathogenomics” method, using RNA-Seq, was applied to compare the historic *Pst* isolates with the recent population. This analysis indicated the possibility of a novel introduction of *Pst* into South Africa in recent years, possibly between 2011 and 2013. Pathotyping of selected *Pst* isolates on supplementary wheat tester genotypes revealed novel variation in infection types that has not been described previously.

This study provides a high resolution, genomic view of the historical and prevailing *Pst* populations and adds valuable information to the potential origin and adaptation of stripe rust in South Africa. The research outcomes provide a genomic base for further investigation of candidate effector genes and the possible recent novel incursion of a pathotype group also seen in Europe, East Africa and New Zealand into South Africa.

Keywords: effector, origin, plant pathology, population genomics, virulence

Contents

Declaration	ii
Acknowledgements	iv
Abstract	vii
List of Figures	xv
List of Tables	xix
List of Abbreviations	xxi
1 General Introduction	1
1.1 Socio-economic importance of wheat	2
1.2 Wheat cultivation in South Africa	2
1.3 Wheat rusts reduce yields	4
1.4 Motivation for this study	6
1.5 Objectives	6
1.6 Thesis outline and approaches	7
2 The Wheat Rusts: Life Histories, Host Response Mechanisms and Ge- nomic Resources	9
2.1 The rusts	9
2.1.1 Filamentous plant pathogens	9
2.1.2 Rusts and their primary host	11
2.1.3 The alternative host	12
2.1.4 Global distribution of stripe rust	13
2.1.5 Favourable conditions for wheat rusts	13
2.1.6 Infection cycle of <i>Puccinia</i> rusts	15
2.1.7 The stripe rust infection process on wheat	19
2.2 Combating wheat stripe rust	21

2.3	Plant defence mechanisms	22
2.3.1	Host-pathogen interaction	23
2.3.2	Other sources of resistance	26
2.4	The <i>Pst</i> genome	26
2.4.1	Genomic variation	26
2.4.2	Rust genomics	27
2.4.3	Challenges in bioinformatics	31
2.4.4	Effector identification	32
3	General Materials and Methods	35
3.1	Preparation and collection of materials	35
3.1.1	Inoculation	35
3.1.2	Protocol for sampling infected wheat tissue	36
3.2	Nucleic acid extraction and quantification	37
3.2.1	Genomic DNA extraction	37
3.2.2	RNA extraction	37
3.2.3	DNA and RNA quantification	38
3.3	Next-generation sequencing and data analysis	39
3.3.1	Library preparation	39
3.3.2	Genomic DNA sequencing	39
3.3.3	RNA sequencing	40
3.3.4	Bioinformatics pipeline	40
3.3.5	Clustering analysis	42
4	Origin of the South African <i>Pst</i> Pathotypes	48
4.1	Introduction	48
4.1.1	Wheat stripe rust in South Africa	48
4.1.2	<i>Pst</i> population diversity	52
4.1.3	Molecular markers and <i>Pst</i>	53
4.1.4	Next-generation sequence analyses of South African <i>Pst</i>	55
4.2	Materials and methods	56
4.2.1	Data description	56
4.2.2	Sample preparation for DNA extraction	57
4.2.3	Genomic DNA extraction and quantification	59
4.2.4	Sequencing and mapping	59
4.2.5	Phylogenetic analysis	60
4.2.6	Population structure analysis	60

4.2.7	Genetic diversity assessment	60
4.3	Results	61
4.3.1	Re-sequencing of South African <i>Pst</i> pathotypes	61
4.3.2	Purity assessment of samples	62
4.3.3	Clustering analyses	62
4.3.4	Phylogenetic analysis	62
4.3.5	Population structure analysis	64
4.3.6	Population differentiation	71
4.3.7	Genetic diversity within and between population clusters	71
4.4	Discussion	73
4.5	Conclusion	77
5	Analyses of Polymorphisms in Historical South African <i>Pst</i> Isolates in Search of Candidate Effector Genes	79
5.1	Introduction	80
5.1.1	The importance of <i>Pst</i> variability	81
5.1.2	Mutations—causes, types and effects	82
5.1.3	Genomic approaches used to identify effectors	85
5.2	Materials and methods	86
5.2.1	SNP analysis	86
5.2.2	Positive selection	87
5.2.3	Presence-absence analysis	87
5.2.4	Comparisons of nonsynonymous SNP sites between isolates	88
5.2.5	Multiple sequence alignments to visualise biallelic SNPs	88
5.3	Results	89
5.3.1	SNP identification in the genomes of the historical South African isolates	89
5.3.2	Assessment of polymorphisms to detect positive selection	93
5.3.3	Presence or absence of genes	98
5.3.4	Investigation of candidate genes that are likely to experience evolutionary changes	105
5.3.5	Candidate effectors with sequence polymorphisms between the South African isolates	106
5.4	Discussion	107
5.4.1	Polymorphic sites	108
5.4.2	STOP codons	110
5.4.3	Transitions and transversions at specific codon positions	111

5.4.4	Stepwise mutations	112
5.4.5	Positive selection	112
5.4.6	Presence-absence analysis	113
5.4.7	Nonsynonymous polymorphisms	114
5.5	Conclusion	114
6	Gene Expression Analysis of Candidate Effectors Identified in South African <i>Pst</i> Isolates	115
6.1	Introduction	115
6.1.1	Regulation of gene expression in eukaryotes	116
6.1.2	Quantification of gene expression	117
6.1.3	Candidate effector features	118
6.1.4	Gene transcription analysis	118
6.2	Methods	120
6.2.1	Inoculation and sampling	120
6.2.2	Tissue disruption and RNA extraction	122
6.2.3	RNA quality control and quantification	123
6.2.4	Complementary DNA synthesis	123
6.2.5	Primer design	123
6.2.6	PCR plate setup	124
6.2.7	Quantitative real-time polymerase chain reaction	126
6.2.8	Reference gene selection	127
6.2.9	Efficiency determination of primers	127
6.2.10	Statistical evaluation of the data	129
6.2.11	Linear mixed effect analysis	129
6.2.12	Relative expression of <i>Pst</i> candidate effector genes	130
6.2.13	Assessment of genes	131
6.3	Results	131
6.3.1	RNA yield, RNA quality scores and cDNA yield	131
6.3.2	Primer design	132
6.3.3	Efficiency determination of primers	134
6.3.4	Statistical analysis of the relative expression of nine <i>Pst</i> candidate effector genes	134
6.3.5	Expression profiles of candidate genes	139
6.3.6	Gene validation using revised gene models and transcript data	139
6.4	Discussion	141

6.5	Conclusion	144
7	Analysis of the Current Stripe Rust Threat in South Africa	145
7.1	Introduction	145
7.1.1	<i>Pst</i> virulence since 2005	145
7.1.2	Global reports on <i>Pst</i> population shifts	146
7.1.3	Objectives	149
7.2	Materials and methods	149
7.2.1	Stripe rust samples used in RNA sequencing analyses . . .	149
7.2.2	Transcriptome sequencing of stripe rust infected wheat leaves	151
7.2.3	<i>Pst</i> pathotype determination	152
7.3	Results	153
7.3.1	Clustering analysis using RNA-Seq and whole genome sequencing data	153
7.3.2	Seedling <i>Pst</i> pathotype testing	162
7.4	Discussion	164
7.5	Conclusion	172
8	General Discussion	173
8.1	The historical South African <i>Pst</i> population	173
8.2	Candidate effector identification and evaluation	175
8.3	The recent South African <i>Pst</i> population	177
8.4	Future work	179
8.5	Conclusion	180
	Appendices	181
A	The Origin of the South African <i>Pst</i> Pathotypes	181
B	Analyses of Polymorphisms in Historical South African <i>Pst</i> Isolates in Search of Candidate Effector Genes	183
B.1	Genes present in the PST130 reference genome but absent in the four historical South African <i>Pst</i> isolates	183
B.2	Annotations of genes homologous to identified PST130 genes . .	185
B.3	Nonsynonymous polymorphisms in candidate genes	193
C	Gene Expression Analysis of Candidate Effectors Identified in South African <i>Pst</i> Isolates	222
C.1	Candidate gene inspection	223

C.2	Additional figures of statistical analyses	232
C.3	Variability in RT-qPCR	239
C.3.1	Variation in the application of treatments to biological replicates	240
C.3.2	Variation introduced by the RNA extraction process	240
C.3.3	Variation introduced by the reverse transcription process	241
C.3.4	Variation introduced by RT-qPCR	242
C.3.5	Variation introduced by primers	242
C.3.6	Choice of reference genes	244
C.3.7	Results of efficiency corrected relative gene expression . .	245
D	Analysis of the Current Stripe Rust Threat in South Africa	248
	Bibliography	254

List of Figures

1.1	Area harvested, production and yield statistics for South African wheat cultivation between 1990 and 2017.	4
2.1	The phylogenetic relationship of plant pathogenic ascomycetes, basidiomycetes and oomycetes following Neighbour Joining analysis.	10
2.2	Taxonomic classification of the wheat rusts.	11
2.3	Global distribution of <i>Puccinia striiformis</i> f. sp. <i>tritici</i> , before and after 2000.	14
2.4	Spore stages and the infection cycle of <i>Pst</i>	16
2.5	A stripe rust uredinium pustule.	16
2.6	Illustration of the infection process of <i>Pst</i>	19
2.7	Illustration of a filamentous plant pathogen haustorium.	20
2.8	The five main classes of plant disease resistant proteins.	25
4.1	Locations of the original detections of South African <i>Pst</i> pathotypes.	49
4.2	Temperature and rainfall measured in 1996 during the wheat-growing season in the Western Cape, compared to the 11 year mean.	50
4.3	Schematic illustration of the increase of <i>Pst</i> virulence in South Africa.	52
4.4	Pathotype identification tests of South African <i>Pst</i> pathotypes.	52
4.5	Read frequency graphs from heterokaryotic SNP sites for SA1–SA4.	63
4.6	The phylogenetic relationship between the South African <i>Pst</i> isolates and European, Asian and East African isolates.	65
4.7	Evaluation of the number of population clusters following STRUCTURE analyses.	66
4.8	Bar charts representing STRUCTURE population clusters.	67
4.9	Discriminant analysis of principal components analysis of 48 <i>Pst</i> isolates.	69
4.10	Bar charts representing DAPC population structure analysis.	70

4.11	Genetic diversity assessed between 10 population clusters.	72
5.1	Nucleotide changes that introduced stop codons.	92
5.2	Distribution of stop codons accross all genes per isolate.	92
5.3	Percentage frequency matrices of transitions and transversions at monoallelic SNP sites.	94
5.4	Percentage occurrence matrices of transitions and transversions at biallelic SNP sites.	95
5.5	Codon positions of nucleotide changes at homokaryotic SNP sites.	96
5.6	Codon positions of nucleotide changes at heterokaryotic SNP sites.	97
5.7	Presence-absence analysis.	103
5.8	Nonsynonymous SNPs in the gene space of the four South African isolates increase over time and with increasing virulence.	106
5.9	Translated sequence alignment of gene PST130_00285.	107
5.10	Over- and underestimates of SNP sites.	109
6.1	Experimental setup for the infection time course experiment.	121
6.2	Plate layouts for RT-qPCR assays.	125
6.3	Linear regression showing estimated efficiency of primers.	135
6.4	Relative gene expression of nine candidate effector genes.	138
7.1	Prevalence of <i>Pst</i> in South Africa between 2008 and 2016.	147
7.2	Locations of <i>Pst</i> collections between 2013 and 2015.	151
7.3	Phylogenetic tree displaying the relationship between <i>Pst</i> isolates.	155
7.4	Relative distance maximum likelihood phylogenetic tree.	156
7.5	Evaluation of number of population clusters following STRUC- TURE analyses.	158
7.6	STRUCTURE histogram plots of population clusters.	159
7.7	Discriminant analysis of principal components analysis of <i>Pst</i> iso- lates.	160
7.8	Histogram plots indicating population structure as inferred by DAPC analysis.	161
7.9	Measurements of genetic diversity by F_{ST} calculation of pairs of population groups.	163
7.10	Infection type comparisons between one historical and one recent <i>Pst</i> isolate.	165
7.11	Number of international tourist arrivals in South Africa between 1995 and 2014.	171

A.1	Read frequency graphs for East African isolates analysed in Chapter 4	182
B.1	Translated sequence alignment of gene PST130_02001.	193
B.2	Translated sequence alignment of gene PST130_02118.	194
B.3	Translated sequence alignment of gene PST130_02403.	195
B.4	Translated sequence alignment of gene PST130_05023.	196
B.5	Translated sequence alignment of gene PST130_05454.	197
B.6	Translated sequence alignment of gene PST130_05944.	198
B.7	Translated sequence alignment of gene PST130_06503.	199
B.8	Translated sequence alignment of gene PST130_06558.	200
B.9	Translated sequence alignment of gene PST130_07448.	201
B.10	Translated sequence alignment of gene PST130_07513.	202
B.11	Translated sequence alignment of gene PST130_07564.	203
B.12	Translated sequence alignment of gene PST130_08031.	204
B.13	Translated sequence alignment of gene PST130_08984.	205
B.14	Translated sequence alignment of gene PST130_09018.	206
B.15	Translated sequence alignment of gene PST130_09275.	207
B.16	Translated sequence alignment of gene PST130_10286.	208
B.17	Translated sequence alignment of gene PST130_12487.	209
B.18	Translated sequence alignment of gene PST130_12491.	210
B.19	Translated sequence alignment of gene PST130_12956.	211
B.20	Translated sequence alignment of gene PST130_13969.	212
B.21	Translated sequence alignment of gene PST130_14091.	213
B.22	Translated sequence alignment of gene PST130_14831.	214
B.23	Translated sequence alignment of gene PST130_16778.	215
B.24	Translated sequence alignment of gene PST130_17605.	216
B.25	Translated sequence alignment of gene PST130_17605.	217
B.26	Translated sequence alignment of gene PST130_07579.	218
B.27	PST130_07579 continued from previous page.	219
B.28	Translated sequence alignment of gene PST130_15131.	220
B.29	PST130_15131 continued from previous page.	221
C.1	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_02001 in SA1 and SA4.	223
C.2	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_02403 in SA1 and SA4.	224

C.3	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_05023 in SA1 and SA4.	225
C.4	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_06503 in SA1 and SA4.	226
C.5	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_07513 in SA1 and SA4.	227
C.6	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_09725 in SA1 and SA4.	228
C.7	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_12487 in SA1 and SA4.	229
C.8	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_12491 in SA1 and SA4.	230
C.9	Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_12956 in SA1 and SA4.	231
C.10	Graphical tests for normality and equal variances of the residuals and random intercepts.	233
C.11	Gene and isolate specific tests for equal variances after the model was fitted to the relative gene expression values.	234
C.12	Gene and isolate specific tests for equal variances after the model was fitted to the relative gene expression values.	235
C.13	Graphical tests for normality and equal variances of the residuals and random intercepts following a $\log_{10}$ transformation.	236
C.14	Gene and isolate specific normal probability plots of the residuals after the model was fitted to the $\log_{10}$ transformed relative gene expression values.	237
C.15	Gene and isolate specific tests for equal variances after the model was fitted to the $\log_{10}$ transformed relative gene expression values.	238
C.16	High inter-run variability in relative expression patterns.	246
C.17	The Pfaffl method of relative gene expression shows the relative gene expression of SA1 to SA4.	247
D.1	Read frequency graphs from heterokaryotic SNP sites for the recent South African field isolates.	249
D.2	Read frequency graphs from heterokaryotic SNP sites for the recent East African field isolates.	250
D.3	Circular relative distance maximum likelihood phylogenetic tree.	251

List of Tables

1.1	Domestic grain consumption of the three highest consumed grains worldwide	3
2.1	Whole genome sequencing projects using next- and third-generation sequencing	29
4.1	Global isolates included in the clustering and genetic diversity analyses	58
4.2	Historical isolates used in re-sequencing and an infection time course experiment	59
4.3	Statistics of read alignment of the historical South African isolates to the PST130 reference genome	63
5.1	Homokaryotic and heterokaryotic SNPs in the South African isolates	90
5.2	The number of SNPs identified in coding regions of the four South African <i>Pst</i> isolates	91
5.3	Polymorphic genes with positive dN values indicating nonsynonymous changes in isolate pairwise comparisons	99
5.4	Polymorphic genes with positive dS values indicating synonymous changes in isolate pairwise comparisons	99
5.5	Number of absent genes in the four South African <i>Pst</i> pathotypes	100
5.6	Potential orthologs of genes absent in all four of the South African isolates	101
5.7	The number of potential paralogs identified in genes absent in all four South African isolates	102
5.8	Potential paralogs of genes absent in the four South African isolates	102
5.9	Potential orthologs of genes absent in three or less of the South African isolates	104
5.10	Number of potential paralogs in PST130	104

5.11	Paralogs of genes that only occurred in one of the South African isolates	105
6.1	Effector features of the identified candidate effectors	119
6.2	Summary statistics describing RNA yield, integrity and cDNA yield as required in the MIQE guidelines	132
6.3	Primer and amplicon specifications for <i>Pst</i> candidate effector gene identification	133
6.4	Significance of the factor “Time Point” in the linear mixed model for those genes where it was significant	139
6.5	Multiple comparisons between time points for each gene that showed significant difference in expression over the time series .	140
7.1	Wheat differential lines used at Agricultural Research Council, Small Grain, South Africa	146
7.2	African isolates collected between 2013 and 2015	150
7.3	Infection type scores used to assess <i>Pst</i> infection on wheat seedlings	153
B.1	PST130 genes (211) that were absent in all four historical South African isolates	184
D.1	Differential testing of South African <i>Pst</i> isolates previously defined as pathotype 6E16A- on an extended set of wheat seedling testers	252
D.2	Differential testing of South African <i>Pst</i> isolates previously defined as pathotype 6E22A+ on an extended set of wheat seedling testers	253

List of Abbreviations

3'	three prime
5'	five prime
A	adenine
ABI	Applied Biosystems Integrated
ADP	Adenosine diphosphate
ACTB	β -Actin
AFLP	Amplified Fragment Length Polymorphism
ANOVA	analysis of variance
ARC-SG	Agricultural Research Council, Small Grain
ARF	ADP ribosylation factors
ATP	Adenosine triphosphate
Avr	avirulence
BAC	bacterial artificial chromosome
BAM	binary alignment map
BBSRC	Biotechnology and Biological Sciences Research Council
BGRI	Borlaug Global Rust Initiative
BIC	bayesian information criterion
bp	base pairs
C	cytosine
CAF	Central Analytical Facilities
cDNA	complementary DNA
CEC	Crop Estimate Committee
CIMMYT	International Maize and Wheat Improvement Center
CTAB	cetyltrimethylammonium bromide
CVEGE	clonal variation in effector gene expression
DA	discriminant analysis

DAPC	discriminant analysis of principal components
DNA	deoxyribonucleic acid
dpi	days post inoculation
ds	double stranded
dsDNA	double stranded DNA
EMS	ethyl methanesulfonate
EST	expressed sequence tag
ETI	effector-triggered immunity
FIR	flanking intergenic regions
G	guanine
GAPs	GTPase activating proteins
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
gDNA	genomic DNA
gene	virus induced
GTR	general time reversible
HCD	hypersensitive cell death
HIGS	host-induced gene silencing
HMC	haustorial mother cell
IH	infection hyphae
IP	infection peg
kbp	kilo base pairs
Lr	wheat leaf rust resistance gene designation
MAMPs	microbe-associated molecular patterns
MAS	marker assisted selection
MBBISP	Monsanto Beachell-Borlaug International Scholars Program
Mbp	mega base pairs
MCMC	Markov Chain Monte Carlo
miRNAs	microRNAs
mRNA	messenger RNA
MSL	Molecular marker Service Laboratory
NB-LRR	nucleotide-binding site (NBS)-leucine-rich repeat (LRR) proteins
NBI	Norwich BioScience Institutes
NGS	next-generation sequencing
NLS	nuclear-localisation signal

NMD	nonsense-mediated mRNA decay
NTC	non template control
oligo-dTs	thymine oligonucleotides
PAMPs	pathogen-associated molecular patterns
PCA	principal component analysis
<i>Pgt</i>	<i>Puccinia graminis</i> f. sp. <i>tritici</i>
PI	phosphoinositide
PR	pathogen-related
PRRs	pathogen receptor proteins
<i>Pst</i>	<i>Puccinia striiformis</i> f. sp. <i>tritici</i>
<i>Pt</i>	<i>Puccinia triticina</i>
PTI	PAMP triggered immunity
qPCR	quantitative or real time PCR
R	resistance or resistant
RAxML	randomized accelerated maximum likelihood
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-Seq	RNA sequencing
ROS	reactive oxygen species
RT	reverse transcriptase
S	susceptible (as in Avocet S)
SAGL	South African Grain Laboratory
SAM	sequence alignment map
SCAR	sequence-characterised amplified region
SCPRID	Sustainable Crop Production Research for International Development
SCR	small and cysteine rich
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SNPs	polymorphisms
<i>Sr</i>	wheat stem rust resistance gene designation
ss	single strand
SSV	substomatal vesicle
T	thymine
tRNA	transfer RNA

TUBB	β -Tubulin
UK	United Kingdom
UKCPVS	UK Cereal Pathogen Virulence Survey
USA	United States of America
UTRs	untranscribed regions
UV	ultraviolet
VIGS	virus induced gene silencing
WC	wheat control
WCT	Winter Cereal Trust
Yr	wheat stripe rust resistance gene designation
Z12	Zadoks growth stage 12

Mathematical notation

C_T	threshold cycle
F_T	fluorescence threshold
R^2	Pearson correlation coefficient

Chapter 1

General Introduction

WHEAT IS A STAPLE CROP in many countries around the globe, including South Africa. In most areas of wheat cultivation, one or more of the three rust diseases have the potential to severely compromise yields (Kolmer, 2005; Huerta-Espino et al., 2011; Shaw and Osborne, 2011; Dean et al., 2012; Beddow et al., 2015). Rusts are specialised in infecting wheat and maintain an obligatory parasitic symbiosis with susceptible hosts throughout their life cycles, using resources predestined for plant growth, maintenance, and grain development to ultimately produce multitudes of spores (Chen, 2005). The continuously growing demand for wheat requires careful consideration of mechanisms to address host resistance to manage these crippling diseases. Management strategies aim to increase crop yields and reduce quantities of inoculum. Smaller rust population sizes reduce the potential of the fungus to gain new pathogenicity through evolutionary machineries such as mutation and somatic and sexual recombination (Hovmøller and Justesen, 2007a; Jin et al., 2010; Zhao et al., 2013; Jiao et al., 2017).

1.1 Socio-economic importance of wheat

Bread wheat, *Triticum aestivum* L., is an important food source making up 20 % of global calories and protein intake (Shiferaw et al., 2014). Recent estimates placed domestic consumption at 736.86 million tons for the 2016/2017 market year (FAS USDA, 2017). The three most prominent staple grains—wheat, maize and rice—are under heavy pressure for increased yields to secure food for the growing world population (Table 1.1; FAS USDA, 2017). By mid-2017, the estimated global population size was 7.6 billion people and predictions estimate increases of up to 9.8 billion by 2050, with a further increase to 11.2 billion by 2100 (United Nations, 2017). The growing population places increased pressure on crop production as a primary source of human nutrition, animal feed and bio-fuel (Edgerton, 2009). Yield improvements of roughly 2.4 % per year are needed to be able to meet the target of doubling global crop production, but currently global average rates are failing to reach this target (Ray et al., 2013). Other sectors also rapidly out-compete the agricultural sector for land, adding to the pressure to produce enough food for the growing population, while acreage continues to diminish.

1.2 Wheat cultivation in South Africa

Wheat was brought to South Africa by the Dutch settlers in 1652 and drought, wind, and disease challenged early wheat production (Du Plessis, 1933), as is still the case today (FAS USDA, 2016). Currently, wheat is the most planted winter cereal crop in South Africa, ranking second to maize for overall crop size (SAGL, 2012) and consumption (FAS USDA, 2017) in the country.

South Africa is the largest consumer of wheat in Sub-Saharan Africa, and population growth and urbanisation will likely continue to increase the demand (ITA USDC, 2017). Most of the crop is cultivated on dry land and grown in

Table 1.1: Domestic grain consumption of the three highest consumed grains worldwide, in million tons, as recorded for 2016/17 (FAS USDA, 2017)

Grain	World	South Africa
Rice	478.46	0.83
Wheat	736.86	3.40
Maize	1 053.85	11.70

the winter-rainfall areas of the Western Cape where there is a Mediterranean climate. Here wheat is planted from mid-April until mid-June, and harvested from October to December. In the Eastern Free State, a summer-rainfall area, wheat is sown from June until August and harvested between November and January. Irrigated wheat cultivation is practised in the Northern Cape, using water from the Orange River (Van Niekerk, 2001; SAGL, 2012).

The trend over the last 20 years indicates a reduction in wheat cultivation (Figure 1.1). This is driven by considerable annual production fluctuations, caused by unpredictable weather patterns and declining profit margins for wheat (AgriOrbit, 2017). The decrease in wheat cultivation—in favour of other, more climate tolerant and often higher value crops such as maize, canola and soybeans—increases the dependence on imports to meet the growing wheat demand in South Africa (ITA USDC, 2017).

The prominent grain industry in South Africa contributes more than 30 % of the total gross value of agricultural production in the country (DAFF, 2015). On average, 63 % of the total demand over the past 10 years was produced domestically, while the remainder was imported (DAFF, 2016). In 2017, 1.8 million tons of wheat were imported (IndexMundi, 2017). South Africa exported 0.2 million tons in 2017 (IndexMundi, 2017). Most exports are destined for neighbouring countries, Zambia, and Mauritius (FAS USDA, 2016).

The increase in yield, despite the reduction in planted area (Figure 1.1), can be attributed to improved agronomic practices and the development of better

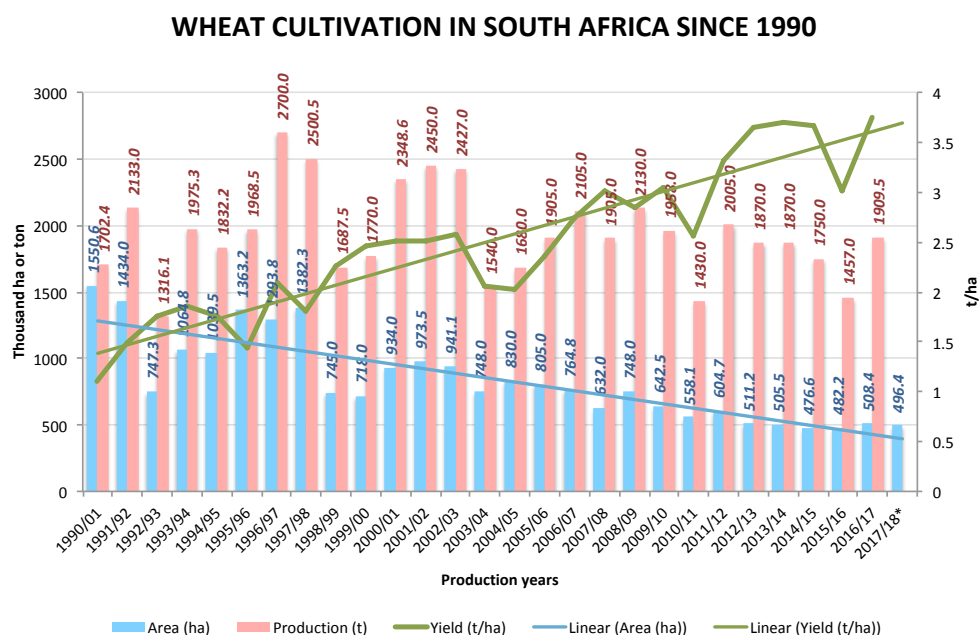


Figure 1.1: Area harvested, production and yield statistics for South African wheat cultivation between 1990 and 2017 (adapted from Production Reports - Crop Estimate Committee (CEC), GRAIN SA, 2017).

varieties. Together and in parallel with global efforts, local research has assisted South African wheat breeders by improving yields, bread making quality, and pest and disease resistance of South African wheat varieties (Smit et al., 2010).

1.3 Wheat rusts reduce yields

Some of the extra demand for wheat has been met by continuing genetic improvement, which leads to the development of high-yielding varieties, but the protection of crops from diseases remains critical to support the higher production requirements (Edgerton, 2009). The wheat rusts—leaf (brown) rust, stem (black) rust and stripe (yellow) rust—occur in most wheat-growing areas around the world and cause widespread disease which is detrimental for yields (Kolmer, 2005; Dean et al., 2012). Rust infection cripples all components of the host, whilst robbing the plant of water and nutrients (Panstruga and Dodds, 2009; Chen et al.,

2015). Rusts further reduce water content in the host through compromising the epidermis. This allows increased water evaporation and renders the plant an easy target for secondary attack by other pests and diseases (Bockus and Wiese, 2010; Malinovsky et al., 2014). Rust infection ultimately results in the death of photosynthetic tissues (Chen et al., 2015). Together, water and green tissue loss decrease the ability of the plant to trap solar energy through photosynthesis for growth and production of grain (Bockus and Wiese, 2010; Chen et al., 2015).

The occurrence of rust on wheat in South Africa was documented in reports dating back to 1726 (Du Plessis, 1933) and today all three rusts occur in South Africa (Pretorius et al., 2007). Pretorius et al. (2007) explain early record-keeping of rust occurrence in South Africa: Improved records became available as structured pathotyping of stem rust started in 1920, and in 1960 regular surveys were introduced. Leaf rust pathotypes were first described in 1937 but were not closely monitored until the 1980s after new pathotypes caused significant yield losses. In contrast, no early official disease reports could be found for stripe rust.

In 1996, however, it was seen on spring wheat in the Western Cape and surveys throughout the growing season revealed stripe rust infections in most of the winter-rainfall wheat cultivating areas. Irrigated wheat in the Northern Cape was also under attack (Pretorius et al., 1997). Mean yield losses attributed to wheat rusts in South Africa were estimated to be between 35 and 65 % (Pretorius et al., 2007). Given the global and local importance of wheat, the detrimental effects of rust pathogens, and the constant emergence of new pathotypes of the pathogen, researchers need to continually monitor the changing rust populations, while searching for new ways and sources of resistance to protect wheat (McIntosh et al., 1995).

1.4 Motivation for this study

The foliar disease stripe rust, caused by the biotrophic fungus *Puccinia striiformis* Westend. f. sp. *tritici* (*Pst*), results in major yield losses annually around the globe (Hovmøller et al., 2010). Growing resistant host varieties has reduced the impact of stripe rust (Hovmøller et al., 2016). However, knowledge of increased aggressiveness and shifts in *Pst* populations (Milus et al., 2009; Rodriguez-Algaba et al., 2014; Hubbard et al., 2015; Hovmøller et al., 2016; Bueno-Sancho et al., 2017) encourages investigation of this pathogen and how it is actively evolving in different geographical areas.

At the start of this project, little was known about the genetic diversity of stripe rust in South Africa. Previous work had genotyped South African *Pst* pathotypes using amplified fragment length polymorphism (AFLP) markers (Hovmøller et al., 2008) and more recently microsatellite markers (Ali et al., 2014; Visser et al., 2016). However, these limited marker systems do not provide a comprehensive genetic picture of the changes that can occur in a *Pst* population. The four pathotypes of *Pst* found in South Africa suggest a clonal lineage, which has evolved within South Africa since its original introduction in 1996 (Visser et al., 2016). In this study, next-generation sequencing and advanced bioinformatic tools were used to answer a number of questions regarding the origins of *Pst* and the evolution of the *Pst* population within South Africa. Through an examination of candidate effector genes, the study also aimed to facilitate a better understanding of the biological interaction between wheat and *Pst*.

1.5 Objectives

Stripe rust first appeared as a significant field disease in South Africa in 1996 (Pretorius et al., 1997). Since its introduction, four distinct pathotypes of *Pst* have

been detected and pathologically confirmed, the last being identified in 2005 (ZA Pretorius, unpublished data). This well-defined and presumed clonal population of *Pst* formed an ideal population to study the genetic evolution of *Pst* within a defined geographical region, addressing the hypothesis of a stepwise gain of virulence within the four South African pathotypes.

The availability of next-generation sequencing datasets of *Pst* isolates from locations in East Africa, South Asia and Europe also allowed a comparative approach to determine where the *Pst* introduction in 1996 may have originated. In addition, the genome sequences obtained for each of the four historical South African *Pst* isolates were used to identify candidate effector proteins that may be associated with avirulence. Lastly, a survey of *Pst* isolates in 2013, 2014, and 2015 within South Africa was undertaken to compare current field isolates to the four historical isolates to assess the stability of *Pst* populations across cropping seasons.

1.6 Thesis outline and approaches

Background information concerning *Pst* can be found in Chapter 2, while detailed methodology is described in Chapter 3 or the relevant research chapters. Five approaches were undertaken to characterise the *Pst* population in South Africa, which are presented across four research chapters. Firstly, the genomes of the four historical South African pathotypes were sequenced using Illumina next-generation sequencing. In Chapter 4, this data was analysed using phylogenetic and statistical clustering analyses to assess the relationship and genetic diversity between isolates and to hypothesise a potential origin of the *Pst* incursion in South Africa in 1996. To further describe the differences between the four South African pathotypes, comparative genomics analyses were performed, as presented in Chapter 5, by investigating signatures of positive selection, as well as the presence

or absence of genes and polymorphisms in genic regions. Chapter 6 reports an RT-qPCR approach that was used to assess candidate effectors showing differential gene expression between different *Pst* pathotypes. To compare the more recent field population of *Pst* with the historical South African isolates and to describe the evolutionary dynamics in the *Pst* population within South Africa, *Pst*-infected wheat leaf tissues from the 2013–2015 seasons were collected and sequenced using an RNA sequencing (RNA-Seq) approach. Pathotyping of a selection of the 2013–2015 field isolates was conducted to link their genotypes to their pathotypes and identify any isolates with profiles distinct to those previously identified in South Africa. This research on the recent population is discussed in Chapter 7. A final discussion of the findings and last remarks for future research conclude the thesis in Chapter 8.

Chapter 2

The Wheat Rusts: Life Histories, Host Response Mechanisms and Genomic Resources

2.1 The rusts

2.1.1 Filamentous plant pathogens

FILAMENTOUS PLANT PATHOGENS are highly specialised and include a wide variety of fungi and oomycetes (Wang et al., 2017). Ascomycota and Basidiomycota are both phyla in the fungi kingdom. The oomycetes include an array of plant pathogens that share many morphological characteristics with fungi, although being distantly related. A representation of the phylogenetic relationships between a number of plant pathogens are illustrated in Figure 2.1 (Fernández-Ortuño et al., 2007). Many of these pathogens have a similar infection process, using haustoria to maintain close interaction with the host (Dodds et al., 2009). There exist thousands of different rust species in the *Pucciniales* order, of which about 4000 species belong to the genus *Puccinia* (Hawksworth et al., 1995; Kirk et al., 2008).

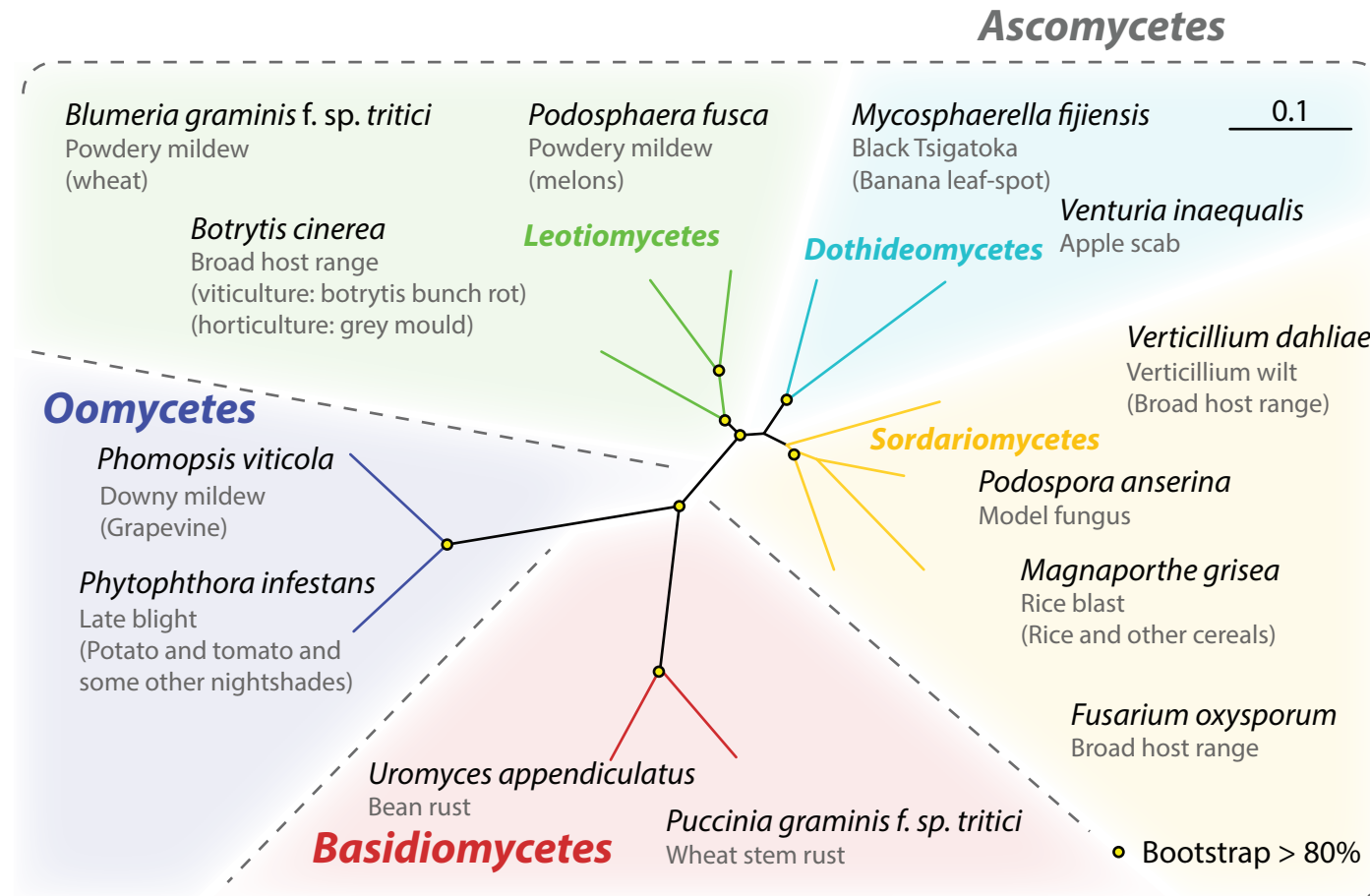


Figure 2.1: The phylogenetic relationship of plant pathogenic ascomycetes, basidiomycetes and oomycetes following Neighbour Joining analysis (adapted from Fernández-Ortuño et al., 2007). Bootstrap values are obtained from 1000 replications. The length of the bar represents 0.1 substitutions per nucleotide. The tree was constructed using nucleotide sequences of nuclear ribosomal DNA internal transcribed spacer regions.

2.1.2 Rusts and their primary host

Rusts are a group of fungi that are harmful to a wide variety of plants with high socio-economic importance such as cereals, legumes, fruit trees, sugarcane, coffee and trees (See taxonomic classification in Figure 2.2; Kirk et al., 2008).

Kingdom:	Fungi
Subkingdom:	Dikarya
Phylum:	Basidiomycota
Class:	Pucciniomycete
Order:	Pucciniales
Family:	Pucciniaceae
Genus:	<i>Puccinia</i>
Species:	<i>P. striiformis</i> <i>P. graminis</i> <i>P. triticina</i>

Figure 2.2: Taxonomic classification of the wheat rusts (Chen, 2005; Kirk et al., 2008).

Within *Puccinia* (*P.*) species, different *formae speciales* (f. sp.) describe specialisation towards specific grass hosts (Anikster, 1984; Wellings, 2007). To date nine f. sp. have been defined (Chen et al., 2017). The three wheat rusts that infect wheat are obligate biotrophs, requiring living plant tissues from which they extract water and nutrients (Dean et al., 2012). Stem rust, also known as black rust, occurs on the leaf and stem surface as oval-shaped brick-red pustules that burst through the host tissue and is caused by the fungus *P. graminis* Pers. f. sp. *tritici*, or *Pgt* (Schumann and Leonard, 2000). Leaf rust, also known as brown rust, caused by *P. triticina* Erikss. (*Pt*) is the most common of the three rusts and the orange to brown spores occur on the leaf surface in round lesions (Bolton et al., 2008). Stripe rust mainly forms yellow to orange lines as pustules occur along leaf veins of adult plants, but it can also infect other parts of the plant such as leaf sheaths, glumes and awns. Stripe rust of wheat, also known as yellow rust, is caused by *P. striiformis* Westend. f. sp. *tritici* (*Pst*; Roelfs and Hettel, 1992).

Each f. s. is further divided into races, strains, or pathotypes (Wellings, 2007), where the ability to infect the host plant depends on the avirulence genes carried by the *Pst* isolate and the resistance genes present in the host plant genotype (Chen, 2005). In the present study, the term “pathotype” is used throughout. To further describe the differences in different rust genotypes, a set of wheat lines with known resistances, is used in infection assays to determine the virulence profile of the isolate. These host plant genotypes form a differential set and the range of *Pst* infection phenotypes seen on each host plant genotype define the pathotype of the *Pst* isolate (Allison and Isenbeck, 1930; Roelfs et al., 1992).

2.1.3 The alternative host

Besides the grass hosts, the rust fungi can also infect a second group of hosts. *Pgt* has been known to infect alternative hosts *Berberis* L. (Jin et al., 2010; Zhao et al., 2011) and *Mahonia* Nutt. (Wang and Chen, 2013), while *Pt* infects *Thalictrum* spp. as alternative hosts (Bolton et al., 2008). Only recently has *Berberis* been confirmed as an alternative host for *Pst*. *Berberis* spp. are not native to South Africa, but are popular ornamentals, commonly stocked by nurseries and are becoming invasive in the wild (Keet, 2015). In South Africa, cultivation of 24 species of Berberidaceae, including 18 *Berberis* and 5 *Mahonia* have been reported (Glen, 2002). Among these are rust susceptible *Berberis holstii*, *Berberis vulgaris*, and *Berberis aristata* (Keet, 2015), but Jin (2011) advised that many more susceptible species could still be discovered. The sexual life cycle of rust fungi is completed in the alternative host (Chen, 2005). Infection of the alternative host has not been reported in South Africa. The rare occurrence thereof globally is fortunate, as it limits the potential for sexual recombination that can lead to faster evolving populations.

2.1.4 Global distribution of stripe rust

Stripe rust exists in most parts of the world where wheat is cultivated and continues to spread (Figure 2.3). In recent years epidemics of stripe rust have been seen in regions of the world where it did not previously occur (Chen, 2005; Milus et al., 2006). In contrast with the other rusts, distant dispersal of *Pst* has only recently been reported (Zadoks, 1961; Hovmøller et al., 2002; Justesen et al., 2002; Hovmøller and Justesen, 2007b; Wellings, 2011). There is evidence that new pathotypes of *Pst* are more aggressive and able to thrive at higher temperatures, showing the ability of this fungus to adapt to new environments (Milus et al., 2006; Markell and Milus, 2008). To date, aggressive pathotypes have not been described in South Africa.

2.1.5 Favourable conditions for wheat rusts

The occurrence of stripe rust on wheat is dependent on climatic and environmental conditions. Compared to leaf and stem rust, stripe rust has lower temperature optima, is prominent in cooler, high altitude and maritime regions and tends to occur earlier in the growing season (Chen, 2005). Stripe rust urediniospore germination is most successful between 9 °C to 13 °C, while stem rust's germination optimum is higher at 15 °C to 24 °C (Roelfs et al., 1992) and leaf rust, the most versatile and common, can infect the host in temperatures ranging between 10 °C to 25 °C (Bolton et al., 2008). Reports of adaptation to higher temperatures in newly emerging *Pst* populations in North America (Milus et al., 2006) show that higher temperatures, while suboptimal, is not insurmountable to *Pst*. Another study suggests that with sufficient light intensity, high temperatures are not necessarily inhibiting to *Pst* infection (de Vallavieille-Pope et al., 2002). However, Chen (2005) reports that temperatures below –10 °C can kill the pathogen in infected leaves.

Free moisture in the form of rain or dew for 3 to 6 hours is essential for germi-

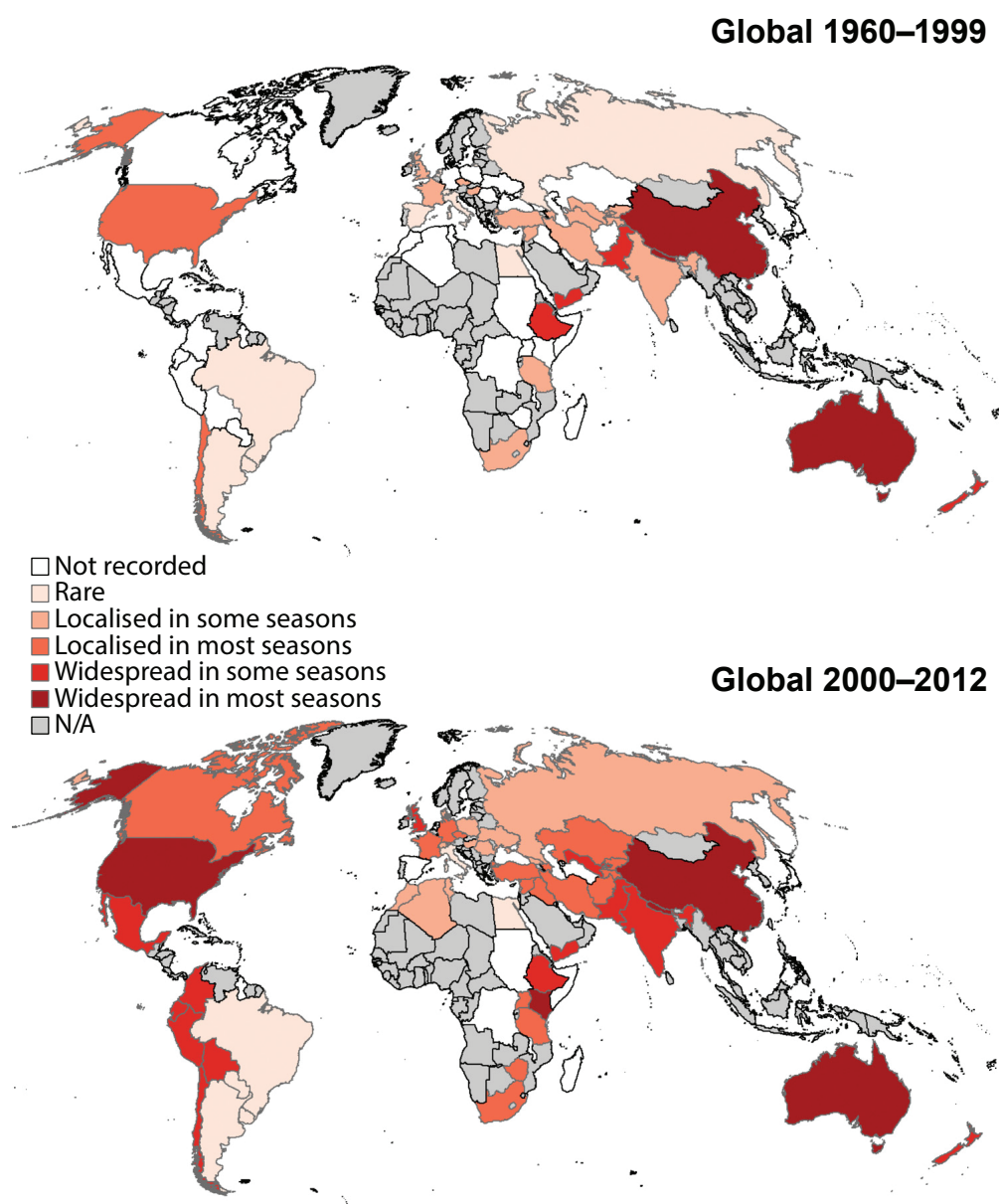


Figure 2.3: Global distribution of *Puccinia striiformis* f. sp. *tritici*, before and after 2000 (from Beddow et al., 2015).

nation of *Pst* urediniospores (Roelfs et al., 1992; Chen, 2005). On the contrary, dry weather and wind, towards the end of the growing season, are favourable for pathogen survival, as dry spores stay viable for longer and are wind dispersed (Zillinsky, 1983; Chen, 2005). Compared to moisture and temperature optima, little work has been done on optimal light requirements during the rust life cycle. There is some evidence that exposure of wheat seedlings to elevated light intensities before inoculation with urediniospores increases infection success (de Vallavieille-Pope et al., 2002). Conversely, compared to stem and leaf rust, *Pst* urediniospores are sensitive to ultraviolet light, and excess exposure reduces long-term viability (Roelfs et al., 1992).

2.1.6 Infection cycle of *Puccinia* rusts

The life cycles of the three wheat rusts are similar. In this section, the *Pst* life cycle is described. There are five spore stages in the life cycle of *Pst*. Three of these—urediniospores, teliospores and basidiospores—occur on wheat and the remaining two—pycniospores and aeciospores—on the alternative host. This is illustrated in Figure 2.4.

Very few cases of sexual reproduction have been reported, leaving the fungus to almost completely rely on asexual reproduction (Jin et al., 2010; Zhao et al., 2013; Chen et al., 2017). In areas where the sexual cycle takes place, aeciospores are formed after infection of the alternative host (Chen, 2005). These spores can infect wheat and result in pustules releasing urediniospores for reinfection. In the majority of regions, where the grass host is the main or only host, only urediniospores are available for host infection. Characteristic of this spore stage, each spore carries two haploid nuclei. About two weeks after urediniospores landed on a leaf and entered the leaf through the stoma, the newly produced, yellow urediniospores erupt through the surface of the leaf (Figure 2.5).

The urediniospores are dispersed by wind, or the mechanical action resulting

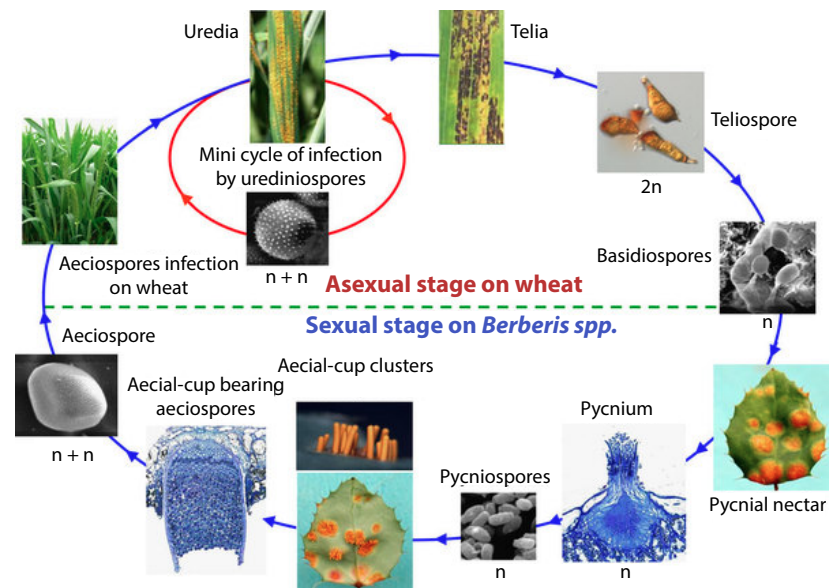


Figure 2.4: Spore stages and the infection cycle of *Pst*. The mini cycle of (re)infection, indicated with red arrows, is the primary source of inoculum for most stripe rust outbreaks in wheat-growing areas worldwide. Only recently, the sexual cycle, indicated with blue arrows, have been observed under natural conditions in China (from Zheng et al., 2013).



Figure 2.5: A stripe rust uredinium pustule. Thousands of yellow spherical echinulated spores, typically 28–34 μm in diameter (Zillinsky, 1983), erupts through the wheat leaf surface (Photo: Kim Findley, John Innes Center, UK).

from raindrops falling onto leaves (Chen, 2005). This phase of *Pst* development constitutes the asexual cycle. This cycle typically takes 12 to 14 days depending on the isolate and environmental conditions (Chen et al., 2014), but Australian studies confirmed a shorter life cycle in aggressive *Pst* pathotypes (Sharma, 2012). The number of infection cycles the pathogen complete in a season determines the severity of the epidemic (de Vallavieille-Pope et al., 2012).

Urediniospores can over summer on voluntary wheat plants and other susceptible grasses. Examples include the wild rye species, *Secale L. strictum* subsp. *africanum*, seen in South Africa (Pretorius et al., 2007, 2015). Alternatively, towards the end of the wheat-growing season, as the wheat plant undergoes senescence, infection sites from some *Pst* isolates can form telia (Chen et al., 2014). The subepidermal telia are present on both sides of the leaf blade and produce dark brown, two-celled, oblong-clavate teliospores (Zillinsky, 1983; Chen, 2005; Chen et al., 2014). Through karyogamy, the nuclei in each of the two cells of the teliospore fuse, resulting in two diploid cells. The diploid nucleus in each cell undergoes meiosis, and the two cells grow into a promycelium of four cells. This develops into a basidium consisting of four cells, each of which releases a haploid basidiospore. These basidiospores can infect an alternative host, initiating the sexual cycle (Chen et al., 2014).

The haploid basidiospores infect the alternative host and forms either pycnia (female) or spermatia (male) on the adaxial side of the leaf. These spore-producing structures contain haploid reproductive structures. Rusts are heterothallic, and spermatia produce pycniospores (the male gametes), which are transferred to pycnia to fertilise receptive hyphae, the female gamete (Rapilly, 1979). Dispersal of pycniospores can be facilitated by precipitation running down the leaf, while the pycnia also produce nectar. It has been described in stem and leaf rust that visiting insects that come into contact with the nectar can act as vectors to spread the spermatia to other pycnia (Leonard and Szabo, 2005; Bolton

et al., 2008). After fertilisation, plasmogamy of compatible mating types develops into a dikaryotic primordium, which matures into an aecium on the abaxial side of the alternative host leaf. The aecium produces dikaryotic aeciospores that can only infect the primary host (wheat), forming an urediospore-producing uredium—the starting material for the roughly 14 day asexual cycle that continues on wheat throughout the growing season (Chen et al., 2014).

Currently, two factors are considered responsible for the rare occurrence of sexual recombination. Firstly, contrasting to other rusts of wheat, teliospores do not enter a dormant phase and readily germinate under prolonged dew conditions (Chen et al., 2014). The time frame in which viable teliospores exist is thus short. Secondly, germination of teliospores requires very specific environmental conditions. The rare occurrence of alternative host infection by *Pst* testifies to the fact that spore availability and lengthy periods of dew formation do not often coincide. Such a natural occurrence has only been recorded twice, both times in China (Zhao et al., 2011, 2013).

Although infection of the alternative host remains rare, these observations explain the increased *Pst* population variation found in the Himalayan region, compared to other regions (Ali et al., 2014). Barberry is also common in these areas, further supporting the hypothesis of genetic recombination through sexual reproduction in the Himalaya region (Ali et al., 2014). Additional evidence based on AFLP and microsatellite markers illustrates the need for further investigation, determining the importance of the sexual stage in *Pst* for the generation of genetic variability (Mboup et al., 2009; Duan et al., 2010; Zheng et al., 2013). Fortunately, in South Africa and most other wheat-growing areas where stripe rust occurs, mutation and somatic hybridisation are believed to be the major sources of variation, theoretically supporting slower evolution. However, in the absence of the sexual cycle, somatic recombination can still contribute to variation leading to the formation of new pathotypes, as described by Lei et al. (2017).

2.1.7 The stripe rust infection process on wheat

Wheat, as the primary host of *Pst*, provides water and photosynthates for urediniospore production, maintaining the dominant asexual stage (Chen et al., 2014). Throughout the wheat-growing season, it repetitively infects the crop while cycling through clonal reproduction (Figure 2.6). *Pst*, as an obligate biotroph, needs to maintain the integrity of the plant cells during this infection process. Resources, predestined for plant growth and grain development, are diverted by the fungus for hyphal growth and spore production. In resistant wheat varieties, the evoking of a cellular hypersensitive response causes necrosis and chlorosis, stopping pathogen development but further compromising the plant's ability to photosynthesise (Chen, 2005).

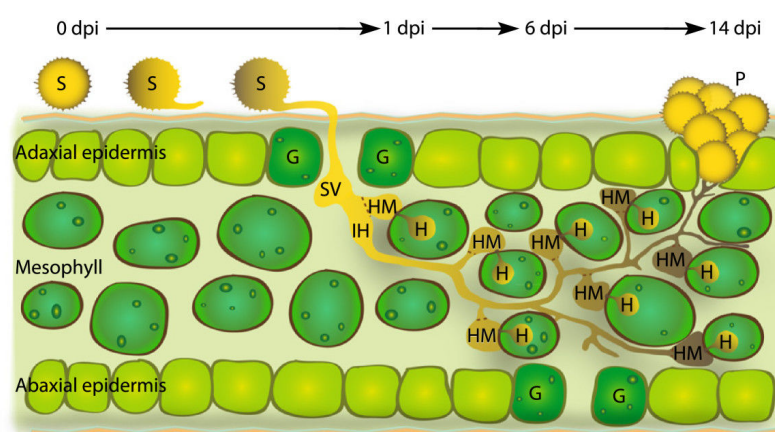


Figure 2.6: Illustration of the infection process of *Pst* (from Cantu et al., 2013). dpi, days post inoculation; S, urediniospore; SV, substomatal vesicle; IH, invasive hyphae; HM, haustorial mother cell; H, haustorium; P, pustule; G, guard cell.

With sufficient moisture on the leaf surface for the urediniospore to germinate, the germ-tube grows across the leaf surface in search of a stoma through which it enters the plant. Unlike *Pgt* and *Pt*, *Pst* does not produce a visible appressorium (Niks, 1989). A substomatal vesicle (SSV) forms within the substomatal cavity from which up to four infection hyphae (IH) develop (Figure 2.6). When an IH

reaches a mesophyll cell, the tip of the IH differentiates a haustorial mother cell (HMC). An infection peg (IP) forms at the tip of the HMC that breaches the cell wall of the plant mesophyll cell (Figure 2.7).

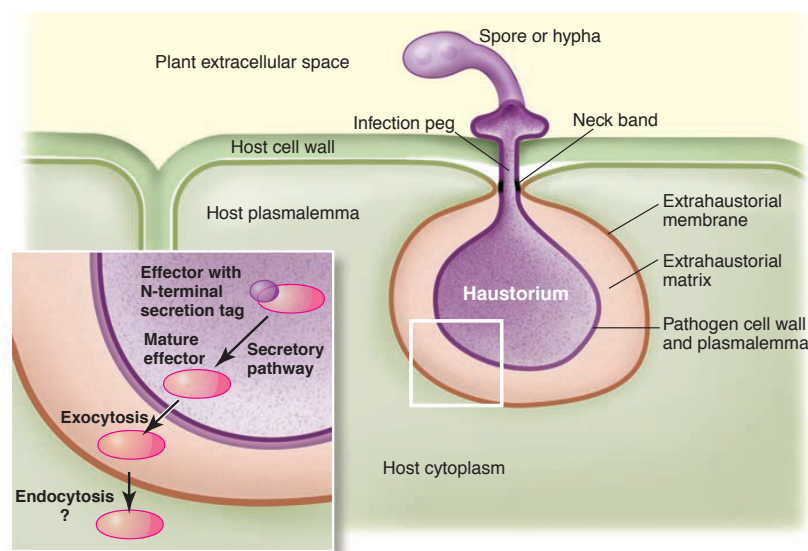


Figure 2.7: Illustration of a filamentous plant pathogen haustorium. Three membranes and the extra haustorial matrix separate the host cytoplasm and the pathogen's haustorium content. The pathogen's cell wall and plasma membrane is situated on the haustorium side. The modified host plasma membrane and neck band seals off the haustorial matrix from the host cytoplasm. Effector delivery is illustrated by the inset (from Panstruga and Dodds, 2009).

Some fungi use mechanical force aided by the turgor of the cell to breach the cell wall, for example in *Magnaporthe oryzae* (Hebert) Barr, or enzymes as in the case of *Pgt* (Duplessis et al., 2011), or a combination, as used by powdery mildew (Pryce-Jones et al., 1999). A different set of enzymes has been found in *Pgt* and other fungi, that likely plays a role in disguising the penetrating hyphae by remodelling of the fungal cell wall (El Gueddari et al., 2002). However, it is currently unknown how the *Pst* IP achieve cell wall penetration (Panstruga and Dodds, 2009).

Having breached the plant cell wall, *Pst* needs to establish a compatible association with the cell, keeping it alive while feeding. From the end of the

IP a haustorium develops that invaginates the plant cell membrane, causing the plant cell membrane to envelope the haustorium (Figure 2.7; Panstruga and Dodds, 2009). Three layers separate the content of the haustorium from the cytosol of the plant cell: the haustorial plasmalemma, the haustorial wall and the extrahaustorial membrane. The haustorial membrane and wall are surrounded by a gel-like layer, called the extrahaustorial matrix (Panstruga and Dodds, 2009). The extrahaustorial membrane is likely derived from the plant cell plasma membrane and is in contact with the cytoplasm of the plant cell (Szabo and Bushnell, 2001).

Due to the biotrophic nature of cereal rust pathogens, it is mostly impossible to culture the fungus artificially. As the multi-layered haustorium cannot be grown *in vitro* (Panstruga and Dodds, 2009), the exact mechanisms of how transport across the membranes is facilitated are currently not confirmed. The haustorium has a dual function, allowing two-way traffic across the membranes (Mendgen et al., 2000). It acts as a feeding structure to take up amino acids and sugars from the host (Panstruga and Dodds, 2009), while at the same time delivering fungal molecules to the plant that enable pathogenicity (Mendgen et al., 2000). Among these are effector proteins that are delivered into the host cytosol and the apoplast, altering plant processes to the advantage of the pathogen, while protecting itself against the host defence systems (Kamoun, 2007; Rovenich et al., 2014; Petre et al., 2016a). Once infection is established, long hyphae branch lengthwise within the leaf, colonising a large area and causing the typical striped pattern of uredinia seen on older plant leaves (Moldenhauer et al., 2006).

2.2 Combating wheat stripe rust

Agronomic management of stripe rust involves both the deployment of host resistance and the application of fungicides. Multiple fungicide applications are often

required during the wheat-growing season, being costly, potentially problematic to the environment, and not always 100 % effective, whereas the right combination of resistance genes can provide complete stripe rust resistance (Boshoff et al., 2003). Despite treatment, significant losses have been recorded (Oerke and Dehne, 2004). Even with resistance breeding and chemical crop protection, yield losses of 14 % to 40 % have been reported (Flood, 2010). Increased success in protection of foliage and ears can, however, be achieved when fungicide application is timed correctly (Boshoff et al., 2003).

Genetic resistance in the host causes selection pressure on the pathogen to overcome that resistance. Strategies to relieve selection pressure include the rotational deployment of resistance genes, regional gene deployment and pyramiding of resistance genes (Chen et al., 2017). Quantitative, polygenic resistance is considered a better choice due to its potential durability and will be discussed later in this chapter.

2.3 Plant defence mechanisms

Plants have passive and active defence mechanisms to protect them from biotic stresses. A compatible interaction between a pathogen and its host is one where the pathogen successfully infects and colonises the host. However, incompatible interactions exist between some combinations of *Pst* pathotypes and wheat genotypes. Different mechanisms contribute to the host being able to withstand a pathogen attack.

Passively, preformed defence mechanisms include the composition of the waxy layers, the cuticle being the first structural barrier to pathogen invasion. Further passive defence is put in place by pre-formed antimicrobial proteins and secondary metabolites, including photoanticipins, inhibitors of essential pathogen enzymatic activities, hydrolytic enzymes, lectins, and defensins (Selitrennikoff,

2001; Egorov et al., 2005; Coram et al., 2008). Passive defence is not pathogen specific, in contrast with many active defence mechanisms that are induced by the presence of the pathogen, and can be either specific or non-specific. Both physical and chemical changes are seen. These include the deposition of callose, cell wall cross-linking and the formation of papillae, changes in membrane permeability, production of reactive oxygen species (ROS), and the synthesis of a whole range of pathogen-related (PR) proteins and secondary metabolites, such as phytoalexins (Malinovsky et al., 2014).

2.3.1 Host-pathogen interaction

The active plant defence mechanisms require very specific pathogen recognition. For biotrophic pathogens, the current model of host-pathogen interactions involves an initial general recognition of a potential pathogen, triggered by plant recognition of conserved pathogen molecular motifs. The conserved pathogen molecular motifs are referred to as pathogen-associated molecular patterns (PAMPs) or microbe-associated molecular patterns (MAMPs), as described by van der Hoorn and Kamoun (2008). These motifs are recognised by transmembrane pathogen receptor proteins (PRRs).

Recognition of pathogen-associated patterns triggers defence responses, which are collectively known as PAMP triggered immunity (PTI; Jones and Dangl, 2006). Many pathogens are able to suppress the defence responses mounted by PTI, leading to successful infection. Proteins, secreted by the pathogen into the plant, facilitate the down-regulation of PTI defence. These proteins, generally considered to be small peptides able to cross membranes, are referred to as effectors (Franceschetti et al., 2017). In addition to down-regulating host defence responses, effectors also have an active role to play in pathogenicity, modifying the plant cellular and molecular environment in such a way that it eventually supports

pathogen growth and reproduction (Rovenich et al., 2014).

The second stage of the host-pathogen interaction involves specific recognition by the plant of specific pathogen effector molecules (Jones and Dangl, 2006). This second layer of defence is referred to as effector-triggered immunity (ETI). This involves recognition of specific pathogen effectors, now termed an avirulence (*Avr*) factor, by a receptor protein in the plant termed an *R* gene (Dangl and Jones, 2001; van der Hoorn and Kamoun, 2008). Alternative models of indirect associations, referred to as the guard and decoy models, have been described (van der Hoorn and Kamoun, 2008). The direct relationship between *R* genes and their corresponding *Avr* genes is known as the gene-for-gene concept described by Flor (1956). This specific plant-isolate recognition enables the plant to trigger a stronger defence response that restricts pathogen growth and reproduction, with the strength of resistance differing with each *R* gene/*Avr* combination. ETI is a specific host-pathogen interaction, depending on the presence of the *R* gene in the plant genotype and the presence of the corresponding avirulence factor in the pathogen isolate. The *R* gene/*Avr* interaction usually results in death of the infected plant cell (and possibly also surrounding plant cells) in a reaction known as hypersensitive cell death (HCD; Jones and Dangl, 2006).

The pathogen can evade *R* gene recognition by selection of mutations within the avirulence effector factor that break the *R* gene/*Avr* interaction (Dodds and Rathjen, 2010). When this occurs, the avirulence factor is subsequently referred to as a virulence factor. The continuing cycle of the pathotype-specific *R* gene/*Avr* interaction breakdown is known in wheat disease breeding as the Boom-and-Bust cycle (Knott, 1989; McDonald, 2004) and is one of the reasons why wheat breeders are interested in characterising and using rust resistance genes that do not fit the *R* gene/*Avr* model (see Section 2.3.2).

Five classes of proteins encoded by plant *R* genes have been modelled, and are illustrated in Figure 2.8. The biggest class characteristically encode for nucleotide-

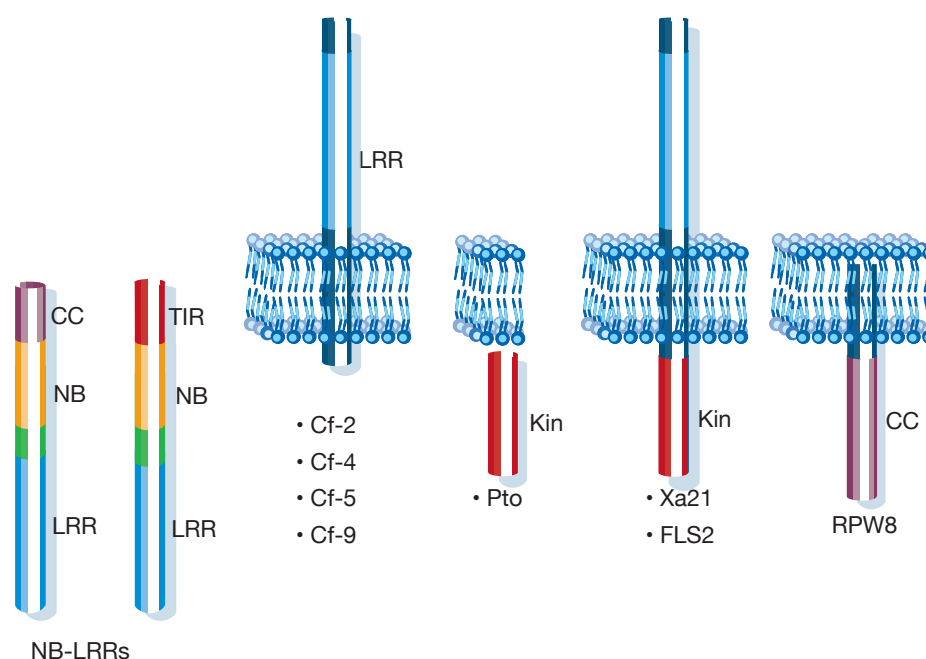


Figure 2.8: The five main classes of plant disease resistant proteins (from Dangl and Jones, 2001). Cytoplasmic nucleotide-binding site leucine-rich repeat proteins are typically not membrane-associated and represent the largest class of resistance proteins. Cf-X and Xa21 typically carry a large transmembrane leucine-rich repeat region. The serine/threonine protein kinase is encoded by the *Pto* gene, with possible membrane association through the N-terminal myristoylation site. A putative N-terminal signal anchor is carried by the RPW8 gene product. CC, coiled-coil domains; NB, nucleotide-binding site; LRR, leucine-rich repeat; TIR, Toll and Interleukin-1 receptor type region; Kin, kinase

binding site (NBS)-leucine-rich repeat (LRR) proteins (NB-LRR; Kolmer, 2005). NB-LRRs are thought to be cytoplasmic and in contrast with the other four classes of R proteins, Xa21 and Cf-X proteins contain transmembrane and extracellular LRR domains, while the *Pto* gene product is membrane-associated with a cytoplasmic kinase. The RPW8 protein has a putative signal anchor at the N-terminus (Dangl and Jones, 2001).

The expression of *R* gene resistance is usually qualitative and expressed at all wheat growth stages (Dangl and Jones, 2001). The profile of *R* gene/*Avr* interactions, tested on a set of wheat lines with known resistance, defines the

pathotype of any given *Pst* isolate. “Yr”, followed by a number, designate genes that confer resistance to stripe rust (McIntosh, 1983).

2.3.2 Other sources of resistance

Other forms of stripe rust resistance have been characterised that are not pathotype-specific. These forms of resistance have remained effective to all *Pst* isolates tested and therefore are termed pathotype-non-specific resistance (Van der Plank, 1968). These forms of resistance are usually quantitative, being partial in effect, expressed more strongly in mature wheat tissues and is therefore also termed adult plant resistance (Simmonds, 1991; Parlevliet, 2002; Mallard et al., 2005). It can further reduce the rate of disease progress, called slow-rusting, partial, or horizontal resistance (Van der Plank, 1968).

2.4 The *Pst* genome

2.4.1 Genomic variation

When point mutations occur in genes, it can change an amino acid which in turn can change the functionality and stability of the protein. If this has a notable impact on the phenotype, it will change the way in which the organism interacts with its environment. Such a change will be under selection to either eliminate it from the population or increase the frequency, depending on the impact of the change on the reproductive ability of individuals with the particular polymorphism.

The variation in *Pst Avr* genes has been evaluated for many years. Pathotype (race) profiling is widely deployed and extremely informative. It has been practised for about 100 years (Thach et al., 2015) and changes in pathotype profiles mostly support a clonal lineage for *Pst*. The ability to genotype isolates to support

pathotypes was a major addition to the development of rust population studies. In the last 30 years the development of molecular markers, which have been used to track global movement, has supported the hypotheses of these clonal population structures. Deployment of molecular marker technologies for *Pst* genotyping have included AFLP markers (Steele et al., 2001; Brown and Hovmøller, 2002; Hovmøller et al., 2008; Mboup et al., 2009), and more recently microsatellite markers (Mboup et al., 2009; Ali et al., 2014; Visser et al., 2016; Walter et al., 2016) and sequence-characterised amplified region (SCAR) markers (Walter et al., 2016) were implemented.

The so-called “genomics era” provides an even higher resolution view of the diversity within and between *Pst* populations. The development of high throughput sequencing techniques provides the opportunity to answer many more questions about the ongoing evolutionary processes in *Pst* and just how far this airborne pathogen can travel.

2.4.2 Rust genomics

Stem rust was the first of the wheat rusts to be sequenced, followed by leaf and stripe rust. The Fungal Genome Initiative at the Broad Institute of Massachusetts Institute of Technology and Harvard University was instrumental in sequencing all three wheat rusts. Genomic research in *Pst* saw fast development as the international community has published a high number of *Pst* next-generation sequencing datasets.

Some of these resources have been applied specifically to develop representative draft reference sequences of *Pst* pathotypes from distinct pathotypes and geographical areas. These are summarised in Table 2.1 and include the North American isolates, PST130 (Cantu et al., 2011) and PST-78 (Cuomo et al., 2017), the Chinese isolate, CY32 (Zheng et al., 2013) and the Indian isolates 46S 119 (Kiran

et al., 2017) and 38S102 (Aggarwal et al., unpublished). The Australian founder pathotype, *Pst* 104E137A-, has recently been assembled using a combination of next-generation Illumina sequencing and third generation sequencing, alternatively termed long read sequencing, on the PacBio platform (Schwessinger et al., 2018). Deployment of such advances in sequencing technology enables comparison of the dikaryotic nuclei in *Pst* to investigate the evolutionary machinery used to drive the development of new *Pst* variation.

Puccinia graminis* f. sp. *tritici

The first rust reference genome and, to date, only *Pgt* reference, was sequenced from the pathotype CRL 75-36-700-3 (Duplessis et al., 2011). The project was led by the Szabo group at the USDA-ARS Cereal Disease Laboratory, University of Minnesota, USA. In 2007 the first 7.88× draft of the genome sequence assembly was released. It was updated in 2010 with a mitochondrial assembly and its accompanying annotation data, and finally in 2011 with an RNA-Seq based annotation. In addition to sequencing the genome, the shotgun fosmid library was used to prepare a physical fingerprint map. To investigate gene expression at various stages of *Pgt* development, complementary DNA (cDNA) libraries were constructed for such tissues. The estimated genome size of *Pgt* is 80 mega base pairs (Mbp). The outbreak of the highly virulent *Pgt* pathotype, Ug99, prompted this research, resulting in the development of many useful markers for pathotype-diagnostic tests since (Godfrey et al., 2010).

Puccinia triticina

Genome sequencing of the *Pt* isolate 1-1 was done using Fosmid-end and bacterial artificial chromosome end (BAC-end) libraries and a hybrid of 454 and Applied Biosystems Integrated (ABI) sequencing technologies, also known as Sanger

Table 2.1: Whole genome sequencing projects using next- and third-generation sequencing. Genomes that were proposed as reference sequences are listed exclusively. Various methodologies have been used for library construction, sequencing and assembly, with varying results. These assemblies are invaluable tools that can be used to reveal genome characteristics of the three wheat rusts (adapted from Kang, 2017 including Cantu et al., 2011, 2013; Cuomo et al., 2017; Schwessinger et al., 2018)

Wheat rust pathogen	Isolate	Genome Size (Mbp)	Protein coding genes	Secreted proteins	No. of contigs* or scaffolds	% TE	Sequencing technology
<i>P. striiformis</i>	PST130	Φ 64.8	18 149	1 088	*22 815	Δ 17.8	Illumina Genome Analyzer II sequencing
<i>P. striiformis</i>	CYR32	110.0	25 288	2 092	12 833	48.9	Fosmid-to-fosmid strategy by Illumina GA paired-end sequencing
<i>P. striiformis</i>	PST-78	117.3	19 542	2 146	9 716	31.5	Roche 454 FLX and Illumina fosmid-end sequencing
<i>P. striiformis</i>	38S102	75.6	–	–	996	–	Illumina NextSeq 500
<i>P. striiformis</i>	Pst-104E	79.8	15 303	–	996	53.7	PacBio RSII
<i>P. triticina</i>	1-1	135.3	14 880	1 358	14 820	50.9	Roche 454 FLX and Sanger fosmid-end and BAC-end sequencing
<i>P. graminis</i>	CRL 75-36-700-3	88.6	15 800	1 106	392	36.5	Sanger sequencing whole-genome shotgun strategy

Φ , 60 % of genome; TE, Transposable and repetitive elements; Δ , Only transposable elements; *, indicate number of contigs if present, otherwise number of scaffolds; –, not available; BAC, bacterial artificial chromosome

sequencing (Cuomo et al., 2017). Considerable advances in characterising *Pt* genes and genomic variation was enabled through the assemblies of two more genomes—the virulent pathotype, Race77, and an older avirulent pathotype, Race106 (Kiran et al., 2016).

Puccinia striiformis* f. sp. *tritici

A number of draft sequences are now available for *Pst*. The PST130 isolate was first identified in Oregon and Washington, USA, in 2007 (Chen et al., 2010). The isolate was chosen to be sequenced for technical reasons and not because it was biologically specifically interesting. Subsequent to genome assembly the PST130 genome has been continually investigated in the research group of Dr Diane Saunders (JIC, UK). PST130 was used as reference genome in the present study as the candidates association with this research group allowed building on and making direct comparisons with previous work in the group.

CYR32 was sequenced as it was a highly prominent pathotype in China. This work confirmed and further emphasised previous reports of high heterozygosity between the two nuclei as a fosmid-to-fosmid sequencing strategy was applied (Zheng et al., 2013). PST-78 was chosen to represent the *Pst* pathotypes virulent to *Yr8* and *Yr9* that were first identified in 2000 (Cuomo et al., 2017). The isolate was collected from the US Great Plains. Incorporating many sequencing platforms, this multi-approach resulted in a high quality genome. Gene annotation was done using transcriptome sequence data and *de novo* gene prediction (Cuomo et al., 2017). The initial approximately 81 × cover assembly of PST-78 was released in 2012, with the RNA-Seq-based annotation containing 19 542 genes. The first genome from an Indian *Pst* isolate was published in 2017 (Kiran et al., 2017). The pathotype 46S 119 has virulence to *Yr9* and emerged and recently spread into the north-western plains of India. The 38S102 pathotype was first isolated from the

Neelgiri Hills in India in 1973 and also has avirulence to *Yr9* (Aggarwal et al., unpublished). These isolates are interesting as many wheat varieties in the north-west of India are protected by the *Yr9* resistance gene (Kiran et al., 2017). The long read assembly of the Australian pathotype, *Pst* 104E137A- (Schwessinger et al., 2018), refined earlier conclusions on genetic diversity that were drawn from short read assessments.

2.4.3 Challenges in bioinformatics

All rust genome sequencing projects have used urediniospores, the major spore stage on wheat. The two nuclei of the dikaryotic urediniospore have been shown to be highly heterozygous (Zheng et al., 2013). A large portion of all genomes was repetitive content and transposable elements. The PST130 genome reference, with 18 % transposable elements, was estimated to include only about 60 % of the genome, although assembly of 95 % of the reads was possible. Highly similar repetitive sequences would be assembled in common contigs, and it was estimated that repetitive content that was misassembled could add an additional 10.6 Mbp to the genome size (Cantu et al., 2011). These repetitive sequences and high density of transposable elements impede the principles assemblers use to reconstruct a genome (Duplessis et al., 2011; Castanera et al., 2016).

Haplotype-phased genomes address this problem to some extent. The first phased *Pst* sequencing effort, (Schwessinger et al., 2018), using long-read DNA sequencing technology, demonstrated the nucleotide and structural differences between the two haploid nuclei. It is expected that single consensus sequences, as generated for all former *Pst* genome sequencing experiments, would be suboptimal in their description of genome diversity and structure.

2.4.4 Effector identification

After assembly and gene annotation, the focus for plant pathogen research is shifted to effector coding gene identification. Investigation of effector proteins is crucial as these proteins are utilised by pathogens to alter biological and metabolic processes in the host (Kamoun, 2007). Resources developed by earlier studies, as the development of cDNA and expressed sequence tag (EST) libraries (Ling et al., 2007; Zhang et al., 2008), and existing knowledge of known effector characteristics of other pathogens, provide resources for the development of bioinformatic pipelines. Using computational methods and gene discovery algorithms, these pipelines facilitate rapid effector gene identification. High throughput sequencing technologies and bioinformatics further relief the challenges of studying effectors of obligate biotrophs by providing a platform to investigate complete transcripts (Joly et al., 2010; Hacquard et al., 2011; Saunders et al., 2012).

Highly conserved motifs have been useful in identifying effector families, such as the RXLR and LXFLAK motifs in oomycetes (Bozkurt et al., 2012). For *Pgt* the [YFW]xC motif has been identified by Godfrey et al. (2010). However, the characteristic of many of the rusts to rarely display conserved motifs known from other plant pathogens makes effector prediction challenging (Hacquard et al., 2011; Saunders et al., 2012; Lorrain et al., 2015). This constraint stresses the need for functional validation that remains a limiting factor due to the relatively low throughput of validation systems that can confirm the pathogen effector targets in the host (Petre et al., 2016a).

Only a few such targets have been identified in hosts of filamentous plant pathogens, among which the dothideomycete (Figure 2.1) *Cladosporium fulvum* Cooke causing tomato leaf mold, the rice blast fungus *Magnaporthe oryzae*, the potato blight fungus *Phytophthora infestans* (Mont.) de Bary and *Ustilago maydis* from the class Ustilaginomycetes, causing corn smut (Rovenich et al., 2014). For

Blumeria graminis (DC.) Speer f. sp. *hordei*, the causal agent of powdery mildew in barley, an ARF-GAP target protein was identified in the host (Rovenich et al., 2014). Adenosine diphosphate (ADP) ribosylation factors (ARF) are important for vesicle trafficking, while its activity is regulated by Guanosine triphosphatase (GTPase) activating proteins (GAPs). The pathogen targets this protein complex to interfere with the host's trafficking of vesicles containing biochemical molecules (Mandiyan et al., 1999). Association of pathogen genes with vesicle trafficking in the host has also been proposed in *Pst*-wheat interaction using RNA-Seq (Dobon et al., 2016).

Genomic resources enabled the use of yeast-two hybrid screens to identify associations between *Pst* and wheat proteins (Lowe et al., 2011). Non-host model plants were further proposed to characterise effector candidates, specifically *Nicotiana benthamiana* Domin, as rust fungi hosts are difficult to manipulate with molecular genetic techniques (Petre et al., 2015). This approach has been instrumental in functional characterisation of a number of *Pst* effectors (Petre et al., 2016a). The authors warn that although the leaf cell environment of *N. benthamiana* is advantageous for protein interaction screens, compared to expression in yeast, false negatives are common due to differences between *N. benthamiana* and the host species (Petre et al., 2016b). A combination of the two approaches can be followed (Liu et al., 2016). Other examples of functional validation include transient expression assays and host-induced gene silencing (HIGS) using RNA interference (Yin and Hulbert, 2015; Liu et al., 2016).

Recent successes in rust effector identification were achieved with the cloning of the two stem rust effectors, *AvrSr35* (Salcedo et al., 2017) and *AvrSr50* (Chen et al., 2017). Variation in *AvrSr35* and loss of heterozygosity in *AvrSr50* resulted in the respective inability of *Sr35* and *Sr50* to recognise specific isolates of the stem rust fungus, resulting in disease. The methodology that was implemented could be transferable to other rust effector searches and is therefore noteworthy.

Candidates were obtained from comparative transcriptomic analysis between wild type and mutant *Pgt* isolates. Validation of candidates included a whole host of techniques including microscopy, transient expression in *N. benthamiana* and *N. tabacum* and yeast-two-hybrid analyses. Transient expression in wheat made use of transforming constructs into *Escherichia coli* (Migula) Castellani and Chalmers and *Agrobacterium tumefaciens* (Smith and Townsend) Conn. strains. Virus-mediated effector expression assays were also performed in wheat using the barley stripe mosaic virus (Lee et al., 2012).

The present study is based on advances in *Pst* bioinformatics regarding *Pst* next-generation sequencing and gene and effector annotations. Annotation procedures considered knowledge of the life history, molecular mechanisms, and complementing computational biology resources in *Pst* and related filamentous plant pathogens. Together, these techniques enabled the identification of genes likely involved in distinct virulence profiles of South African *Pst* pathotypes. Additional functional validation methods discussed in this review would add value in future studies to further investigate the identified candidate effector genes. Furthermore, genomic and transcriptomic *Pst* resources allowed predictions to be made regarding the relatedness of different *Pst* isolates to one another, based on genetic proximity when single nucleotide polymorphisms (SNPs) were evaluated in population analyses. This provided valuable insights into the global prevalence of specific genetic groups to better understand their potential movement and the risks it may involve.

Chapter 3

General Materials and Methods

3.1 Preparation and collection of materials

3.1.1 Inoculation

THE FOLLOWING STANDARD *Pst* inoculation protocol, developed and performed at the University of the Free State (UFS), South Africa, was performed to obtain urediniospores for genomic DNA (gDNA) extraction used for next-generation sequencing (NGS) and total RNA extraction of infected tissue used for analyses of gene expression through RT-qPCR.

For multiplication of urediniospores for sequencing purposes (Chapter 4), as well as the time course (Chapter 6), and infection assays (Chapter 7), the wheat variety Morocco was used as a susceptible host. The time course itself was performed on Avocet S (susceptible), and the infection assay varieties are listed in Chapter 7. Seedlings were grown for seven days until two unfolded leaves developed (Zadoks growth stage 12 (Z12); Zadoks et al., 1974). For initial multiplication, urediniospores, previously dried on silica gel and stored at -80°C were suspended in Soltrol® 130 Isoparaffinic Solvent oil (Chevron Phillips Chemical Company, USA), at 5 mg/ml, upon retrieval from the freezer. Several rounds of multiplication were performed for the sequencing experiment (see Chapter 4).

Inoculations of the time course and the infection assays were done with fresh spores harvested from initial multiplication. Seedlings of seven-day-old wheat (Z12), grown in Mikskaar Professional Potting Soil 70 (Mikskaar, Estonia) in 10 cm diameter plastic pots, were lightly sprayed with the spore-oil suspension. Inoculated plants were dried in a growth cabinet at 25 °C for about 45 minutes. Custom-made incubation chambers (755 × 500 × 300 mm) made from galvanised metal sheeting, with a 30 mm raised grid at the bottom, were filled with hot tap water to just below the grid level. Seedlings were then placed on the grid, and the chambers were immediately sealed to capture maximum water vapour and maintain saturated conditions. The chambers were housed in a cold room at 11 °C, after which plants were incubated for 24 hours at 11 °C, in total darkness. These conditions simulate high atmospheric moisture levels and low temperatures resulting in dew formation, usually during night time, in natural conditions. Next, inoculated plants were transferred to a growth chamber at 17 °C for 1.5 days, with a 14 hour day and 10 hour night cycle. Daylight was simulated with a light intensity of 200 $\mu\text{mol}/(\text{m}^2 \text{ s})$. Plants were then moved to a glasshouse with natural light and a day-night temperature cycle set to 20 °C (06:00–18:00) and 15 °C (18:00–06:00), respectively.

3.1.2 Protocol for sampling infected wheat tissue

Infected wheat leaf samples that were used for RNA-Seq discussed in Chapter 7 were collected in wheat fields in South Africa. For every sample, an area of approximately 20 mm of the leaf covered in *Pst* pustules was cut into small segments of roughly 7 mm and placed in a 5 ml tube with RNeasy lysis solution (Thermo Fisher Scientific, USA), immediately after sampling from the wheat plant. RNeasy lysis solution was used to preserve RNA integrity as advised by Taylor et al. (2010). The same procedure was used to collect material from the time course for gene expression analysis (Chapter 7).

3.2 Nucleic acid extraction and quantification

3.2.1 Genomic DNA extraction

Genomic DNA was extracted from urediniospores using the cetyltrimethylammonium bromide (CTAB) extraction method of Chen et al. (1993). Beforehand, CTAB was heated to 65 °C, and 70 % ethanol was prepared and chilled at –20 °C. Spores were frozen using liquid nitrogen and ground using a pestle and mortar. Silicon dioxide (SiO₂; Sigma-Aldrich, USA) was used to aid in tissue disruption, using 100 mg of spores with 600 mg of sand. The disrupted material was transferred to a 15 ml Falcon tube. In a separate tube, 2 ml of pre-warmed CTAB buffer was added to 5 µl Proteinase K 10 mg/ml, mixed, and incubated at 65 °C for 2 hours. After incubation, 1 volume of chloroform:isoamylalcohol (24:1, v/v) was added to the previous mixture and vigorously mixed followed by centrifugation at 12 000 g for 10 minutes. The aqueous, upper phase was transferred to a fresh tube, and 20 µl of RNaseB 10 mg/ml was added after which samples were incubated at room temperature (20 °C) for 1 hour. The chloroform step was repeated and the supernatant transferred to a fresh tube again. Pre-chilled isopropanol was added (1 volume), followed by gentle inversion to precipitate the gDNA. Samples were incubated at –20 °C overnight. The next day, samples were centrifuged at 12 000 g for 10 minutes. The pellet was washed in 1 ml to 2 ml of the pre-chilled 70 % ethanol. The ethanol was decanted without disturbing the pellet, which was subsequently allowed to dry at room temperature (20 °C) and dissolved in 50 µl 1 % TE buffer [10 mM Tris-Cl (pH 8.0); 1 mM Ethylenediaminetetraacetic acid (EDTA) (pH 8.0)].

3.2.2 RNA extraction

Total RNA was extracted from *Pst* inoculated leaf tissue, non-inoculated wheat and germinated fungal spores using the RNeasy Plant Mini Kit (Qiagen, Ger-

many) according to the manufacturer's instructions. Tissue was disrupted with a pestle and mortar. To promote tissue disruption, SiO₂ was added to the mortar. All instruments used, including the mortar and pestle and the spatula used to scrape the homogenised tissue from the mortar, were washed with detergent, ethanol, and RNase AWAY Decontamination Reagent (Thermo Fisher Scientific, USA) between extractions. All instruments were cooled in liquid nitrogen or on dry ice to prevent degradation of RNA due to ubiquitous RNase activity (Holland et al., 2003). The dry mortar and pestle were placed on dry ice in a polystyrene box, and further cooled with liquid nitrogen. Approximately 100 mg SiO₂ was added to the mortar with the liquid nitrogen before the leaf sample was added.

Forceps were used to move the preserved sample material from the tubes to a clean paper towel, where samples were tapped dry to prevent the RNAlater solution from forming ice crystals when the sample came into contact with the liquid nitrogen. Samples were then placed in the mortar with liquid nitrogen and SiO₂, followed by homogenisation of the sample into a fine powder. The ground sample was scraped with a cooled spatula into a 2.2 ml safe-lock microcentrifuge tube without allowing it to thaw. The tube with the ground sample was kept on dry ice until extraction buffer was added.

The procedure was concluded followed the optional step in the protocol. To prevent degradation RNase inhibitor (0.5 µl) was added to each sample. Aliquots of 3 µl were prepared for RNA quantification and quality control. Extracted RNA samples were stored at -80 °C.

3.2.3 DNA and RNA quantification

Extracted gDNA was quantified using the Qubit 2.0 Fluorometer (Invitrogen/Thermo Fisher Scientific, USA). The rationale behind the method is that it detects dyes that only fluoresce when bound to a specific substrate, in this case, double

stranded (ds) DNA. The intensity of the fluorescence is indicative of the amount of dsDNA in the sample (Simbolo et al., 2013). Assays were performed at room temperature (20 °C) as recommended. The instrument was calibrated with the Quant-iT dsDNA BR Assay according to the manufacturer's instructions, and DNA concentrations quantified for all samples.

The Agilent 2100 Bioanalyzer (Agilent Technologies, USA) was used to assess the quality and quantity of the extracted RNA. The reaction kit was stored at 4 °C. A gel-dye mix was first prepared according to the manufacturer's instructions. The quality of RNA samples was assessed within one to three days after preparation and RNA was converted into cDNA within one to three days after an aliquot passed the quality assessment. Aliquoting prevented multiple freezing and thawing cycles, as this imposes a risk of degradation of RNA (Taylor et al., 2010). RNA stocks were stored at -80 °C between extraction and being used for cDNA synthesis.

3.3 Next-generation sequencing and data analysis

3.3.1 Library preparation

A sequencing library was prepared from raw extracted nucleic acids. DNA fragmentation was followed by size selection and the addition of oligonucleotide adapters to fragments, for the sequencer to process the library.

3.3.2 Genomic DNA sequencing

Libraries for gDNA sequencing were prepared by the Earlham Institute, UK, using the Illumina TruSeq DNA Sample Preparation Kit (Illumina, UK), according to the manufacturer's instructions. To assess library quality before sequencing, a High Sensitivity DNA analysis assay was performed on the Agilent 2100 Bioana-

lyzer. Quantification of libraries was conducted with the Qubit 2.0 Fluorometer. One lane of the Illumina flow cell was used for a pool of 10 libraries diluted to a concentration of 12.71 nM. Sequencing was performed on the Illumina HiSeq 2500 platform at the Earlham Institute, UK, where after adapter and multiplexing barcode oligonucleotide sequences were removed. Upon receipt of the data, read quality was assessed using FastQC software (version 0.10.1; Andrews, 2010).

3.3.3 RNA sequencing

Sequencing of messenger RNA (mRNA) extracted from *Pst* infected wheat samples was performed at Earlham Institute, UK. The mRNA was reversed transcribed to cDNA. Sequencing libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, UK). The RNA 6000 Nano kit was used to assess the library quality on the Agilent 2100 Bioanalyzer. Libraries were sequenced using the Illumina HiSeq 2500 platform, and adapters and barcodes were removed from the resulting sequences.

3.3.4 Bioinformatics pipeline

Mapping of gDNA samples

The 100 bp Illumina paired end reads were filtered using a Perl script to discard reads containing N calls where nucleotides could not be determined by the sequencer (Cantu et al., 2013; Hubbard et al., 2015). After filtering, each gDNA sample was independently aligned to the PST130 reference genome (Cantu et al., 2011) implementing Burrows-Wheeler Alignment tool (BWA version 0.7.7; Li and Durbin, 2009) with parameters set to the default setting.

Mapping of cDNA (RNA-Seq) samples

Similar to the gDNA samples, the 100 bp Illumina paired end reads were filtered to discard reads containing nucleotides that could not be determined by the sequencer (Cantu et al., 2013; Hubbard et al., 2015). The alignment of cDNA samples was carried out using the Bowtie alignment program (version 0.12.7; Langmead et al., 2009) from the TopHat package (version 1.3.2; Trapnell et al., 2012), again aligning to the PST130 reference genome (Cantu et al., 2011), using the parameter `-r` set to 200 to accommodate the mate pair sequences with 50 bp ends.

Identifying single nucleotide polymorphisms

Resulting sequence alignment map (SAM) format files from the gDNA and RNA-Seq mapping, were converted to binary alignment map (BAM) format with the software package SAMtools (version 0.1.19; Li et al., 2009). SAMtools sort, SAMtools index and SAMtools mpileup were used to identify SNPs. Custom Perl scripts were used to extract allele counts at each position of the genome. A depth of coverage threshold was set for polymorphic sites, and gDNA SNPs with a minimum depth of coverage of $10\times$ were extracted, while for RNA-Seq data minimum depth coverage of $20\times$ were required.

Allele frequencies between 0.2 and 0.8 were classified as heterokaryotic sites, whereas sites with allelic frequencies above 0.8 were classified as homokaryotic sites (Cantu et al., 2013). SnpEff (version 3.6; Cingolani et al., 2012) was used to annotate polymorphisms, to indicate whether they resulted in synonymous or nonsynonymous substitutions, or whether a stop codon was gained or lost in coding regions. SnpEff further displayed the codon position of polymorphisms. Polymorphisms in intergenic regions were also indicated.

Quality assessment of samples through sequence data

Each of the two haploid nuclei in the dikaryotic urediniospore is assumed to contribute a maximum of one allele to each nucleotide site. A variant site is described as a homokaryotic SNP when both alleles are identical, but different from the reference PST130 nucleotide. A heterokaryotic SNP describes the situation where two different alleles occur at the nucleotide site. These alleles may both be different from the reference, or only one, while the other would be identical to the reference (Hubbard et al., 2015).

To ensure that the genomic data was in each case derived from a single genotype, the allelic distribution at heterokaryotic sites was assessed across the genome. It is important to note that the reference genome is not phased. Implications are discussed in Chapter 5. When a single genotype is present, it is expected that the frequency plot, exhibiting both alleles at the heterokaryotic SNP sites, will form a distribution with a mode of 0.5 due to the equal contribution of both nuclei (Yoshida et al., 2013).

In this analysis, the number of heterokaryotic SNP sites were plotted on the y -axis, and the proportion of alleles across reads at each site, ranging between 0 and 1, on the x -axis, as explained in the supplementary documents of Cantu et al. (2013). Read frequency graphs of isolates unique to the current study are summarised in Appendices A and D.

3.3.5 Clustering analysis

Clustering analyses are grouping algorithms that operate in such a way that individuals placed in the same group are more similar compared to individuals in other groups. Genomic and transcriptomic data were used for phylogenetic clustering and population cluster analyses. As the transcriptomic data does not include intergenic regions, only the coding regions of the gDNA samples

were considered in this analysis. Brief descriptions of the different underlying statistical and genetic models deployed in the analyses follow in the next sections.

Phylogenetic analysis

A “Randomized Axelerated Maximum Likelihood” (RAxML) phylogenetic approach was used to determine the genetic relationships between South African *Pst* and to compare them to *Pst* isolates from other countries.

First a subset of sites in each gene in the PST130 gene models was used to construct synthetic genes. Sites identical to the PST130 reference genome were only included when a minimum of 2× depth of coverage was reached. Variant sites were included when coverage depths of 10× for gDNA samples or 20× for cDNA samples were reached. Introducing placeholders at sites where the required depth of coverage was not achieved preserved codon positions. Then a phylip file was prepared as input to RAxML software (version 8.0.20; Stamatakis, 2014) to construct the phylogenetic tree.

Accurate nucleotide substitution models are required in most phylogenetic analyses as the rate of nucleotide substitution varies in molecular evolution (Jia et al., 2014). To account for the fact that all sites do not evolve at an identical rate, codon positions and the model used to determine phylogenetic clades were considered. Due to the degeneracy of codons, there is redundancy in the genetic code that can cause the occurrence of synonymous substitutions. Substitutions at the third position are more often synonymous, and therefore less likely to influence the phenotype and be a target for positive or negative selection, than at the first and the second codon position. Nucleotide changes at the third codon positions can, for this reason, be considered to evolve at a higher rate (Rambaut and Grass, 1997). The third codon position further shows less nucleotide bias and a more homogenous rate of evolution when compared to the first and second

codon position (Bofkin and Goldman, 2006).

Nonsynonymous sites are not evolutionary neutral and, depending on the effect of the resulting phenotype, can experience high levels of selection pressure resulting in gene specific evolution. Phylogenetic trees derived from such data can be misleading when convergent evolution of such genes in different populations are present. The phylip input file was therefore prepared containing the third codon positions of synthetic genes to illustrate the evolutionary history of the populations without being influenced by gene specific evolutionary development. The third codon position of those synthetic genes that had a minimum of 80 % breadth of coverage of the original reference gene length in at least 80 % of isolates were included in the phylogenetic analysis to ensure that only genes with high coverage were included.

In addition, the General Time Reversible (GTR) model of nucleotide substitution under the Gamma (Γ) model of rate heterogeneity was selected for the RAxML model parameter (-m GTRGAMMA). The GTR model parameters account for unequal frequencies for the four nucleotides and the unique rate of each of the possible six nucleotide substitutions. Furthermore, the Γ model uses a discrete Γ distribution to assign different rates of heterogeneity to different sites (Stamatakis, 2014). Reproducibility was ensured by specifying an initialising value for the pseudo-random number generator (-p 100) and the process was parallelised on 10 threads (-T 10). To demonstrate the reliability of the inferred tree, bootstrapping was applied by generating 100 (-N 100) alternative runs on distinct starting trees (-b 12345). Bootstrap values were added to the maximum likelihood tree with the -f b parameter to generate the bipartition tree where after MEGA (version 6.06; Tamura et al., 2013) was used to visualise the phylogenetic tree (Cantu et al., 2013; Hubbard et al., 2015).

Population structure analyses

Two methods were used to predict population structure: STRUCTURE (version 2.3.4; Pritchard et al., 2000) and Discriminant analysis of principal components (DAPC; Jombart et al., 2010). STRUCTURE is a model-based approach, whereas DAPC does not make any assumptions about the biological processes that influenced and shaped the dataset. Both methods have limitations and benefits, and these are discussed in the relevant research chapters. The same depths of coverage minima as for the phylogenetic tree were required: $10\times$ coverage for gDNA samples and $20\times$ coverage for cDNA samples. The SNP data was prepared using BEDTools (version 2.17.0; Quinlan and Hall, 2010) for variant site annotation in SnpEff.

Sites where a synonymous substitution was introduced in at least one isolate were extracted. These, together with sites identical to the reference with at least $2\times$ coverage, were repositioned according to their position in the reference genome. From these files, a data matrix was generated using a custom python script. The software, STRUCTURE, was used to assign isolates to specific population groups and to determine the number of these groups, or clusters (K), due to genetic differentiation. For this analysis, nonsynonymous SNPs were excluded as these sites are more likely involved in fitness traits and under selection and STRUCTURE relies on neutral substitution models. Furthermore, different populations could have evolved convergently. Such similarity would falsely deduce that individuals are related.

Analyses consisting of five independent runs for each value of K were carried out. The “admixture” model was used, and each run was set to a burn-in period of 110 000 iterations. Thereafter, 200 000 Markov Chain Monte Carlo (MCMC) generations for each value of K , ranging from 1 to 15, were carried out. K values were evaluated in two ways: the Evanno method (Evanno et al., 2005) and by

calculating the log probability, referred to as $\text{LnP}(D)$, of each K value (Pritchard et al., 2000). STRUCTURE assumes a population that is under Hardy-Weinberg equilibrium, and the *Pst* data does not fit this assumption. Therefore the multi-variate DAPC analysis within the adegenet R software package (Jombart et al., 2010), was carried out on the same dataset used with STRUCTURE. Principal component analysis (PCA) summarised genetic variation in the dataset by reducing the dataset to include only the most impactful loci. The lowest Bayesian information criterion (BIC) suggested the optimum number of population clusters (K), thereafter discriminant analysis (DA) was used to divide samples into subgroups of population clusters.

Differentiation between and within population clusters

In a segregated population, individuals that aggregate into a subpopulation tend to interbreed more than what is expected under random mating of the whole population under Hardy-Weinberg equilibrium. When assessing a dataset, groups with low levels of heterozygosity among individuals within groups allow the identification of genetic structure in a global population from which biological interpretations can be made. To quantify the variation between subpopulations, the general reduction in heterozygosity H_X is assessed by evaluating the observed heterozygosity H_{obs} against the expected heterozygosity H_{exp} using the equation

$$H_X = \frac{H_{exp} - H_{obs}}{H_{exp}}. \quad (3.1)$$

Three specific inbreeding coefficients need consideration to take into account heterozygosity observed in individuals, subpopulations and the whole population, substituting H_X in Eq. (3.1) with H_I , H_S and H_T , respectively.

Reduction in heterozygosity that is due to the population structure can then be evaluated using the so called “F-statistics”

$$\begin{aligned} F_{IS} &= \frac{H_S - H_I}{H_S}, \\ F_{IT} &= \frac{H_T - H_I}{H_T}, \\ F_{ST} &= \frac{H_T - H_S}{H_T}, \end{aligned}$$

with the relationship

$$F_{ST} = 1 - \frac{1 - F_{IT}}{1 - F_{IS}}.$$

The proportion of the genetic variance assigned to the differences between subpopulations, evaluated in Section 4.2.7 and Section 7.3.1, were calculated using GenePop (version 4.2; Rousset, 2008) to estimating Wright’s F_{ST} statistic (Hubbard et al., 2015). The F_{ST} values varied from zero to one, where zero indicated the absence of differentiation and one complete differentiation (Hartl and Clark, 1998).

To assess the genetic diversity within each of the *Pst* population clusters identified herein, the population diversity parameter theta (θ) was estimated in Section 4.2.7 and Section 7.3.1. Theoretically, θ estimates genetic differentiation amongst subpopulations depending on the number of reproducing individuals in the population and the mutation rate. Different empirical approximations of θ exist. In this study, Watterson’s theta, $\hat{\theta}_W$, was reported as it takes into account the number of segregating sites—SNPs in the current case—to estimate the mutation rate of the population.

The degree of polymorphism between genes in individuals of a subpopulation was calculated using DnaSP (version 5.10.1; Librado and Rozas, 2009) as suggested by Hubbard et al. (2015).

Chapter 4

Origin of the South African *Pst* Pathotypes

4.1 Introduction

4.1.1 Wheat stripe rust in South Africa

IN MOST WHEAT CULTIVATION REGIONS globally, *Puccinia striiformis* f. sp. *tritici* prevails and is a threat to wheat production (Brown, 2003; Hovmøller et al., 2010; Sharma-Poudyal et al., 2013). Wind dispersal of the asexual urediniospores enables *Pst* to travel thousands of kilometres (Kolmer, 2005; Hovmøller et al., 2008; Ali et al., 2014). Foreign incursions can become established in new geographical regions, completely shifting the pathotype profile of the *Pst* population in a single season. In addition to wind dispersal, *Pst* can be transmitted via anthropogenic activities such as human travel. For instance, Wellings et al. (1987) considered that the introduction of *Pst* into Australia in 1979 could easily have been facilitated by human-assisted movement. With increases in global travel and freight movement in recent years, multiple destinations are now within easy reach of many pathogens in a single day (Parker and Gilbert, 2004), regardless of wind dispersal patterns. In South Africa, the first verified identification and characterisation of



Figure 4.1: Locations of the original detections of South African *Pst* pathotypes. Stripe rust was first detected near Moorreesburg in the Western Cape in 1996. It occurred throughout the wheat breeding regions of the southwestern part of South Africa during the season. The pathotype 6E16A- was designated. New pathotypes (6E22A-, 7E22A- and 6E22A+) observed in following years were first detected in the Eastern Free State and Lesotho.

stripe rust was in the Western Cape in 1996 (Figure 4.1; Pretorius et al., 1997), making it a relatively new disease compared to leaf rust and stem rust that were already recorded in the 1700s (Du Plessis, 1933).

Subsequent surveys in 1996 confirmed that the disease was well established throughout the winter rainfall regions of the Western, Northern and Eastern Cape (Pretorius et al., 1997). Traces were also found on irrigated wheat in summer rainfall regions. As stripe rust has a lower temperature optimum (Roelfs and Hettel, 1992), the lengthy cool and wet conditions in the Western Cape in 1996 (Figure 4.2), likely contributed to the rapid spread and development of *Pst* epidemics (Boshoff et al., 2002).

The first *Pst* pathotype was confirmed as pathotype 6E16A- through testing of 32 *Pst* isolates on 17 standard stripe rust wheat differential lines and seven supplementary tester lines with known resistance genes (Pretorius et al., 1997).

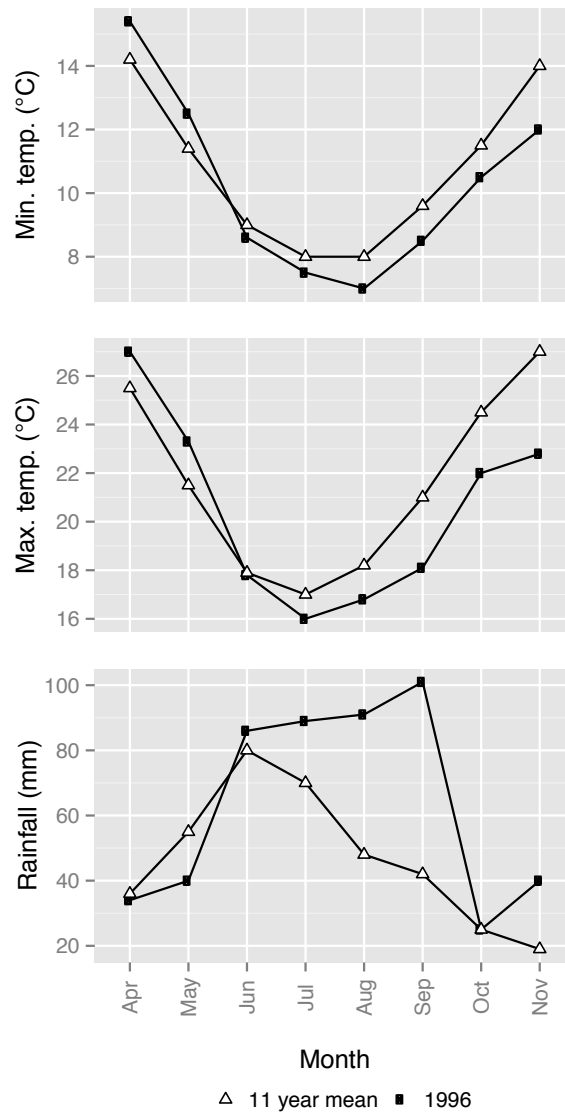


Figure 4.2: Temperature and rainfall measured in 1996 during April to November in the Western Cape compared to the 11 year mean (from Boshoff et al., 2002). Max. temp., maximum temperatures; Min. temp., minimum temperatures.

This pathotype was similar to the stripe rust pathotype 6E16 found in the Mediterranean region in the 1970s (Wahl et al., 1984). A similar pathotype, 6E16 was also detected in East and North Africa, the Middle East and Western Asia (Stubbs, 1988; Badebo et al., 1990; Pretorius et al., 1997). The “A-” added to the pathotype name of the South African isolate expanded on the notation protocol developed by Johnson et al. (1972) by adding testing for virulence to *YrA*, as described by Wellings et al. (1988).

In 1998 another stripe rust epidemic occurred in South Africa, this time in the Eastern Free State. The wheat varieties Huguenoot and Carina, that were resistant to 6E16A-, were widely and severely affected (Boshoff and Pretorius, 1999). Frequent cases of severe *Pst* infection were observed, often colonising 100 % of wheat leaves. Virulence tests on an expanded wheat differential set confirmed a virulence gain for *Yr25*, defining a new pathotype, 6E22A- (Figure 4.3; Boshoff and Pretorius, 1999). Pathotype 6E22 has since been reported in Iran in 2009 and 2010 (Elyasi-Gomari and Petrenkova, 2011).

In 2001 yet another new pathotype, 7E22A- (Figure 4.3), was detected on the wheat variety Chinese 166 in trap nurseries in Makobateng, Lesotho (Pretorius et al., 2007). This pathotype contained additional virulence to *Yr1*, but although Lesotho neighbours the Eastern Free State, an important wheat cultivation area in South Africa, the pathotype was not considered a threat to the South African wheat industry, as *Yr1* did not occur in local wheat varieties (Pretorius et al., 2007).

In 2005 a fourth new pathotype, 6E22A+ (Figure 4.3), was detected near Clocolan in the Eastern Free State. This pathotype was virulent to *YrA*, but avirulent to *Yr1* (Visser et al., 2016). The phenotypic characterisation of the four *Pst* pathotypes is indicated in Figure 4.4.

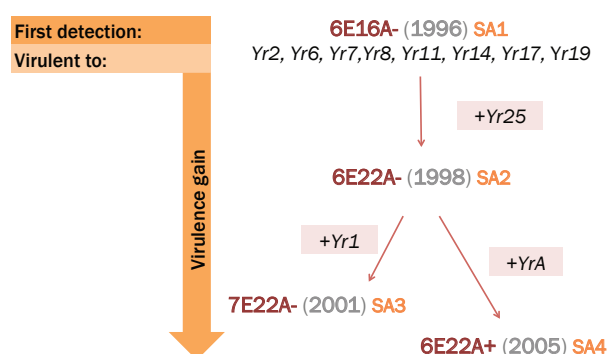


Figure 4.3: Schematic illustration of the increase of *Pst* virulence in South Africa. Gain of virulence in South African *Pst* populations, based on traditional pathotype analysis, between 1996 and 2016 (Pretorius et al., 1997; Boshoff et al., 2002; Pretorius et al., 2007; ZA Pretorius, unpublished data). Pathotypes analysed in this study that represent the identified pathotypes were named SA1—SA4.

4.1.2 *Pst* population diversity

Sufficient genetic diversity in a population increases the likelihood that some individuals will have superior fitness in changing environmental conditions (Hartl and Clark, 1998). Due to the stepwise gain in virulence together with molecular evidence (Visser et al., 2016), *Pst* likely reproduces clonally in South Africa. Factors that can increase genetic diversity in asexual *Pst* populations are mutations and gene flow, and although not considered to occur frequently, somatic recombination. Newly introduced alleles—that can be slightly deleterious, neutral, or slightly advantageous—can stay in the population just by chance, called genetic drift. When new alleles provide a fitness incentive, positive selection can

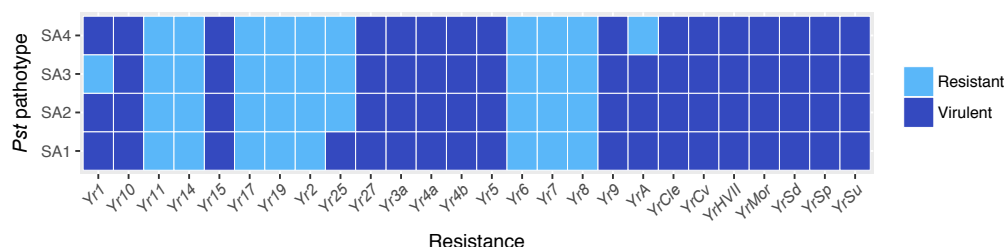


Figure 4.4: Pathotype (race) identification tests of South African *Pst* pathotypes. Pathotypes were defined by compatibility with wheat hosts possessing indicated sources of resistance (data from Visser et al., 2016).

fix such alleles in the population, while negative selection will remove deleterious mutations. Such selective evolutionary forces can result in an erosion in genetic diversity, dominating the direction of change in allele frequencies, which can, in turn, be counteracted by balancing and diversifying selection to increase diversity again (Hartl and Clark, 1998).

Allele frequencies in a population can be influenced by multiple biotic and abiotic factors. Clustering analyses can be implemented to illustrate the genetic relationship between individuals, define the number of populations, and assign isolates within these populations. This population structure indicates the evolutionary history through alleles present in samples (McDonald and Linde, 2002). To quantify the genetic diversity between individuals and populations a wide range of molecular markers have been developed and deployed over the past 37 years (Schlötterer, 2004).

4.1.3 Molecular markers and *Pst*

Molecular markers improved the traceability of *Pst* considerably, enabling refinement of dispersal distance approximations and population dynamics. Population studies based on AFLP molecular markers (Vos et al., 1995) were first applied (Hovmøller et al., 2002). A widely inclusive population study, analysing isolates from North America, Australia, Europe, Western and Central Asia, the Red Sea Area, East Africa and South Africa provided the first genotyping information for the South African pathotypes (Hovmøller et al., 2008). These 876 *Pst* isolates, collected over a period of 30 years between 1975 and 2005, were pathotyped on a set of 30 wheat differential lines including at least 17 stripe rust resistance genes. A subset containing 151 of the collected isolates, which represented the diversity with respect to virulence phenotypes, region, and sampling year, were then genotyped using AFLP molecular markers (Hovmøller et al., 2008), identify-

ing presence-absence polymorphisms. This subset contained South African *Pst* isolates representative of the pathotypes 6E16A-, 6E22A-, and 7E22A- that were sampled between 1996 and 2001. The subset was screened with 117 informative AFLP makers, however, these markers did not show any differentiation between the South African isolates. This analysis indicated that the South African *Pst* isolates were closely related to isolates detected in Central (sampled in 2003) and Western Asia (sampled in 2005), and Southern Europe (sampled in 1997 and 1998).

Differential testing showed that 6E16A- is similar to the pathotype 6E16, also called *Pst*S3 (Hovmøller et al., 2016), that was identified in Southern Europe since 1985 (Enjalbert et al., 2005). In the south of France, a stable divergent subpopulation was described using AFLP markers, also comparable to an Italian isolate sampled in 1998 (Enjalbert et al., 2005). Pathotypes similar to the South African pathotypes (Figure 4.3) have also repeatedly been detected in Northern Europe since 2004 (Hovmøller et al., 2008). Ali et al. (2014) concluded similar results using 20 microsatellite markers (Vieira et al., 2016), identifying the Mediterranean region and Central Asia as the probable origin of the South African *Pst* pathotypes. In these two studies, seven and six South African isolates were used, respectively. Pathotype 6E22A+, detected in South Africa in 2005, was not included in these analyses. Similar to AFLP markers, microsatellite markers reported low levels of genetic diversity in the South African population and could not differentiate between pathotypes.

Since 1996 characterisation of the *Pst* population in South Africa has largely been carried out through traditional pathotype analysis methods (see Figure 7.1). More recently 17 microsatellite markers were used to genetically characterise the South African *Pst* pathotypes (Visser et al., 2016), confirming previous findings of low genetic variability between pathotypes (Hovmøller et al., 2008; Ali et al., 2014). These markers were however able to distinguish between the South

African pathotypes. Through network analysis Visser et al. (2016) proposed seven hypothetical intermediates between the four South African pathotypes, indicating a model for the establishment of *Pst* in South Africa.

4.1.4 Next-generation sequence analyses of South African *Pst*

Along with the cost and time limitations in the development of traditional marker systems such as microsatellites and AFLPs, genotyping samples with traditional marker panels—even with a large marker selection—will only provide a low resolution view of the genetic diversity between samples (Davey et al., 2011). This can be especially problematic when aiming to distinguish between samples with low genetic variability. Next-generation sequencing relieved this limitation of traditional molecular markers by facilitating the limitless identification of markers in a multitude of samples (Davey et al., 2011). As is the case with AFLP markers, another advantage is that no prior knowledge of the target is needed (Naccache et al., 2014). The extensive datasets generated from this technology across species' genomes, enable searches for diversity at nucleotide level that traditional marker systems will never generate. It allows addressing of population structure questions with a level of detail and improved accuracy that ordinary markers have not achieved.

To add to the traditional pathology and marker work carried out on the South African *Pst* pathotypes whole genome sequencing of four *Pst* isolates was undertaken. These isolates represent the major pathotypes following the first confirmed incursion of stripe rust into South Africa in 1996. Data from the four representative isolates of the identified South African pathotypes, together with available data from global isolates, were used to (i) re-evaluate the potential origin of the South African pathotypes using a comparative genomics approach and to (ii) assess the genetic diversity within the South African population. The

South African pathotypes identified between 1996 and 2005 will be referred to as the historical South African population. Specific isolates analysed in this study that represent the identified pathotypes were named SA1—SA4.

4.2 Materials and methods

Work done by co-workers is indicated in the relevant sections. The methodology followed the field pathogenomics approach described in (Hubbard et al., 2015). See Chapter 3 for detailed descriptions.

4.2.1 Data description

Four isolates representing the four pathotypes observed in South Africa to date have been sequenced in this study. Hubbard et al. (2015) reported an in-depth analysis of the UK population comparing several UK *Pst* isolates, collected between 1974 and 2013. A subset of the data used by Hubbard et al. (2015) was included in the present study to draw comparisons between the South African isolates and other available *Pst* datasets.

The UK *Pst* population in 2013 showed high diversity and differed to the pre-2011 population Hubbard et al. (2015). Population genetic analysis defined this 2013 population into four distinct genetic groups. Notable features of these four groups were that UK Group II was detected on triticale and UK Groups I and II were genetically less diverse compared to Groups III and IV.

Sequence data of the South African historical isolates, together with sequence data of 44 other isolates including 32 isolates from Europe (Table 4.1) that were sequenced and described before (Hubbard et al., 2015), five isolates from Pakistan (Bueno-Sancho et al., 2017) and seven isolates from East Africa, including three isolates from Ethiopia, two from Kenya and two from Eritrea, were used in this chapter to determine the relationship of the South African isolates with the

available data from other wheat-growing areas where stripe rust occurs. The East African isolates were obtained from Mogens Hovmøller. The isolate ET03b/10, that was assigned to the pathotype group *PstS2*, and ET08/10 were included in previous analysis by Ali et al. (2017). Isolates KE74217, KE89069 (V23) and ET87094 are part of the Stubbs collection and were described by Thach et al. (2015, 2016).

4.2.2 Sample preparation for DNA extraction

The urediniospores used for extraction of gDNA were purified and multiplied at UFS, South Africa. The isolates that were sequenced were representative of the identified pathotypes. Table 4.2 lists the UFS stocks collection identities and the collection date of the *Pst* isolates that were used for multiplication of the urediniospore samples that were sequenced.

To obtain single pustule isolates for genome sequencing, seeds of the susceptible wheat variety, Morocco, were planted and grown for seven days to the two leaf stage (Z12; Zadoks et al., 1974). Urediniospores of the four pathotypes were previously dried on silica gel and kept at -80°C in storage. Inoculations were performed where after plants were moved to a glasshouse with natural light and a day—night temperature cycle set to 20°C (06:00-18:00) and 15°C (18:00-06:00), respectively. When flecks appeared, all plants were cut away to leave only half a leaf with a single infection site, the result of infection by a single spore. Due to the systemic nature of the infection, the entire leaf segment eventually sporulated from the single infection site. For each isolate urediniospores were collected from one actively sporulating lesion and increased twice on Morocco seedlings to produce several grams of spores. The final spore harvest was desiccated for five days on silica gel and used to extract the DNA for sequencing. To maintain isolate purity, multiplication of the different isolates were spatially or temporally separated in the glasshouse.

Table 4.1: Global isolates included in the clustering and genetic diversity analyses

Isolate number	Isolates	Country of isolation	Year of isolation	Type of data	References
1	88.55S1	UK	Pre 2011	gDNA	Hubbard et al. (2015)
2	03/7	UK	Pre 2011	gDNA	Hubbard et al. (2015)
3	08/21	UK	Pre 2011	gDNA	Hubbard et al. (2015)
4	88.45SS	UK	Pre 2011	gDNA	Hubbard et al. (2015)
5	78.66SS1	UK	Pre 2011	gDNA	Hubbard et al. (2015)
6	88.44SS3	UK	Pre 2011	gDNA	Hubbard et al. (2015)
7	J0085F	France	Pre 2011	gDNA	Hubbard et al. (2015)
8	J01144Bm1	France	Pre 2011	gDNA	Hubbard et al. (2015)
9	J02-022	France	Pre 2011	gDNA	Hubbard et al. (2015)
10	J02055C	France	Pre 2011	gDNA	Hubbard et al. (2015)
11	11/13	UK	2011	gDNA	Hubbard et al. (2015)
12	11/75	UK	2011	gDNA	DGO Saunders & S Holdgate
13	11/128	UK	2011	gDNA	Hubbard et al. (2015)
14	11/140	UK	2011	gDNA	Hubbard et al. (2015)
15	11/08	UK	2011	gDNA	Hubbard et al. (2015)
16	11/08	UK	2011	RNA-Seq	Hubbard et al. (2015)
17	13/19	UK	2013	RNA-Seq	Hubbard et al. (2015)
18	13/15	UK	2013	RNA-Seq	Hubbard et al. (2015)
19	13/123	UK	2013	RNA-Seq	Hubbard et al. (2015)
20	13/27	UK	2013	RNA-Seq	Hubbard et al. (2015)
21	CL1	UK	2013	RNA-Seq	Hubbard et al. (2015)
22	T13/2	UK	2013	RNA-Seq	Hubbard et al. (2015)
23	T13/3	UK	2013	RNA-Seq	Hubbard et al. (2015)
24	T13/1	UK	2013	RNA-Seq	Hubbard et al. (2015)
25	13/38	UK	2013	RNA-Seq	Hubbard et al. (2015)
26	13/21	UK	2013	RNA-Seq	Hubbard et al. (2015)
27	13/33	UK	2013	RNA-Seq	Hubbard et al. (2015)
28	13/182	UK	2013	RNA-Seq	Hubbard et al. (2015)
29	13/25	UK	2013	RNA-Seq	Hubbard et al. (2015)
30	13/29	UK	2013	RNA-Seq	Hubbard et al. (2015)
31	13/71	UK	2013	RNA-Seq	Hubbard et al. (2015)
32	13/40	UK	2013	RNA-Seq	Hubbard et al. (2015)
33	SA1	SA	1996	gDNA	Pretorius et al. (1997)
34	SA2	SA	1998	gDNA	Boshoff and Pretorius, (1999)
35	SA3	SA	2001	gDNA	Pretorius et al. (2007)
36	SA4	SA	2005	gDNA	Pretorius, (Unpublished)
37	KE74217	Kenya	1974	gDNA	Thach et al. (2015; 2016)*
38	KE89069	Kenya	1989	gDNA	Thach et al. (2015; 2016)*
39	ET87094	Ethiopia	1987	gDNA	Thach et al. (2015; 2016)*
40	ET08/10	Ethiopia	2010	gDNA	Ali et al. (2017)**
41	ET03b/10	Ethiopia	2010	gDNA	Ali et al. (2017)**
42	ER179b/11	Eritrea	2011	gDNA	Ali et al. (2017)**
43	ER181a/11	Eritrea	2011	gDNA	Ali et al. (2017)**
44	Qld-1	Pakistan	2014	gDNA	Bueno-Sancho et al. (2017)
45	Qld-2	Pakistan	2014	gDNA	Bueno-Sancho et al. (2017)
46	ATR-1	Pakistan	2014	gDNA	Bueno-Sancho et al. (2017)
47	ATR-2	Pakistan	2014	gDNA	Bueno-Sancho et al. (2017)
48	ATR-3	Pakistan	2014	gDNA	Bueno-Sancho et al. (2017)

*Isolates KE74217, KE89069, and ET87094 were provided by Aarhus University, Denmark, and Plant Research International, Wageningen, The Netherlands, maintaining the Global Yellow Rust Gene Bank of the late ir. RW Stubbs up to 25-01-2010. ** Provided by MS Hovmøller. Personal communication with MS Hovmøller confirmed inclusion in the listed studies.

Table 4.2: Historical isolates used in re-sequencing and an infection time course experiment (Chapter 6)

Pathotype	First occurrence	Alias	Isolate ID	Collection date
6E16A-	1996	SA1	Isolate 49	2003
6E22A-	1998	SA2	Isolate 3	2001
7E22A-	2001	SA3	Isolate 27	2004
6E22A+	2005	SA4	Isolate 35	2011

4.2.3 Genomic DNA extraction and quantification

Genomic DNA was extracted from urediniospores using the CTAB extraction method described by Chen et al. (1993) and quantified using the Qubit 2.0 Fluorometer (Invitrogen/Thermo Fisher Scientific, USA).

4.2.4 Sequencing and mapping

Sequencing libraries were prepared, quality assessed, quantified and sequenced by the Earlham Institute. Sequences containing missing data indicated with “N” were discarded (Cantu et al., 2013; Hubbard et al., 2015). The 100 bp paired end reads were aligned to the PST130 draft reference genome (Cantu et al., 2011) using BWA (version 0.7.7; Li and Durbin, 2009) with default parameters producing sequence alignment map (SAM) format files. SAMtools (version 0.1.19; Li et al., 2009) was used, to identify variant sites. SnpEff (version 3.6; Cingolani et al., 2012) was used to identify whether homokaryotic SNPs resulted in synonymous or nonsynonymous substitutions similar to the procedures in Cantu et al. (2013). Based on the rationale explained in Yoshida et al. (2013), the read frequency graph of each isolate was assessed to determine whether the starting material could be considered uncontaminated containing predominantly a single genotype (Cantu et al., 2013; Hubbard et al., 2015). Read frequency graphs of other isolates used in this chapter that have not been published before are displayed in Appendix A, Figure A.1.

4.2.5 Phylogenetic analysis

A maximum likelihood phylogenetic approach was used to determine the genetic relationships amongst the South African *Pst* isolates and to compare them with isolates from elsewhere. Synthetic genes were prepared, and the third codon positions of these genes were used to determine the phylogeny. Due to the degeneracy of the genetic code, this will include mostly nucleotide changes that do not result in amino acid changes resulting in more evolutionary neutral positions. The RAxML software (version 8.0.20; Stamatakis, 2014) was used. One hundred iterations of bootstrapping were performed to assess the reliability of the maximum likelihood dendrograms (Cantu et al., 2013; Hubbard et al., 2015).

4.2.6 Population structure analysis

The genetic differentiation of the 48 isolates (Table 4.1) was assessed by two population-clustering methods: (i) STRUCTURE (version 2.3.4; Pritchard et al., 2000) was used to assign isolates to subpopulation clusters (K) based on genetic differentiation at nearly neutral or neutral SNP sites, and (ii) Multivariate DAPC within the Adegenet package (Jombart et al., 2010) was carried out in the R environment on the same dataset as STRUCTURE.

4.2.7 Genetic diversity assessment

Inter-cluster variance

The SNP dataset used in STRUCTURE and DAPC analyses containing only biallelic synonymous SNPs was converted to the applicable format for the program Genepop (version 4.2.2; Rousset, 2008) using a Perl script. The dataset was split into population clusters as differentiated by DAPC. The between population differentiation was then determined by calculating the special case of Wright's

F-statistic (F_{ST}) to describe the repartition of allelic frequency between subpopulations.

Intra-cluster variance

Synthetic genes, containing SNP sites and sites identical to the reference that passed the respective coverage thresholds, were used to quantify the genetic diversity in subpopulations that were determined by clustering analysis. The program DnaSP (version 5.10.1; Librado and Rozas, 2009) was used to compare loci between individuals within each cluster. The average and standard deviation of the Watterson theta estimate ($\hat{\theta}_W$) across all sites were calculated to obtain the genetic diversity estimate within each cluster. A characteristic of DnaSP is that it cannot differentiate between intra-individual (between haplotypes) and inter-individual diversity (between isolates). It means that when the diversity of a population is computed, it actually considers the haplotype diversity. Every haplotype is considered as one “isolate”. Generally speaking, *Pst* contains two haplotypes, therefore one can compute the diversity with only one isolate. This was not the main focus of this analysis but was conducted on the isolate that was on its own in a genetic group. Haplotype diversity in *Pst* is generally considered to be high and was confirmed by the phased haplotype sequencing effort of Schwessinger et al. (2018).

4.3 Results

4.3.1 Re-sequencing of South African *Pst* pathotypes

To investigate variation in the South African population, whole genome, next-generation sequencing of four historical South African isolates (SA1–SA4) was performed. More than 20 million reads were generated for each isolate using the

Illumina HiSeq2500 platform (Table 4.3). Reads were filtered and subsequently mapped to the PST130 reference genome (Cantu et al., 2011). The average genome depth of coverage across the PST130 genome for SA1–SA4 was between 25 and 39 $\times$ (Table 4.3). All four alignments spanned 97 % of the breadth of the reference genome with at least 2 $\times$ coverage depth.

4.3.2 Purity assessment of samples

To assess whether the urediniospores used as starting material consisted of a single genotype, allele frequencies for each of the historical South African isolates were analysed. The resulting plots displayed clear peaks at 0.5 (Figure 4.5) and a fairly bell-shaped distribution. Although a pattern such as seen in SA4 is more desirable, SA1–SA3 still followed the expected trend that supports that samples consisted predominantly of a single genotype.

4.3.3 Clustering analyses

Three methods of data clustering were implemented to infer population structure. First, a maximum likelihood RAxML phylogenetic tree was generated, using the third codon position of the synthetic genes. Next, STRUCTURE and DAPC were used to assign isolates to population clusters.

4.3.4 Phylogenetic analysis

To determine the relationship of the historical South African *Pst* isolates to available isolates from the UK, France, Pakistan, Eritrea, Ethiopia and Kenya, phylogenetic analyses using available genomic and transcriptomic data from 48 *Pst* isolates (Table 4.1) were carried out. To characterise the genetic relationship between these isolates, a maximum likelihood approach was used. The third codon position across 5844 predicted genes, including 2 437 462 sites, were used

Table 4.3: Statistics of read alignment of the historical South African isolates to the PST130 reference genome. An average of $85.2 \pm 4.0\%$ of filtered reads mapped to the reference genome

Lab code	Platform	Pathotype	Total number of reads	Filtered reads	Percent discarded	Number of reads aligned	Unmapped reads	Average depth of coverage
SA1	Illumina Hi-Seq	6E16A-	23 031 402	22 827 102	0.89 %	20 131 984	2 695 118	30
SA2	Illumina Hi-Seq	6E22A-	22 628 648	22 433 194	0.86 %	16 490 301	5 942 893	25
SA3	Illumina Hi-Seq	7E22A-	26 876 262	26 637 762	0.89 %	23 960 896	2 676 866	36
SA4	Illumina Hi-Seq	6E22A+	30 300 476	30 056 556	0.81 %	26 751 160	3 305 396	41

69

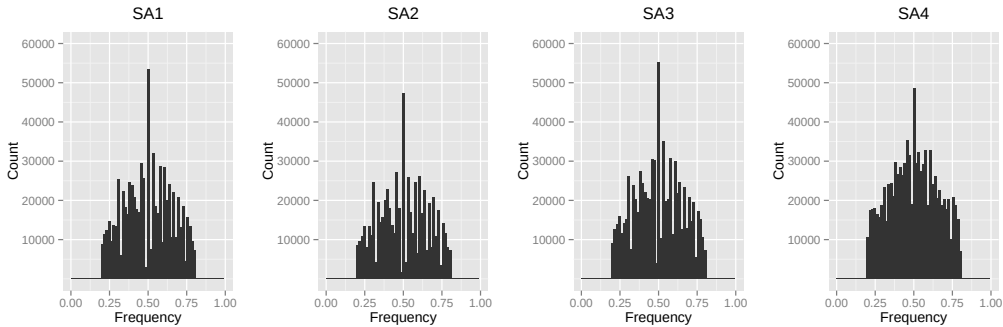


Figure 4.5: Read frequency graphs from heterokaryotic SNP sites for SA1–SA4.

to generate the phylogenetic tree (Figure 4.6), including those sites in genes that had 80 % breadth of coverage in 80 % of the isolates.

From the phylogenetic tree (Figure 4.6), it can be concluded that the South African isolates a) are closely related to one another, and b) are most closely related to isolates from Kenya and Ethiopia. This is indicative of either (i) southward movement of inocula, with the South African pathotypes being derived from East African isolates, or (ii) that the South African and the identified East African isolates may share a common origin.

4.3.5 Population structure analysis

STRUCTURE

To assign individual *Pst* isolates to population groups the Bayesian model based clustering method STRUCTURE (Pritchard et al., 2000) was applied to the 146 400 biallelic synonymous SNP sites that were identified across the 48 isolates.

The log probability plot in Figure 4.7(i) confirmed the optimum number of population clusters as 4, with the graph reaching a plateau parallel to the x -axis for 4 or more population clusters (Pritchard et al., 2000). The number of population clusters was also evaluated using the Evanno method of population cluster analysis (Evanno et al., 2005). This method, based on the second order derivation of the maximum likelihood estimation of the model given a specific K , suggested the population number $K = 2$ (Figure 4.7(ii)). From these two estimates of K , STRUCTURE suggests the number of population clusters is either $K = 2$ or $K = 4$. Figure 4.8 displays bar charts representing STRUCTURE population clusters. To further assess population structure, STRUCTURE results were compared to DAPC clustering that does not assume Hardy-Weinberg equilibrium.

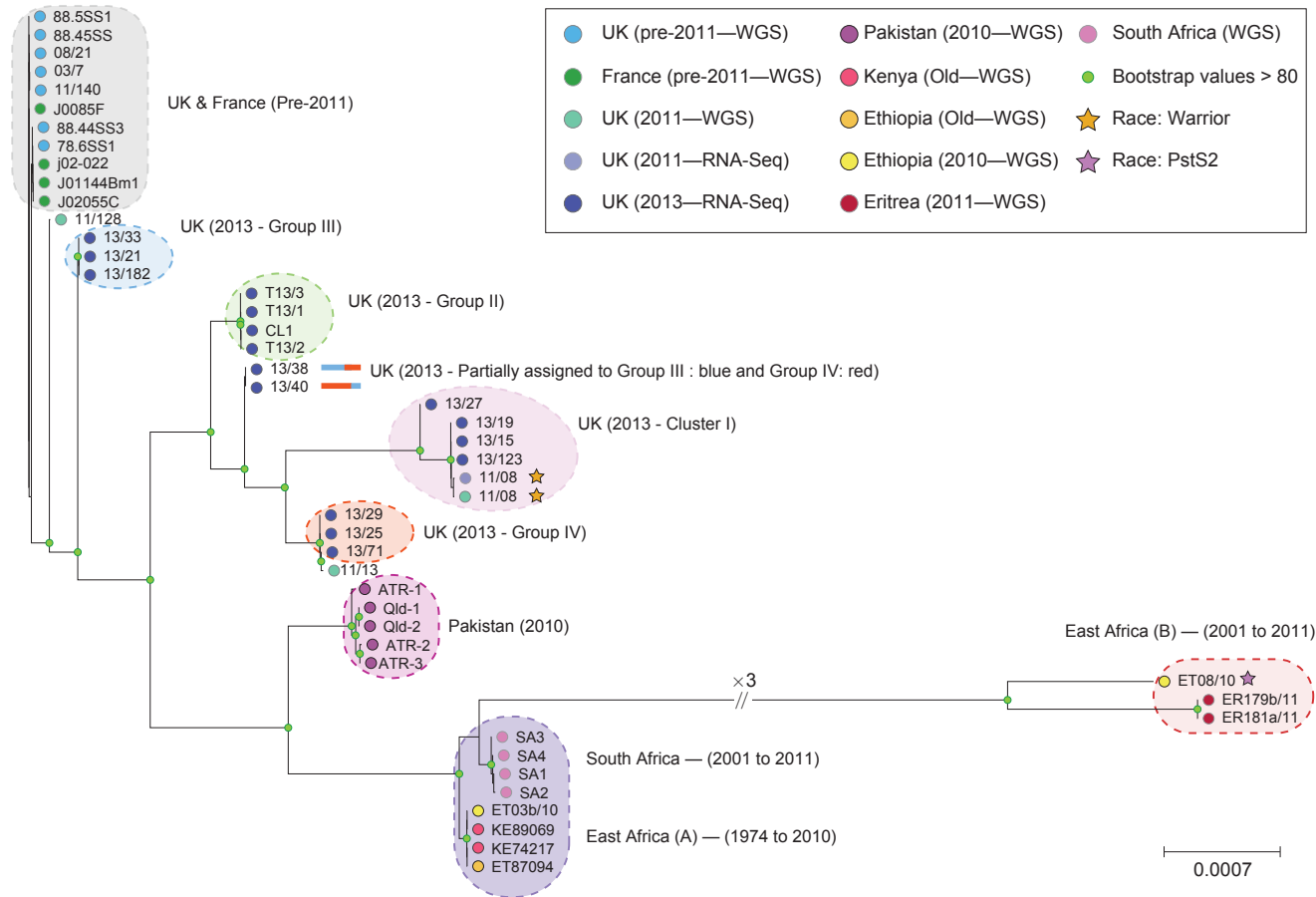
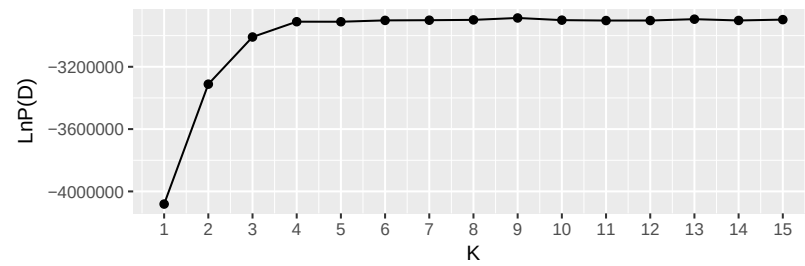
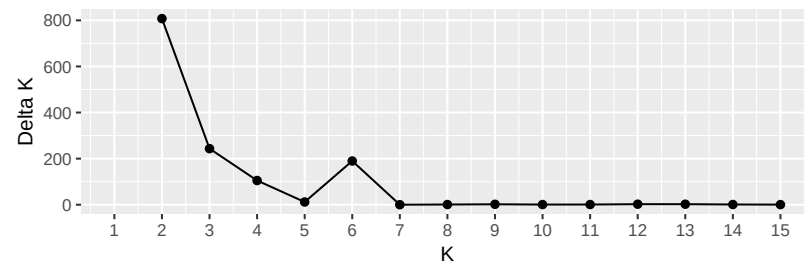


Figure 4.6: The phylogenetic relationship between the South African *Pst* isolates and European, Asian and East African isolates. South African *Pst* isolates are closely related to isolates from East Africa. RAxML non-rooted phylogenetic analysis were performed assessing four South African and 44 global *Pst* isolates using the third codon position of 5844 PST130 gene models. Only those genes that had 80 % coverage in 80 % of the isolates were included, resulting in the inclusion of 2 437 462 sites to construct the tree. Clades are supported by evaluation of 100 bootstrap iterations. Bootstrap values of greater than 80 are indicated with green dots on applicable nodes.



- (i) Log probability of data $L(K)$ as a function of K to identify the optimal amount of clusters. The population structure of *Pst* inferred by model based Bayesian cluster analysis of genome-wide SNP data indicate the optimum number of clusters $K = 4$.



- (ii) The Evanno method of inferring the number of STRUCTURE populations (K) from the modal value of ΔK . A strong signal was detected for $K = 2$ with where ΔK was at a maximum. Δ , Delta.

Figure 4.7: Evaluation of the number of population clusters following STRUCTURE analyses.

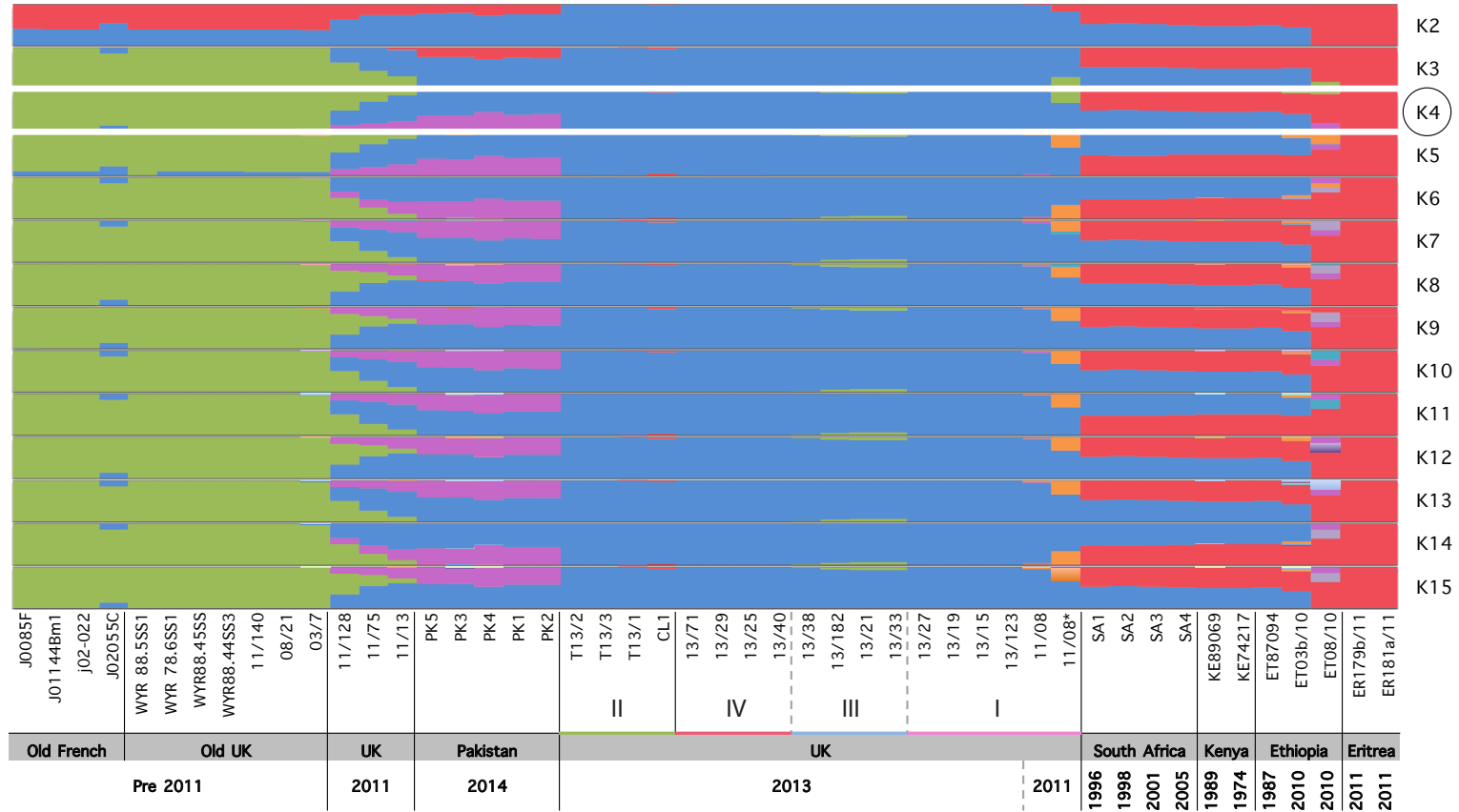
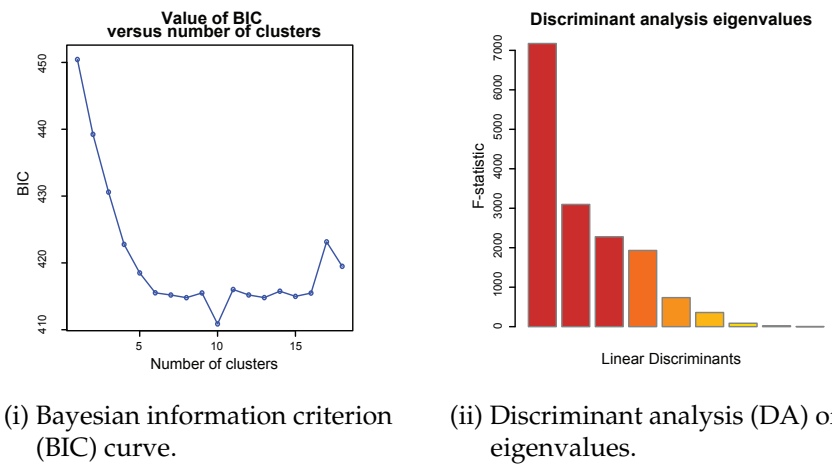


Figure 4.8: Bar charts representing STRUCTURE population clusters, with colour representing a group and each bar indicating the fraction of sites assigned to a specific group representing estimated membership fractions for each individual isolate. The UK 2013 population is divided in subgroups: green (UK Cluster II), red (UK Cluster IV) blue (UK Cluster III) and pink (UK Cluster I) as previously described by Cantu et al. (2013). Asterisk (*) indicates genomic data of isolate 11/08, while no asterisk indicates RNA-Seq data for 11/08. $K = 4$ was proposed as the optimal population number (see Figure 4.7(i)). K_2 to K_{15} indicate the number of clusters individuals in the population were assigned to in each cluster number evaluation.

Discriminant analysis of principal components

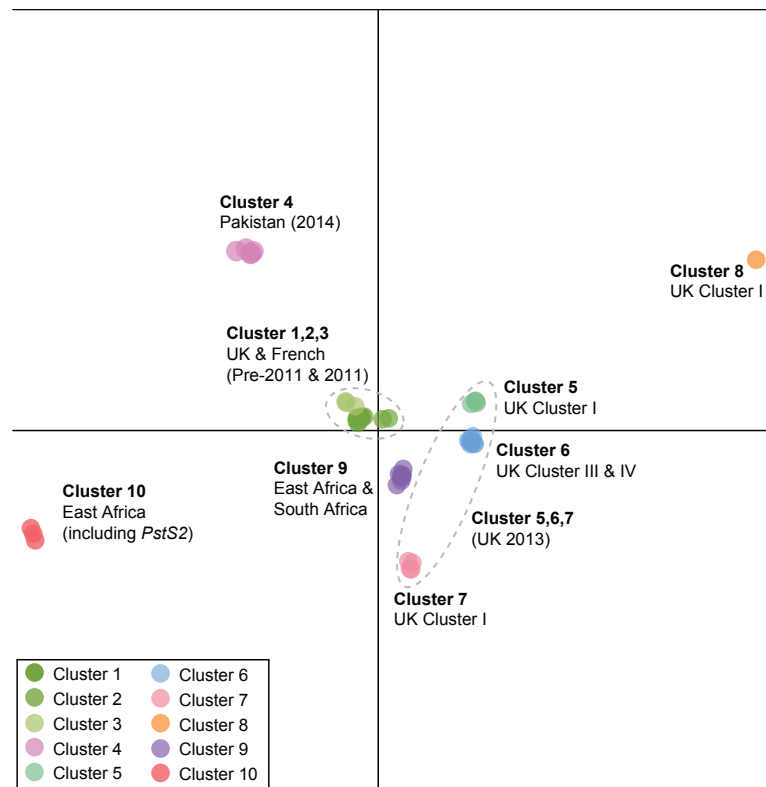
The same 146 400 synonymous biallelic SNP sites were used as input for the analysis. Genetic variation within and between population clusters was then summarised using PCA. The elbow of the Bayesian Information Criterion (BIC) curve formed at 6 and a minimum was observed at 10 (Figure 4.9(i)), indicating the optimum number of clusters ranged between 6 and 10. Discriminant analysis (DA) of eigenvalues was performed to assign individuals to population clusters. The bar-plot in Figure 4.9(ii) represents the DA of eigenvalues for the main principal components. The scatterplot (Figure 4.9(iii)) uses the first two principal components (the y -axis and x -axis, respectively) of the DAPC of the synonymous SNP sites. Each circle represents a single *Pst* isolate.

The non-parametric DAPC of the *Pst* isolates identified at most ten clusters ($K = 10$), as supported by the BIC curve (Figure 4.9(i)). Some similarities between the STRUCTURE groups and the DAPC groups can be seen (Figure 4.8 and Figure 4.10). The elbow of the BIC curve suggests six populations (Figure 4.9(ii)) (Jombart et al., 2010). The bar charts corresponding to $K = 6$ has similarity to the STRUCTURE bar chart for $K = 4$. Differences between STRUCTURE and DAPC included that UK Cluster I was the fifth cluster to differentiate in DAPC analysis, while the post 2011 UK clusters did not show clear differentiation in the STRUCTURE analysis. Pakistan isolates differentiated at $K = 4$ in the STRUCTURE analysis and only differentiated at $K = 7$ in the DAPC analysis. Due to *Pst* predominantly reproducing asexually, specifically in regions where isolates were obtained from, DAPC is more suitable for the specific dataset. Subsequent analyses were based on DAPC results.



(i) Bayesian information criterion (BIC) curve.

(ii) Discriminant analysis (DA) of eigenvalues.



(iii) Relative proximity of *Pst* population clusters.

Figure 4.9: Discriminant analysis of principal component (DAPC) analysis of 48 *Pst* isolates. (i) Bayesian Information Criterion (BIC) curve suggesting the minimum number of clusters (K) required to explain variation between pathotype clusters to be between 6 and 10. The first nine eigenvalues components from the DAPC analysis (ii), supported the maintenance of three discriminant functions in the DAPC analysis indicated with red bars. (iii) DAPC for 48 *Pst* isolates.

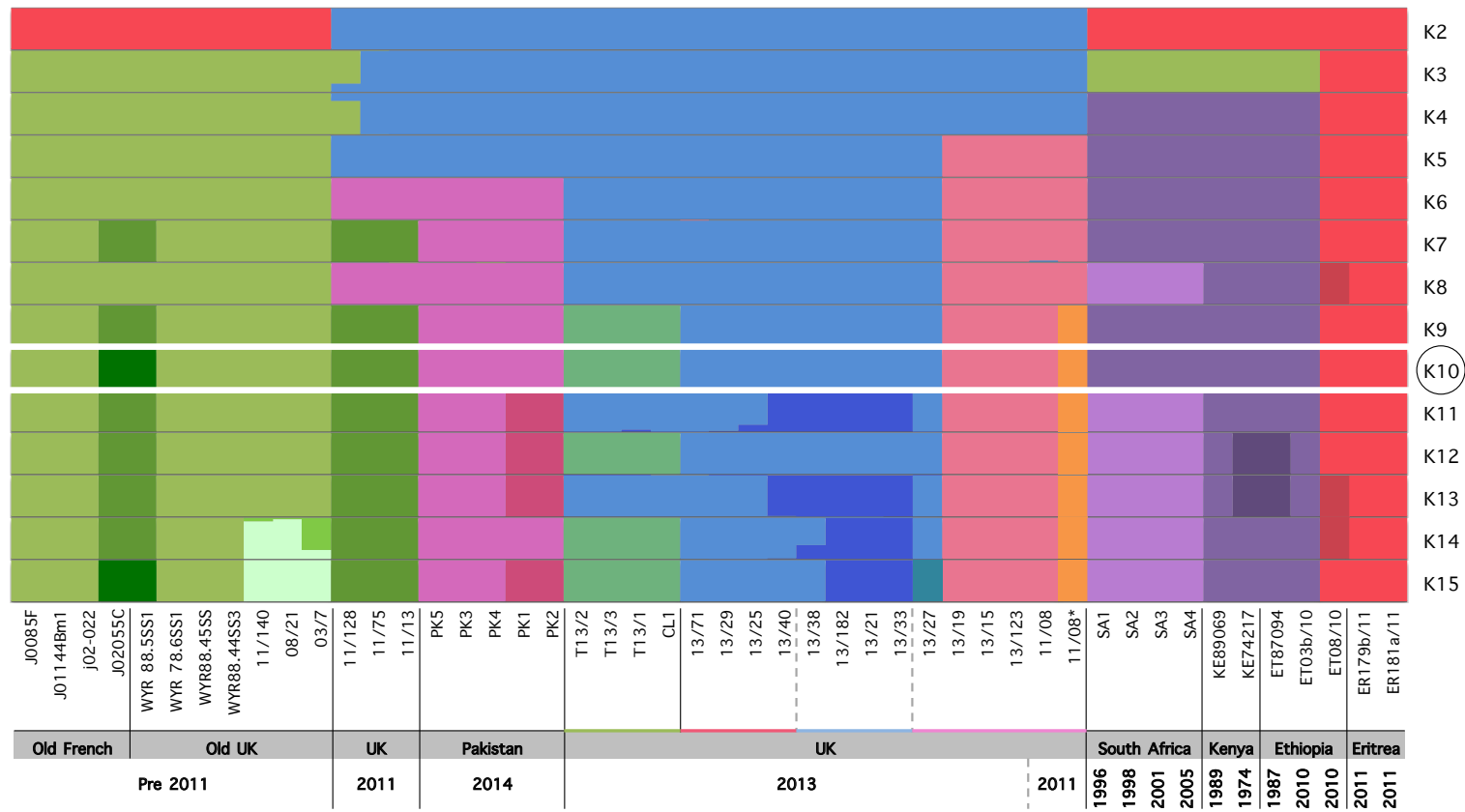


Figure 4.10: Bar charts represent DAPC population structure analysis, with each bar estimating the proportion ascription of each isolate to a population cluster. UK clusters are indicated similar to Figure 4.8. Asterisk (*) indicates genomic data of isolate 11/08, while no asterisk indicates RNA-Seq data for 11/08. K2 to K15 indicate the number of clusters individuals in the population were assigned to in each cluster number evaluation.

4.3.6 Population differentiation

F_{ST} values were calculated using the software Genepop (version 4.2.2; Rousset, 2008) to show genetic differentiation between clusters. Pairwise comparisons of biallelic SNP data were assessed for each group comparison. This analysis quantifies the correlation of alleles within a subpopulation comparing to all subpopulations. Some clusters were very similar and others more divergent with F_{ST} values ranging between 0.08 and 0.86 across the 10 *Pst* clusters (Figure 4.11). The biggest genetic differentiation (0.37 to 0.86) was seen when Group 10—East Africa (B)—was compared to other groups. Comparison between Group 9 and Group 7 showed the highest genetic differentiation involving the South African isolates. Group 7 comprised of the UK Cluster I isolates.

In addition to calculating the F_{ST} values for the population groups as defined by DAPC, this diversity statistic was calculated among the historical South African isolates and isolates from East Africa that were co-arranged by the phylogenetic tree (Figure 4.6) and the clustering analysis (Figure 4.10). The high similarity between these two groups (Group A: SA1-SA4 and Group B: KE89069, KE74217, ET87094 and ET03b/10) was quantified by a very low F_{ST} of 0.08. In contrast, a second group of East African isolates containing two isolates from Eritrea and one Ethiopian isolate, was generally the most genetically diverse from all other *Pst* isolates maintaining high F_{ST} values throughout all comparisons. This genetic difference was also reflected by their position in the distantly related clade in the phylogenetic analysis (East Africa (B); Figure 4.6).

4.3.7 Genetic diversity within and between population clusters

To estimate the genetic variation within the subpopulations the Watterson estimator was used as described in Chapter 3. The Watterson estimator incorporates the number of SNPs and the population size of each population cluster. The

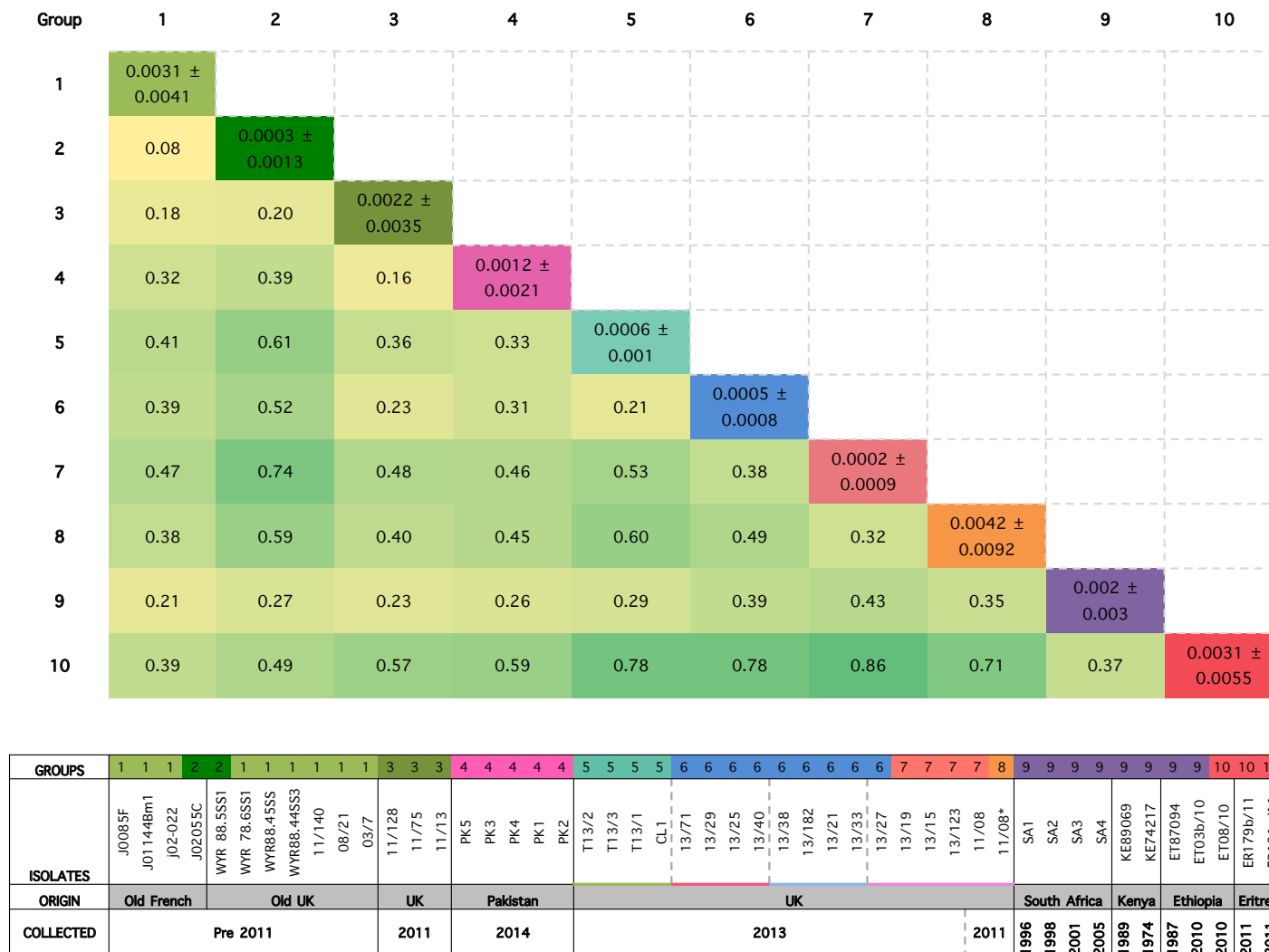


Figure 4.11: Genetic diversity assessed between 10 population clusters derived from DAPC analysis of biallelic SNP data. F_{ST} values are indicated in the lower diagonal matrix, with the diversity in the groups indicated on the diagonal. Group 8 contains one isolate indicating haplotype diversity in this isolate on the diagonal. Isolate information is displayed in the key. The East African isolates in group 9 (purple) are referred to as East Africa I, while group 10 (red) is referred to as East Africa II in the text. Asterisk (*) indicates genomic data of isolate 11/08, while no asterisk indicates RNA-Seq data for 11/08.

degree of polymorphism in the gene set of each subpopulation was calculated by evaluating SNPs across isolates in a population cluster, gene-by-gene. Θ of different clusters, as shown on the diagonal of the matrix in Figure 4.11, can subsequently be compared to assess the relative nucleotide diversity in the different clusters. This metric of Group 8 was calculated on a single isolate and indicates haplotype diversity of this isolate. The highest intra-cluster variability was computed for Groups 1 and 10 and the lowest for Group 7.

4.4 Discussion

To test prevalence and identify new pathotypes of *Pst*, surveys are routinely carried out in South Africa by seasonal phenotyping of rust isolates on a differential set of wheat lines that possess an array of rust resistance genes. Pathotype names, such as 6E16A-, are based on such traditional pathology screens on differential sets. In addition to the pathotype description pathologists often report the virulence profile of specific isolates that show virulence to additional resistance genes not represented in the differential set. These descriptions are complementary, but not necessarily identical across all isolates of a specific pathotype. For example, Ethiopian wheat varieties resistant to *Pst* isolates of pathotype 6E16A- and 6E22A- from South Africa were susceptible to a 6E22 isolate from Germany (Hussein and Pretorius, 2005; Denbel, 2014). Also, different isolates of the 0E0 *Pst* pathotype, showing avirulence to all wheat genotypes with known *Yr* genes, were suggested to be genetically different using microsatellite marker screens (Hovmöller et al., 2016).

In addition to these phenotypic markers, genotyping, using molecular markers has aided in a more detailed description of *Pst* isolates. For instance, South African pathotypes have been genotyped using AFLP markers (Hovmöller et al., 2008) and phylogenetic analysis using these markers indicated that the South

African isolates were related to isolates from Western and Central Asia and Southern Europe. However, the seven isolates, belonging to the pathotype groups 6E16A-, 6E22A-, and 7E22A-, collected between 1996 and 2001 could not be differentiated using AFLP markers. A subsequent study using microsatellite markers that genotyped South African isolates collected between 1996 and 2004 also indicated a close relationship with Central Asian and Mediterranean *Pst* isolates (Ali et al., 2014). Only a single genotype was recorded for the six South African samples tested. More recently the pathology characterisation of the virulence profiles of the South African isolates has been complemented with genotype information from microsatellite markers. The diversity in these molecular markers successfully distinguished the South African isolates (Visser et al., 2016). The close relationship of the South African pathotypes and the stepwise development of new pathotypes were confirmed through this analysis.

Further to this work, the current study implemented a next-generation sequencing approach to determine the possible origin and characterise the genetic relatedness of the four historical South African *Pst* pathotypes identified in 1996, 1998, 2001 and 2005, through investigation of isolates SA1–SA4. First, population substructure was assessed based on allele frequencies at multiple loci of neutral or nearly neutral alleles. After that, the F_{ST} was calculated to quantify genetic variation between the predefined population clusters (Pritchard et al., 2000) and the diversity amongst isolates in a group was assessed. Knowledge of population structure is valuable in the study of emerging and re-emerging pathogens as it reports the dynamics of subpopulations with distinct pathogenicity (Hubbard et al., 2015). In this study, the Bayesian clustering method STRUCTURE (Pritchard et al., 2000) and multivariate DAPC (Jombart et al., 2010) were used to identify genetic clusters.

It is often hard to meet the assumptions analysis methods rely on. STRUCTURE is one of the most popular methods to infer population structure. It was

developed to be applied to various markers that are not closely linked, and assumes Hardy-Weinberg equilibrium (Pritchard et al., 2000). The high marker density obtained from re-sequencing data, together with the asexual reproduction of *Pst*, resulted in violation of this prerequisite, making STRUCTURE less appropriate for analysing clonal populations. An additional shortcoming of STRUCTURE is that the complex models include many parameters to estimate, causing lengthy runtimes when assessing large data sets (Jombart et al., 2010), as is the case with sequence data. In contrast, DAPC first transforms the data using PCA to prepare the input variables to the DA to be uncorrelated principle components. The DA then predicts a grouping variable using one or more of the principle components. This approach is time efficient, and can easily be applied to large re-sequencing datasets. In DAPC, like in STRUCTURE, *K*-means clustering is run with different numbers of clusters (*K*). The clustering models resulting from each chosen *K* can be assessed by their likelihood. DAPC uses BIC to determine the model that fits the data best and by implication the number of clusters (Jombart et al., 2010). After assessment of population structure, the genetic differentiation between and within proposed clusters can be calculated to quantify the diversity between and within groups.

In the pairwise comparisons of clusters, lower F_{ST} values indicate groups that are closely related, while groups distant from each other have high F_{ST} values. Phylogenetic and clustering analysis illustrated that from the isolates evaluated in this study, the historical South African isolates were most closely related to isolates from East Africa (A), also confirmed by the low F_{ST} of 0.08. Higher genetic differentiation between East African and South African isolates ($F_{ST} = 0.23$) was previously reported using microsatellite markers (Ali et al., 2014). In the present study, high differentiation was observed between East African isolates, with an F_{ST} of 0.37 observed between Group 9 (containing East Africa (A)) and group 10 (East Africa (B)). Group 9 and Group 10 included isolates from Ethiopia

sampled in 2010. This indicates high diversity in the *Pst* population in Ethiopia and that the South African isolates is closely related to some of the East African isolates, but different to others.

Diversity calculations amongst isolates assigned to groups using DAPC indicated that groups 2, 5, 6 and 7 were less diverse by one order of magnitude compared to groups 1, 3, 4, 9 and 10. Group 8 consists of a single isolate and the diversity calculation represent the haplotype diversity for this isolate. This high haplotype diversity is a characteristic of *Pst*. Schwessinger et al. (2018) describe the haplotype diversity measured in *Pst*-104E higher than a number of plant pathogens, including *Puccinia coronata* Corda f. sp. *avenae*, *Zymoseptoria tritici* (Desm.) Quaedvl. & Crous and *Verticillium dahliae* Kleb., and associates this diversity with long-term asexual reproduction.

One isolate in Group 10, ET08/10, has previously been assigned to the pathotype *PstS2* (Ali et al., 2017). This aggressive pathotype possibly originated in East Africa and quickly spread to the Middle East, Australia, and Europe. In aggressive pathotypes like *PstS2* generation time is shortened and it is able to infect in spite of relatively warm and dry climates (Hovmøller et al., 2008; Walter et al., 2016).

From this analysis, it was concluded that the closest relatives of the South African isolates were a group of isolates from East Africa. As the East African isolates included historical isolates that date back to the 1970s and 1980s, this result supports the hypothesis that inoculum could have moved southwards from East Africa with subsequent introduction to South Africa. The East African isolates showing high similarity to the South African isolates also included a more recent isolate from 2010, indicating that the historical pathotypes are likely still occurring in Ethiopia. Alongside these pathotypes, new pathotypes have clearly developed, as reported for the aggressive pathotypes *PstS1* and *PstS2*, for example. Group 10 included two isolates from Eritrea that was sampled

in 2011 and the *Pst*S2 2010 isolate from Ethiopia. The historical South African isolates showed significant differentiation from this group. Previous studies that speculated about the origin of South African *Pst* excluded East Africa as a possible origin based upon the diversity observed between South African and Eritrean isolates (Hovmøller et al., 2008; Ali et al., 2014). These studies did not include isolates from Ethiopia.

Apart from considering an incursion from East Africa, the South African and some East African isolates could also share a similar origin. To assess their relationship with isolates from Central and Western Asia and the Mediterranean, suggested to be the origin of the South African isolates (Hovmøller et al., 2008; Ali et al., 2014), the same resolution of variation assessment would be needed for historical isolates from these regions. Currently, molecular marker work suggests East African isolates to have originated from the Middle East (Ali et al., 2014) and isolates sampled from this region, at different time points in the past, should also be considered to unravel possible origins.

From this study, it cannot be confirmed that the South African isolate SA1 is closely related to the 6E16 pathotype found in Southern and Northern Europe (Enjalbert et al., 2005; Hovmøller et al., 2008). Although samples from the same regions and possibly the same time frame were considered, the samples did not overlap between the current study and the work of Enjalbert et al. (2005) and Hovmøller et al. (2008).

4.5 Conclusion

Based on genomic analysis, this study confirms the association between the South African and East African *Pst* populations previously proposed through pathotype analysis (Pretorius et al., 1997; Boshoff et al., 2002; Pretorius et al., 2007). In future, similar next-generation sequencing analysis of Central and

Western Asian, Mediterranean and Middle Eastern isolates would fill in the missing information to be able to draw parallels between the traditional marker work and the next-generation sequencing data analysis included in this work.

From the samples analysed in this work, it was demonstrated that the South African isolates are closely related to one another, which supports the findings of the microsatellite marker work of Visser et al. (2016) that stepwise evolution is likely responsible for the consecutive pathotypes. This hypothesis is further assessed in Chapter 5 when polymorphisms in the South African isolates will be analysed in search of the evolutionary changes that gave rise to subsequent pathotypes of *Pst* in South Africa.

Chapter 5

Analyses of Polymorphisms in Historical South African *Pst* Isolates in Search of Candidate Effector Genes

MANY FILAMENTOUS PLANT PATHOGENS, such as *Pst*, use effector proteins to manipulate their hosts (Kamoun, 2007). These proteins also put the pathogen at risk of being recognised by the host via the resistance (R) proteins leading to an incompatible interaction (Rovenich et al., 2014). A change in amino acid sequences could lead to the host defence mechanisms not being able to recognise the pathogen. This inability results in a compatible interaction where the pathogen is virulent on host genotypes that were previously able to detect the attack and restrict or stop infection. In this study, *Pst* isolates collected from a wide geographical area were assessed using different clustering analysis methods to assign isolates to population clusters (discussed in Chapter 4). It was concluded that the historical South African isolates that were collected between 2001 and

2011 (Table 4.2) are closely related, while their closest relatives outside South Africa are isolates from East Africa. In this chapter differences and similarities among these South African isolates were further explored. In particular, to gain an understanding of how the different pathotypes became established in the population. In accordance, a search for candidate genes that could be involved in the specific virulence of individual isolates was conducted using three approaches: i) polymorphisms in the genomes were evaluated to determine whether selection pressure could be detected, ii) the presence or absence of selected genes and the impact that such inclusion or exclusion could bring about was investigated and iii) genes of interest with regards to virulence were identified through isolate specific nonsynonymous polymorphisms in putative effector coding genes.

5.1 Introduction

To obtain nutrients from the host for its own development, *Pst* must grow in infection structures able to bridge host structural barriers, while simultaneously trying to avoid recognition by the host's molecular defence mechanisms (Garnica et al., 2014). To achieve this, *Pst*, like other filamentous plant pathogens, makes use of a diverse set of proteins called effector proteins which the pathogen uses to manipulate host metabolism for its own advantage in cases where it can escape the host's ETI (see Section 2.3.1). These proteins have critical roles during the infection process and fulfil specific tasks with accurate timing at particular locations inside the host (Hogenhout et al., 2009; Stergiopoulos and de Wit, 2009).

Two major groups of effectors exist, namely apoplastic and cytoplasmic effectors. Among the apoplastic effectors are toxins and cell wall degrading proteins, which are important for necrotrophs, freeing up nutrients by degrading plant tissues. For hemibiotrophic and biotrophic pathogens, a more subtle approach is needed, in which the integrity of the host cell is preserved, allowing the pathogen

to obtain nutrients from living tissues. These groups of pathogens rely more on intracellular effector proteins to modify the host cellular environment (Dou and Zhou, 2012; Stotz et al., 2014). Biotrophic fungi, like rusts, make use of haustoria to deliver fungal effectors into the plant's living cells (Garnica et al., 2014). Some genotypes of the host have the ability to recognise these cytoplasmic effector proteins that activate ETI, triggering a cascade of defence processes that reduce or completely halt ingress of the pathogen. Genetic changes within the plant, or elimination or modification of effector genes by the pathogen, can prevent recognition of pathogen invasion by the host's defence system (Dodds et al., 2006). This underpins what is commonly known as resistance gene mediated, pathotype-specific resistance. This type of resistance leads to the classic "Boom-and-Bust" cycle described for *R-Avr* interactions in phytopathology (McDonald, 2004).

5.1.1 The importance of *Pst* variability

Isolates with a pathotype that enables the pathogen to remain undetected, or which is able to overcome plant defence systems, will become established in the population. In addition to gene flow, genetic recombination and mutation can introduce genetic variability within the population that enables *Pst* pathotypes to continue to evolve and overcome host resistance.

Genetic recombination can occur during sexual reproduction in the form of sexual recombination, or in asexual populations through somatic recombination. Somatic recombination is believed to be rare in *Pst* (Little and Manners, 1969), however recent evidence in the stem rust gene, *AvrSr50* indicates somatic recombination as the mode of action to overcome *Sr50* (Chen et al., 2017). This illustrates the significance of somatic recombination to be responsible for new variation. Sexual recombination in *Pst* requires an alternative host to wheat.

Although *Pst* susceptible *Berberis* and *Mahonia* species have been found in South Africa, providing the opportunity for sexual reproduction, infection by *Pst* has not been observed in nature (Visser et al., 2016). The apparent stepwise changes in virulence seen in South African *Pst* isolates further confirms the absence of sexual recombination in South African *Pst* populations, suggesting that variation in the *Pst* population in South Africa might be mostly due to mutations (Visser et al., 2016).

5.1.2 Mutations—causes, types and effects

Mutations occur naturally due to errors in DNA replication (Griffiths et al., 2015), spontaneous DNA lesions (Bienko et al., 2005) and by the action of mobile elements within the genome, called transposons (Klug, 2012). The mutation rate is the number of mutations that occur in a gene or organism in a given time period. Natural mutations vary between genes within an organism and occur at different rates across species (Drake et al., 1998; Scally, 2016). In general mutation rates are low in most organisms, but this depends on evolutionary forces, the life history of the organism and chance events (Drake et al., 1998). Agents called mutagens can accelerate the rate of mutation. A wide variety of mutagens exist, and they induce different types of mutations. Physical mutagens such as radiation from the invisible light spectrum can cause chromosomal aberrations, including chromosomal inversions, chromosomal arm deletions, duplications and repeat expansions, for example, ultraviolet light can cause various types of mutations with distinct properties for each wavelength component UVA (320—400 nm), UVB (280—320 nm), and UVC (200—280 nm) (Pfeifer et al., 2005).

Some chemicals react directly with DNA, for example, ethyl methanesulfonate (EMS) and sodium azide induce SNPs in the form of random point mutations (Rao and Sears, 1964; Olsen et al., 1993). Mutagens can cause diseases such as

cancer in mammals (Ames, 1979), but are also used in functional genomic studies to develop populations used in reverse genetic techniques such as targeting induced local lesions in genomes, or TILLING (Henikoff et al., 2004). A common challenge in these approaches is that many of the mutated individuals are in a compromised condition, highlighting that beneficial mutations are rare. Most mutations are either neutral or deleterious and not conserved in the population (Kimura and Ohta, 1969).

In the absence of gene flow and genomic recombination, mutations are the main source of genetic variation. Natural selection removes harmful mutations from the population through a reduced ability of affected individuals to grow and reproduce. A carrier of a beneficial mutation will have enhanced fitness traits and therefore will be able to pass the mutation on to the next generation. Such a mutation will likely become fixed in the population (Hartl and Clark, 1998). Mutations that are passed on to the next generation increase gene polymorphisms, for example, multiple alleles of the same gene in the species (Salemi et al., 2009). In a deterministic model of evolution, changes in allele frequency depend on fitness and selection, assuming an infinitely large population size. On the other hand, the stochastic model acknowledges the influence of genetic drift, that increases as the effective population size decreases. Depending on the phenotype of the mutation—whether it is advantageous, neutral or deleterious—and the effective population size, population evolution is more influenced by either drift or natural selection (Salemi et al., 2009).

Polymorphisms outside coding regions are not usually under strong selection pressure, however, depending on where SNPs occur in intron splice sites of pre-mRNA, they may interfere with alternative splicing operations during or shortly after transcription. This can lead to altered levels of mRNA, modified mRNA, or a complete shift in the reading frame. Additionally, in coding regions, synonymous SNPs can occasionally have functional consequences due to alterations in the

structure and stability of the translated protein, but are generally considered to maintain the integrity and function of the protein. Nonsynonymous SNPs, however, result in amino acid changes, which can significantly change the protein. These SNPs can have an effect on the function of the resulting protein and the phenotype.

Mutations within genes

When a mutation results in a purine (nucleotides G and A) being substituted with another purine, or a pyrimidine (nucleotides C and T) with another pyrimidine, it is called a transition, while substitution of a purine with a pyrimidine, or vice versa, is called a transversion (Salemi et al., 2009). Although there are twice as many possible transversions mutations compared to transitions, transitions are 10 times more common than transversions because of chemical and steric properties (Klug, 2012; Griffiths et al., 2015).

Mutations do not always result in a functional change in the protein encoded by the gene. A silent mutation or synonymous mutations describes a codon change that does not alter the amino acid in the encoded protein due to degeneracy in the genetic code. Most synonymous mutations are considered to be selectively neutral, but may alter RNA secondary structure and stability (Salemi et al., 2009). In addition, tRNA molecules can vary in abundance, which is important for the success of translation. Mutations in genic regions can be missense or non-sense. Missense being single point mutations that result in amino acid changes, while non-sense mutations introduce early stop codons that truncate proteins. Some consider conservative missense mutations synonymous, as in the case where similar chemical properties or structures are encoded by the new amino acid, for instance, leucine and isoleucine that are both aliphatic. Nonsynonymous mutations describe a mutation where the new codon specifies an amino

acid with different chemical properties from the amino acid it replaces.

In this study, silent mutations are regarded as synonymous mutations and missense and non-sense mutations as nonsynonymous mutations (Miyata and Yasunaga, 1980; Li et al., 1985; Nei and Gojobori, 1986). Mutations can also interfere with gene expression if they occur in a promotor region of a gene or at the splice site of an intron. Mutations in these regions of the gene are not considered in this study.

5.1.3 Genomic approaches used to identify effectors

Effector annotation used in this chapter relied on the bioinformatics pipeline developed by Saunders et al. (2012). The pipeline provides a basis for candidate effector gene identification. It first clusters secreted proteins into protein families and classifies and ranks these protein families for their likeliness to be effectors. Using a modified version of this pipeline, Cantu et al. (2013) annotated the PST130 transcriptome, identifying genes encoding candidate effectors and ranking these to generate a top 100 tribe list that contained high priority candidate effector genes. Due to the biotrophic nature and the infection structures produced by *Pst*, effector proteins are likely to be secreted. Therefore, at first, the pipeline screened the predicted proteome for candidates with secreted signals. Markov clustering was then used to group secreted and non-secreted proteins into protein families using sequence similarity with secreted proteins. Thirdly, tribe annotation was carried out based on sequence homology, after which a search for conserved motifs was performed. Individual members of secreted protein families were annotated based on features they share with known effectors. Through hierarchical clustering of tribes, a priority list was compiled for functional validation of candidates that were most likely effectors.

In this chapter, the focus was on the investigation of SNPs found between

the genomes of the four historic South African *Pst* pathotypes, with specific concentration on the protein coding regions of predicted effector genes to link specific *Pst* virulence profiles with nucleotide polymorphisms within these effector genes. The effector feature annotations, ranking protein tribes according to their probability of containing effectors, were used (Saunders et al., 2012; Cantu et al., 2013).

5.2 Materials and methods

The genomes of four South African historical isolates, representing the four pathotypes found in South Africa, were sequenced, mapped to the PST130 reference genome, and polymorphisms were identified, as described in Chapter 3.

5.2.1 SNP analysis

From the SAMtools mpileup files, with coverage information of each position, Perl and Python scripts were used to find SNPs with at least 10× depth of coverage and to identify homokaryotic and heterokaryotic SNPs (see Chapter 3).

SNP effect prediction

SnEff software (version 3.6; Cingolani et al., 2012) was used to predict the effects of the polymorphisms and to investigate the frequency of transitions and transversions in the gene space. SnEff distinguishes SNP location and type, including characterisation of nonsynonymous and synonymous SNPs in coding regions, which indicates introduced or lost stop codons, lost start codons and changes in splice sites and introns. For this analysis, a bed format file of each isolate's SNP set was prepared using BEDTools (version 2.17.0; Quinlan and Hall, 2010) and the annotation information of the PST130 genome. The bed file was converted into a SnEff input file using a Perl script. The predicted effects of SNPs in the gene space were evaluated with specific focus on the introduced stop

codons and synonymous and nonsynonymous polymorphisms. Codon positions of SNP sites that introduced stop codons were evaluated, and the gene positions where stop codons occurred were considered to evaluate any biases that could indicate the effect on the resulting protein. The frequency of specific nucleotide changes resulting in transitions and transversions were determined and further evaluated to determine biases in codon positions for specific nucleotide changes.

5.2.2 Positive selection

The program Yn00 (Yang and Nielsen, 2000), which is part of the PAML package (Yang, 2007), was used to assess genetic diversity through polymorphism and positive selection analysis using the synthetic genes described in Chapter 3. A pairwise comparison that yielded a nonsynonymous substitution rate or dN value of more than zero indicated a polymorphic gene, while positive selection was considered when a dN/dS value that indicates the rate of nonsynonymous vs synonymous polymorphisms, also called the omega value, of more than one was observed. Perl scripts were used to enable the automated use of Yn00 on the PST130 gene set (Cantu et al., 2013).

5.2.3 Presence-absence analysis

Unique presence and absence of genes were investigated to identify possible associations between specific genes and a gain in virulence in the four South African isolates. The read coverage of each gene was calculated using BEDTools (version 2.17.0; Quinlan and Hall, 2010). Genes with zero coverage were considered absent from the specific isolate (Cantu et al., 2013). The nucleotide and amino acid sequences of these genes were used to query publicly available databases using the basic local alignment search tool (BLAST version 2.6.0; Altschul et al., 1997) to find homologous genes in related species and orthologs in the PST130 reference genome.

5.2.4 Comparisons of nonsynonymous SNP sites between isolates

An additional method to investigate polymorphisms across isolates was used. Polymorphic sites predicted to cause nonsynonymous changes were identified and nucleotides at these positions, across isolates, were compared in a pairwise manner. The number of nucleotide sites at which a difference in nucleotides between two isolates was observed was used as a distance statistic in an unweighted pair group method with arithmetic mean (UPGMA) tree, indicating the relationship of isolates to one another in terms of the number of nonsynonymous changes. The list of genes showing differences between each pairwise comparison was compared to the list of candidate effector genes and the list of secreted proteins generated by Cantu et al. (2013). These lists were generated as described in Section 5.1.

5.2.5 Multiple sequence alignments to visualise biallelic SNPs

A custom Python script was developed to visualise translated proteins of candidate genes indicating the presence of alternative amino acids due to nonsynonymous polymorphisms. Where coverage was lower than $2\times$ at nonpolymorphic sites or $10\times$ at polymorphic sites, manual inspection of the genome was done using Integrative Genomics Viewer (IGV version 2.3.91; Thorvaldsdóttir et al., 2013). In cases where the low coverage sequence was the same as in the other South African isolates, these nucleotide sequences were included in the figure, but indicated with lighter shading. Blank spaces indicate isolates with no sequence information. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), indicating specific categories.

5.3 Results

5.3.1 SNP identification in the genomes of the historical South African isolates

Polymorphism data provides information on how a population is evolving. After filtering the Illumina paired end reads and independent mapping of each of the four South African isolates to the PST130 draft reference genome (as described in Chapter 3), SNPs were identified across the whole genome, using SAMtools mpileup. Variant sites were only taken into account in cases where a coverage depth of 10 reads or more was seen.

The four isolates displayed similar SNP frequencies with $0.62 \pm 0.12\%$ of the genomes containing polymorphisms when compared to the PST130 reference, resulting in an average rate of heterozygosity of 6.25 ± 1.15 SNPs/kbp. Heterokaryotic SNPs were polymorphic to the reference, being biallelic or multi-allelic, while homokaryotic SNPs were monoallelic. Heterokaryotic SNPs were in the majority and averaged $92.96 \pm 0.18\%$ of all variant sites across the four isolates, with a SNP density of 5.81 ± 1.06 SNPs/kbp, a high number comparing to the 1.51 SNPs/kbp found on *Melampsora larici-populina* Kleb., the sexually reproducing poplar rust fungus (Persoons et al., 2014). The remaining $7.04 \pm 0.18\%$ of variant sites comprised of homokaryotic sites occurring at a frequency of 0.44 ± 0.09 SNPs/kbp (Table 5.1).

Determining the genetic impact of polymorphisms

Information regarding polymorphisms in genes can be used to determine the impact of the variant on the resulting protein. Identifying the nature and location of SNPs show how the pathogen changes on the genetic level, including changes related to its pathogenicity phenotype. To determine the nature and genome position of polymorphisms, the SNPs identified in SA1 to SA4 (Table 5.1) were

Table 5.1: Homokaryotic and heterokaryotic SNPs in the South African isolates

Isolate	PST130 reference sites	Total number of SNPs	% of reference	SNPs/kb	Homokaryotic			Heterokaryotic					
					Monoallelic			Biallelic	Biallelic	Multiallelic		Total	
					One alternative allele			One alternative allele	Two alternative alleles	Three or four alternative alleles			
					Number	%	SNPs/kbp	Number	Number	Number	Number	%	SNPs/kbp
SA1	64 782 816	378 259	0.58	5.84	25 975	6.87	0.40	351 719	228	337	352 284	93.13	5.44
SA2	64 782 816	324 200	0.50	5.00	22 788	7.03	0.35	300 839	211	362	301 412	92.97	4.65
SA3	64 782 816	414 489	0.64	6.40	28 853	6.96	0.45	384 958	275	403	385 636	93.04	5.95
SA4	64 782 816	501 728	0.77	7.74	36 588	7.29	0.56	464 344	334	462	465 140	92.71	7.18
Average			0.62	6.25	7.04			0.44			92.96		5.81
Standard deviation			0.12	1.15	0.18			0.09			0.18		1.06

annotated using SnpEff. Across the four isolates $29.93 \pm 0.20\%$ of SNPs were within genes, of which 52.74% resulted in synonymous substitutions, while 47.26% represented nonsynonymous substitutions. Loss or gain of start and stop codons can also have major effects on translation, resulting in complete loss of translation or truncated peptides. Table 5.2 describes the major predicted effects of polymorphisms in genic regions in the four South African isolates.

Table 5.2: The number of SNPs identified in coding regions of the four South African *Pst* isolates

Isolate	Location of polymorphism		
	Synonymous coding	Nonsynonymous coding	Stop gained
SA1	58 868	52 499	3 347
SA2	50 008	44 829	2 933
SA3	59 595	53 481	3 380
SA4	71 140	64 278	3 992

Between about 3000 and 4000 SNPs resulted in stop codons (Table 5.2). The three stop codons are TAA, TAG, and TGA. C to T mutations in the first codon position often introduces stop codons in the gene space (Hane and Oliver, 2010). In the second and third codon position, SNP sites where changes to an A or G occur, are responsible for the introduction of stop codons. The majority (99.4%) of SNPs that introduced stop codons were biallelic/heterokaryotic. G to Y (C or T) mutations occurred most frequently (29.2%), followed by C to R (A or G) at 17.5% at the second codon position and 14.7% at the third codon position. Biases in SNP type at codon positions were assessed in Figure 5.1. Patterns of nucleotide changes were conserved between isolates.

To identify the impact of introduced stop codons, the gene positions where a stop codon was introduced were evaluated for possible patterns in occurrence (Figure 5.2). No distinct trend was observed, and it appears that stop codons are introduced with no particular preference, randomly appearing in the gene.

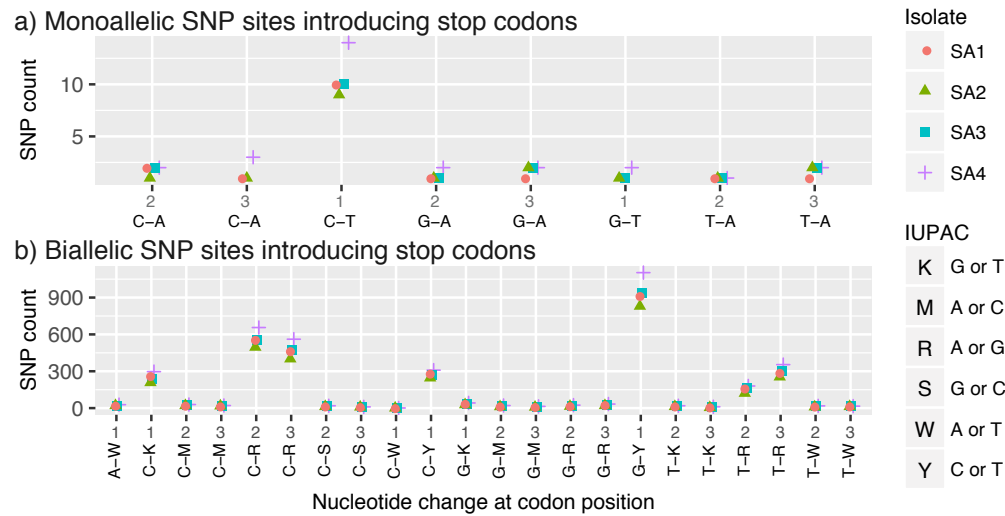


Figure 5.1: Nucleotide changes that introduced stop codons were highly conserved between isolates. A small number of monoallelic SNPs (0.6 %) were responsible (a), but 99.4 % of stop codons were introduced at biallelic SNP positions (b). Numbers indicate codon positions 1, 2 or 3. Nucleotide changes are indicated underneath the codon position, the first nucleotide indicating the reference nucleotide and the second, the polymorphism nucleotide(s).

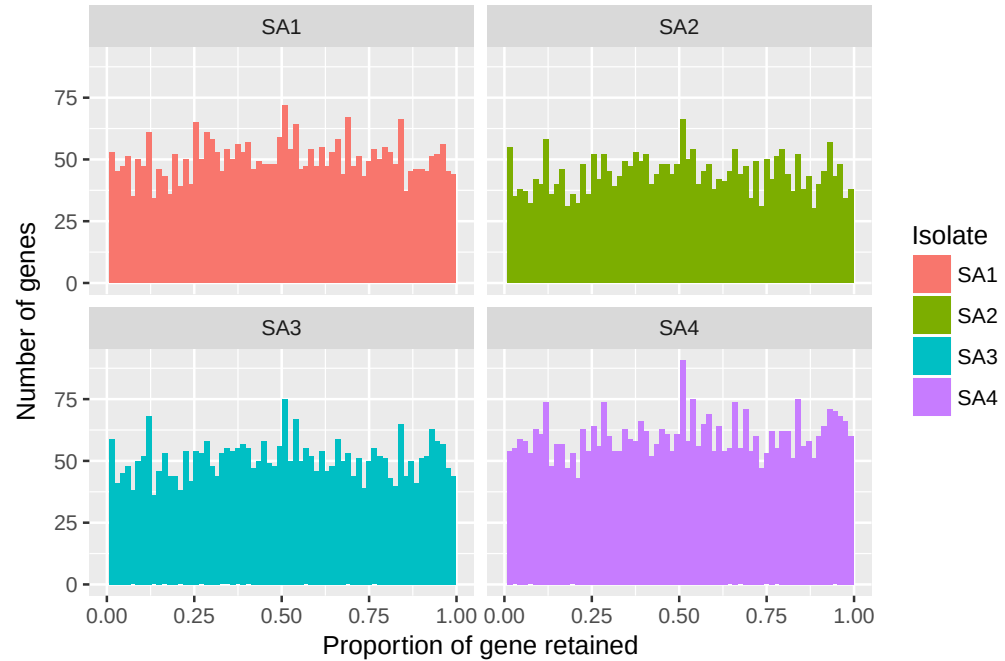


Figure 5.2: Distribution of introduced stop codons across all genes per isolate. The bar charts show the number of genes with a specific gene proportion retained after a stop codon was introduced.

Frequency of transitions and transversions at polymorphic sites

The SNPeff information was used to determine whether mutations represented transversions or transitions. When considering the frequency at which transitions occurred in comparison with transversions, more transitions than transversions occurred at SNP sites, as expected. At synonymous SNP sites, C to T transitions were most common, while A to G transitions occurred most frequently in non-synonymous SNPs at homokaryotic sites. At homokaryotic SNP sites, where synonymous substitutions were observed, a transition to transversion ratio of 2 : 1 was displayed, while a 3.5 : 1 ratio was observed for nonsynonymous substitutions (Figures 5.3 and 5.4). Similar to the finding in Figure 5.1, Figures 5.5 and 5.6 indicated conserved patterns in the specific nucleotide changes at codon positions 1, 2 and 3, respectively.

5.3.2 Assessment of polymorphisms to detect positive selection

This SNP data reveals information about how the population is evolving. Highly polymorphic genes are more likely linked with improved fitness and being under positive selection. The dN/dS statistic, which assesses the ratio of non-synonymous polymorphisms to synonymous polymorphisms, was evaluated to identify genes that are under selection. The term “ dN ” describes nonsynonymous polymorphisms that replace an amino acid and “ dS ” describes synonymous polymorphisms where the amino acid remains unchanged. SNPs within all genes annotated within the PST130 reference genome (18 023 genes) were compared in a pairwise isolate analysis. It is commonly expected that synonymous sites will evolve more neutrally and that changes in allele frequencies would be due to random chance (genetic drift). In contrast, a polymorphism that affects fitness will evolve more rapidly due to its selective advantage.

Synthetic, consensus genes were created for each isolate that incorporated SNPs that had a 10× or higher coverage and where nonpolymorphic sites had

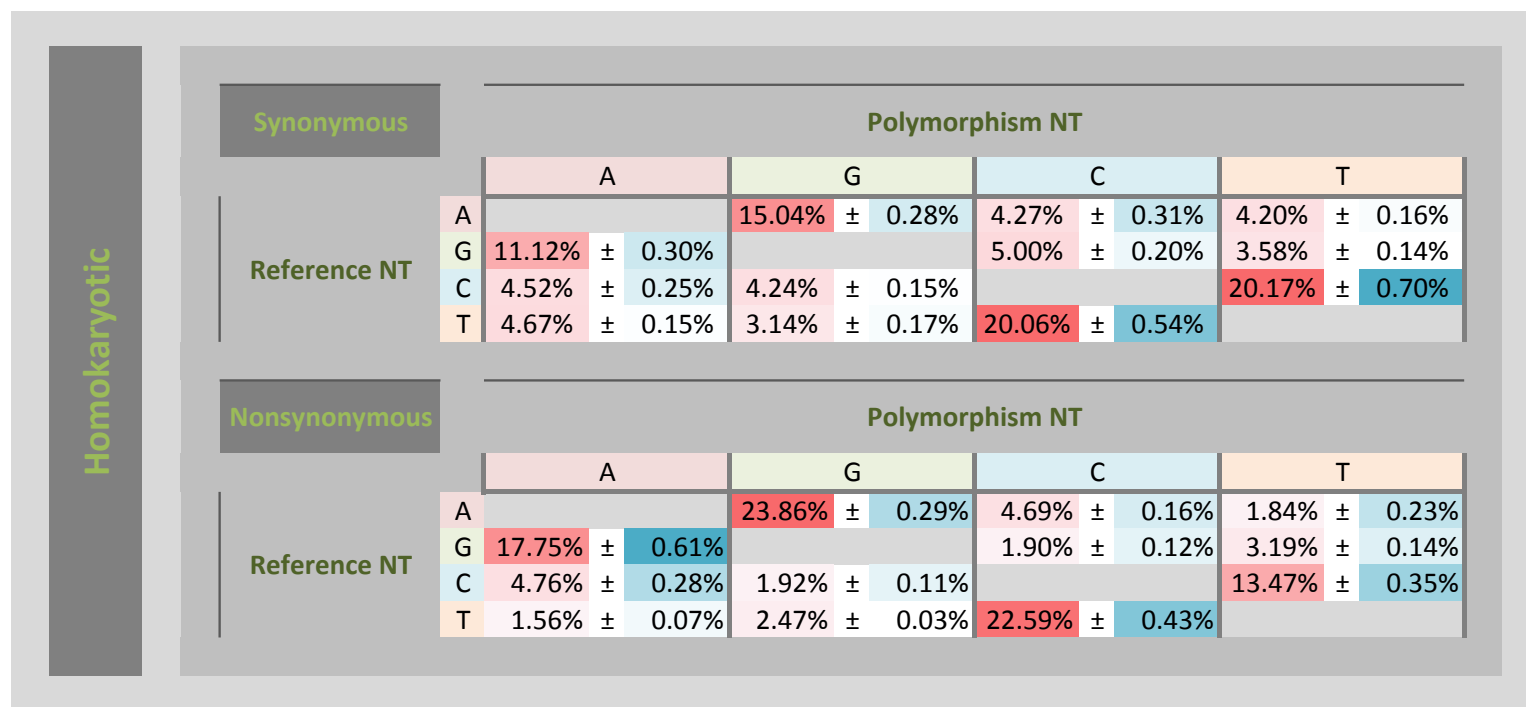


Figure 5.3: Percentage frequency matrices of transitions and transversions at monoallelic SNP sites. In both synonymous and nonsynonymous substitutions, transitions were more frequent compared to transversions. Darker red indicates a higher percentage and darker blue a higher standard deviation.

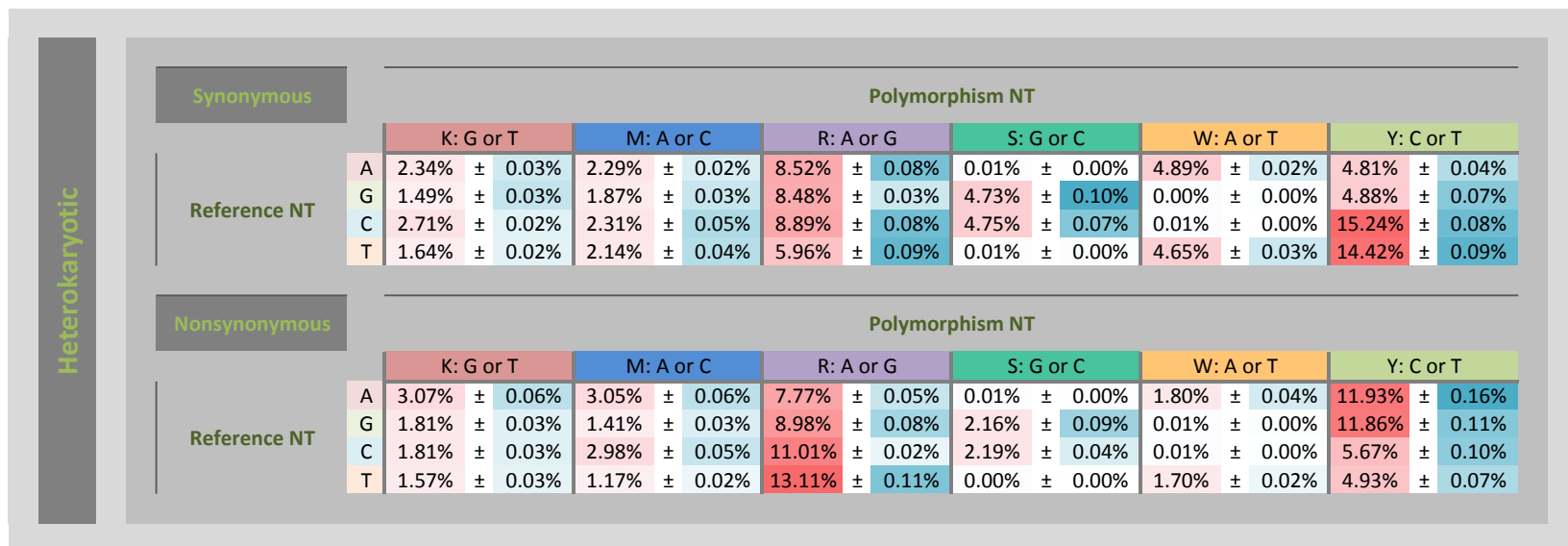


Figure 5.4: Percentage occurrence matrices of transitions and transversions at biallelic SNP sites. Biallelic SNP sites showed a high transition frequency of 14 % to 15 % for C and T to Y (C or T), and 8.5 % for A and G to R (A or G) at synonymous sites. For nonsynonymous sites transition occurrences were still fairly high with an average of 6.84 % across all possible transitions. However, transversion occurrences were more frequent at 11.98 %. Darker red indicates a higher percentage and darker blue a higher standard deviation.

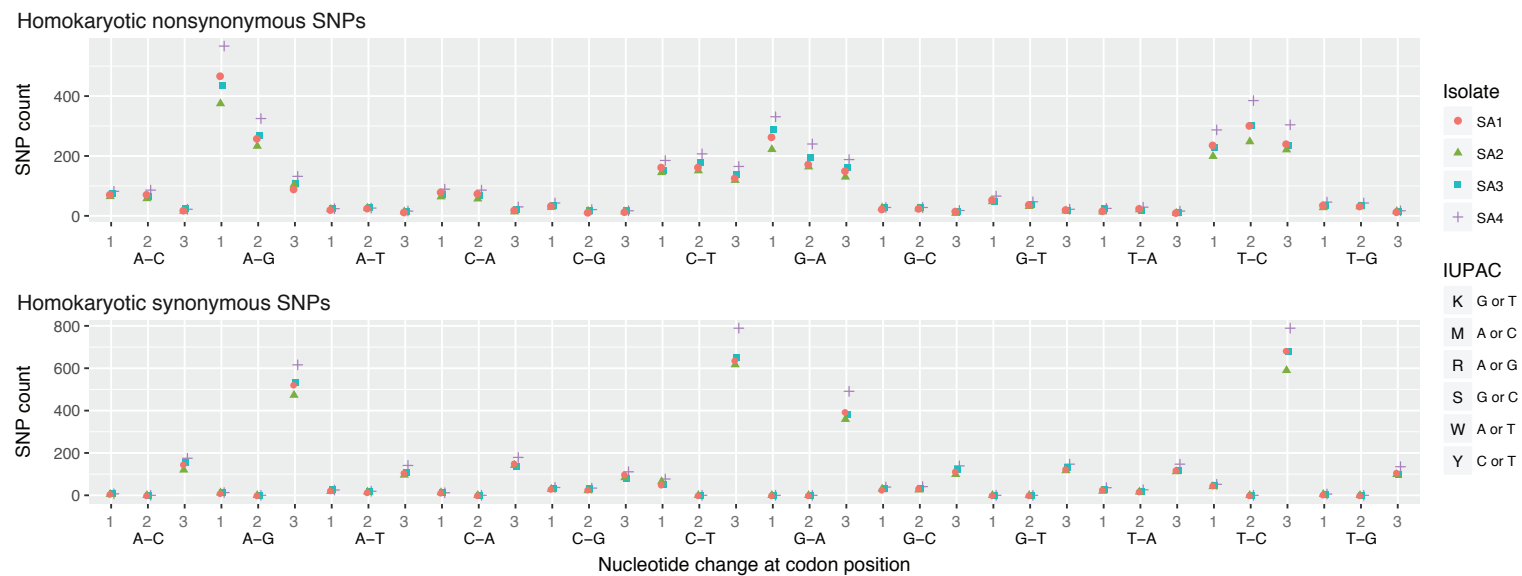


Figure 5.5: Codon positions of nucleotide changes at homokaryotic SNP sites explained broadly in terms of transitions and transversion in Figures 5.3 and 5.4.

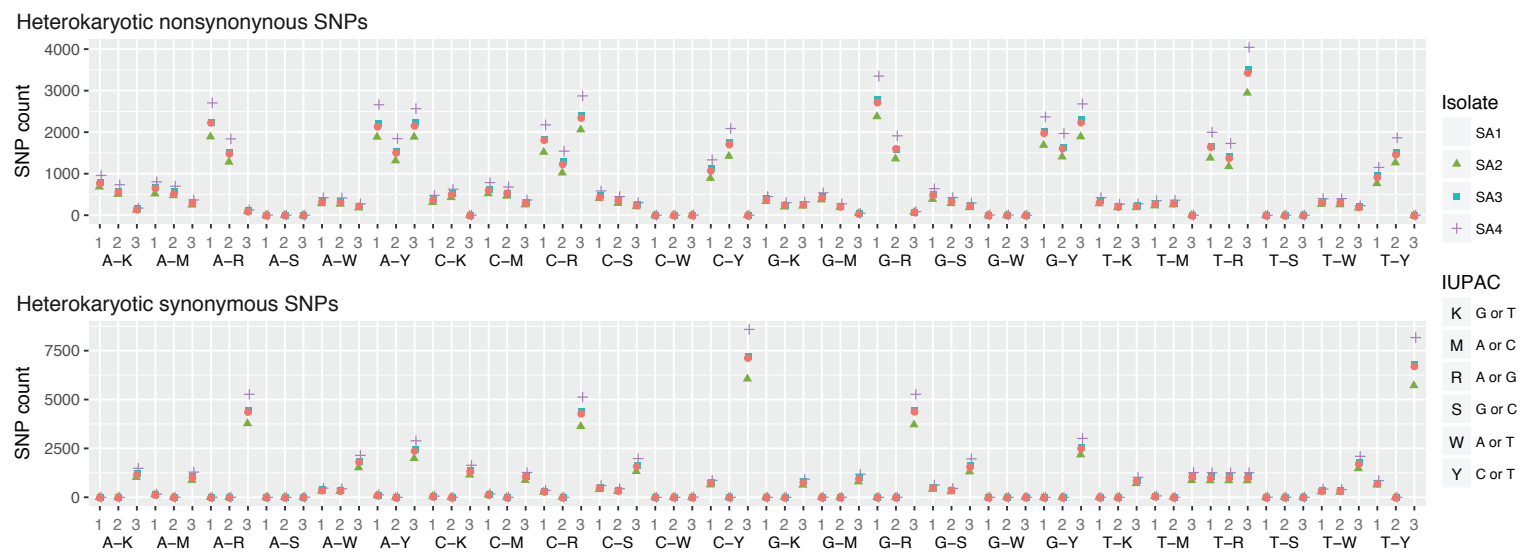


Figure 5.6: Codon positions of nucleotide changes at heterokaryotic SNP sites explained broadly in terms of transitions and transversion in Figures 5.3 and 5.4.

at least $2\times$ coverage of the PST130 reference gene (see Section 3.3.5). Pairwise isolate comparisons of each consensus gene were carried out using the YN00 program in the PAML package. Pairwise comparisons yielding positive dN values indicated that the specific gene under investigation was polymorphic between the two isolates. Alternatively, where positive dS values were obtained, genes were considered to have evolved more neutrally.

No signals for positive selection were detected, as no dN/dS values, also known as omega values, of greater than 1.0 were observed. Only seven genes were given a positive dN value in the pairwise comparisons of the South African isolates, while positive dS values were computed for two genes. There were no genes in common and therefore all dN/dS values were undefined. These nine genes (Tables 5.3 and 5.4) were not investigated further as they did not display characteristics of genes coding for secreted proteins or putative effectors, as identified in the lists reported by Cantu et al. (2013), and were therefore not considered likely candidates for pathogenicity factors.

5.3.3 Presence or absence of genes

Elimination of an effector gene and its resulting protein could aid the pathogen to escape host recognition. Similarly, specific genes may enhance the pathogenicity and reproducibility of the pathogen. Therefore, in addition to point mutations, inclusion or exclusion of entire genes was also assessed to look for associations of genes with virulence phenotypes. After monitoring whether there were genes in the PST130 reference genome that were not covered by read sequences from the South African isolates, 211 genes were found to be absent in all four the South African isolates. In addition, there were 36 genes that were absent in three or fewer of the South Africa isolates, in different combinations, that were present in the reference genome of PST130 (Table 5.5).

Table 5.3: Polymorphic genes with positive dN values indicating nonsynonymous changes in isolate pairwise comparisons

Gene	SA1 vs SA2	SA1 vs SA3	SA2 vs SA3	SA1 vs SA4	SA2 vs SA4	SA3 vs SA4
PST130_03694	0	0	0	0.0009 ± 0.0009	0.0009 ± 0.0009	0.0009 ± 0.0009
PST130_07979	0.0002 ± 0.0002	0	0.0002 ± 0.0002	0	0.0002 ± 0.0002	0
PST130_09146	0	0	0	0.0013 ± 0.0013	0.0013 ± 0.0013	0.0013 ± 0.0013
PST130_10326	0	0.0013 ± 0.0013	0.0013 ± 0.0013	0.0013 ± 0.0013	0.0013 ± 0.0013	0
PST130_10374	0.0036 ± 0.0036	0	0.0036 ± 0.0036	0	0.0036 ± 0.0036	0
PST130_11223	0	0	0	0.0017 ± 0.0017	0.0017 ± 0.0017	0.0017 ± 0.0017
PST130_17618	0	0	0	0.0006 ± 0.0006	0.0006 ± 0.0006	0.0006 ± 0.0006

Table 5.4: Polymorphic genes with positive dS values indicating synonymous changes in isolate pairwise comparisons

Gene	SA1 vs SA2	SA1 vs SA3	SA2 vs SA3	SA1 vs SA4	SA2 vs SA4	SA3 vs SA4
PST130_00923	0.0037 ± 0.0037	0	0.0037 ± 0.0037	0	0.0037 ± 0.0037	0
PST130_04022	0	0	0	0.0015 ± 0.0015	0.0015 ± 0.0015	0.0015 ± 0.0015

Table 5.5: Number of absent genes in the four South African *Pst* pathotypes. This includes a total of 247 genes, where 211 genes were absent in all four isolates and 36 genes that were absent in one to three isolates

Isolate	Pathotype	Number of absent genes
SA1	6E16A-	211 + 11
SA2	6E22A-	211 + 13
SA3	7E22A-	211 + 19
SA4	6E22A+	211 + 18

Figure 5.7 displays genes that are absent in the South African isolates. Presence-absence genes may be involved in virulence of the pathogen, however, none of these genes was on the list of putative effector genes (Cantu et al., 2013). The number of genes absent in a single isolate increased with the increase in virulence. (See Appendix B, Table B.1, for gene names of the 211 genes that were absent in all four South African isolates).

A BLAST search against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide databases using default parameters, revealed homology in other plant pathogens for eight of the 211 genes absent in all South African isolates (Table 5.6). Investigations of the functionality of the *Pgt* homologs were undertaken, and characteristics are listed in Appendix B, Section B.2. As redundancy often exists in genomes of filamentous plant pathogens (Dangl and Jones, 2001) a BLAST search of the 211 genes against the PST130 transcriptome was performed. Of the 211 genes, 152 had one or more potential paralogs within the PST130 genome (Table 5.7).

Table 5.6: Potential orthologs of genes absent in all four South African isolates. All orthologs identified were from fungi, besides one ortholog from the oomycete, *Albugo laibachii*

PST130 gene	Homolog	PST130 length	Match length
PST130_00159	<i>Pgt</i> isoleucyl-tRNA synthetase (PGTG_09131), mRNA	252	252
PST130_07080	<i>Pgt</i> hypothetical protein (PGTG_01952), mRNA	252	133
PST130_16763	<i>Pgt</i> hypothetical protein (PGTG_02128), mRNA	798	431
PST130_17182	<i>Pgt</i> hypothetical protein (PGTG_02971), mRNA	270	247
PST130_17354	<i>Pgt</i> hypothetical protein (PGTG_20899), mRNA	1 188	345
	<i>Pgt</i> glycogen [starch] synthase (PGTG_07651), mRNA	1 188	562
PST130_17620	<i>Pgt</i> hypothetical protein (PGTG_15464), mRNA	141	136
PST130_17815	<i>Pgt</i> 1,3-beta-glucan synthase component FKS1 (PGTG_00125), mRNA	666	535
PST130_06262	<i>Albugo laibachii</i> Nc14, genomic contig CONTIG_2252_NC14_v4_941_117	210	175
	<i>Rhynchosporium orthosporum</i> mitochondrion, complete genome	210	196
	<i>Rhynchosporium secalis</i> mitochondrion, complete genome	210	196
	<i>Rhynchosporium commune</i> mitochondrion, complete genome	210	196
	<i>Rhynchosporium agropyri</i> mitochondrion, complete genome	210	196

Table 5.7: The number of potential paralogs identified in genes absent in all four South African isolates

Number of genes	Number of potential paralogs in PST130
107	1
22	2
9	3
9	4
2	7
2	5
1	10

In the group of 211 genes that were absent in all four South African isolates, only five genes were coding for secreted proteins according to the lists in Cantu et al. (2013). They were PST130_01946, PST130_03059, PST130_03060, PST130_06608 and PST130_08220. These five genes returned no hits in a BLAST search against NCBI non-redundant nucleotide databases. Two of the genes, PST130_01946 and PST130_03059, had potential paralogs within the PST130 transcriptome with higher than 80 % identity and E-values lower than 0.01 (Table 5.8).

The PST130 paralogs identified in these BLAST hits did not appear in the original list of 247 genes absent across the four South African isolates and therefore were present in the South African isolates. PST130_01946 had four paralogs, while PST130_03059 had one paralog, highlighting the occurrence of redundancy in the *Pst* genome that could be the result of duplication events.

Table 5.8: Potential paralogs of genes absent in the four South African isolates

qseqid	sseqid	% Identity	Length	Mismatch	Gaps	E-value	Bit score
PST130_01946	PST130_00235	95.745	329	11	1	1.04E-15	527
	PST130_10569	89.362	235	25	0	3.21E-80	296
	PST130_08196	87.179	234	30	0	2.51E-71	267
	PST130_11479	87.342	158	20	0	9.36E-46	182
PST130_03059	PST130_02767	92.958	142	9	1	7.54E-53	206

qseqid, query sequence ID; sseqid, subject sequence ID

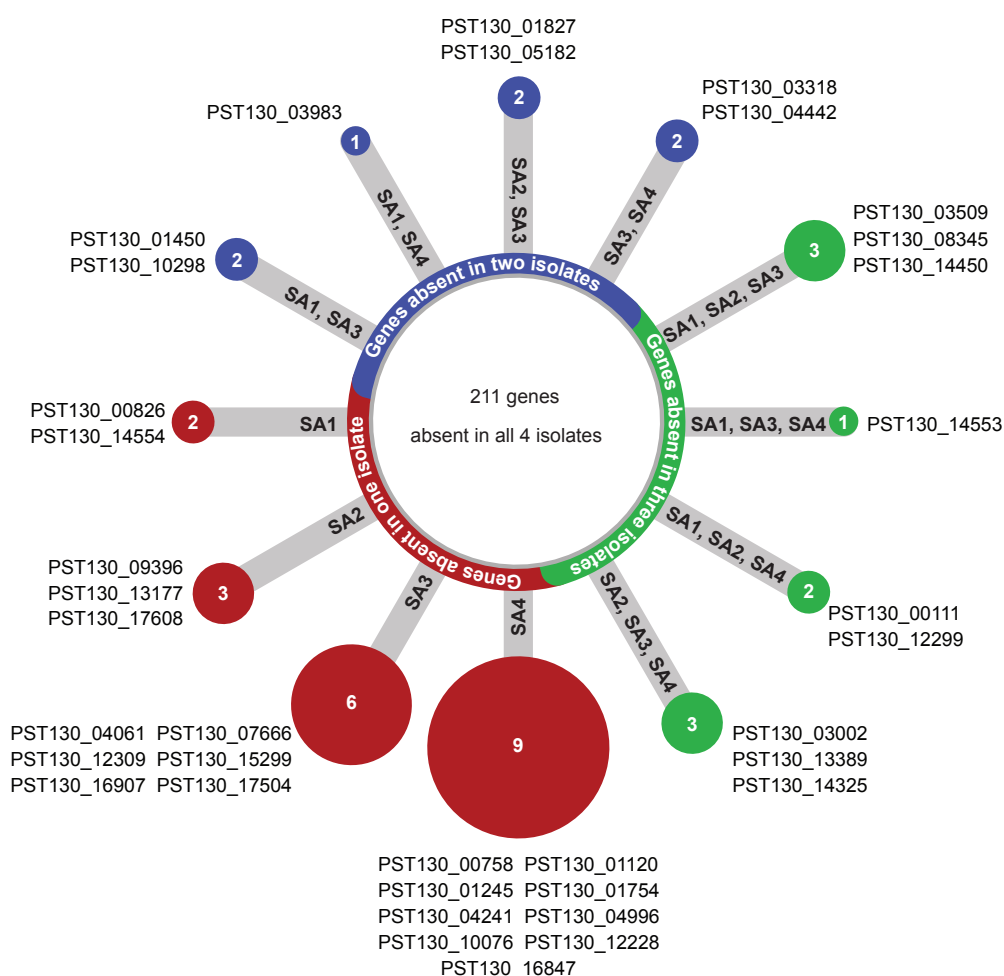


Figure 5.7: Presence-absence analysis revealed 211 genes absent in all four South African isolates and an additional 36 genes absent in some isolates.

Of the 36 genes that were absent in three or less of the South African isolates (Figure 5.7), three had highly similar nucleotide sequences in NCBI non-redundant nucleotide databases with more than 80 % identity and E-values smaller than 0.01 in BLAST searches (Table 5.9). PST130_00758, PST130_08345 (only present in SA4) and PST130_12299 (only present in SA3) had hits with PGTG_02401, PGTG_03886 and PGTG_14583 respectively. However, these three *Pgt* proteins are uncharacterised to date. Conserved domains of the *Pgt* proteins are listed in Appendix B, Section B.2.

Table 5.9: Potential orthologs of genes absent in three or less of the South African isolates

PST130 gene	Homolog	Species
PST130_00758	<i>Pgt</i> hypothetical protein (PGTG_02401)	Fungi
PST130_08345	<i>Pgt</i> hypothetical protein (PGTG_03886)	Fungi
PST130_12299	<i>Pgt</i> hypothetical protein (PGTG_14583)	Fungi

Of the 36 genes absent in three or less of the isolates, nine genes were present in only one of the South African isolates. These nine genes included three genes in SA1: PST130_03002, PST130_13389 and PST130_14325, one in SA2: PST130_14553, two in SA3: PST130_00111 and PST130_12299 and three genes in SA4: PST130_03509, PST130_08345 and PST130_14450. Notable BLAST hits for two of these genes, PST130_12299 and PST130_08345, were obtained showing high similarity with *Pgt* genes as shown in Table 5.9, where they were identified according to their absence in one or more of the isolates. Conserved domains are listed in Appendix B, Section B.2. Of these 36 genes absent in three or fewer isolates, 24 displayed potential paralogs in the PST130 genome (Table 5.10).

Table 5.10: Number of potential paralogs in PST130. Of the 36 genes that were absent in three or less of the South African isolates, 24 had potential paralogs in the PST130 genome. All potential paralog genes were present in all isolates

Number of genes	Number of potential paralogs in PST130
14	1
3	3
3	2
2	4
1	7
1	10

Two potential paralogs were identified in the PST130 genome for PST130_00111 (SA1) and one for PST130_03002 (SA1), PST130_14325 (SA1), PST130_12299 (SA3) and PST130_08345 (SA4) as summarised in Table 5.11.

To investigate possible functions of the present and absent genes, functional

annotation of possible orthologs were assessed (see Appendix B, Section B.2).

Table 5.11: Paralogs of genes that only occurred in one of the South African isolates

qseqid	sseqid	% Identity	Length	Mismatch	Gaps	E-value	Bit score
PST130_00111	PST130_12514	95.78	332	14	0	2.00E-156	536
	PST130_15801	92.77	332	24	0	3.00E-140	481
PST130_03002	PST130_08845	89.36	235	23	2	2.00E-84	294
PST130_08345	PST130_11503	96.73	275	9	0	2.00E-133	459
PST130_12299	PST130_05481	96.21	396	15	0	0	649
PST130_14325	PST130_00979	95.93	246	10	0	1.00E-115	399

qseqid, query sequence ID; sseqid, subject sequence ID

5.3.4 Investigation of candidate genes that are likely to experience evolutionary changes

By comparing heterokaryotic SNPs in the four South African isolates in a pairwise manner, all genes with unique nonsynonymous changes in the four South African isolates were identified. It was found that the number of genes with nonsynonymous mutations increased with an increase in virulences as indicated by the UPGMA dendrogram in Figure 5.8(a). This supports the previous hypothesis of stepwise evolution, with each pathotype derived from the preceding pathotype through single-step mutation events (Visser et al., 2016).

Nonsynonymous heterokaryotic biallelic SNPs that differed between isolates (11 185 SNPs) were observed in 2689 genes. According to the gene annotation of Cantu et al. (2013), 138 of these were predicted to encode secreted proteins (613 SNPs), of which 27 were putative effector proteins (106 SNPs) that could be involved in the specific virulence phenotypes of the four South African *Pst* pathotypes. Figures 5.8 (b), (c) and (d) display the pairwise comparison of isolates, with the number of genes that show nonsynonymous SNPs in each gene set comparison, for example, proteomes, secretomes and effectomes.

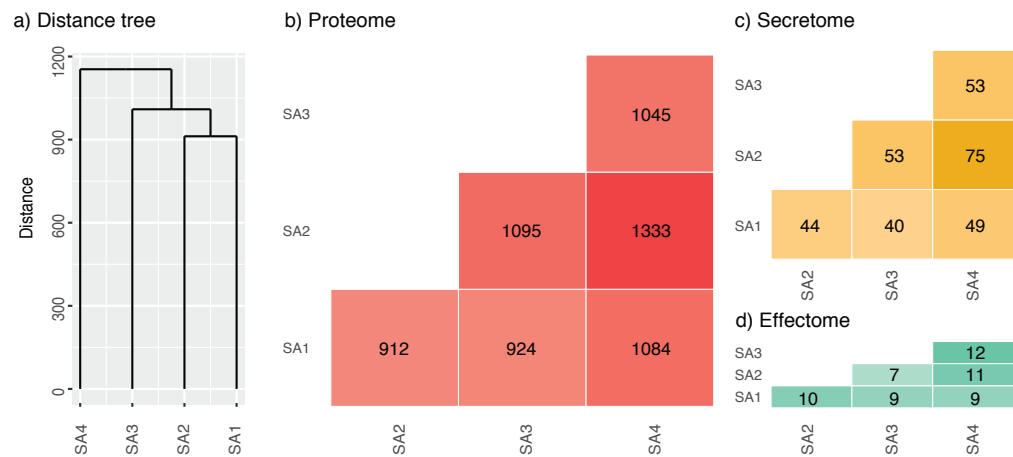


Figure 5.8: Nonsynonymous SNPs in the gene space of the four South African isolates increase over time and with increasing virulence. The branch lengths of the UPGMA distance tree (a) is derived from the distance matrix in (b) and illustrates the progressive accumulation of genes with nonsynonymous mutations over time as new pathotypes developed, given that the population evolved stepwise through mutations. Heat maps indicate frequencies of unique nonsynonymous substitutions in the *Pst* Proteomes (b), secretomes (c) and effectomes (d). UPGMA, unweighted pair group method with arithmetic mean.

5.3.5 Candidate effectors with sequence polymorphisms between the South African isolates

After applying the three assessment methods (positive selection analysis, presence-absence analysis and nonsynonymous polymorphism analysis) to the polymorphic datasets, only genes that were members of the top 100 ranking protein families for effectors as described in Cantu et al. (2013), were considered for further investigation to identify candidate genes that could explain gain-of-virulence. The justification of selection of these 27 candidate genes (Figures 5.8), is shown in Section 6.2, Table 6.1. As an example, Figure 5.9 illustrates five nonsynonymous changes due to heterokaryotic SNPs in one of the 27 candidate genes, PST130_00285. Please consult the Appendix B, Section B.3, for changes in the remaining 26 genes.

A molecular analysis, focussed on a selection of the 27 polymorphic candidate

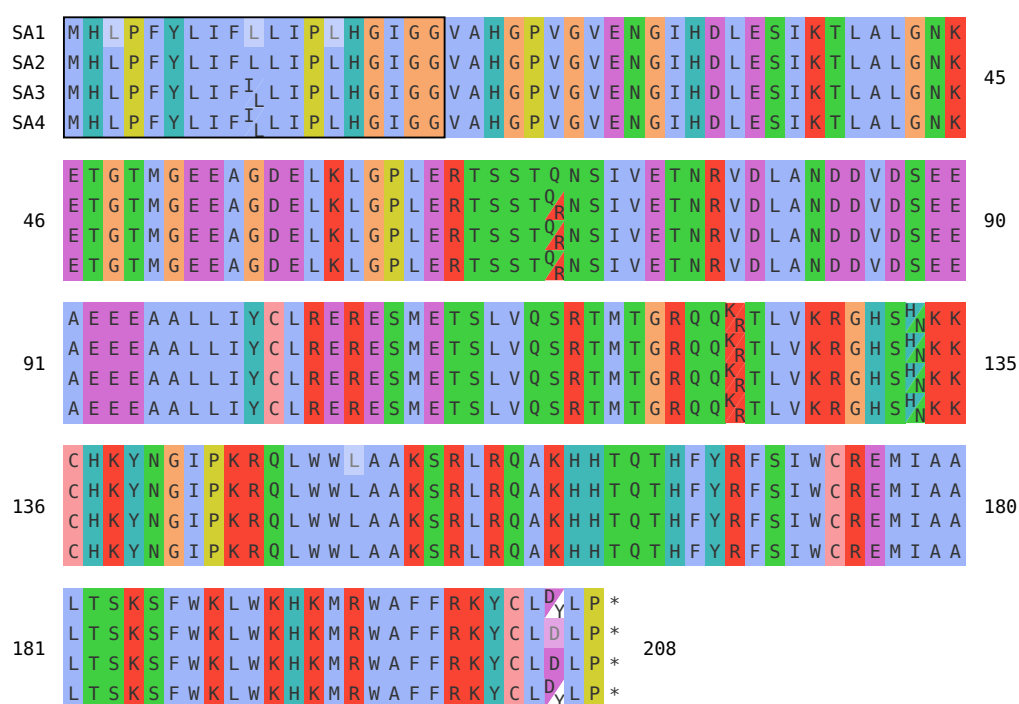


Figure 5.9: Translated sequence alignment of gene PST130_00285. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Please consult the appendix for the sequence alignments of the remaining 26 candidates. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

effector genes that were identified was the next step of investigation and is reported in Chapter 6.

5.4 Discussion

The present study implemented the gene models developed for the PST130 draft genome sequence (Cantu et al., 2011). These gene models have been further assessed for various effector features to create a subset of genes that could likely be involved in pathogenicity (Cantu et al., 2013).

In a clonal population, mutations are the main source of genetic variation. In

this study, the focus was on point mutations causing SNPs—other DNA aberrations were not investigated. Characterisation of SNPs was undertaken to understand how the pathogen changes at the genetic level to achieve changes in its pathogenicity phenotype. SNPs that result in nonsynonymous amino acid changes present an allelic pool of protein variation upon which selection pressures can impact, leading to changes in allelic frequencies within the pathogen population.

5.4.1 Polymorphic sites

SNP analysis showed a higher frequency of SNPs in isolate SA4 compared to isolates SA1, SA2 and SA3. This is expected as the biggest time span between the collection of these isolates was between SA3 and SA4 (seven years), while only one to two years passed between collection of SA1 to SA3 and progressive accumulation of mutations is expected over time (Salemi et al., 2009). The density at which homokaryotic (0.44 ± 0.09 SNPs/kbp) and heterokaryotic (5.81 ± 1.06 SNPs/kbp) SNPs occurred in the South African isolates mapped against the PST130 reference were comparable to SNP densities described by Cantu et al. (2013). The authors investigated five isolates with distinct virulence profiles, two from the UK and three from the USA. These displayed a homokaryotic SNP density of 0.41 ± 0.28 SNPs/kbp, and 5.29 ± 2.23 SNPs/kbp heterokaryotic SNP density (Cantu et al., 2013). Using similar methods similar to Cantu et al. (2013), Kiran et al. (2017) reported SNP densities of 1.90 ± 1.27 SNPs/kbp at homokaryotic sites and 4.67 ± 1.17 SNPs/kbp at heterokaryotic sites for three Indian isolates from different epidemiological regions, sequenced and mapped against each other.

An average rate of heterozygosity of 6.25 ± 1.15 SNPs/kbp was computed in the South African isolates. This is slightly higher when compared to the average

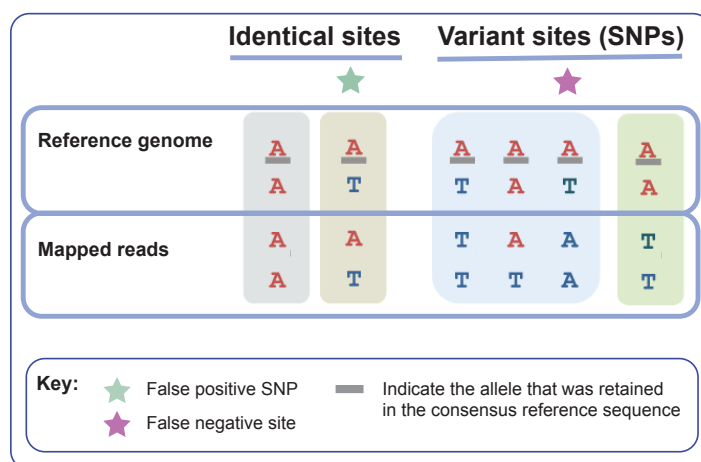


Figure 5.10: Over- and underestimates of SNP sites. Overestimation of heterokaryotic SNP sites is indicated with a green star and underestimation of homokaryotic SNP sites with a pink star. These misinterpretations occur due to un-phased reference genomes (adapted from Cantu et al., 2013).

between PST-21 (USA), PST-43 (USA), PST130 (USA), PST-87/7 (UK) and PST-08/21(UK) (5.70 ± 2.47 SNPs/kbp) (Cantu et al., 2013). Increased heterozygosity was seen in intergenic regions compared to genic regions in the South African isolates, as also reported by Cantu et al. (2013) and Cuomo et al. (2017). This is expected as selection acts more strongly on coding regions.

Next-generation sequencing approaches for sequencing *Pst* have only recently implemented long read information to produce phased genomes where the genomes of the two haploid nuclei are separated (Schwessinger et al., 2018). Due to this constraint, it is expected that homokaryotic SNPs will be underestimated and heterokaryotic SNPs will be overestimated using short read assembly reference genomes such as PST130 (Cantu et al., 2011), CY32 (Zheng et al., 2013), PST-78 (Cuomo et al., 2017) and 46S 119 (Kiran et al., 2017).

Every position in the reference genome represents only one allele at that position, although for genetic material present in both nuclei, two alleles (identical or not) would be present in the genome (Figure 5.10). At nucleotide bases where the reference would have two different alleles, such as heterozygous sites, only

one allele would be in the consensus reference sequence used to align reads in re-sequencing. The mapped isolate identical to the biallelic reference site will appear to be a heterokaryotic SNP site. For example, when the reference is a heterokaryotic site (AT) and the mapped isolate is identical (AT) and the chosen reference site is either A or T, it would indicate a polymorphism causing an overestimation of heterokaryotic SNP sites. It is however expected that heterokaryotic SNPs will be in the majority as mutations are expected to be random and independent between nuclei. True variant sites for single isolates that contain only one genotype would have an allele frequency of one over all aligned reads at monoallelic sites. When the consensus reference sequence contains the allele at a biallelic site that is the same base in the mapped isolate in all alleles in the mapped reads, it would not be known that the mapped isolate was not identical to the reference genome. For example, when the reference is a heterokaryotic site (AT) and the mapped isolate is homokaryotic (AA) and the chosen reference site is A, it would underestimate homokaryotic sites. The availability of a high quality phased reference genome (Schwessinger et al., 2018) allows the improvement of accuracy of current polymorphism classification.

5.4.2 STOP codons

This study focused on polymorphisms in genic regions. SNP analysis revealed the introduction of multiple stop codons. These stop codons appeared at similar frequencies across genic sites in all four isolates. This is of interest as premature stop codons can cause gain in virulence when it causes loss of an avirulence effector function (Dong et al., 2015). The majority (99.4 %) of the SNP sites that introduced stop codons were biallelic. This result will be interesting to re-evaluate using a phased *Pst* genome to account for the overestimation in heterokaryotic SNPs identified when using an unphased genome.

5.4.3 Transitions and transversions at specific codon positions

A transition mutation does not alter the amino acid encoded by that codon, while a transversion would incorporate a different amino acid into the peptide. Due to the degeneracy of the genetic code, the third codon position can be changed for 12 of the 20 amino acids, without altering the amino acid. This is displayed in the biallelic SNP data, where nonsynonymous biallelic SNP sites displayed more transversions, while synonymous biallelic SNPs mainly displayed transitions. At synonymous biallelic SNPs sites, transitions were most frequent with the highest SNP frequencies in C ↔ T and G ↔ A mutations, while nonsynonymous biallelic sites (excluding sites that induced stop codons) displayed higher number of transversions, with C and T to R (A or G) and A and G to Y (C or T) occurring most frequently. Transition:transversion biases occurred at different levels at the three codon positions due to variability in physical constraints that in turn caused variability in selection for or against a specific nucleotide change (Bofkin and Goldman, 2006).

G to A and C to T changes have been described as the most frequent mutations induced by long wave ultraviolet A (UVA) and short wave ultraviolet B (UVB) irradiation in mouse embryo fibroblasts (Pfeifer et al., 2005). The exposure of urediniospores to solar radiation and short wave ultraviolet (UV) light are suggested to reduce viability (Sharp, 1967; Maddison and Manners, 1972). It has also been hypothesised that the distance of dispersal of *Pst* is shorter in comparison with *Pgt* and *Pt*, likely due to its sensitivity to UV light (Rapilly, 1979). Further investigation is needed to draw more parallels between the effect UV irradiation has on mammalian cells, as explained by Pfeifer et al. (2005), and urediniospores, or whether the phenomenon is mostly due to the stronger selective pressure in favour of transitions compared to transversions (Bofkin and Goldman, 2006). Nonetheless, multiple studies have shown the mutagenic effect

of UV light on urediniospores of *Pst* (Johnson, 1978; Cheng et al., 2014), while in this study biases were observed in the frequency of nucleotide changes at specific codon positions.

5.4.4 Stepwise mutations

It is hypothesised that the stepwise changes in virulence seen in South African *Pst* pathotypes have resulted from mutations within a fairly static *Pst* population (Visser et al., 2016). Establishment of new alleles in the population is due to the unique combination of selection and genetic drift in the population (Salemi et al., 2009). In the gene space, selection pressure acts on mutations that cause changes in the function and stability of the gene or the resulting protein, ultimately changing the manner in which the organism interacts with its environment. Genotype frequencies depend on selection that is driven by fitness traits. Genes that are highly polymorphic are thus likely to be involved in fitness traits that enable the genotype to contribute to the next generation.

5.4.5 Positive selection

The YN00 software package was implemented to investigate the presence of signatures of selection by comparing synonymous and nonsynonymous substitution rates. The dN/dS statistics were computed. New alleles introduced by random mutations that evolve neutrally will change in frequency in the population only due to genetic drift and not because it has an effect on fitness. This is generally expected for synonymous SNPs. In contrast, a nonsynonymous polymorphism that affects fitness will evolve more rapidly.

Comparing synonymous and nonsynonymous substitution rates can reveal whether a specific allele at a locus is under positive or negative selection. No omega values greater than 1 were obtained in this analysis. The inability to

identify genes under selection could indicate that genes under strong selection pressure in the South African isolates do not exist in the PST130 reference genome. However, trade-offs exist between statistical robustness and power. It is known that dN/dS methods often fail to detect signals of selection (Salemi et al., 2009). The stringency of dN/dS methods could therefore fail to detect selection pressure between the four clonally derived, and therefore relatively similar, pathotypes. The McDonald-Kreitman test (McDonald and Kreitman, 1991) is often considered more powerful to detect positive selection. It compares dN/dS intra-species against a sister species to remove the demographic background.

5.4.6 Presence-absence analysis

In addition, the South African pathotypes were compared on the basis of genes that were uniquely present in, or absent from the South African pathotypes. This method has shown changes in virulence in other pathogens (Bubić et al., 2004; Yoshida et al., 2009; Gilroy et al., 2011).

Homology with genes of known functions was investigated to determine whether genes could play a specific role in pathogenicity. BLAST searches in public databases provided homology information for 11 genes. Gene ontology of characterised identified genes suggested that these homologs were involved in protein translation, sugar transport, metabolism and components of the fungal cell wall. Postulated gene function did not indicate a role of the homologs in host manipulation or the escape of host recognition, as expected for virulence factors. Biological validation of suggested functionality is needed to draw clearer conclusions.

BLAST searches against the PST130 transcriptome revealed putative paralogs that could indicate functional redundancy for many of the genes shown as absent from the South African isolates, where these paralogs could functionally replace

the absent gene. Such redundancy has been described as genetic buffering (Dangl and Jones, 2001). However, genes that were absent or uniquely present between pathotypes did not fit effector protein characterisation and were not in the putative effector subset of Cantu et al. (2013). Therefore these genes were not considered as candidate genes involved in pathogenicity dynamics.

5.4.7 Nonsynonymous polymorphisms

Lastly, pairwise comparisons of the South African pathotypes, evaluating nonsynonymous differences gene-by-gene, were performed similarly to Cantu et al. (2013). A total of 2689 genes showed nonsynonymous differences in pairwise comparisons between the four pathotypes. Of these genes, 138 carried a secretion signal, of which 27 were in the subset of putative effector genes.

5.5 Conclusion

After characterisation of the polymorphisms across the genomes of the four South African isolates, three methods were used to identify differences in the gene space of the four South African pathotypes. Where applicable, results were compared to lists containing genes that encode secreted proteins and putative effectors (Cantu et al., 2013), to further narrow down the list of candidate genes. Of the three methods followed, namely to search effector candidates with signatures of positive selection, to evaluate the complete exclusion or the unique inclusion of effector candidates and to evaluate nonsynonymous polymorphisms between effector candidates between isolates, only the latter included genes that were previously identified as effector candidates. Different methods exist for validation of candidates, although limitations exist due to the biotrophic nature of *Pst*. One example of validation is to test expression of genes at infection stages using time course experiments. Candidate genes will be further investigated in Chapter 6.

Chapter 6

Gene Expression Analysis of Candidate Effectors Identified in South African *Pst* Isolates

6.1 Introduction

THE OBLIGATE BIOTROPHIC NATURE of rust prevents *in vitro* functional validation. In addition, rust cereal hosts are difficult to transform, which makes *in vivo* functional characterisation challenging (Petre et al., 2016b). While *in planta* studies have been undertaken using techniques as for example virus induced gene silencing (VIGS), they are difficult and time consuming (Panwar and Bakkeren, 2017). Recent successes in stem rust effector identification are reviewed in Chapter 2 Section 2.4.4. As an early step in functional validation, effector gene function in the infection process can be predicted by evaluating gene expression at specific developmental stages of the fungus (Wang et al., 2007, 2009; Sørensen et al., 2012; Cantu et al., 2013). These gene expression levels can be evaluated using methods including microarrays, transcriptome sequencing, and RT-qPCR that was used in this chapter.

6.1.1 Regulation of gene expression in eukaryotes

Gene expression differs throughout development, between cell types and in response to different environmental stimuli. Regulation of gene expression is an intricate, multi-stage process. Transcription regulatory processes occur in the nucleus, while regulation of pre- and post-translation occur in the cytosol. This ability to selectively express genes is essential for the development and survival of a complex organism (Bustin and Nolan, 2004). For transcription factor proteins to access genes and initiate transcription the chromatin must be remodelled through a process of acetylation. Acetylation opens up nucleosomes and allows transcription factor proteins access to gene promoter sites.

The transcription process is further regulated by the assembly and arrangement of the transcriptional machinery enzymes that initiate transcription of RNA from the DNA template. Processing of the pre-mRNA molecule prepares it as a template for protein synthesis. A methylated cap is added to the 5' end soon after transcription starts, while at the 3' end a poly-adenylated tail is added upon completion of transcription. Introns, if present, are then spliced from the pre-mRNA molecule. Binding sites for microRNAs (miRNAs) and regulatory proteins are often found in the 3' untranscribed regions (UTRs) that can down-regulate gene expression or degrade the mRNA molecule.

Double stranded, small interfering RNA (siRNA) can also modulate gene expression at the post-transcription stage. After maturation, the mRNA molecule leaves the nucleus through a nuclear pore and enters the cytosol. Stable mRNA molecules can now be translated into peptides. Post-translational modifications may also be required to transform gene products into functionally active proteins (Klug, 2012). These processes happen in various different organelles depending on the protein, and determine whether a functional gene product is produced.

6.1.2 Quantification of gene expression

Several approaches can be taken to assess the different stages of gene expression. These include validating protein levels, transcription of genes and the effectiveness of small interfering RNAs (siRNAs; Schmittgen and Livak, 2008). Different methods have been developed for these multiple approaches (Speed, 2004; Mehta et al., 2010). One such tool used to measure gene transcript levels is quantitative or real time PCR (qPCR). The first form of qPCR was developed by Higuchi et al. (1993). It measures the level of gene transcription by quantifying the amount of a specific RNA (Schmittgen and Livak, 2008). Quantitative PCR is a powerful tool, with its strength lying in its ability to detect DNA sequences with high specificity, for a wide range of concentrations. In addition, qPCR also eliminates downstream processing that is needed by some other assays using a camera that can detect fluorescence (Higuchi et al., 1993). The fluorescent dye intercalates with the double stranded DNA (dsDNA) as it is synthesised, so that, as dsDNA accumulates the fluorescence increases. The rate at which the fluorescence increases (kinetics) is directly proportional to the original amount of target cDNA. The fluorescent signal is observed by the camera in the qPCR instrument at each annealing/extension phase during thermocycling (Higuchi et al., 1993).

Different methods have been developed to study relative gene expression, e.g. the comparative C_T method, the simulated kinetic model (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008) and the efficiency correction method (Pfaffl, 2001). The efficiency correction method of relative gene expression was used for the analyses in this chapter. This method accounts for differences in the efficiencies of the PCR reaction (see Section 6.2.9) when amplifying the target regions of the test and reference genes, in contrast with the comparative C_T method (Livak and Schmittgen, 2001) that assumes equal amplification efficiencies between the two compared gene products. This is however only possible for

small experiments, with a limited number of genes. Both the efficiency corrected and simulated kinetic model approaches aim to improve the accuracy of the comparative C_T method. The simulated kinetic model is the best for studying large numbers of genes (Schmittgen and Livak, 2008) as the efficiency correction method is a relatively costly and time consuming process (VanGuilder et al., 2008).

6.1.3 Candidate effector features

In Chapter 5, 27 candidate effector genes that displayed nonsynonymous SNPs between the historical South African isolates were identified. These genes, based on the PST130 gene models, were previously identified as putative effectors (Cantu et al., 2013) using a modified version of the effector identification pipeline developed by Saunders et al. (2012). For the 27 candidate effector proteins, annotation and tribe rankings, as taken from Cantu et al. (2013) are listed in Table 6.1. None of the 27 candidate genes had flanking intergenic regions (FIR) of 10 kbp (kilo base pairs) or more. Only PST130_05944 had a nuclear-localisation signal (NLS) at amino acid position 238, and only PST130_07564 was classified as a small and cysteine rich (SCR) protein.

6.1.4 Gene transcription analysis

In this chapter gene transcription is measured as an indication of gene expression, although it is clear from the preceding explanations that many regulatory steps need to be successfully completed to yield a functional protein. When gene expression is studied under different conditions, or at different time points in a developmental time series, spatial and temporal patterns of gene expression show differential accumulation of gene products that are associated with treatment or the specific stages of development (Tomancak et al., 2007). Ideally time points

Table 6.1: Effector features of the identified candidate effectors. Identified candidate effectors were secreted proteins in tribes ranking within the top 100 potential effector tribes as described by Cantu et al. (2013)

Gene ID	Isolate pairs with nonsynonymous substitutions	Tribe no.	Tribe ranking	Length (amino acids)	Similarity to HESPs or fungal AVR	No. of repeat units	Effector motifs (amino acid position)	PFAM mapping	Expressed in infected material	Expressed in Haust.
PST130_06558	SA2 & SA3; SA3 & SA4	9	6	341	No	9		No	No	No
PST130_12487	SA1 & SA2; SA1 & SA3; SA1 & SA4 ; SA2 & SA3; SA2 & SA4; SA3 & SA4	31	7	197	No	0		No	Yes	Yes
PST130_14091	SA1 & SA2; SA2 & SA4	11	14	167	No	0	Y/F/WxC(85);LIAR(32)	Yes	Yes	Yes
PST130_17605	SA2 & SA4; SA3 & SA4	11	14	239	Yes	7	Y/F/WxC(103)	No	Yes	Yes
PST130_05454	SA1 & SA2; SA2 & SA3; SA2 & SA4	68	15	266	No	0		Yes	Yes	No
PST130_09275	SA1 & SA2; SA1 & SA3; SA1 & SA4	134	16	210	Yes	0		Yes	Yes	No
PST130_12491	SA1 & SA4	8	17	182	No	13		No	No	No
PST130_05023	SA1 & SA4 ; SA3 & SA4	351	22	281	No	6		Yes	Yes	Yes
PST130_13969	SA3 & SA4	437	23	394	No	0		No	Yes	Yes
PST130_00285	SA1 & SA3; SA3 & SA4	317	25	207	No	0		Yes	Yes	Yes
PST130_14831	SA2 & SA4	596	31	139	No	0		No	Yes	Yes
PST130_10286	SA3 & SA4	54	33	254	No	0	LIAR(96)	Yes	Yes	Yes
PST130_16778	SA3 & SA4	409	40	172	No	0		No	Yes	Yes
PST130_06503	SA1 & SA4 ; SA2 & SA3; SA3 & SA4	120	41	292	No	9		No	Yes	Yes
PST130_05944	SA2 & SA4; SA3 & SA4	320	49	318	No	0	LIAR(10)	No	Yes	Yes
PST130_07579	SA2 & SA4	170	68	926	No	0		Yes	Yes	Yes
PST130_09018	SA1 & SA3	289	69	430	No	0		No	Yes	Yes
PST130_08031	SA1 & SA2; SA2 & SA4	162	77	206	No	0	LIAR(18)	No	Yes	Yes
PST130_02403	SA1 & SA4 ; SA2 & SA3; SA3 & SA4	21	83	215	No	8		No	Yes	Yes
PST130_02001	SA1 & SA2; SA1 & SA3; SA1 & SA4 ; SA2 & SA4	65	84	148	No	0		Yes	Yes	Yes
PST130_08984	SA2 & SA4	65	84	116	No	0		Yes	Yes	Yes
PST130_07564	SA1 & SA2; SA1 & SA3	482	86	145	No	10		No	Yes	Yes
PST130_15131	SA1 & SA2; SA1 & SA3; SA2 & SA3	186	87	546	No	2		No	Yes	Yes
PST130_02118	SA1 & SA2; SA2 & SA3; SA2 & SA4	92	88	187	No	0	Y/F/WxC(21)	No	Yes	Yes
PST130_07513	SA1 & SA2; SA1 & SA3; SA1 & SA4	128	95	154	No	0		Yes	No	Yes
PST130_12956	SA1 & SA4	128	95	156	No	0		Yes	Yes	Yes
PST130_07448	SA1 & SA3; SA3 & SA4	192	100	191	No	0	Y/F/WxC(73)	No	Yes	Yes

HESPs, Haustorial expressed secreted proteins; AVR, proteins encoded by avirulence genes ;Haust, Haustorial library. PFAM, Protein family database.

Genes in **boldface** had nonsynonymous substitutions between SA1 and SA4 and their expressions were evaluated over a time series. Genes marked in grey were also nonsynonymous between PST-87/7 and PST-08/21 (refer to Cantu et al., 2013). PST130_14091 also known as PST21_19014 and

PST130_13696 also known as PST21_18360.

would be chosen that capture gene expression during early infection processes, at various stages of haustorial and hyphae network development, and sporulation. Comparisons were drawn from histological evaluation of *Pst* haustorial development (Sørensen et al., 2012) and gene expression studies in seedlings (Wang et al., 2007). Spore germination, formation of the substomatal vesicle, development of infection hyphae, the formation of the haustorial mother cells, and haustoria formation are all apparent within the first 24 hours after inoculation. Hyphae and haustoria continue to develop in the host tissue until roughly 5 days post inoculation (dpi). Sporogenous cells become visible at about 7 dpi. By 12 dpi to 14 dpi, depending on the experimental setup, visibly sporulating pustules are usually apparent.

Two of the historical South African *Pst* isolates were further investigated for gene expression using a selection of the 27 candidate effectors. The isolates that were used are representatives of the first *Pst* pathotype detected in South Africa in 1996: 6E16A- (SA1), and the most recent pathotype, 6E22A+ (SA4), that was identified in 2005. These two isolates are the furthest apart in terms of time of collection and pathogenicity as they differ in virulence for three *Yr* resistance genes and were collected seven years apart (Table 4.2). They were chosen to improve the chances of identifying a virulence-related effector candidate. This chapter focuses on further investigation of these candidates using RT-qPCR gene expression analysis. The nine genes selected were those polymorphic between SA1 and SA4 (Table 6.1 boldface).

6.2 Methods

6.2.1 Inoculation and sampling

Seedlings of the stripe rust susceptible wheat variety, Avocet S, were inoculated with urediniospores of the *Pst* South African pathotypes 6E16A- (SA1) and

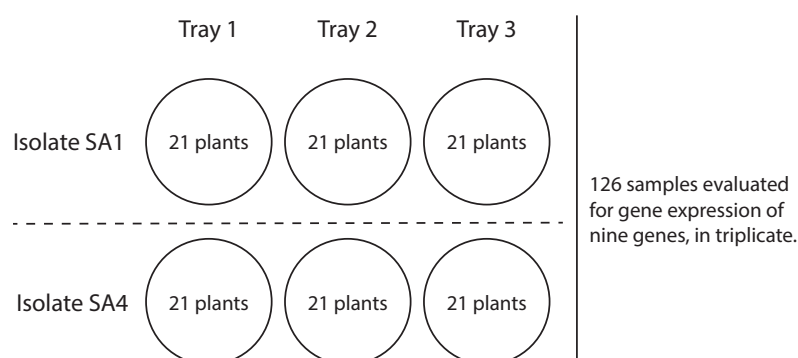


Figure 6.1: Experimental setup for the infection time course experiment.

6E22A+ (SA4) (see Section 3.1.1), using an inoculation concentration of 5 mg/ml. As each seedling can only be sampled once, nine plants—subsequently referred to as biological replicates—were sampled for each treatment (isolate SA1 or SA4) at each time point (Figure 6.1). The 63 plants (9 seedlings $\times$ 7 time points) were equally divided between three trays (21 plants per tray). The three trays for each of the treatments were inoculated independently, which introduced a blocking variable to test reproducibility. Inoculated leaf samples were taken at 0, 1, 2, 3, 5, 9 and 12 dpi, taking three seedlings, per time point, from each of the three trays. Samples were taken about 8 cm from the tip of each leaf, cut into shorter pieces and immediately stored in the RNA stabilising agent, RNAlater (Thermo Fisher Scientific, USA; Taylor et al., 2010). Scissors used to cut inoculated leaf samples were wiped clean with ethanol between sample collections.

Fresh spores of both isolates were germinated and used as positive, fungal controls. The germinated spore samples were prepared in a laminar flow cabinet. Spores were sprinkled on a thin layer of autoclaved double distilled water in a sterilised Petri dish, comparable to the method of Zhang et al. (2008), and kept overnight in a dark room at 11 °C. After 8–12 hours a thick mat of intertwined germination tubes was collected from the surface of the water with a spatula and stored in RNAlater. The preserved samples in RNAlater were kept at room

temperature for 20 days before RNA was extracted.

Caution was taken throughout the experiment to control and define conditions to minimise external stimuli that could interfere with the sensitive process of mRNA transcription (Taylor et al., 2010). A detailed explanation of the inoculation protocol can be found in Chapter 3.

6.2.2 Tissue disruption and RNA extraction

Total RNA was extracted from the inoculated leaf tissue, non-inoculated wheat and germinated fungal spore controls using the Qiagen RNeasy Plant Mini Kit according to the manufacturer's instructions. To minimise the time between subsequent sampling events, the sample processing steps that follow were performed on small batches of 12–24 samples as recommended by Taylor et al. (2010). Tissue was disrupted using a mortar and pestle and the addition of extraction sand (SiO₂). All instruments used were washed with detergent, ethanol and RNase AWAY decontamination reagent between samples, and cooled down in liquid nitrogen, or on dry ice, to prevent degradation of RNA due to ubiquitous RNases activity (Holland et al., 2003). The dry mortar and pestle were placed on dry ice in a polystyrene box and was further cooled with liquid nitrogen. About 100 mg of extraction sand was added to each sample.

Forceps were used to move the preserved sample material from the tubes and tapped dry on a clean paper towel to prevent the stabilising solution from forming ice crystals when the sample comes in contact with liquid nitrogen. Samples were then placed in the mortar, along with liquid nitrogen and extraction sand, and homogenised into a fine powder. Without letting it thaw, the powder was scraped with a cooled spatula into a 2.2 ml safe lock microcentrifuge tube. The ground sample was kept on dry ice until extraction buffer was added.

6.2.3 RNA quality control and quantification

Automated capillary-electrophoresis systems are popular for generating accurate profiles for RNA quality assessment (Fleige and Pfaffl, 2006). The Agilent 2100 Bioanalyzer (Agilent Technologies, USA) was used to assess the quality and quantity of the extracted RNA. The reaction kit was stored at 4 °C. A gel-dye mix was first prepared according to the manufacturer's instructions. The quality of samples was assessed within 1 to 3 days after RNA extraction. RNA samples were appropriately aliquoted to prevent multiple freezing and thawing steps that impose the risk of RNA degradation (Taylor et al., 2010). RNA stocks were stored at –80 °C.

6.2.4 Complementary DNA synthesis

The SuperScript IV First-Strand Synthesis System (Invitrogen/Thermo Fisher Scientific, USA) was used for the conversion by reverse transcription of mRNA to cDNA according to the manufacturer's instructions. Excess RNA was removed by adding 1 µl of *E. coli* RNase H to the synthesised cDNA. An aliquot of 3 µl of cDNA was prepared and quantified on the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) at the Central Analytical Facilities (CAF) at Stellenbosch University, South Africa. cDNA was diluted to approximately 12.5 ng/µl for use in PCR reactions, and cDNA was stored at –20 °C (Taylor et al., 2010).

6.2.5 Primer design

Primers for RT-qPCR were designed using the compiled Illumina sequences obtained of the two *Pst* isolates, SA1 and SA4 respectively (Chapter 4). The PrimerQuest Tool by Integrated DNA Technologies¹ was used to design primers for the nine *Pst* genes of interest. Primers were designed that would amplify the

¹<http://eu.idtdna.com/scitools/Applications/RealTimePCR/>

respective gene from both SA1 and SA4, and produce gene amplicons between 84 bp to 129 bp in length (see Section 6.3.2), as the kinetics of the PCR reaction are influenced by the length of the resulting amplicon.

The primer sequences were evaluated in NCBI BLAST (version 2.6.1; Altschul et al., 1997) homology searches to ensure that they would not amplify sequences within the wheat genome. The likelihood of the primers to form secondary structures, such as primer dimers and hairpins were also assessed, and absence of SNPs in primer sequences was confirmed (Derveaux et al., 2010). Primers were manufactured by Integrated DNA Technologies, USA. Primers were empirically tested for a negative result in a reaction with wheat template DNA and for specificity to amplify the desired amplicon with *Pst* cDNA by evaluating the melt curve of the RT-qPCR, followed by gel electrophoresis to confirm amplicon length. Primer efficiencies were determined using C_T (threshold cycle) of serial dilutions (Derveaux et al., 2010).

6.2.6 PCR plate setup

Complementary DNA templates were used to evaluate transcription levels of the nine *Pst* genes of interest. Only one target gene and the reference gene, which is expected to be expressed constantly over the infection time course, were evaluated on each PCR plate. The same isolates as used in sequencing in Chapter 4 were used for inoculation. Three controls were included: two positive controls—SA1 and SA4—in duplicate, a negative wheat control (WC) from the same wheat variety, Avocet S, and a Non Template Control (NTC; Figure 6.2). Quantitative PCRs of each cDNA sample were performed in triplicate for each gene assay and time point measured in days post inoculation.

cDNA		rep 1	rep 2	rep 3	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3
		1	2	3	4	5	6	7	8	9	10	11	12
SA1: 0-3 dpi	A	SA1:0dpi	SA1:0dpi	SA1:0dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:3dpi	SA1:3dpi	SA1:3dpi
SA4: 0-3 dpi	B	SA4:0dpi	SA4:0dpi	SA4:0dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:3dpi	SA4:3dpi	SA4:3dpi
SA1: 0-3 dpi	C	SA1:0dpi	SA1:0dpi	SA1:0dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:3dpi	SA1:3dpi	SA1:3dpi
SA4: 0-3 dpi	D	SA4:0dpi	SA4:0dpi	SA4:0dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:3dpi	SA4:3dpi	SA4:3dpi
SA1: 5-12 dpi	E	SA1:5dpi	SA1:5dpi	SA1:5dpi	SA1:9dpi	SA1:9dpi	SA1:9dpi	SA1:12dpi	SA1:12dpi	SA1:12dpi	SA1 rep 1	SA4 rep 1	WC
SA4: 5-12 dpi	F	SA4:5dpi	SA4:5dpi	SA4:5dpi	SA4:9dpi	SA4:9dpi	SA4:9dpi	SA4:12dpi	SA4:12dpi	SA4:12dpi	SA1 rep 2	SA4 rep 2	NTC
SA1: 5-12 dpi	G	SA1:5dpi	SA1:5dpi	SA1:5dpi	SA1:9dpi	SA1:9dpi	SA1:9dpi	SA1:12dpi	SA1:12dpi	SA1:12dpi	SA1 rep 1	SA4 rep 1	WC
SA4: 5-12 dpi	H	SA4:5dpi	SA4:5dpi	SA4:5dpi	SA4:9dpi	SA4:9dpi	SA4:9dpi	SA4:12dpi	SA4:12dpi	SA4:12dpi	SA1 rep 2	SA4 rep 2	NTC

Primers		rep 1	rep 2	rep 3	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3
		1	2	3	4	5	6	7	8	9	10	11	12
REF	A	SA1:0dpi	SA1:0dpi	SA1:0dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:3dpi	SA1:3dpi	SA1:3dpi
	B	SA4:0dpi	SA4:0dpi	SA4:0dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:3dpi	SA4:3dpi	SA4:3dpi
GOI	C	SA1:0dpi	SA1:0dpi	SA1:0dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:1dpi	SA1:3dpi	SA1:3dpi	SA1:3dpi
	D	SA4:0dpi	SA4:0dpi	SA4:0dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:1dpi	SA4:3dpi	SA4:3dpi	SA4:3dpi
REF	E	SA1:5dpi	SA1:5dpi	SA1:5dpi	SA1:9dpi	SA1:9dpi	SA1:9dpi	SA1:12dpi	SA1:12dpi	SA1:12dpi	SA1 rep 1	SA4 rep 1	WC
	F	SA4:5dpi	SA4:5dpi	SA4:5dpi	SA4:9dpi	SA4:9dpi	SA4:9dpi	SA4:12dpi	SA4:12dpi	SA4:12dpi	SA1 rep 2	SA4 rep 2	NTC
GOI	G	SA1:5dpi	SA1:5dpi	SA1:5dpi	SA1:9dpi	SA1:9dpi	SA1:9dpi	SA1:12dpi	SA1:12dpi	SA1:12dpi	SA1 rep 1	SA4 rep 1	WC
	H	SA4:5dpi	SA4:5dpi	SA4:5dpi	SA4:9dpi	SA4:9dpi	SA4:9dpi	SA4:12dpi	SA4:12dpi	SA4:12dpi	SA1 rep 2	SA4 rep 2	NTC

Figure 6.2: Plate layouts for RT-qPCR assays. **Template cDNA layout:** Plate layout for DNA of each biological replicate and gene assay. Nine biological replicates were assessed for each gene assay. **Primer layout:** Plate layout for PCR reaction mix. Nine genes were assessed in total. Each plate assessed transcript levels of one target, candidate Pst effector gene and one reference gene. REF, reference gene; GOI, Gene of interest.

6.2.7 Quantitative real-time polymerase chain reaction

Reactions were set up manually. Plates and accompanying seals were manufactured by Thermo Fisher Scientific, USA. Transcript levels of nine candidate effector genes (Chapter 5; Cantu et al., 2013) were assessed using RT-qPCR. A fully skirted, 96 well PCR plate was prepared with 2 μ l (approximately 25 ng) of template cDNA. The 8 μ l reaction mix consisted of 2.4 μ l of double distilled water, 5 μ l of BioRad Precision Melt Supermix and 3 pmol of each forward and reverse primer. The plate with template cDNA was kept on an Eppendorf PCR Cooler block (Sigma-Aldrich, USA) while the reaction mix was added. The plate was sealed, briefly centrifuged and ran on the BioRad CFX96 Touch Real-Time PCR System. The first part of the PCR program included the following steps: An initiation step of 5 minutes at 95 °C, followed by 40 cycles of a 15 second denaturation step at 95 °C, a 20 second primer annealing step at 60 °C and a 20 second primer extension step at 72 °C.

The second part of the PCR program was included to generate a dissociation curve as an indication of the amplification specificity of the primers and to evaluate the formation of primer dimers. High specificity was expected as the fluorescent dye, EvaGreen, is known for its high sequence specificity, and allowing a robust PCR with less PCR inhibition than SYBR Green I, due to its thermal and hydrolytic stability (Mao et al., 2007). The following steps were included in the program: 1 minute at 95 °C to denature the double stranded DNA (dsDNA) with fluorescent intercalating dye to single strand DNA (ssDNA). No fluorescence is expected after this step. To induce the formation of dsDNA, the temperature was lowered for 10 seconds to 40 °C. A ramped step with a 0.2 °C/s incremental increase of temperature starting at 60 °C and stopping at 90 °C was used to denature the dsDNA incrementally. Fluorescence decreases as the dye dissociates. A final cooling step of 10 seconds at 40 °C was added, where after

the reaction was kept at 15 °C.

6.2.8 Reference gene selection

Examples of genes that are often used as internal references in qPCR are 18S rRNA, 7S rRNA, U6 RNA, β -actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH; Schmittgen and Livak, 2008). Three genes were assessed for use as standards of gene expression: *P. striiformis* elongation factor 1 (PST_EF1; Ling et al., 2007), β -Actin (ACTB) and β -Tubulin (TUBB; Huang et al., 2012). Amplification signals in the negative wheat control occurred in multiple qPCR reactions with the primer pair for PST-EF1. However, no amplification with the wheat DNA control was observed with the PST-ACTB and PST-TUBB primers. Both genes would therefore be suitable to use as references. Due to limited wells on the PCR plate, only one reference gene was used, and PST-TUBB was arbitrarily chosen as the reference gene in this study.

6.2.9 Efficiency determination of primers

The BioRad² Precision Melt Supermix contains hot-start iTaqTM DNA polymerase, dNTPs, MgCl₂, EvaGreen dye, enhancers and stabilisers. The polymerase enzyme is responsible for producing amplicons using primers, dNTPs, and template cDNA, with the help of magnesium as a cofactor and optimal temperature cycles. PCR efficiency describes the rate of action of the polymerase and indicates the fold increase of the target DNA per thermocycle (Ruijter et al., 2013). Full efficiency would mean that there is a 2-fold increase of amplicon with every thermocycle during the exponential phase (Yuan et al., 2006). Efficiencies between 90 % and 110 % are acceptable. Poorly calibrated pipettes are often the reason for efficiencies to fall outside of this range. Additionally, low efficiency

²<http://www.bio-rad.com/webroot/web/pdf/lsr/literature/10022094.pdf>

can be caused by suboptimal temperatures, the presence of inhibitors or inactive polymerase, poor primer design or amplicons with secondary structures, while overly high efficiencies result from primer dimers or nonspecific amplicon amplification (Taylor et al., 2010). Efficiency is also not constant throughout the PCR reaction, and low levels of DNA template can result in inaccurate determination of efficiency (Karlen et al., 2007).

The efficiencies of primers were estimated by calculating the slope of the standard curve of a serial dilution of template DNA. Two 2-fold serial dilutions were made by adding RNase free water to the DNA sample, with PCR reactions being done in duplicate. For each DNA concentration, in each dilution series, the mean of the C_T values of the two replicate PCRs was plotted against the base-10 logarithmic transformation of the dilution factor. The data was fitted to a linear regression model and the Pearson correlation coefficient (R^2) was assessed. The amplification efficiency E is theoretically expected to be between 0 and 1, and was calculated with

$$E = 10^{(-1/s)} - 1, \quad (6.1)$$

where the s is the gradient of the linear regression line (Kubista et al., 2006).

The obtained efficiencies were used in the efficiency corrected method to obtain the expression pattern of each gene of interest. The relative expression, R , of the candidate genes to the reference gene was first determined with

$$R = E^{C_T} / E'^{C'_T},$$

where E and E' are the efficiencies as calculated in Eq. (6.1) for the gene of interest and reference gene, and C_T and C'_T are the cycle threshold values for the gene of interest and reference, respectively. The cycle threshold indicates the number of cycles it took to reach the fluorescence threshold, F_T . It is important that

this threshold value is set to fall in the exponential phase of the amplification process (Karlen et al., 2007). This is the earliest phase, with ample reagents, and is followed by the linear phase as reagents decrease and finally reach the plateau phase where reagents become depleted (Yuan et al., 2006). Default F_T was used for all PCR runs.

Transcript levels of the candidate genes were expressed as relative expression to the reference gene *P. striiformis* β -tubulin (TUBB; Huang et al., 2012).

6.2.10 Statistical evaluation of the data

The treatments applied were SA1 and SA4 inoculations. These were applied three times in three independent tray inoculations. For each tray inoculation, three seedlings were prepared for each of the seven time point sampling efforts (7 time points $\times$ 3 = 21 seedlings). The three treatment applications were used as a grouping variable in a linear mixed model. Each of the nine biological replicates per time point (3 plants $\times$ 3 trays) was assessed on a different plate. Inter-plate variability was not corrected for. Intra-plate variability was addressed by performing three technical PCR replications of each biological replicate (plant) per plate (Schmittgen and Livak, 2008). Grubbs' test (Grubbs, 1969) was applied to identify outliers as suggested by Burns et al. (2005). The relative expression values obtained by using the efficiency corrected method were statistically analysed.

One-way analyses of variance (ANOVAs) were performed to assess the variation within and between the groups of biological replicates at different time points in each gene expression assay for both isolates, SA1 and SA4.

6.2.11 Linear mixed effect analysis

The R package, lme4, was used for statistical evaluation of the data (Bates et al., 2014). To determine the relationship between the time that elapsed after inocula-

tion and the relative expression of the candidate genes in each isolate, a linear mixed model with random intercepts was fitted for the data generated for each gene:

$$y_{ij} = \beta_0 + \beta_1 x_{Tij} + \beta_2 x_{Iij} + \beta_3 x_{Tij} x_{Iij} + b_{0j} + e_{ij}, \quad (6.2)$$

where β_0 is the fixed intercept; β_1, β_2 , and β_3 are fixed effects for time, isolate, and interaction, respectively; b_{0j} is a random intercept for each tray j ; the x_T and x_I terms are independent variables for time point and isolate, respectively; and e_{ij} is error. The model was fitted, and assumptions that linear mixed models are based on were assessed. These assumptions include equal variances, and normality of the residuals and random intercepts. The tests were repeated, and re-evaluated after a $\log_{10}$ transformation (Burns et al., 2005) of the relative expression values.

A likelihood ratio test of the full model against the model without the effect (x_I (Isolate), x_T (Time Point), or $x_T x_I$ (Isolate $\times$ Time Point)) in question were performed (Winter, 2013) to assess which model fits the data best. A p -value lower than 0.001 were considered statistically significant, providing evidence for inclusion of the effect in the model. Such a high significance threshold was used to account for the expected high variability in RT-qPCR data. Tukey multiple comparison post-hoc tests were used to indicate where the significant differences in effects were (Section 6.3.4).

6.2.12 Relative expression of *Pst* candidate effector genes

The *Pst* gene expression fold difference between the standardised expression levels of SA1 and SA4 was estimated using the method proposed in Pfaffl (2001) taking primer amplification efficiency into account:

$$R = E^{\Delta C_t(SA1-SA4)} / E'^{\Delta C'_t(SA1-SA4)},$$

where E and E' are the efficiencies as calculated in Eq. (6.1) for the gene of interest and reference gene, respectively, and ΔC_T and $\Delta C'_T$ are the difference between the two isolates (SA1 and SA4) in cycle threshold values for the gene of interest and reference, respectively.

This method is similar to the $2^{-\Delta\Delta C_t}$ method (Schmittgen and Livak, 2008) for determining linearised values, with the difference that the $2^{-\Delta\Delta C_t}$ method assumes that primers have 100 % efficiency causing a two-fold increase of the replicated amplicon in every thermocycle.

6.2.13 Assessment of genes

BLAST searches were performed to assess whether genes were present in both the PST130 gene models and the revised gene models (Dobon et al., 2016). The original PST130 gene discovery was done using the machine learning algorithm geneid³ and *Pgt* gene annotations as training set, followed by filtering for transposable elements (Cantu et al., 2011), while the revised annotation made use of the 2013 UK *Pst* RNA-Seq data and the annotation tools cufflinks, trinity, stringtie and portcullis (D Bunting, personal communication). BLAST searches of the nine candidates against *Pst* transcript data sequenced from the 2013 UK *Pst* population were used to evaluate the occurrence of alternative splicing.

6.3 Results

6.3.1 RNA yield, RNA quality scores and cDNA yield

The integrity of each RNA sample was evaluated on the Agilent 2100 Bioanalyzer producing gel-like visuals, RNA integrity number (RIN) scores, RNA concentrations, and ratios between ribosomal units. Summary statistics were performed on the RNA yields, RIN scores and the reverse transcribed cDNA yields as required

³<http://genome.crg.es/software/geneid/>

Table 6.2: Summary statistics describing RNA yield, integrity and cDNA yield as required in the MIQE guidelines (Bustin et al., 2009). Yield was measured in ng/ μ l

	n	Median	IQR	Mean	SD
RNA_Yield	128	786.00	382.50	793.81	297.84
RIN	128	6.10	0.50	6.06	0.77
cDNA_Yield	128	151.50	137.00	178.26	90.98

RIN: RNA integrity number, n: number of samples, IQR: Inter-quartile range, SD: standard deviation.

for reporting qPCR experiments (Table 6.2; Bustin et al., 2009). RIN scores had a satisfactory mean of 6.06, while the respective means for total RNA and cDNA were 786 and 151 ng/ μ l.

6.3.2 Primer design

Unique primers to each of the nine *Pst* candidate effector genes were designed using PrimerQuest (Table 6.3). The NCBI databases were used in a BLAST (version 2.6.1) search to test uniqueness of primers (Altschul et al., 1997). In no case was a sequence similarity found that spanned 100 % of the primer length. Primer lengths ranged from 19 to 23 nucleotides, and GC content was between 41 and 58 %. Amplicon size affects the number of amplicon copies at the threshold fluorescence (Rutledge and Cote, 2003), so primers were designed to amplify amplicons of identical size to ensure equal specificities in the two treatments (SA1 and SA4; Karlen et al., 2007). Amplicons were between 84 and 129 bp in length (Table 6.3). Melting temperatures were optimised at 60 °C. Primers were tested and dissociation curves were evaluated for specificity in the positive control *Pst* cDNA. The negative control, wheat variety Avocet S gDNA and the NTCs did not show any amplification. Further details on primer design, the location of amplicons and the depth of coverage of the sequence data used to design primers can be found in Appendix C, Figures C.1 to C.9.

Table 6.3: Primer and amplicon specifications for *Pst* candidate effector gene identification

Gene name	Primer sequence	Primer length	Amplicon sequence	Amplicon length	GC content	Efficiency %
PST130_02001	GTGGCCCTAGTGTACCAATTAT CTCTCCTGGATTGAGAGTTTGG	22 22	GTGGCCCTAGTGTACCAATTATCTGGCATCAATGCCAACTCGATCGTCTCGCCTAAGCCCAACCAAA CTCTCAATCCAGGAGAG	84	50 50	88
PST130_02403	CGAGGAACCCAAATATGCTAGT GACGGTAGCCGTCTTTCTTT	22 20	CGAGGAACCCAAATATGCTAGTCCAAATATGATSCGCCCTACGAGAAGACCCCTGATGAAGAGCCA AAATACTCGGCCCAAGCTACGATTACAATCCACCAAAGAAAGACGGCTACCGTC	122	45 50	107
PST130_05023	ACTTGGTACGGTGGACATTC CCTTGGCTTCCTTGCTCTTA	20 20	ACTTGGTACGGTGGACATTCGGCTGTGGCCAGGTTTTTTCGCGCGCTTGTTAATTACTTTACCCAA GAAAGATGAGTAAGAGCAAGGAAGCCAAGG	97	50 50	97
PST130_06503	CAGCGGTGTCATTGCTTTAC TGTATTTCGGAAGAGGCGTATTT	20 22	CAGCGGTGTCATTGCTTTACCTACTTCCAACCAAGCACAAATCGAAACTCGGGCCGAGAAGACCCGT TCCAGCGACAAATACGCCTCTTCCGAATACA	98	50 41	107
PST130_07513	GTACCGAGCAGGACGAATTATG GTATACGGCCATCCTTCCATTT	22 22	GTACCGAGCAGGACGAATTATGTGCCGAGCATTTACTTCCAAGTTACCCAACCTCTCAAGGTGTTTT CAAATGGAAGGATGGCCGTATAC	89	50 45	94
PST130_09275	GAGCGAACTCAACCGCTAATA CAGCCGTACCCGAGTTATATTT	21 22	GAGCGAACTCAACCGCTAATACCCCTGCTGCAAGTACTCCTGTCGCTAACACGACCTCCCCGACCCA ATCCACATCCTCCACTGGTGACCA	92	48 45	101
PST130_12487	CTACCATCATTAGACGGCACAT GCACTTGCTTCCACCATAAAC	22 21	CTACCATCATTAGACGGCACATTGTGCAATGCCCCATCACCTTCGTGGCAACTGACTATTGACAAT GGTCAAATCAGGAACCGTAGGTTTATGGTGGAAGCAAGTGC	107	45 48	90
PST130_12491	CAGAGCACTTCGCCTTAC CGAGAGGGCAATGTTGAGAA	19 20	CAATTTTCGAGAAGCGTGCCGAGACTGAAGGCACCGGAAAAGGTGAATCAAGCTCCCCTCCTTAG GTGGCTGCAGCAACCAAGTTGGCC	90	58 50	90
PST130_12956	TGTTTGCCCTAGCTTCTTCTATC GGTGTCGAAGTTCTCTGATGTC	23 22	TGTTTGCCCTAGCTTCTTCTATCCATGCCGACGCAGGACTCAACCCCAATGACGCTCCAGATGACGT CATCGAATTGACATCAGAGAACTTCGACACC	98	43 50	92

6.3.3 Efficiency determination of primers

Primer efficiency was evaluated using the standard curve method. The C_T values of a cDNA dilution series were plotted, with $\log_{10}$ dilution fold on the x -axis and C_T on the y -axis. A linear regression was fitted to the data and the Pearson correlation coefficient (R^2) calculated (Figure 6.3). This indicated how well the data fitted a linear model, with $R^2 = 1$ being a 100 % fit. A high R^2 is needed to accurately determine the efficiency of primers. It is recommended that the efficiency of primers should be within 10 % of each other when a relative gene amplification comparison is to be made. Less optimisation is required when the efficiencies are taken into account as in the efficiency correction methodology used in this work (Schmittgen and Livak, 2008). R^2 values of greater than 0.95 were achieved for all *Pst* gene primers except for PST130_12491, which had a R^2 value of 0.81.

6.3.4 Statistical analysis of the relative expression of nine *Pst* candidate effector genes

Relative expression values were calculated using the method proposed in Pfaffl (2001, See Section 6.2.12). To determine the relationship between the time that has elapsed since *Pst* inoculation and the relative expression of the candidate genes in each South African isolate, a linear model with mixed effects was fitted to the data with “Gene” and “Time Point” and their interaction as fixed effects, and “Tray” as a blocking variable or random intercept. This approach was taken as sampling was not random as is expected in a simple linear model (Fitzmaurice et al., 2008). The model explains the relationship between the independent and dependent variables. An error term is used where the model does not fully represent the data. It is expected that the three plants that were inoculated together, placed in the same tray, will be more similar to each other. The mixed model therefore

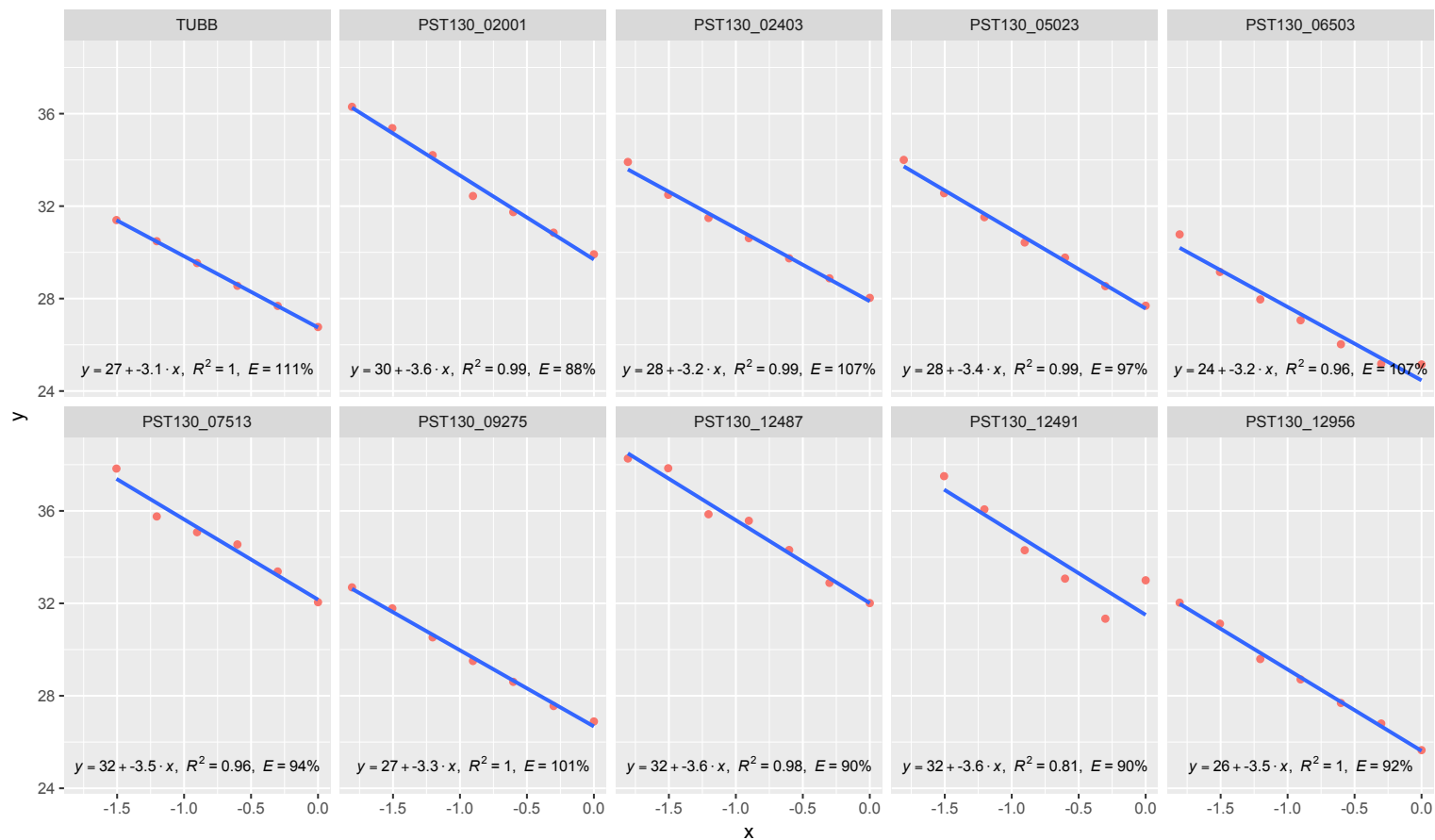


Figure 6.3: Linear regression figures indicate estimated efficiency of primers for nine *Pst* Candidate gene assays and the reference gene, β -tubulin (TUBB). The threshold cycle number is indicated on the y -axis and plotted against the $\log_{10}$ dilution fold (x -axis). The Pearson correlation coefficient, R^2 , indicate how well the data fitted a linear model. Values over 0.95 are desired.

reduces the error term by introducing the variable “Tray”. At every time point, three samples (seedlings) were taken from each of the three trays. The same sampling procedure was applied for both SA1 and SA4. The mixed model shown in Equation 6.2 was at first fitted to the data.

Evaluations of the assumptions of linear mixed models were performed for the relative expression dataset. The residuals of relative gene expression and the random intercept (the grouping variable “Tray”) did not fit a normal distribution. The residuals also did not scatter equally around the $y = 0$ horizontal line as expected when variances are equal and showed clear fan-like patterns in some cases. Appendix C, Figures C.10(i), (ii), and (iii) illustrate the graphical tests for the whole dataset, while Figures C.11 and C.12 show assessments for each isolate and gene.

Due to the use of the grouping variable, the normal probability plot of the random intercepts was constructed from limited points as the intercepts per gene only consisted of six data points at each time point, three per isolate. This data was therefore only plotted for the whole dataset and not by gene.

The relative expression data did not follow a normal distribution and a $\log_{10}$ transformation was applied. Graphical tests for normality and equal variances of the residuals were repeated. The $\log_{10}$ transformed data fitted the assumptions of a linear mixed model considerably better and it was concluded to proceed using the transformed data in the linear mixed model (Appendix C, Figures C.13, C.14, and C.15). As equal variances and normality are assumed for the residuals of the $\log_{10}$ transformed data, parametric tests can be applied. Variability of the data across different trays was assessed by using a one-way ANOVA with “Trays” as fixed effect on subsets of the data that included expression data of one isolate and one gene at a specific time point (nine data points for each of the time points). Time points 0 and 1 were excluded from this evaluation due to too many missing values. This resulted in analysing the effect “Trays” on nine genes at five time

points and was done for two isolates (90 ANOVAs). The effect of “Trays” over these 90 cases were quantified.

The between group variance (between the three trays) was in only 15 % of the cases more than the within groups variance (plants per tray). This showed that there existed a high level of variability in the data. Such variation is often accumulated from the multiple steps in RT-qPCR, described by some as a “fragile assay” (Bustin and Nolan, 2004), due to its sensitivity to inevitably accumulate technical noise. This result should be considered in further interpretation of the data.

To assess the significance of the fixed effects (“Time Point” and “Isolate”) in the model, likelihood ratio tests were performed on two linear mixed models, one including the effect in question (“Time Point” or “Isolate”) and one without. Because of the high variability in RT-qPCR data, a *p*-value was only considered significant if it was smaller than 0.001. A significant *p*-value obtained indicated that the fixed effect term was significant to include in the model. The factor “Time Point” was significant for seven *Pst* genes (Table 6.4). For PST130_12956 and PST130_02403 the term “Time Point” was not significant. Figure 6.4 further revealed relatively stable expression for PST130_12956, while PST130_02403 showed large error bars, especially at early time points. Variability in the data makes it difficult to conclude a change in expression for PST130_02403. The fixed effect “Isolate” was not statistically significant with any of the nine *Pst* genes, both isolates displaying a similar expression profile across all time points (Figure 6.4).

Multiple comparisons were done using the Tukey test to determine between which time points significant differences in gene transcription occurred (Table 6.5). As the term “Isolate” and the interaction term “Isolate $\times$ Time Point” were not significant for any of the nine *Pst* genes, this showed that SA1 and SA4 have a similar expression profile across all time points, for all genes (Figure 6.4).

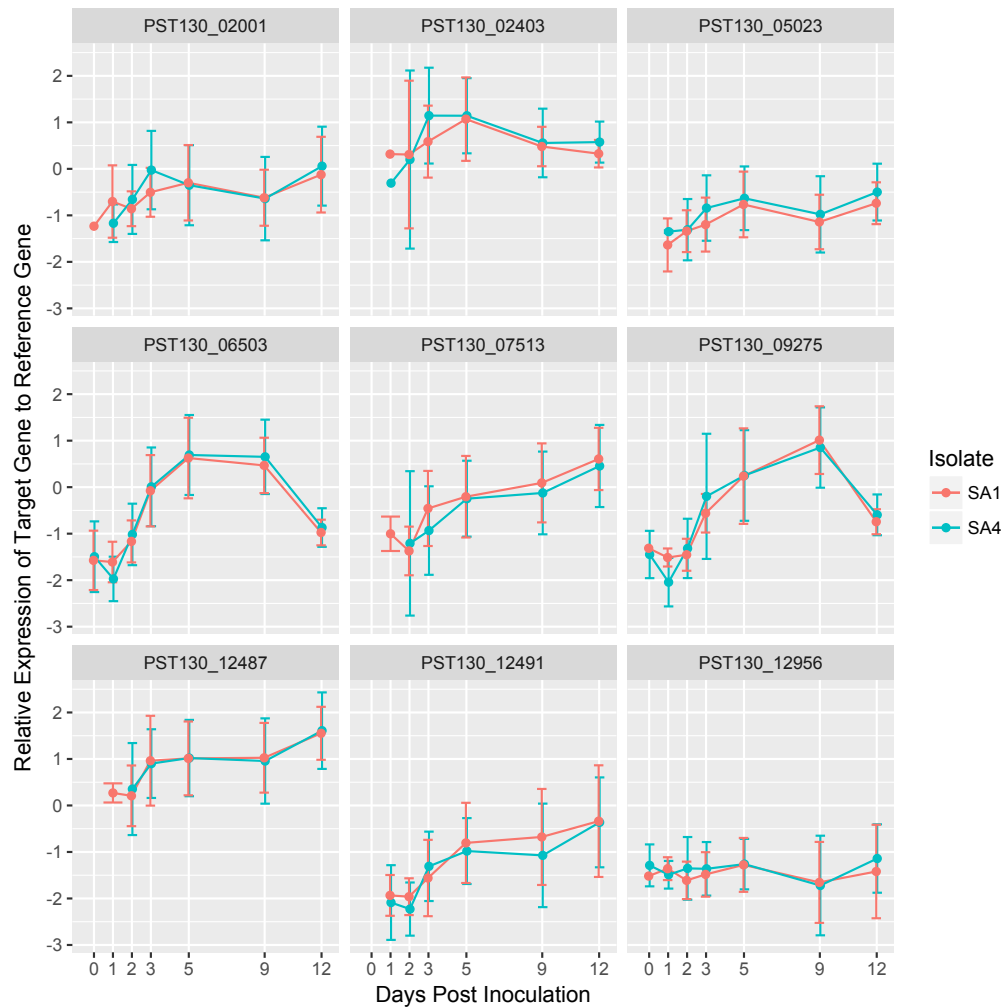


Figure 6.4: Relative gene expression (log₁₀ transformed) of nine candidate effector genes expressed in the *Pst* isolates SA1 and SA4 measured at different time points after inoculation. Significant changes in expression across the time series were seen in all genes, except PST130_02403 and PST130_12956. PST130_06503 and PST130_09275 showed the most dynamic expression patterns, while other genes showed smaller differences in gene expression across time points. The gene, β -tubulin, was used as reference gene.

Table 6.4: Significance of the factor “Time Point” in the linear mixed model for those genes where it was significant

Gene	Chi-squared	Df	<i>p</i> -value
PST130_02001	22.542	6	0.000 965 4
PST130_05023	22.919	5	0.000 349 8
PST130_06503	113.71	6	$< 2.2 \times 10^{-16}$
PST130_07513	31.358	5	7.96×10^{-6}
PST130_09275	173.93	6	$< 2.2 \times 10^{-16}$
PST130_12487	23.837	5	0.000 233 4
PST130_12491	27.644	5	4.27×10^{-5}

6.3.5 Expression profiles of candidate genes

Significant changes in expression across the time series were seen in all genes, except PST130_02403 and PST130_12956. PST130_06503 and PST130_09275 showed similar and the most dynamic expression patterns. The remaining five genes showed smaller differences in gene expression across time points. Expression profiles of PST130_02001 and PST130_05023 were comparable, while PST130_07513, PST130_12491 and PST130_12487 followed a similar trend. (Compare Figure 2.6 that broadly illustrates the infection process and describes the physical processes during the time course of infection in *Pst.*).

6.3.6 Gene validation using revised gene models and transcript data

The nine genes were assessed for alternative splicing using transcript data. The quality of the PST130 gene models, specifically for the nine genes evaluated were also assessed using improved PST130 gene models (Dobon et al., 2016). PST130_07513 and PST130_12491 lacked high sequence similarity with predicted genes in the revised gene models. The remaining seven gene sequences had high (roughly 95 %) similarity and reasonable coverage with the revised predicted genes. In four of the seven genes, PST130_02001, PST130_05023, PST130_06503 and PST130_09275, no evidence for alternative splicing was found. PST130_02001, PST130_05023, PST130_06503 and PST130_09275 are therefore most likely

Table 6.5: Multiple comparisons between time points for each gene that showed significant difference in expression over the time series. Differences with a p -value of <0.001 were considered significant. From this data and Figure 6.4 it was clear that PST130_06503 and PST130_09275 displayed a much more dynamic expression pattern across time points compared to the other genes tested

Gene	Time Point comparison	z value	$Pr(> z)$
PST130_02001	3 - 1	3.002	0.03673
	12 - 1	3.77	0.00266
	12 - 2	3.343	0.01265
	12 - 9	3.076	0.02929
PST130_05023	5 - 1	3.613	0.00396
	12 - 1	3.933	0.00113
	12 - 2	3.242	0.01432
PST130_06503	3 - 0	6.876	< 0.001
	5 - 0	9.337	< 0.001
	9 - 0	9.008	< 0.001
	12 - 0	3.532	0.0074
	3 - 1	7.08	< 0.001
	5 - 1	9.671	< 0.001
	9 - 1	9.293	< 0.001
	12 - 1	3.578	0.00616
	3 - 2	5.257	< 0.001
	5 - 2	8.409	< 0.001
	9 - 2	7.963	< 0.001
	5 - 3	3.153	0.02647
	12 - 3	-4.403	< 0.001
	12 - 5	-7.603	< 0.001
	12 - 9	-7.137	< 0.001
PST130_09275	3 - 0	4.295	< 0.001
	5 - 0	6.541	< 0.001
	9 - 0	8.595	< 0.001
	12 - 0	3.305	0.0157
	3 - 1	5.607	< 0.001
	5 - 1	8.297	< 0.001
	9 - 1	10.763	< 0.001
	12 - 1	4.466	< 0.001
	3 - 2	4.685	< 0.001
	5 - 2	8.064	< 0.001
	9 - 2	11.296	< 0.001
	12 - 2	3.335	0.0142
	9 - 3	5.498	< 0.001
	12 - 5	-5.077	< 0.001
	12 - 9	-8.217	< 0.001
PST130_12487	12 - 1	2.99	0.03029
	12 - 2	4.44	< 0.001
	12 - 3	3.45	0.00676
	12 - 5	3.38	0.00852
	12 - 9	3.53	0.00495
PST130_12491	12 - 1	4.158	< 0.001
	12 - 2	3.864	0.00147
	12 - 3	4.486	< 0.001

correctly annotated and low risk sequences for alternative splicing. Significant alternative splicing was revealed for PST130_02403. PST130_12487 displayed two retained introns, while two overlapping genes in the new gene models mapped to PST130_12956.

6.4 Discussion

Early time points yielded little fungal RNA due to the low *Pst* biomass in infected wheat tissues. This was also the case in the RNA-Seq study of Dobon et al. (2016). This is unfortunate as multiple effector proteins are known to be deployed during the first 24 hours after inoculation. Consequently, amplification failed in samples that were collected early after inoculation, mostly at 0 and 1 dpi, and occasionally at 2 dpi, as the copy number of target sequences was not sufficiently high.

Statistical evaluation using a linear mixed model revealed that expression patterns between the two isolates did not vary significantly. Differences in gene expression across different time points were significant for most genes, with some genes showing a dynamic expression pattern over the course of the time series. However, considerable inter-plate variation was detected, and the relative gene expression determination with efficiency correction did not correct for inter-plate variability. One option of standardisation is to include a calibration sample in multiple wells across all plates to correct for plate technical variation. Such a sample can be prepared for each gene in the experiment to allow sufficient quantities for all inter-plate comparisons.

The possibility of high biological variance in expression patterns of effectors cannot be excluded. In the rice blast fungus *Magnaporthe oryzae*, clonal variation in effector gene expression (CVEGE) has been suggested as a mechanism to escape host recognition, a different suite of effector genes being expressed in individual blast lesions (Mark Farman, University of Kentucky, personal commu-

nication). If this was the case in *Pst*, different seedlings, or even infection sites on a single seedling, inoculated with the same isolate might exhibit differences in effector gene expression profiles. The discovery in *M. oryzae* establishes a new paradigm for plant-microbe recognition wherein resistance involves detection of deterministic Avr effectors which are layered over suites of effectors that are variably expressed among individuals. Consequently, tracing the expression of such effector genes in host-microbe interaction studies becomes a more difficult proposition and would require a different approach to RT-qPCR analysis in whole seedling leaves.

Pst gene expression early in the infection process, between 0 dpi and 1 to 2 dpi, needs further investigation to draw sound conclusions. For later time points, PST130_05023 and PST130_02001 displayed a similar expression pattern, showing an increase in expression early in the infection process that differed between SA1 and SA4, although it was not statistically different, but had nearly identical expression patterns at the later time points. This could indicate that both these genes are functional in the same or co-occurring infection processes. PST130_05023 was the only gene that was assessed in the current study as well as in the RT-qPCR evaluation of Avocet S inoculated with PST-08/21 (Cantu et al., 2013). In Cantu et al. (2013) it was found that PST130_05023 expression peaked at sporulation (14 dpi), similar to the result in the current study, where the expression peak was observed at sporulation (12 dpi).

The main differences in the evaluated gene expression profiles were between 5 and 9 dpi, and 9 and 12 dpi. Genes can be placed in three groups according to their expression profiles.

Group 1 PST130_02001 and PST130_05023 shared an increase in expression up to 3 dpi to 5 dpi, followed by a decrease in expression from 5 dpi to 9 dpi and another increase from 9 dpi to 12 dpi. This could indicate that the gene

is involved in the early establishment of the *Pst* colony, and then functional again during the sporulation processes, such as the formation of vertical hyphae and spores. These genes all contained a PFAM domain (PFAM, Protein family database), and were expressed in both infected material and haustoria (Cantu et al., 2013).

Group 2 PST130_07513, PST130_12491 and PST130_12487 exhibited an expression pattern of initial increase up to 3 dpi, followed by a relatively stable expression, showing a slight increase all the way up to 12 dpi. This could indicate some functionality during the early stages of colony establishment, plus a constant requirement for the protein throughout the asexual lifecycle.

Group 3 PST130_06503 and PST130_09275 showed a similar expression profile. A steep increase in gene expression was observed from 2 dpi to 5 dpi, with maximum expression at 9 dpi falling off at 12 dpi. From the expression profile, one can speculate that these genes have their main function in establishment and maintenance of the *Pst* colony, and do not have a role in sporulation. In Cantu et al. (2013) PST130_06503 was expressed in the haustoria, while PST130_09275 was expressed in the infected material, but not in the haustoria.

No statistically significant change was identified in PST130_02403 and PST130_12956 expression over the time course of this study. For PST130_02403 this could be due to high variability in the data, as illustrated by the error bars at early time points. Further investigation of the expression profile of this gene is needed to draw conclusions. Variation in the data for PST130_12956 is smaller, and a fairly stable expression across the infection process for this gene is concluded. Some similarities can be drawn between the expression profiles of PST130_12956 and genes in Group 1 in the previous paragraph.

The nine candidate effector genes were assessed for alternative splicing using *Pst* transcript data. The genes were further verified by evaluating whether the

candidate effector genes were included in both PST130 gene annotations. This analysis revealed no evidence of alternative splicing for PST130_02001, PST130_05023, PST130_06503 and PST130_09275. High sequence similarity was also found in the new gene models for these four genes. PST130_07513 and PST130_12491 did not have good hits in the new gene models and could have been misidentified, in either attempt to predict genes. Although primers were not designed to amplify fragments across splice sites, underestimation of gene expression could have resulted in alternatively spliced genes if the exon containing the amplicon was excluded during splicing.

In a functional study using heterologous expression screens in *Nicotiana benthamiana*, accumulation patterns of PST130_05023 were observed in endomembranes that are suspended in the cytoplasm of leaf cells (Petre et al., 2016b).

6.5 Conclusion

Clear conclusions regarding gene expression could not be drawn from the RT-qPCR experimental procedure applied in this chapter. Interesting questions arise from the variability in the relative expression data. Future work addressing these questions should involve the inclusion of different biological replicates in one PCR run to investigate reproducibility. Other methods, such as RNA-Seq could be explored, but as shown, does not address the problem of low fungal transcripts at early time points.

In retrospect, it could be argued that the method would only work if genes had no homologs and if they were absent from one of the isolates. If primers were designed across SNP sites, they could have been more successful in displaying the differences between the isolates for the nine candidate genes. Further discussion on the qPCR experimental procedure outlining pitfalls and precautions taken is included in Appendix C.3.

Chapter 7

Analysis of the Current Stripe Rust Threat in South Africa

7.1 Introduction

7.1.1 *Pst* virulence since 2005

THE FIRST DISCOVERY of *Puccinia striiformis* f. sp. *tritici* in South Africa was in 1996 (Pretorius et al., 1997), with three subsequent pathotypes that appeared to have evolved in a clonal, stepwise manner (Visser et al., 2016). Previous analysis that compared the virulence profiles of the historical and current *Pst* populations suggested that the population has stayed fairly consistent, with routine, traditional pathology testing on wheat differential sets (Table 7.1) reporting no additional virulences since 2005 (Agricultural Research Council, Small Grain (ARC-SG), personal communication).

The prevalence of *Pst* pathotypes in South Africa during the growing seasons of 2008 to 2016 is shown in Figure 7.1(i). Data was obtained from the South African *Pst* virulence survey undertaken by ARC-SG, South Africa. The SA2 pathotype, 6E22A- (detected in 1998), and the SA4 pathotype, 6E22A+ (detected in 2005), were present in all eight seasons. Pathotype 6E16A- (SA1), which

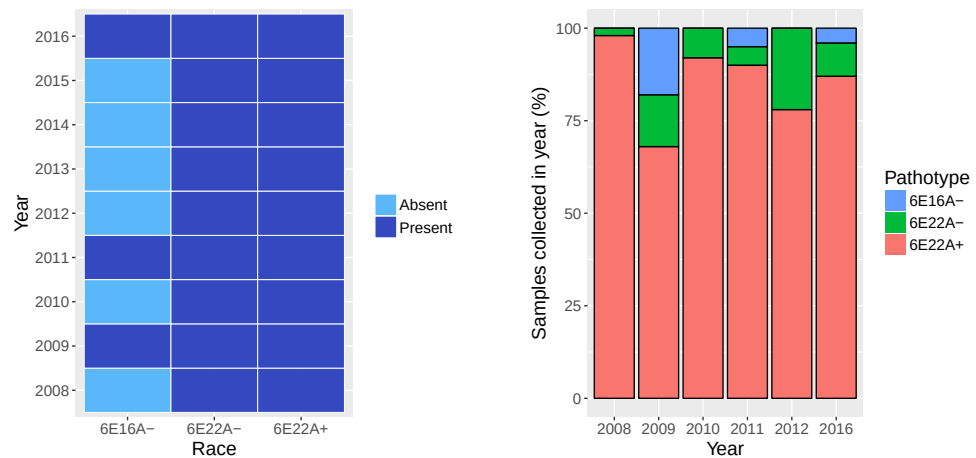
Table 7.1: Wheat differential lines used at Agricultural Research Council, Small Grain, Bethlehem, South Africa to identify *Pst* pathotypes. Standard world (1 to 7) and European (10 to 17) differential sets are listed. Lines 9, 8 and 18, containing resistance genes *Yr5*, *Yr9* and *YrA* respectively, are used as supplemental lines

No	Line/variety	<i>Yr</i> gene
1	Chinese 166	1
2	Lee	7,22,23
3	Heines Kolben	2,6
4	Vilmorin 23	3a,4a
5	Moro	10,Mor
6	Strubes Dickkopf	25,Sd
7	Suwon 92/Omar	Su,4
8	Clement	2,9,25,Cle
9	Triticum spelta	5
10	Hybrid 46	4b
11	Reichersberg 42	7,25
12	Heines Peko	2,6,25
13	Nord Desprez	3a,4a
14	Compair	8,19
15	Carstens V	25,32,Cv
16	Spaldings Prolific	Sp,25
17	Heines VII	2,25,HVII
18	Avocet R	A

was first detected in 1996, only occurred in samples collected in 2009 and 2011. Figure 7.1(ii) displays the percentage of *Pst* samples, classified by pathotype, collected between 2008 and 2012, and in 2016, and Figure 7.1(iii) shows the corresponding sampling sites of each isolate by pathotype from 2008 to 2012. Information about the number of samples collected per year per location could not be obtained. The available survey data indicate that pathotype 6E22A+ was the most prevalent, followed by regular occurrence of 6E22A-, at a lower frequency. It seems that 6E16A- has mostly been replaced by the 6E22 pathotypes, with the pathotype 6E22A+, virulent to *YrA*, predominating.

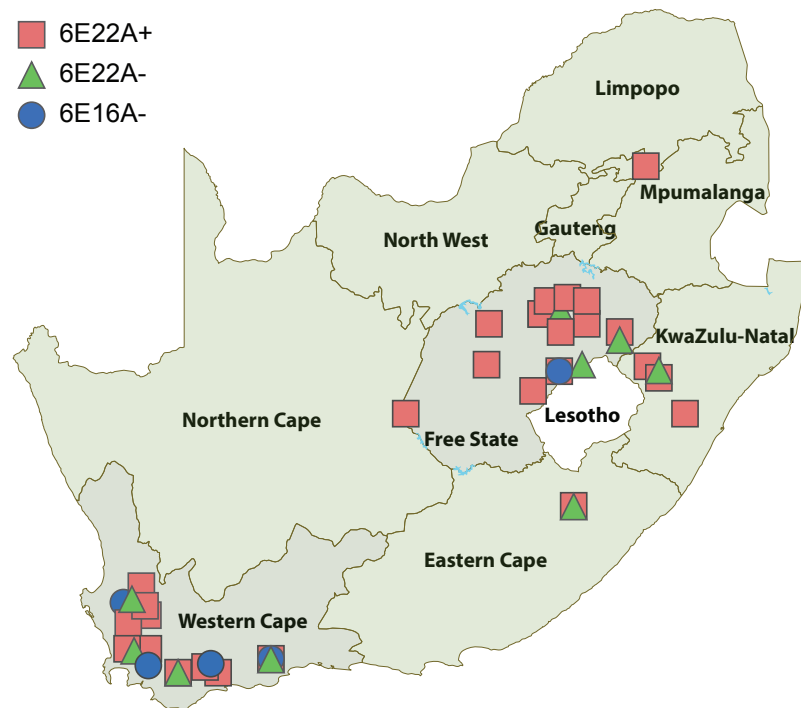
7.1.2 Global reports on *Pst* population shifts

The dynamics and demographics of several *Pst* populations have been described. Wellings (2007) described three reasons for a change in population demography in clonal populations as seen in Australia. Firstly, increased pathogen virulence



(i) South African *Pst* pathotypes observed between 2008 and 2016.

(ii) Percentage of *Pst* isolates, by specific pathotypes, found between 2008 and 2012, and 2016.



(iii) Collection sites and pathotypes of *Pst* isolates between 2008 to 2012 in South Africa.

Figure 7.1: Prevalence of *Pst* pathotypes in South Africa between 2008 and 2016. Data was made available by the Agricultural Research Council, Small Grain (ARC-SG) of South Africa (map adapted from SENSAGO's oral presentation during the Borlaug Global Rust Initiative (BGRI), New Delhi, 2013).

through mutation and selection following resistance gene deployment are common mechanisms (Brown and Hovmøller, 2002; McDonald and Linde, 2002; Milus et al., 2009; de Vallavieille-Pope et al., 2012). Secondly, exotic incursions have been shown to occur over long distances, causing sudden unsuspected epidemics and shifts in the pathogen population dynamics. This has included *Pst* pathotypes with increased aggressiveness (Milus et al., 2009). The establishment of such incursions seems to depend on the host population and possible abiotic stressors (Wellings, 2007). Lastly, the survival of *Pst* mutations by genetic drift, during unfavourable conditions, can totally change the following season's re-emerging population. Such population bottlenecks can lead to a severe shift in allele frequencies.

Exotic incursions in the USA in 2000 and Australia in 2002 have shown relatively homogeneous incursions suggesting that a single genotype of *Pst* was introduced (Wellings, 2007; Milus et al., 2009; Hovmøller et al., 2016). In Europe, a major population shift was seen in 2011 that included several *Pst* pathotypes, some of which could infect the wheat variety, Warrior. Through pathotyping in subsequent years, these newly introduced *Pst* pathotypes were shown to be diverse. A method to rapidly genotype and compare field samples was developed by Hubbard et al. (2015). Using next-generation sequencing data, it was confirmed that the older UK *Pst* population was replaced by a new, much more diverse population. UK and French *Pst* isolates pre-2011 were closely related, with low genetic diversity, while isolates from 2011 and 2013 formed a distinct, more diverse population. Isolates collected post-2011, included the *Pst* pathotype virulent on the wheat variety Warrior and three more genetic groups. Hubbard et al. (2015) also found historical and new *Pst* isolates with different genetic profiles, but the same virulence profile. This radical population shift in 2011 was also confirmed by Hovmøller et al. (2016), and the authors suggested that the two new pathotypes, "Warrior" and "Kranich" carried characteristics that suggested that they might have originated from a sexual population possibly from the near-Himalayan region in Asia.

7.1.3 Objectives

Stripe rust is a global problem of increasing proportion (Hovmøller et al., 2010) and migration of spores over long distances was repeatedly reported (Ali et al., 2017). Furthermore, the existence of recombinant *Pst* populations increases the risk for new variants appearing in each new season (Rodriguez-Algaba et al., 2014). In Chapter 4, four historical South African *Pst* isolates were analysed in context with other global isolates. In this chapter, changes seen in the current field population of *Pst* in South Africa were characterised in context with the global isolates examined in Chapter 4.

7.2 Materials and methods

7.2.1 Stripe rust samples used in RNA sequencing analyses

Field samples of stripe rust were collected in South Africa during the 2014 and 2015 wheat growing seasons. Twenty-five single lesion leaf samples of *Pst*-infected wheat leaves were collected from various locations (Figure 7.2; Table 7.2). In 2013, a *Puccinia* sample was collected on wild rye and found to be virulent on wheat with the pathotype classification 6E16A- (Pretorius et al., 2015), similar to SA1. This isolate was included in this analysis and named 13/SAZP1. Microsatellite markers have also been used to describe this isolate, also known as Sutherland (Visser et al., 2016). In addition, four *Pst* isolates were collected from Ethiopia and 14 isolates from Kenya during the 2014 growing season. All stripe rust infected leaf samples were stored in RNA stabilising solution (RNAlater, Life Technologies, UK). Selected samples (44) that passed quality assessments as explained in Chapter 3 were included in the analysis.

Table 7.2: African isolates collected between 2013 and 2015. Read frequency graphs of these isolates are displayed in Appendix D, Figures D.1 and D.2

Isolate number	Isolates (year/code)	Country of isolation	Year collected	Type of data
49	14/SADL1	South Africa	2014	RNA-Seq
50	14/SADL2	South Africa	2014	RNA-Seq
51	14/SADL3	South Africa	2014	RNA-Seq
52	14/SADL4	South Africa	2014	RNA-Seq
53	14/SADL5	South Africa	2014	RNA-Seq
54	14/SADL6	South Africa	2014	RNA-Seq
55	14/SATT1	South Africa	2014	RNA-Seq
56	14/SATT2	South Africa	2014	RNA-Seq
57	14/SATT3	South Africa	2014	RNA-Seq
58	14/SATT4	South Africa	2014	RNA-Seq
59	14/SATT5	South Africa	2014	RNA-Seq
60	13/SAZP1	South Africa	2013	RNA-Seq
61	14/SAZP2	South Africa	2014	RNA-Seq
62	14/SAZP3	South Africa	2014	RNA-Seq
63	15/SAZP1*	South Africa	2015	RNA-Seq
64	15/SAZP2	South Africa	2015	RNA-Seq
65	15/SAZP3	South Africa	2015	RNA-Seq
66	15/SAZP4	South Africa	2015	RNA-Seq
67	15/SAZP5	South Africa	2015	RNA-Seq
68	15/SAZP6	South Africa	2015	RNA-Seq
69	15/SAZP7	South Africa	2015	RNA-Seq
70	15/SAZP8	South Africa	2015	RNA-Seq
71	15/SAZP9	South Africa	2015	RNA-Seq
72	15/SAZP10	South Africa	2015	RNA-Seq
73	15/SAZP11	South Africa	2015	RNA-Seq
74	15/SAZP12	South Africa	2015	RNA-Seq
75	14/ET2	Ethiopia	2014	RNA-Seq
76	14/ET3	Ethiopia	2014	RNA-Seq
77	14/ET4	Ethiopia	2014	RNA-Seq
78	14/ET5	Ethiopia	2014	RNA-Seq
79	14/K2	Kenya	2014	RNA-Seq
80	14/K4	Kenya	2014	RNA-Seq
81	14/K5	Kenya	2014	RNA-Seq
82	14/K6	Kenya	2014	RNA-Seq
83	14/K7	Kenya	2014	RNA-Seq
84	14/K8	Kenya	2014	RNA-Seq
85	14/K9	Kenya	2014	RNA-Seq
86	14/K10	Kenya	2014	RNA-Seq
87	14/K11	Kenya	2014	RNA-Seq
88	14/K12	Kenya	2014	RNA-Seq
89	14/K13	Kenya	2014	RNA-Seq
90	14/K14	Kenya	2014	RNA-Seq
91	14/K15	Kenya	2014	RNA-Seq
92	14/K16	Kenya	2014	RNA-Seq

*also known as Sutherland (Visser et al., 2016); 14/ET2-5 (Bueno-Sancho et al., 2017) obtained from D Hodson; Kenyan field samples provided by DGO Saunders (14/K2-16) obtained from R Wanyera; South African field samples collected by D Lesch (SADL), T Terefe (SATT), and ZA Pretorius (SAZP).(15/SAZP2 was not used in the analyses due to poor read frequency graph.)

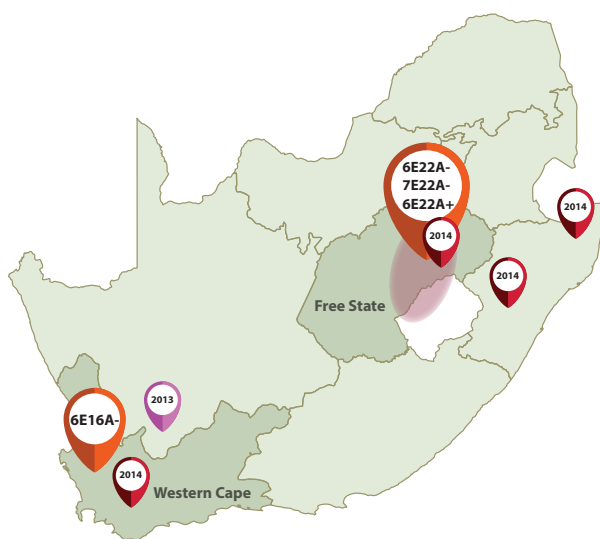


Figure 7.2: Locations of *Pst* collections between 2013 and 2015 for RNA sequencing and historical isolate collection sites.

7.2.2 Transcriptome sequencing of stripe rust infected wheat leaves

Total RNA was extracted using the Qiagen RNeasy Mini kit (Qiagen, Germany). RNA integrity and quantity were assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) as explained in Chapter 3. RNA was reverse transcribed to cDNA using the Illumina TruSeq RNA sample preparation kit (Illumina, UK). Transcriptome sequencing was performed on the Illumina HiSeq instrument at the Earlham Institute, UK. Bowtie software (version 0.12.7; Langmead et al., 2009) from the TopHat package (version 1.3.2; Trapnell et al., 2012) was used to align the pair-end reads of each transcriptome independently to the PST130 reference genome (Cantu et al., 2011). Purity of isolates was confirmed using the method described in Chapter 3. Phylogenetic and population structure analyses, followed by F_{ST} calculations and the Watterson estimator of population diversity ($\hat{\theta}_W$), were used to describe genetic variation in population clusters in a similar manner to the methodology followed in Chapter 4 and described in Chapter 3. These analyses were performed on the field isolates listed in Table 7.2 and the

48 isolates in Chapter 4 in Table 4.1, resulting in the assessment of 92 isolates in total.

7.2.3 *Pst* pathotype determination

Roelfs et al. (1992) explained that the infection types given in Table 7.3 “are often refined by modifying characters as follows: = means uredinia at lower size limit for the infection type; " means uredinia somewhat smaller than normal for the infection type; + means uredinia somewhat larger than normal for the infection type; ++ means uredinia at the upper size limit for the infection type; C means more chlorosis than normal for the infection type; and N means more necrosis than normal for the infection type.

Discrete infection types on a single leaf when infected with a single biotype are separated by a comma (e.g., 4, ; or 2=, 2+ or 1,3C). A range of variation between infection types is recorded by indicating the range, with the most prevalent infection type listed first (e.g., 23 or ;1C or 31N) (Roelfs and Hettel, 1992).”

Fresh inoculum was prepared by inoculating seedlings of the susceptible wheat variety Morocco. Four cultures were prepared: two cultures of the historical South African isolates, SA1 and SA4 and two more recently collected isolates, 13/SAZP1 and 15/SAZP4. The isolate 13/SAZP1 was previously tested and identified to be pathotype 6A16A- (Pretorius et al., 2015), while 15/SAZP4 was identified as 6E22A+ on the standard differential sets, using the scoring system in Table 7.3 (ZA Pretorius, unpublished data).

An extended set of wheat differential lines were inoculated with each *Pst* isolate after growing seedlings for 7 to 8 days as explained in Chapter 3. Infection types were evaluated 21 days after inoculation and reported in Appendix D, Tables D.1 and D.2 (UK differential lines were obtained from S Holdgate, National Institute of Agricultural Botany (NIAB), UK and DGO Saunders, John Innes Centre (JIC), UK).

Table 7.3: Infection type scores used to assess *Pst* infection on wheat seedlings (adapted from Roelfs et al., 1992 and McIntosh et al., 1995)

Host response (class)	Infection type ^a	Disease symptoms
Immune	0	No visible uredia
Very resistant	;	Necrotic flecks
Resistant	;N	Necrotic areas without sporulation
Resistant	1	Necrotic and chlorotic areas with restricted sporulation
Moderately resistant	2	Moderate sporulation with necrosis and chlorosis
Moderately susceptible	3	Sporulation with chlorosis
Susceptible	4	Abundant sporulation without chlorosis

7.3 Results

7.3.1 Clustering analysis using RNA-Seq and whole genome sequencing data

To investigate the pathotype and genetic profile of the current *Pst* population in South Africa, stripe rust infected wheat samples were collected from wheat fields between 2013 and 2015 (Figure 7.2). The interaction transcriptomes of these *Pst* infected wheat samples were sequenced along with similar field isolates from Kenya and Ethiopia. Cluster analysis was carried out using SNP datasets to assess the existence of population structure in the *Pst* population.

Phylogeny

A phylogenetic tree (Figure 7.3) was constructed using the randomized accelerated maximum likelihood (RAxML) method as described in Section 3.3.5, to determine the genetic relationship among samples (Table 7.2). Isolates examined in Chapter 4 (Table 4.1) were included in the analysis of the field samples. The tree

illustrates a well-defined shift in the genetic structure of the South African *Pst* population, with the recent samples collected between 2013 and 2015 clustering distantly from earlier collected isolates. Field isolates from Ethiopia and Kenya were more closely related to the historical East African and South African populations, while the South African field isolates clustered together with a group of isolates found in the UK in 2013 on triticale, called UK Group II (Hubbard et al., 2015). The relative distances tree was also constructed, (Figure 7.4) excluding isolates from the East Africa (B) group in the interest of legibility of the figure. The UK 2013 Group II isolates clusters distantly from the other 2013 UK isolates. The 2013 - 2015 South African isolates cluster with these UK isolates, away from the historical South African isolates.

Population structure analysis

To assess population structure, STRUCTURE software (version 2.3.4; Pritchard et al., 2000) was applied to analyse a dataset of 112 180 synonymous biallelic SNPs. Both the log probability plot (Figure 7.5(i)) from Pritchard et al. (2000) and the plot of ΔK (Figure 7.5(ii)), based on the method described by Evanno et al. (2005), suggested that the population could be grouped into five subclusters.

The histogram plots of the data, with K estimated between 2 and 15 (Figure 7.6), describe each isolate's cluster allocation given a certain number of clusters (K). No additional information regarding population differentiation was gained when K was increased above five.

STRUCTURE assumes that the population is under Hardy-Weinberg equilibrium: Equilibrium of allelic and genotypic frequency with infinite size population, diploid and sexual reproducing species, no migration and panmixia (random crossing among isolates). As some of these criteria are violated by our data (asexual reproduction, small populations and no panmixia) the STRUCTURE result

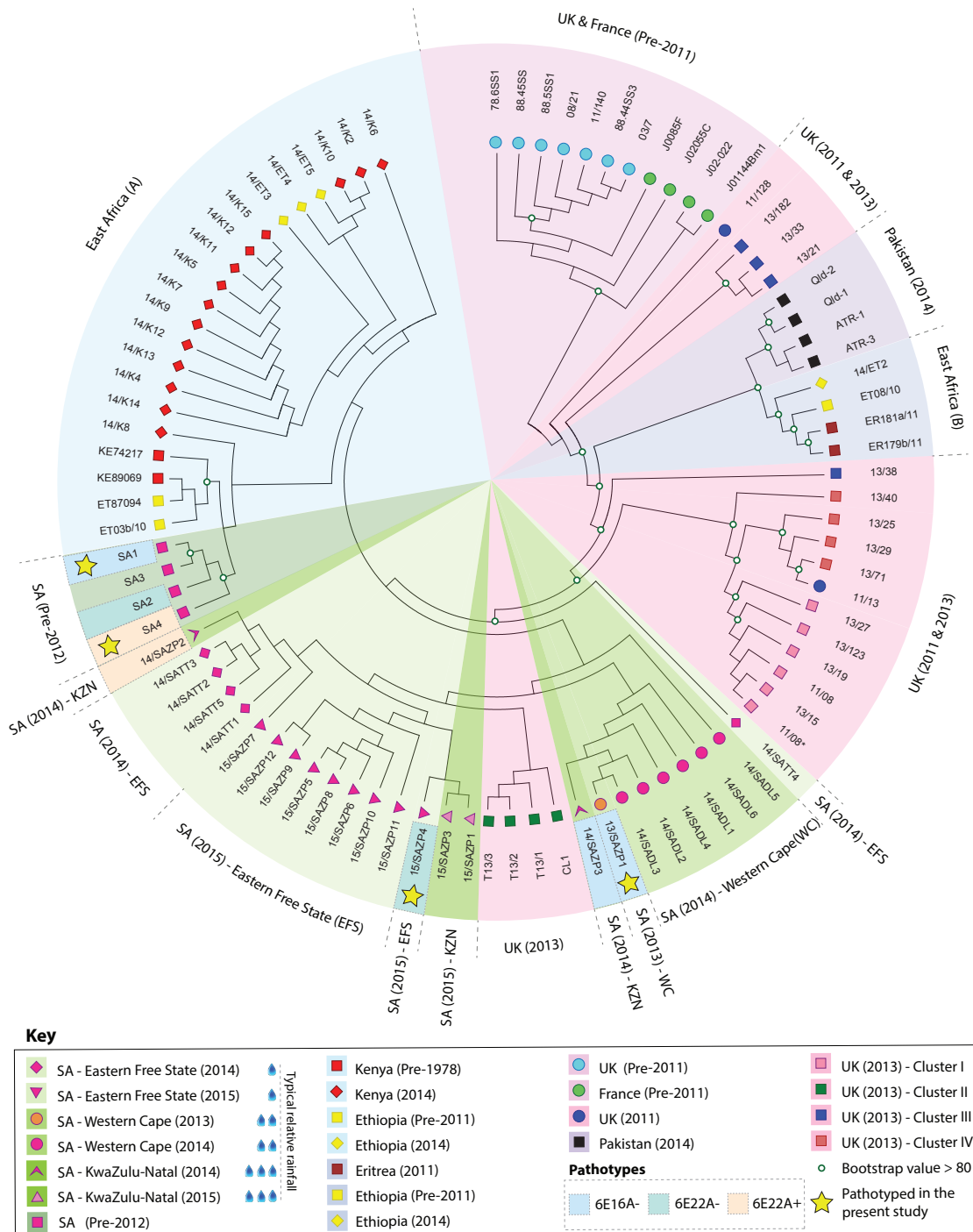


Figure 7.3: Phylogenetic tree displaying the relationship between *Pst* isolates. Samples representative of older *Pst* populations and more recent populations were compared. The maximum likelihood phylogenetic tree was obtained using the RAxML method. The relationship between samples was determined using those *Pst* genes that had 80 % breadth of coverage in 80 % of the samples. This included 2597 genes and a total of 792 535 third codon sites. Only the topology is indicated here, while Figure 7.4 displays relative distances. Both dendrograms were visualised using MEGA software (version 6.06). Asterisk (*) indicates genomic data of isolate 11/08, while 11/08 without an asterisk indicates RNA-Seq data. RAxML, Randomized Axelerated Maximum Likelihood; EFS, Eastern Free State; KZN, KwaZulu-Natal; WC, Western Cape

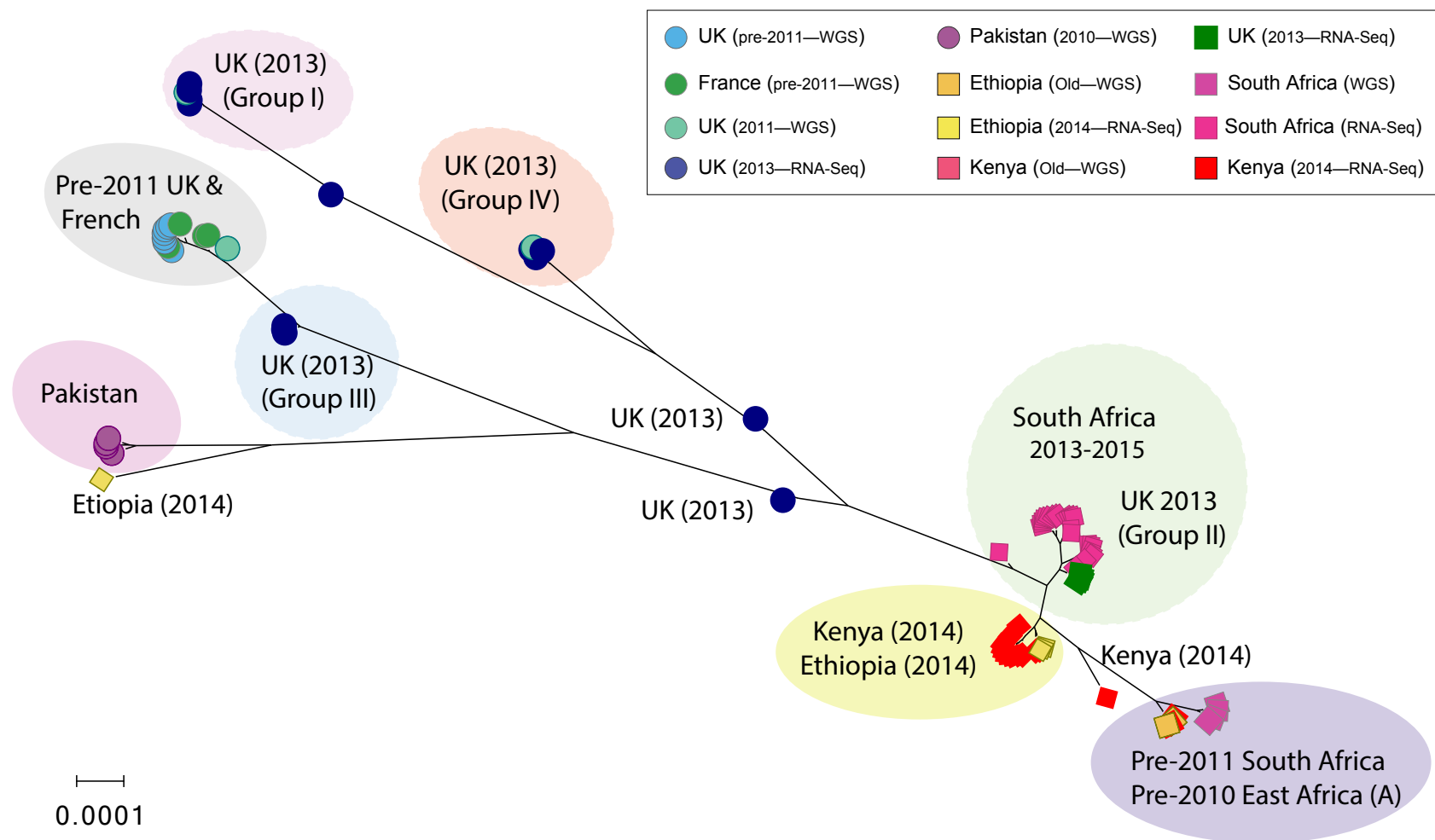
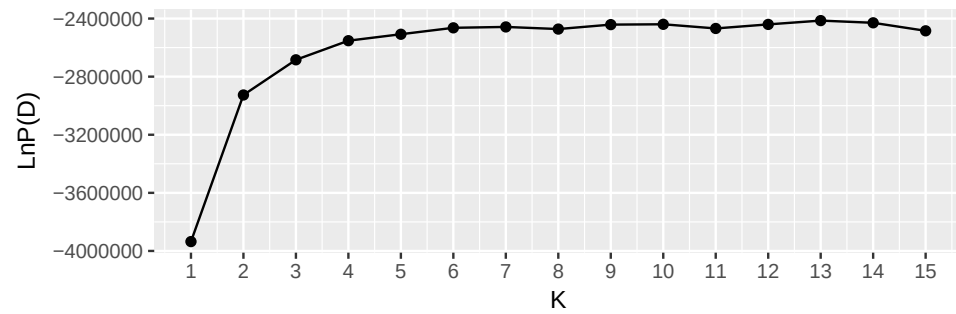


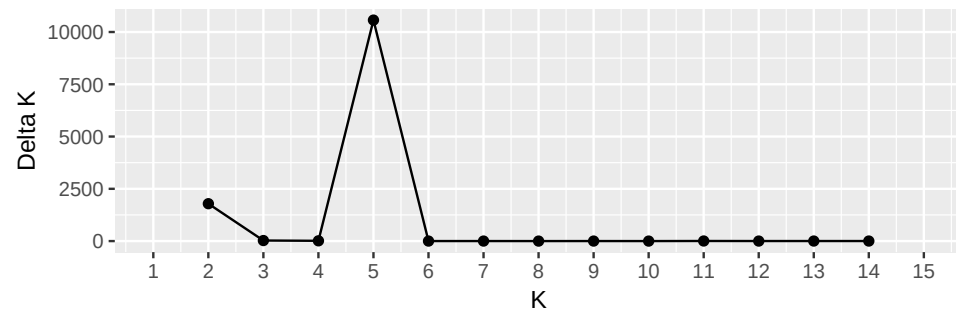
Figure 7.4: Relative distance maximum likelihood phylogenetic tree describes the relative relationship between isolates described in Figure 7.3 where branch lengths were ignored and only topology was considered. In this dendrogram, East Africa (B) was not shown. Compare Appendix D, Figure D.3, that includes the East Africa (B) group.

was compared with the non-parametric method DAPC (Jombart et al., 2010). The biallelic synonymous SNP dataset used in the STRUCTURE analysis was used to summarise genetic variance within and between populations by PCA. The BIC graph (Figure 7.7(i)) illustrates an elbow at $K = 7$ to $K = 8$, while an absolute minimum was observed at $K = 11$. This indicated that the optimum number of population clusters falls between 7 and 11.

Individual isolates were assigned to population clusters by DA of eigenvalues (Figure 7.7(ii)). According to the DA, the first two PCAs explained most of the genetic variability seen in the data. The histogram plots (Figure 7.8) at different values of K showed an increase in differentiation from $K = 7$ to $K = 11$. The gain of differentiation from 10 to 11, shown in the South African isolates form 2014, is lost at $K = 12$. Taking this and the BIC graph into account, $K = 10$ was concluded to be the optimal estimate of population clusters. The first two principal components of the DAPC analysis of the synonymous SNP sites are shown in the scatter plot (Figure 7.7(iii)). The distances between groups are representative of the relative differentiation between population groups, taking the first two principle components into account.



- (i) Log probability of the data $L(K)$ as a function of K to estimate the optimal number of population clusters as identified by STRUCTURE. The optimum number of clusters (K) inferred by the model-based Bayesian cluster analysis of genome-wide SNP data is 5.



- (ii) The Evanno method of inferring the number of STRUcTURE populations (K) from the modal value of ΔK . A strong signal was detected for $K = 5$ where ΔK was at a maximum. Δ , Delta

Figure 7.5: Evaluation of number of population clusters following STRUcTURE analyses.

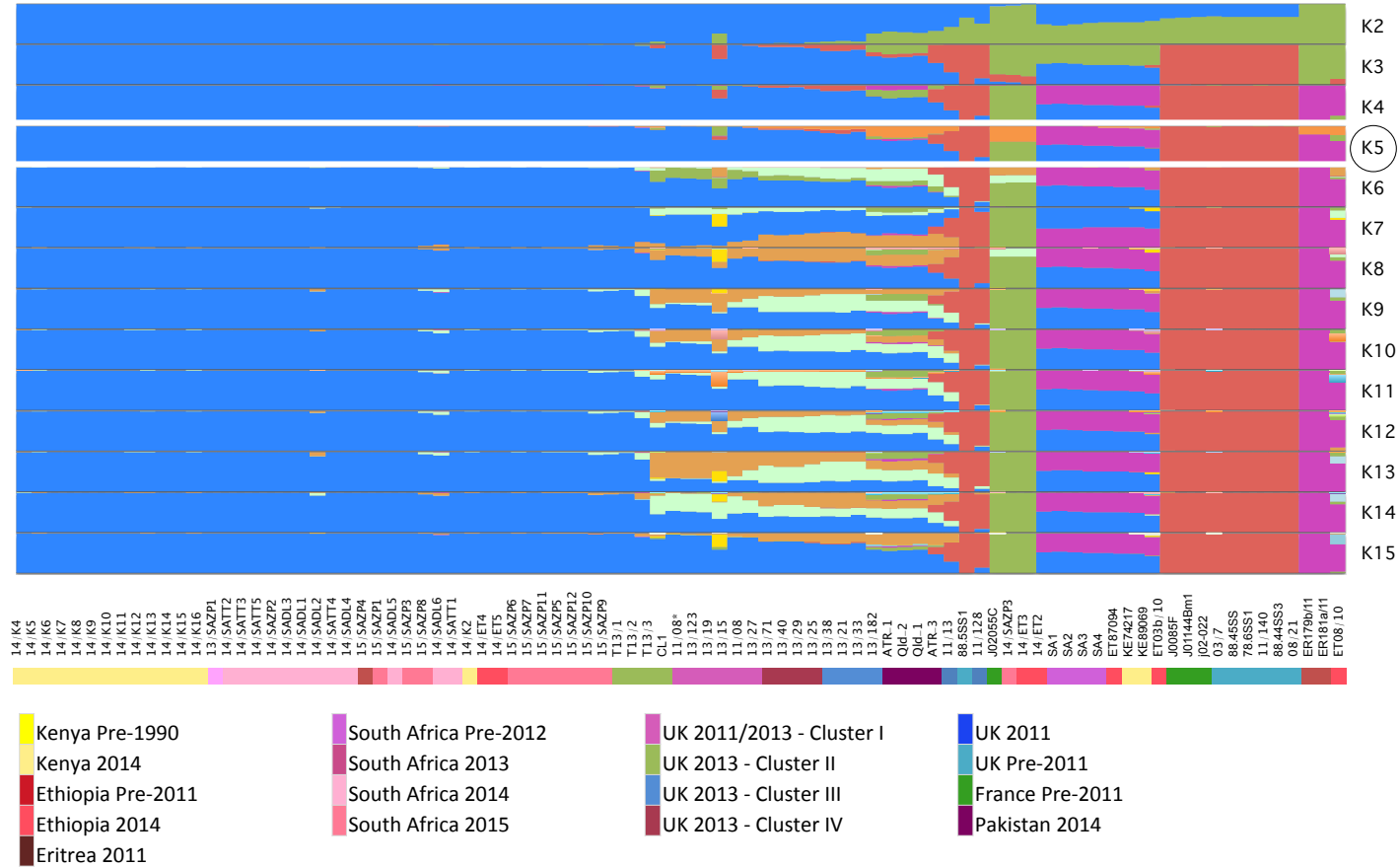


Figure 7.6: Histogram plots of population clustering with K between 2 and 15 as obtained from STRUCTURE analyses. Each bar represents estimated membership fractions for each *Pst* isolate. No further differentiation was observed after $K = 5$. Asterisk (*) indicates genomic data of isolate 11/08, while no asterisk indicates RNA-Seq data for 11/08. ATR-2 and 11/75 (Table 4.2) were not used in this analysis.

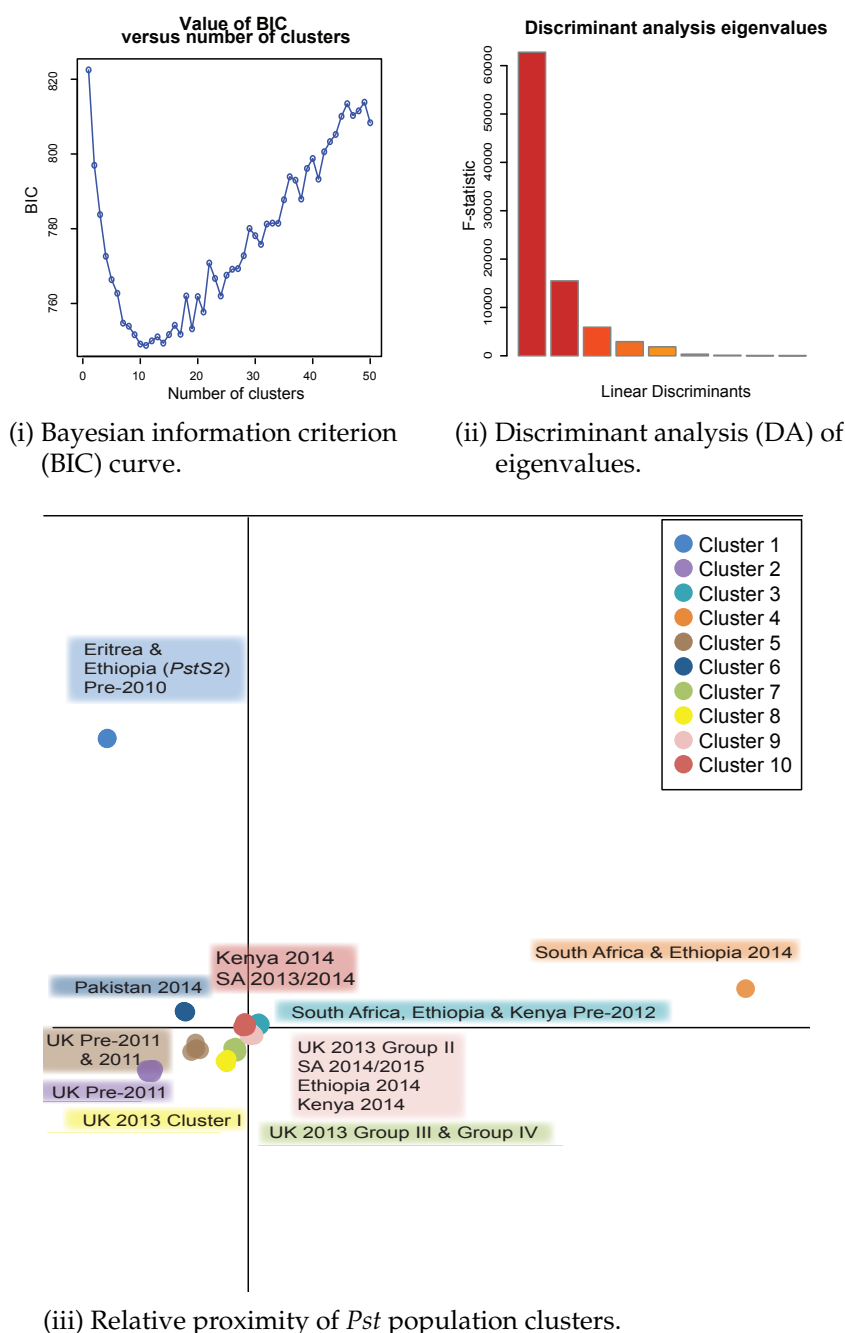


Figure 7.7: Discriminant analysis of principal components (DAPC) analysis of *Pst* isolates. Panel (i) shows the Bayesian information criterion (BIC) curve suggesting the minimum number of clusters (K) required to explain the variation between pathotype clusters. An elbow is observed at $K = 7$ and a minimum at $K = 11$. From this result it can be derived that the optimal predicted number of population clusters (K) for the dataset falls between 7 and 11. Panel (ii) shows a bar plot representing discriminant analysis (DA) of eigenvalues for main principal component functions. This indicates that most of the variation in the dataset can be explained by the first two principle components. Panel (iii) shows a scatter plot indicating the relative proximity of *Pst* population clusters following DAPC analysis.

Figure 7.8: Histogram plots indicating population structure as inferred by DAPC analysis. Each bar indicates the group an isolate is assigned to. Field samples from Africa, collected between 2013 and 2015 were assigned to three groups, coloured orange, light red and red. The light red group contains South African isolates from 2014 and 2015, two Ethiopian isolates and one Kenyan isolate from 2014 and groups with the UK 2013 Cluster II, containing triticale field isolates. The red cluster contains field samples from 2014 collected in Kenya and South Africa, and one sample from 2013 that was collected from wild rye in South Africa. The orange group differentiated earlier (K6) than the red (K8) and light red groups (K9). This small group contains field samples collected in 2014 from Ethiopia and South Africa. From these three groups it is evident that the recent *Pst* population in South Africa is fairly diverse and that South African isolates share similarities with the Kenyan and Ethiopian populations. The Ethiopian population shows higher diversity compared to the Kenyan population. Asterisk (*) indicates genomic data of isolate 11/08, while no asterisk indicates RNA-Seq data for 11/08. ATR-2 and 11/75 (Table 4.2) were not used in this analysis

Differentiation within and between population clusters

Differentiation between groups was calculated through pairwise comparisons of the 10 population clusters identified by the DAPC analysis. F_{ST} statistics for all pairwise comparisons are indicated in Figure 7.9 in the lower diagonal matrix. Highest F_{ST} values were observed in comparisons of Group 1 (≥ 0.37) and Group 4 (≥ 0.58). These groups were also positioned most distantly from the other eight groups in the DAPC scatter plot (Figure 7.7(iii)). Comparing the diversity between the two groups resulted in a high F_{ST} of 0.8, further indicating that the two groups differentiated distinctly from all other groups and from each other. These high values of F_{ST} also confirmed the importance of asexual reproduction known to increase the differentiation among populations by the absence of genetic mixing. Group 1 contained an Ethiopian isolate, ET08/10, previously identified as *PstS2*, as stipulated in Chapter 4 (Hovmøller et al., 2008; Walter et al., 2016; Ali et al., 2017; M Hovmøller, personal communication). This isolate grouped with two isolates from Eritrea collected in 2011. Group 4 included two isolates from Ethiopia and one isolate from South Africa, all collected in 2014. Group 4 was distinctly different to Groups 9 and 10, containing the remaining South African and East African isolates collected from 2013 to 2015. Groups 9 and 10 had a low F_{ST} of 0.12, indicating that these two groups are closely related. Besides the recent African field samples, Group 9 also contained samples collected on triticale in the UK in 2013. Low variability within the three groups (Groups 4, 9 and 10) that contained the post-2012 African samples was observed as indicated on the matrix diagonal (Figure 7.9).

7.3.2 Seedling *Pst* pathotype testing

To compare the virulence profiles of the historical *Pst* isolates to isolates collected from the field between 2013 and 2015, seedling inoculation tests were performed

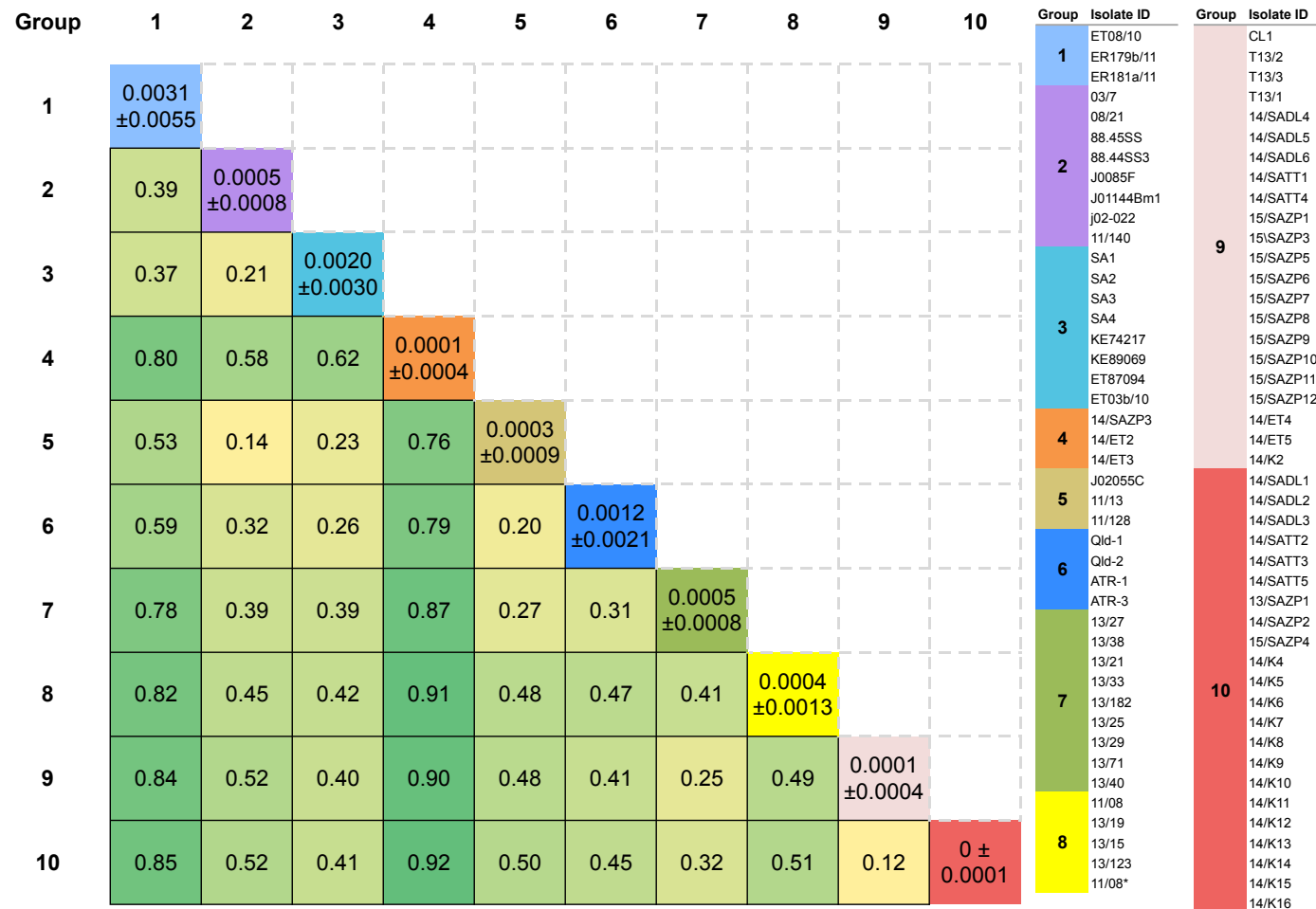


Figure 7.9: Measurements of genetic diversity by F_{ST} calculation of pairs of population groups indicated by the lower triangular matrix. The Watterson estimator of population diversity is given on the diagonal of the matrix. Colours of subpopulations is as shown in the DAPC population structure analysis bar plots (Figure 7.8). Comparisons with Group 4 (orange), and often Group 1, showed high F_{ST} values indicating that these groups were genetically very different from the other samples. Asterisk (*) indicates genomic data of isolate 11/08, while no asterisk indicates RNA-Seq data for 11/08.

on an extended set of wheat differential lines. The wheat differential set contained varieties with known *Yr* resistance genes, as well as unidentified sources of stripe rust resistance. Seedlings of 56 wheat varieties were tested under controlled environmental conditions with the historical South African isolates SA1 and SA4 and two field isolates, 13/SAZP1 and 15/SAZP4, collected in 2013 and 2015, respectively. To determine whether the genetic variance displayed in the phylogenetic analyses was linked to changes in the virulence profiles of these isolates the differential set was expanded by including additional wheat lines from the UK and Australia.

In the comparison between the isolates SA1 and 13/SAZP1 significant variability was not observed. The SA4 and the 2015 isolate, 15/SAZP4, displayed slight differences in infection types, with most prominent differences after infection of the wheat varieties Monterey (;cn versus 2cn) and Heines VII (;1+cn versus 3c), and a smaller, but observable difference on Kranich (;cn versus 1cn), Solstice (; versus ;c) and Selkirk (2cn versus 3=cn). These differences are visually displayed in Figure 7.10. Detailed results of all infection assays are listed in Appendix D, Tables D.1 and D.2.

7.4 Discussion

The field *Pst* population in South Africa was assessed at transcriptome level using 25 samples collected between 2013 and 2015. Along with these, four *Pst* isolates collected in 2014 in Ethiopia and 14 isolates collected in Kenya in 2014 were also evaluated.

Phylogenetic analysis placed the 2014 East African isolates in close proximity to one another. Additionally, it revealed patterns of high similarity between the field *Pst* population in South Africa, collected between 2013 and 2015, and the UK Cluster II triticale field isolates (T13/1, T13/2, T13/3 and CL1) described

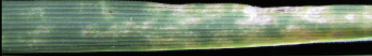
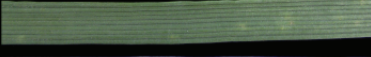
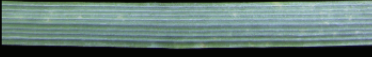
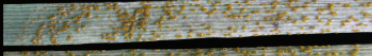
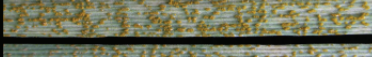
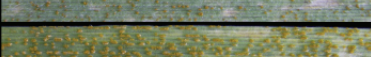
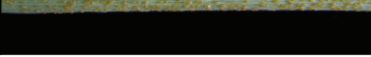
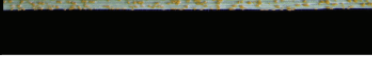
Wheat Differentials	SA4	15/SAZP4	Infection Types	
	6E22A+	6E22A-	SA4	15/SAZP4
Carstens V			;cn	;cn
Mosaic			;cn	;cn
Talon			;cn	;cn
Warrior			;	;
Vilmorin 23			;1=c	;1=c
Heines VII			;1+cn	3c
Heines Peko			;1=c	;c1=c
Reichersberg 42			3++c	3++c
Selkirk			2cn	3=cn
Cadenza			3c	3+c
Kranich			;cn	1cn
Monterey			;cn	2cn
Solstice			;	;c
Vuka			3=c	3++
Avocet S			4	4

Figure 7.10: Infection type comparisons between one historical and one recent *Pst* isolate. Infection types of SA4, from the historical population, and isolate 15/SAZP4 collected in 2015 are shown. Highly similar phenotypes were observed on wheat Warrior, Vilmorin 23, Heines Peko, Reichersberg 42. Differences in UK testers, including Kranich, Monterey and Solstice, were observed. The outcome of the remaining differential tests are summarised in Appendix D, Tables D.1 and D.2.

by Hubbard et al. (2015). Isolates collected in the same region of South Africa commonly clustered together. South African isolates from corresponding geographical and, by implication, climatic regions were often grouped together in the phylogenetic tree. Although this result has to be further investigated to draw further conclusions, pathotyped isolates with different virulence profiles were positioned on different branches. Genotyping of the *Pst* isolates indicated that a shift occurred in the South African *Pst* population, with the current *Pst* population being clearly differentiated from the earlier isolates sampled before 2012, and assessed in Chapter 4. The unexpected genetic relationship with the 2013 UK isolates, found on triticale, suggests a potential recent incursion of *Pst* into South Africa.

Results from the DAPC analysis mostly correlated with the phylogenetic findings and revealed signs of population structure, with three distinct groups containing field samples from South Africa. The historical South African isolates were placed in a separate, fourth group. All three South African field sample groups also included 2014 isolates from Kenya and/or Ethiopia. This supports the hypothesis raised in Chapter 4 stating the potential exchange of inoculum between South Africa and East Africa, although the East African isolates did not show as close resemblance to the UK Cluster II isolates as the co-clustering South African field isolates did (Figure 7.3). This indicates that inoculum was somehow spread between these locations or could be derived from the same progenitor. The South African and UK populations remained more similar, but share similarity with the East African population.

Regarding the DAPC analysis outcome of the South African field population, isolate 15/SAZP4, exhibiting a partially successful infection on Monterey that was not seen in the compared earlier isolate (SA4), was placed in Group 10, that was in addition to its high similarity to Group 9 ($F_{ST} = 0.12$), also very homogeneous according to the differentiation calculation within groups (0.0000 ± 0.0001). Both

Group 4 and Group 9 also containing South African field isolates, indicated low diversity amongst isolates (0.0001 ± 0.0004). The DAPC clusters only differed from the clades in the phylogenetic analysis by one isolate (14/SAZP3) when $K = 10$ is considered. The placement of 14/SAZP3 together with two isolates from Ethiopia, namely 14/ET2 and 14/ET3, forming Group 4 (orange) in the DAPC analysis is noteworthy. In the phylogenetic tree, sample 14/ET2 is the only isolate from the Ethiopian field isolates that show similarity to the East Africa (B) group, which contains isolate ET08/10, that was identified to be of the aggressive *PstS2* type (Hovmøller et al., 2008; Walter et al., 2016; Ali et al., 2017; M Hovmøller, personal communication). Grouping of these three isolates was however not displayed in the inferred phylogeny where 14/SAZP3 grouped with the other South African isolates. However, in the DAPC analysis, Group 4 differentiated early ($K = 6$) as displayed by the orange bars, compared to the red group ($K = 9$), containing the rest of the East African and South African field samples.

The high diversity that is shown when Group 4 (orange) is compared to Group 9 (light red, F_{ST} value 0.90) and Group 10 (red, F_{ST} value 0.92) indicate that Group 4, containing 14/SAZP3, differentiates considerably from Groups 9 and 10. This isolate, carrying the 6E16A- pathotype, was previously evaluated using microsatellite markers, and differentiated from other South African 6E16A- isolates (B Visser, personal communication). In previous infection assays, this isolate had a typical 6E16A- pathotype. This isolate was not evaluated on the extended differential set used in this study. This could be similar to the case discussed by Hubbard et al. (2015), where phenotypically similar isolates were genetically distinct and belonged to different populations. This further highlights the importance of genotyping along with differential testing in seasonal surveys. The genetic diversity between Groups 9 and 10 (F_{ST} value 0.12) was low in comparison to their diversity with Group 4. Further investigation is needed to

conclude that isolates related to the aggressive pathotype *PstS2* are present in South Africa.

Genetic change revealed by the phylogenetic and DAPC analyses in the South African *Pst* post-2012 population does not support stepwise mutational adaptation, but lead to speculations that an introduction of new *Pst* isolates occurred after 2012. This introduction could have occurred either through natural means where urediniospores could have been transferred by wind, or by human movement. According to the virulence profiles of the South African field isolates, pathological support for a new incursion is limited since new virulence has not been described in routine surveys. Ali et al. (2014) describe how such surveys can be biased as sampling is often done from wheat varieties that carry resistance genes that have been overcome, and usually not from field isolates. After evaluating some historical isolates with more recent phenotypic counterparts, SA1 and 13/SAZP1 had near to identical seedling phenotypes on all the wheat lines. Genotypic differences were however observed between SA1 and 13/SAZP1 using molecular marker analysis (Visser et al., 2016), as well as in this study.

Newly introduced *Pst* populations may also carry avirulences not inspected in local differential sets, as seen in the differentiation in infection types between SA4 and 15/SAZP3. The most notable difference between infection by these two isolates were on Monterey. It is a winter wheat cultivar bred in the UK by the company Senova. It is not known what stripe rust resistance genes are present in this variety, but it shows moderate levels of resistance in the UK. For instance, it was listed by the UK Cereal Pathogen Virulence Survey (UKCPVS) as being “susceptible as an adult plant to one or more of the current stripe rust pathotypes” and scored 7.3 on the stripe rust resistance rating in 2014, where possible scores ranged from 1 to 9; 1 = highly susceptible, 9 = resistant (Hubbard et al., 2014). Monterey has the pedigree Istabraq x Robigus. Robigus is fully susceptible to all UK *Pst* pathotypes, whereas Istabraq has the pedigree Consort

x Claire which both provide some resistance to UK *Pst* pathotypes, although resistance in Claire has been eroded over the past few years. The aggressive *Pst* Kranich pathotype was first detected in the UK on Monterey (S Holdgate, personal communication). The wheat variety Kranich has the pedigree Heines-2167-50/Heines-VII//Merlin/Deu. Small pustules were observed on Kranich inoculated with 15/SAZP4, while the SA4 inoculation resulted in flecks only, with signs of chlorotic and necrotic tissue. Taken together, this may indicate that a source of stripe rust resistance from Heines VII, present in Kranich and Monterey has become less effective towards *Pst* isolate 15/SAZP4.

The role that the host plays in shaping the characteristics of the pathogen has not been addressed in this study. Wheat breeding in South Africa has generally relied on selection for resistance in the field and information about stripe rust resistance genes deployed in commercial wheat over the past 20 years is not obtainable, as reviewed by Pretorius et al. (2007). Only as recent as 2012, marker assisted selection (MAS) has been incorporated in breeding programmes with the establishment of the Molecular marker Service Laboratory (MSL) for wheat breeding in South Africa (Prins and Agenbag, 2013). In the past, germplasm from the International Maize and Wheat Improvement Center (CIMMYT) has been the origin of valuable resistance complexes. The presence of the slow rusting complex *Lr34/Yr18/Sr57*, present in South African spring wheat cultivar, Kariega (Ramburan et al., 2004; Prins et al., 2011), was likely introduced by a CIMMYT source. The lack of structured molecular breeding efforts incorporating rust resistance in the past makes it difficult to track specific selection pressures on *Pst* imposed by host resistance. No connection could be made between the Monterey and the South African germplasm.

The widely homogenous nature of the *Pst* population could be due to the introduction of a relatively small amount of inoculum displaying the founder effect, where genetic variation is lost when a small number of individuals es-

establish the population. Additionally or alternatively, a population bottleneck could have occurred where a limited amount of genotypes, able to sustain for example environmental conditions, survived from one wheat-growing season to the next. Environmental factors, either directly, or indirectly through effects on the host, can increase stress on the pathogen, acting as a force for adaptation. Severe droughts have been experienced in both main wheat growing regions in South Africa. This could have contributed to the low occurrence of stripe rust in recent years. It is possible that these non-optimal conditions encountered by the *Pst* population, during and between wheat seasons, may have also contributed to a population bottleneck. The majority of the 2014 South African field samples differentiated from the 2015 field samples, indicating that the population evolved from one growing season to the next, or it could again indicate the influx of new alleles into the population. The low occurrence of stripe rust in South Africa could have led to a change in allele frequencies in the population, similar to the “chance events” described by Wellings (2007). Relatively low numbers of spores may survive during the non-crop seasons, possibly on alternative grass species (Boshoff et al., 2002; Pretorius et al., 2015), resulting in such allele frequency shifts.

Anthropogenic movement in and out of South Africa has drastically increased since the change in the country’s political system in 1994. Tourism and trade act as passages for pathogens to travel long distances much more quickly and frequently than through migration via animal vectors or storms (Anderson et al., 2004). The increase in number of international arrivals indicated by the World Tourism Organisation (Figure 7.11¹) demonstrates the increase in the potential for exotic incursions by pathogens via human movement. Anderson et al. (2004) considered this as the major driver of emerging infectious diseases.

¹<https://data.worldbank.org/indicator/ST.INT.ARVL?contextual=default&end=2014&locations=ZA&start=1995&view=chart>

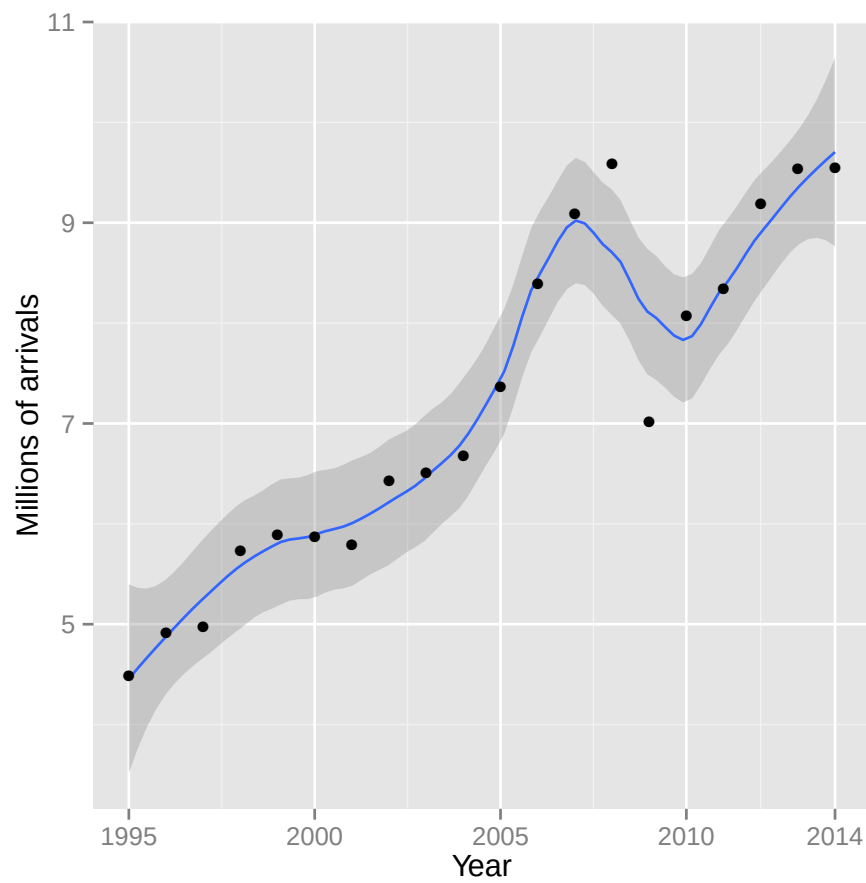


Figure 7.11: Number of international tourist arrivals in South Africa between 1995 and 2014.

7.5 Conclusion

No new virulence profiles for stripe rust have been reported in routine surveys in South Africa since 2005. In addition, no exclusive correlation could be seen between the genotypic change observed in South African field isolates and their virulence profiles, as shown in the SA1 vs 13/SAZP1 comparison in this study. The 2015 field isolate 15/SAZP4 was found to be partially virulent on the UK winter wheat cultivar, Monterey, but no change in virulence was seen with isolate 13/SAZP1 on the extended differential wheat set. The differentiation from SA1 that was observed in this study for 14/SAZP3, also carrying the 6E16A-pathotype, was also characterised using microsatellite markers (Visser et al., 2016). The microsatellite marker research supports the fact that a genetically diverse population carries the pathotype 6E16A-. No evidence could be found of common parentage between Monterey and South African wheat varieties that could account for selection for virulence in South Africa to the stripe rust resistance currently present in Monterey. Further investigation would be needed to identify which source of resistance in Monterey was challenged by 15/SAZP4.

The data discussed in this chapter shows evidence of a definite change in the South African *Pst* population between 2013 and 2015. It is likely that this is due to an exotic incursion of *Pst* from outside South Africa. The *Pst* population also showed an allele frequency change between 2014 and 2015. It is possible that a population bottleneck, due to unfavourable environmental conditions, was responsible for this shift. Further research is required to determine which scenario has contributed to the changes in the *Pst* population in South Africa, including a systematic collection of stripe rust infected wheat leaf samples throughout the growing season, and wild grass between seasons.

Chapter 8

General Discussion

THIS STUDY SET OUT to examine the genetic structure of the *Pst* population in South Africa, with specific focus on the genetic variation related to pathotype variation. Previous descriptions made use of traditional pathotyping and molecular marker technologies (Pretorius et al., 1997; Boshoff et al., 2002; Pretorius et al., 2007; Ali et al., 2014; Hovmøller et al., 2008; Visser et al., 2016). In this study, characterisation was undertaken using next-generation Illumina sequencing of *Pst* genomes and transcriptomes, and bioinformatics analyses to extend our knowledge of the South African *Pst* population and its evolutionary dynamics. Specific interests included the origin of *Pst* introduced into South Africa, the relationship between the four pathotypes identified so far, identification of effector coding genes possibly responsible for distinct virulences, and genomic and pathological investigation of recent field *Pst* populations.

8.1 The historical South African *Pst* population

Phylogenetic and clustering analyses, supported by evaluation of the genetic diversity, reinforced previous findings which stipulated that the historical *Pst* population in South Africa, represented by the *Pst* isolates collected between 2001

and 2011, had a close relationship to each other despite their distinct differences in virulence. Data from this study supports previous reports that the four pathotypes were derived from one another through stepwise evolution (Visser et al., 2016).

Analysis of the relationship of the historical South African isolates with available foreign isolates indicated a possible origin from Kenya and Ethiopia, or a common progenitor from elsewhere. Significant diversity was observed in the East African isolates, which formed two distinct groups, one closely related to the South African isolates and one distant from all other isolates assessed in this study. The East African isolates (Group A) that clustered with the historical South African isolates contained three isolates collected in the 1970s and 1980s and one isolate collected in 2010 (Figure 4.6 and 4.10). We therefore confirm associations based on pathotype analysis that the South African *Pst* incursion of 1996 had a high probability of originating from East Africa (Pretorius et al., 1997; Boshoff et al., 2002; Pretorius et al., 2007).

These conclusions were supported by previous pathotype analyses that showed the presence of 6E16A- in East Africa (Badebo et al., 1990). However, similar pathotype designations may be shared between distinct isolates, for example the Ethiopian wheat variety Et-13 A2 was resistant to 6E16A and 6E22A isolates from South Africa, but susceptible to 6E22 isolates from Germany (Hussein and Pretorius, 2005; Badebo et al., 2008; Denbel, 2014). Genetic evidence from microsatellite marker analysis indicated 48 % similarity between South African isolates and the Kenyan isolates KE 10/09 and KE 12/09 (Visser et al., 2016). Differences could be due to virulence for *Yr9* and *Yr27* that is frequently observed in East Africa but absent in South Africa¹. It was unfortunate that the present study did not include isolate data from additional locations south of Kenya which would have enabled the tracking of the putative southward spread of *Pst* into

¹<http://rusttracker.cimmyt.org>

South Africa. Earlier reports state that stripe rust is not a major problem in these regions (Stubbs, 1985), however, analysis of samples from Rwanda and Tanzania suggests that collections from more Southern African countries could be included in on-going work to monitor gene flow (Ali et al., 2017).

Previous studies that included *Pst* isolates from Eritrea, indicated Central and Western Asia, and the Mediterranean as possible origins of South African isolates (Enjalbert et al., 2005; Hovmöller et al., 2008). Ali et al. (2014) further reported that the South African isolates (collected between 1996 and 2004) grouped with the older, aggressive group known as *PstS3* often seen in Southern Europe. There is agreement between these studies and the present study with regards to the South African isolates not showing close relationships with isolates from Eritrea, however, isolates from Ethiopia and Kenya were not included in these studies. It would be interesting to assess more South African isolates collected between 1996 and 2011, and also to compare the South African isolates to *Pst* isolates from other Eastern and Southern African countries, as well as Asian and Mediterranean isolates, using the field pathogenomics approach as method of investigation. Such analyses would be subject to the availability of historical samples, but would enable inspection of the different hypotheses regarding the origin of South African *Pst*.

8.2 Candidate effector identification and evaluation

Nonsynonymous polymorphism analysis aided in identifying candidate genes possibly involved in virulence. The analysis relied on available effector gene annotations and made use of the initial gene models developed for the PST130 reference genome. It is widely argued that high throughput effector gene annotation protocols are difficult to develop for the rusts as they do not exhibit many of the common features that are known to be characteristic of other, more

thoroughly described pathogens (Dodds et al., 2009; Saunders et al., 2012). It is therefore accepted that any computational protocol, despite its best efforts, would likely misidentify some effector genes. New research findings and tools allow constant refinement of gene predictions, as was the case for the PST130 reference, where gene annotations have been improved since the start of this study.

To evaluate candidate effector gene expression during *Pst* infection, RT-qPCR was used. This methodology has been used in a number of published studies, but many of these lack detail on experimental procedures. It is often seen that best practices, as advised by developers and supporters of the technology, are not followed or not reported, misleading newcomers to the field. Greater efforts are needed to ensure that published work using RT-qPCR follow The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Huggett et al., 2013).

In this study, the consistent expression patterns shown by the two South African isolates across all genes indicated a low level of technical variation seen between individual assays within a PCR plate. However, variation between plates hindered the formulation of confident conclusions from these experiments. In addition, evaluations of early time points were not informative using this method due to low concentrations of fungal transcripts. Continued efforts are needed to enable evaluation of gene expression from the moment of inoculation up to around two dpi to capture expression profiles of genes involved in the early processes of infection. Four candidate effector genes overlapped between this study and time course evaluations of two UK *Pst* isolates (Cantu et al., 2013). Future research should prioritise investigation of these four candidate genes. As a start, heterologous expression screens in *Nicotiana benthamiana* could be performed to add to the available information gained from this system about one of the four candidates, PST130_05023 (Petre et al., 2016b).

8.3 The recent South African *Pst* population

Surprisingly, analysis of RNA-Seq data of recent field isolates indicated an allele frequency shift in the South African *Pst* population. Previously this population was thought of as fairly stable because of the lack of detection of additional virulences between 2005 and 2015, when the last field isolates were sampled. These field isolates showed a close relationship to UK *Pst* isolates collected on triticale (UK Group II; Hubbard et al., 2015 and Bueno-Sancho et al., 2017). Whether or not these UK isolates were able to infect wheat is not known as they were not successfully cultured and have been lost (S Holdgate, personal communication).

Compared to the 2013–2015 South African isolates, field isolates collected in Kenya and Ethiopia in 2014 were more similar to the pre-2011 East African and South African isolates, as indicated by the phylogenetic analysis. This analysis used the third codon position of genes with 80 % breadth of coverage in 80 % of isolates. DAPC clustering analysis used sites where a polymorphism resulting in a synonymous substitution in at least one isolate was recorded. In this analysis, the 2014 East African isolates did not group with the pre-2011 East African and South African isolates, but with the recent South African and UK Group II isolates. Two groups, namely Group 1—also described as East Africa (B)—indicated in blue in Figure 7.7(iii), and Group 4, indicated in orange and containing three 2014 isolates, two from Ethiopia and one from South Africa, included in the dataset in Chapter 7, showed high diversity, clustering away from the rest of the isolates considered in the DAPC analyses. This diversity could result in the software having difficulty to separate more similar isolates into population clusters. The two results differ primarily in their indication of the closest relatives of the 2014 East African isolates. There is however consensus between the two analyses with regards to the recent South African isolates, showing closer similarity to

the UK Group II isolates than the historical South African isolates. Comparative re-evaluation of selected recent South African isolates to pre-2011 isolates on an extended wheat differential set confirmed previous findings in two isolates. The 6E16A- pathotype was confirmed in isolates SA1 and 13/SAZP1 with nearly identical infection types. However, evaluation of SA4 and 15/SAZP4 revealed diverging infection types.

Disagreement exists between studies regarding similarities in the European and Ethiopian populations. Using virulence phenotyping together with AFLP, microsatellite and SCAR marker information, Ali et al. (2017) described a diverse *Pst* population with more than four pathotype groups in East Africa, collected between 2009 and 2015, that were distinct from the assessed European isolates. Among these East African isolates were samples from Ethiopia. In contrast, support for a close relationship between the UK Group II isolates and Ethiopian isolates from 2014 was reported where a number of Ethiopian isolates were assigned to this group, along with isolates collected in Europe in 2014 (Bueno-Sancho et al., 2017). The authors further revealed the assignment of historical *Pst* isolates from New Zealand that were collected between 2006 and 2012, to this group.

Taken together this data provides evidence that a new incursion may have occurred in South Africa, possibly between 2011 and 2013, and the commonalities with UK Group II *Pst* indicate the possible spread of this *Pst* group over vast distances. These findings should alert the research and agricultural community that the *Pst* population in South Africa could be more dynamic than is currently thought to be the case. However, similar infection types in historical and recent isolates tested on existing differentials gave rise to scepticism. Further investigation of East African and UK Group II *Pst* isolates is needed to support the current findings and track the global movement of this group. Sequencing of field isolates to monitor new incursions complementary to virulence profiling of *Pst* across

cropping seasons would be beneficial to facilitate comprehensive surveys. The cost of implementing the field pathogenomics approach (Hubbard et al., 2015) is unfortunately a major limiting factor to deployment of this technology in routine pathotype surveys in South Africa.

8.4 Future work

Effective, long term rust resistance in wheat can be implemented by pyramiding resistance genes. Ideally, breeders should combine major, *R* gene type and APR genes. This relaxes selection pressure on the pathogen population that can normally rapidly overcome singly deployed *R* genes. Understanding the mechanisms of *R* genes and their corresponding *Avr* genes, as in the case of the recently published *AvrSr35* (Salcedo et al., 2017) and *AvrSr50* (Chen et al., 2017) studies, can help breeders to track high-risk pathotypes to help tailor the deployment of resistance genes. Another approach would be to identify the target “susceptibility” genes of *Pst* effectors, such as the barley powdery mildew susceptibility gene *Mlo*. Mutations in *Mlo* created the recessive *mlo* allele that has provided broad-spectrum resistance against the fungus *Blumeria graminis* f. sp. *hordei* for many years (Büschges et al., 1997). Targeted mutation breeding of *Pst* effector target genes in wheat, using DNA-editing technologies such as CRISPR/Cas9 (Kim et al., 2018), could generate suites of mutant genes conferring resistance to *Pst*. Identifying the mechanisms, both in the host and the pathogen, that provide durable resistance is the aim of many future studies (Harris et al., 2015). Advances in research that enable understanding of how effectors function include protein interaction assays such as yeast-two-hybrid screens, gene expression knock-downs, for example using virus-mediated host-induced gene silencing and heterologous expression of effector genes in easily transformed host plants such as *N. benthamiana* (Liu et al., 2016; Petre et al.,

2016b). Other delivery systems such as the type III secretion system in bacteria have also been proposed to deliver specific proteins into host cells (Ma et al., 2009; Upadhyaya et al., 2013). Using these technologies, refinement of *Pst* gene annotations and the first available *Pst* haplotype-phased genome (Schwessinger et al., 2018) all provide promising potential resources to further assess wheat-*Pst* interactions in the search for long lasting resistance to improve wheat yields and reduce the evolutionary potential of rust pathogens by reducing inoculum.

8.5 Conclusion

In conclusion, although there remains a significant gap in our understanding of genes that are responsible for the virulence gain in the historical South African population, this study showed that, contrary to conclusions from previous studies, novel genetic variation that has not been described previously, is indeed present in the recent South African population. For the first time, according to our knowledge, the *Pst* populations of Ethiopia, Kenya and South Africa were linked using high-resolution genomic and transcriptomic data. This confirms earlier associations between pathotypes from eastern Africa and South Africa and verifies the risk for the introduction of more aggressive pathotypes into South Africa. Further characterisation of isolates that are associated with the UK Group II isolates, with specific focus on their pathogenicity, will aid in understanding the risks involved in long distance movement of *Pst* and ultimately help producers to decrease the incidence of disease and increase crop yields, which will in turn relieve the pressure on global food production to meet rising demands.

Appendix A

The Origin of the South African *Pst* Pathotypes

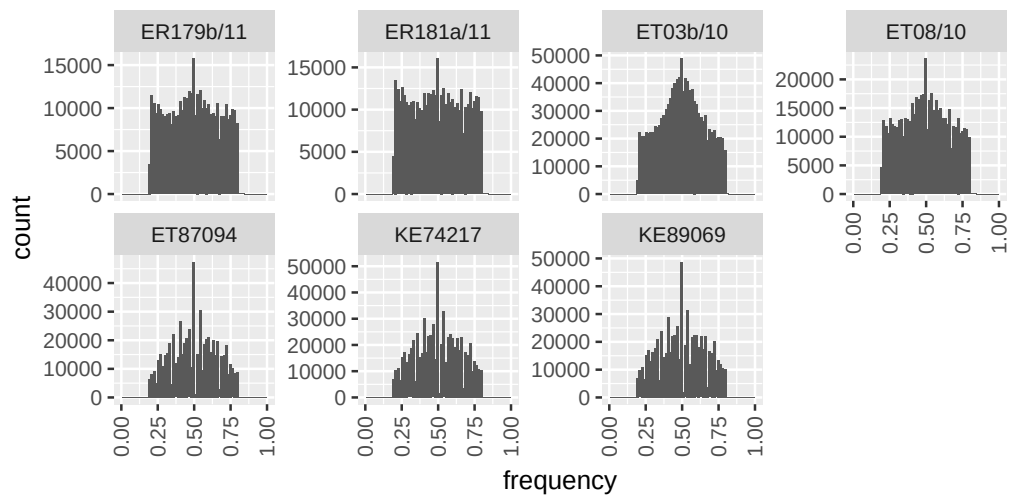


Figure A.1: Read frequency graphs for East African isolates analysed in Chapter 4, that have not been similarly assessed in published studies (Cantu et al., 2013; Hubbard et al., 2015; Bueno-Sancho et al., 2017). See Table 4.1 for further identification purposes.

Appendix B

Analyses of Polymorphisms in Historical South African *Pst* Isolates in Search of Candidate Effector Genes

B.1 Genes present in the PST130 reference genome but absent in the four historical South African *Pst* isolates

Table B.1: PST130 genes (211) that were absent in all four historical South African isolates

PST130_00014	PST130_03142	PST130_08220	PST130_14020
PST130_00053	PST130_03351	PST130_08341	PST130_14034
PST130_00147	PST130_03414	PST130_08456	PST130_14069
PST130_00148	PST130_03415	PST130_08466	PST130_14429
PST130_00159	PST130_03429	PST130_08469	PST130_14430
PST130_00173	PST130_03543	PST130_08470	PST130_14605
PST130_00227	PST130_03607	PST130_08628	PST130_14606
PST130_00246	PST130_03762	PST130_08645	PST130_14653
PST130_00348	PST130_03775	PST130_08669	PST130_14781
PST130_00404	PST130_03798	PST130_08880	PST130_14925
PST130_00445	PST130_03847	PST130_08891	PST130_14963
PST130_00483	PST130_04103	PST130_09448	PST130_14964
PST130_00611	PST130_04396	PST130_10110	PST130_15027
PST130_00612	PST130_04591	PST130_10111	PST130_15648
PST130_00656	PST130_04612	PST130_10209	PST130_15841
PST130_00812	PST130_04613	PST130_10271	PST130_16094
PST130_00848	PST130_05005	PST130_11019	PST130_16216
PST130_00945	PST130_05050	PST130_11064	PST130_16356
PST130_00950	PST130_05150	PST130_11200	PST130_16357
PST130_00989	PST130_05183	PST130_11219	PST130_16435
PST130_01030	PST130_05199	PST130_11289	PST130_16508
PST130_01031	PST130_05303	PST130_11403	PST130_16509
PST130_01079	PST130_05357	PST130_11404	PST130_16568
PST130_01080	PST130_05569	PST130_11537	PST130_16737
PST130_01081	PST130_05640	PST130_11550	PST130_16763
PST130_01082	PST130_05683	PST130_11607	PST130_16764
PST130_01107	PST130_05804	PST130_11862	PST130_16830
PST130_01143	PST130_06069	PST130_11902	PST130_16914
PST130_01368	PST130_06079	PST130_11946	PST130_16963
PST130_01388	PST130_06120	PST130_11947	PST130_17078
PST130_01690	PST130_06121	PST130_11948	PST130_17111
PST130_01696	PST130_06122	PST130_12027	PST130_17182
PST130_01697	PST130_06123	PST130_12084	PST130_17218
PST130_01825	PST130_06147	PST130_12310	PST130_17238
PST130_01826	PST130_06262	PST130_12311	PST130_17253
PST130_01847	PST130_06356	PST130_12346	PST130_17316
PST130_01859	PST130_06479	PST130_12435	PST130_17354
PST130_01946	PST130_06533	PST130_12436	PST130_17435
PST130_02005	PST130_06608	PST130_12446	PST130_17515
PST130_02139	PST130_06609	PST130_12481	PST130_17560
PST130_02140	PST130_06687	PST130_12509	PST130_17599
PST130_02142	PST130_06741	PST130_12825	PST130_17620
PST130_02153	PST130_06775	PST130_12971	PST130_17812
PST130_02289	PST130_07080	PST130_12992	PST130_17815
PST130_02406	PST130_07081	PST130_13083	PST130_17898
PST130_02413	PST130_07180	PST130_13431	PST130_17956
PST130_02482	PST130_07220	PST130_13432	PST130_17990
PST130_02770	PST130_07285	PST130_13436	PST130_17991
PST130_02826	PST130_07330	PST130_13455	PST130_17992
PST130_03059	PST130_07486	PST130_13530	PST130_18018
PST130_03060	PST130_07943	PST130_13926	PST130_18083
PST130_03094	PST130_07959	PST130_13932	PST130_18108
PST130_03099	PST130_08034	PST130_13936	

B.2 Annotations of genes homologous to identified PST130 genes

PST130_00159 Accession gi|403167846| ref|XM_003327549.2|

Homolog *Pgt* isoleucyl-tRNA synthetase (PGTG_09131)

UniProtKB/ TrEMBL ID E3KG82

Protein name Isoleucyl-tRNA synthetase

Associated Function or Cellular location

Enzyme involved in protein biosynthesis during translation. Present in cytoplasm.

GO terms

GO:0002161 (aminoacyl-tRNA editing activity), GO:0005524 (ATP binding), GO:0004822 (isoleucine-tRNA ligase activity), GO:0000049 (tRNA binding). GO:0006428 (isoleucyl-tRNA aminoacylation)

Conserved domains

PLN02882: aminoacyl-tRNA ligase; cd07961: Anticodon-binding domain of archaeal, bacterial, and eukaryotic cytoplasmic isoleucyl tRNA synthetases; cd00818: catalytic core domain of isoleucyl-tRNA synthetases

PST130_07080 Accession gi|403159121| ref|XM_003319730.2|

Homolog *Pgt* hypothetical protein (PGTG_01952)

UniProtKB/ TrEMBL ID E3JT80

Protein name Uncharacterised protein

Associated Function or Cellular location

Helicases are ATPase enzymes that catalyse the unwinding of double-stranded nucleic acids. Involved in processes such as DNA replication, recombination, and nucleotide excision repair, as well as RNA transcription and splicing.

GO terms

GO:0009055 (electron transfer activity), GO:0016491 (oxidoreductase activity), GO:0035091 (phosphatidylinositol binding), GO:0009061 (anaerobic respiration), GO:0022900 (electron transport chain)

Conserved domains cd06869: The PX domain is a phosphoinositide (PI) binding module involved in targeting proteins to PI-enriched membranes. Diverse functions such as cell signalling, vesicular trafficking, protein sorting, lipid modification, cell polarity and division, activation of T and B cells, and cell survival.; pfam12825: Domain of unknown function in PX-proteins.; pfam12828: PX-associated

PST130_16763 Accession gi|403160602| ref|XM_003321038.2|

Homolog *Pgt* hypothetical protein (PGTG_02128)

UniProtKB/ TrEMBL ID E3JX92

Protein name Uncharacterised protein

Associated Function or Cellular location

Location associated with P-body and nucleolus. Cytoplasmic stress granule.

GO terms

GO:0005524 (ATP binding), GO:0004004 (ATP-dependent RNA helicase activity), GO:0003676 (nucleic acid binding), GO:0033962 (cytoplasmic mRNA processing body assembly), GO:0006417 (regulation of translation), GO:0010501 (RNA secondary structure unwinding)

Conserved domains

COG0513: Superfamily II DNA and RNA helicase [Replication, recombination and repair], cd00079. Helicase superfamily c-terminal domain; cl21455. P-loop containing Nucleoside Triphosphate Hydrolases. Involved in diverse cellular functions

PST130_17182 Accession gi|403160450| ref|XM_003320901.2|

Homolog *Pgt* hypothetical protein (PGTG_02971)

UniProtKB/ TrEMBL ID E3JWV5

Protein name Uncharacterised protein

Associated Function or Cellular location

Location associated with cytoplasm, endoplasmic reticulum, membrane component.

GO terms

GO:0016491 (oxidoreductase activity); GO:0016627 (oxidoreductase activity, acting on the CH-CH group of donors); GO:0042761 (very long-chain fatty acid biosynthetic process)

Conserved domains

PLN02560: enoyl-CoA reductase; cl00155: Ubiquitin homologs. Ubiquitin-mediated proteolysis is part of the regulated turnover of proteins required for controlling cell cycle progression. cl21511: The *Saccharomyces cerevisiae* Meyen ex EC Hansen phospholipid methyltransferase (EC:2.1.1.16) has a broad substrate specificity of unsaturated phospholipids.

PST130_17354 — A Accession gi|403161086| ref|XM_003890392.1|

Homolog *Pgt* hypothetical protein (PGTG_20899)

UniProtKB/ TrEMBL ID H6QPU7

Protein name Glycogen [starch] synthase

Associated Function or Cellular location

Enzyme that catalyse the transfer of glycosyl (sugar) residues to an acceptor, both during degradation (cosubstrates= water or inorganic phosphate) and during

biosynthesis of polysaccharides, glycoproteins and glycolipids.

GO terms

GO:0004373 (glycogen (starch) synthase activity); GO:0005978 (glycogen biosynthetic process)

Conserved domains

cl10013: Glycosyltransferases catalyse the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds.

PST130_17354 — B Accession gi|403166809| ref|XM_003326625.2|

Homolog *Pgt* glycogen [starch] synthase (PGTG_07651)

UniProtKB/ TrEMBL ID E3KCW8

Protein name Glycogen [starch] synthase

Associated Function or Cellular location —

GO terms

GO:0004373 (glycogen (starch) synthase activity); GO:0005978 (glycogen biosynthetic process)

Conserved domains

cd03793: Glycogen synthase, catalyses the transfer of a glucose molecule from UDP-glucose to a terminal branch of a glycogen molecule, a rate-limit step of glycogen biosynthesis.; pfam05693: Glycogen synthase. It is the rate limiting enzyme in the synthesis of the polysaccharide, and its activity is highly regulated through phosphorylation at multiple sites and also by allosteric effectors, mainly glucose 6-phosphate (G6P).

PST130_17620 Accession gi|403174779| ref|XM_003333656.2|

Homolog *Pgt* hypothetical protein (PGTG_15464)

UniProtKB/ TrEMBL ID E3KYK8

Protein name Uncharacterised protein

Associated Function or Cellular location —

GO terms

GO:0003824(Catalysis of a biochemical reaction at physiological temperatures.);

GO:0009058 (The chemical reactions and pathways resulting in the formation of substances; typically the energy-requiring part of metabolism in which simpler substances are transformed into more complex ones.)

Conserved domains

cd00609: Aspartate aminotransferase family. This family belongs to pyridoxal phosphate (PLP)-dependent aspartate aminotransferase superfamily (fold I). Pyridoxal phosphate combines with an alpha-amino acid to form a compound called a Schiff base or aldimine intermediate, which depending on the reaction, is the substrate in four kinds of reactions (1) transamination (movement of amino groups), (2) racemisation (redistribution of enantiomers), (3) decarboxylation (removing COOH groups), and (4) various side-chain reactions depending on the enzyme involved.; COG0436: Amino acid transport and metabolism. linked to 3D-structure.

PST130_17815 Accession gi|403157775| ref|XM_003307127.2|

Homolog *Pgt* 1,3-beta-glucan synthase component FKS1 (PGTG_00125)

UniProtKB/ TrEMBL ID E3JR07

Protein name 1,3-beta-glucan synthase component FKS1

Associated Function or Cellular location

Component of the plasma membrane.

GO terms

GO:0003843 (1,3-beta-D-glucan synthase activity); GO:0006075 ((1->3)-beta-D-glucan biosynthetic process).

Conserved domains

pfam02364: 1,3-beta-glucan synthase component. 1,3-beta-glucan synthase EC:2.4.1.34 also known as callose synthase catalyses the formation of a beta-1,3-glucan polymer that is a major component of the fungal cell wall/

PST130_00758 Accession gi|403160953| ref|XM_003321311.2|

Homolog *Pgt* hypothetical protein (PGTG_02401)

UniProtKB/ TrEMBL ID E3JY15

Protein name Uncharacterised protein

Associated Function or Cellular location

P-body: A focus in the cytoplasm where mRNAs may become inactivated by decapping or some other mechanism. Protein and RNA localized to these foci are involved in mRNA degradation, nonsense-mediated mRNA decay (NMD), translational repression, and RNA-mediated gene silencing.

GO terms

GO:0003729 (mRNA binding) GO:0030371 (translation repressor activity - Antagonises ribosome-mediated translation of mRNA into a polypeptide) GO:0017148 (negative regulation of translation) GO:0000289 (nuclear-transcribed mRNA poly(A) tail shortening).

Conserved domains

smart00454. Sterile alpha motif. Widespread domain in signalling and nuclear proteins.; cl15755. SAM (Sterile alpha motif) is a module consisting of approximately 70 amino acids. This domain is found in the Fungi/Metazoa group and in a restricted number of bacteria. SAM domains have diverse functions and locations. They can interact with proteins, RNAs and membrane lipids, contain site of phosphorylation and/or kinase docking site, and play a role in protein homo and hetero dimerisation/oligomerisation in processes ranging from signal

transduction to regulation of transcription. Mutations in SAM domains have been linked to several diseases.

PST130_08345 Accession gi|403162070| ref|XM_003322301.2|

Homolog *Pgt* hypothetical protein (PGTG_03886)

UniProtKB/ TrEMBL ID E3K0V5

Protein name Aconitate hydratase, mitochondrial

Associated Function or Cellular location

Associated with the mitochondrion. Protein which binds at least one iron atom, or protein whose function is iron-dependent. Involved in metabolic processes that result in cell growth.

GO terms

GO:0051539 (4 iron, 4 sulfur cluster binding); GO:0003994 (aconitate hydratase activity); GO:0046872 (metal ion binding); GO:0032543 (mitochondrial translation); GO:0006099 (tricarboxylic acid cycle).

Conserved domains

TIGR01340: aconitate hydratase, mitochondrial. [Energy metabolism, TCA cycle]; cl00215. Aconitase swivel domain. Aconitase (aconitate hydratase) catalyses the reversible isomerisation of citrate and isocitrate as part of the TCA cycle. cl00285. Aconitase catalytic domain. Both cl00215 and cl00285 are present in enzymes involved in biosynthesis of leucine.

PST130_12299 Accession gi|403173188| ref|XM_003332239.2|

Homolog *Pgt* hypothetical protein (PGTG_14583) **UniProtKB/ TrEMBL ID** E3KU93

Protein name Uncharacterised protein

Associated Function or Cellular location

Associated with the cytosol, nucleus and membrane.

GO terms

GO:0003723 (RNA binding) GO:0043130 (binding ubiquitin, involved in proteolytic degradation) GO:0031081 (nuclear pore distribution) GO:0016973 & GO:0006606 (poly(A)+ mRNA export / protein import from nucleus into the cytoplasm / vice versa) GO:0000972 & GO:0000973 (transcriptional & posttranscriptional tethering of RNA polymerase II gene DNA at nuclear periphery) GO:2000728 (regulates mRNA export from nucleus in response to heat stress) GO:0006405 (RNA export from nucleus to the cytoplasm).

Conserved domains

COG2319: WD40 repeat [General function prediction only] sd00039: WD40 repeats in seven bladed beta propellers. The WD40 repeat is found in a number of eukaryotic proteins that cover a wide variety of functions including adaptor/regulatory modules in signal transduction, pre-mRNA processing, and cytoskeleton assembly; cl02567: WD40 Superfamily.

B.3 Nonsynonymous polymorphisms in candidate genes

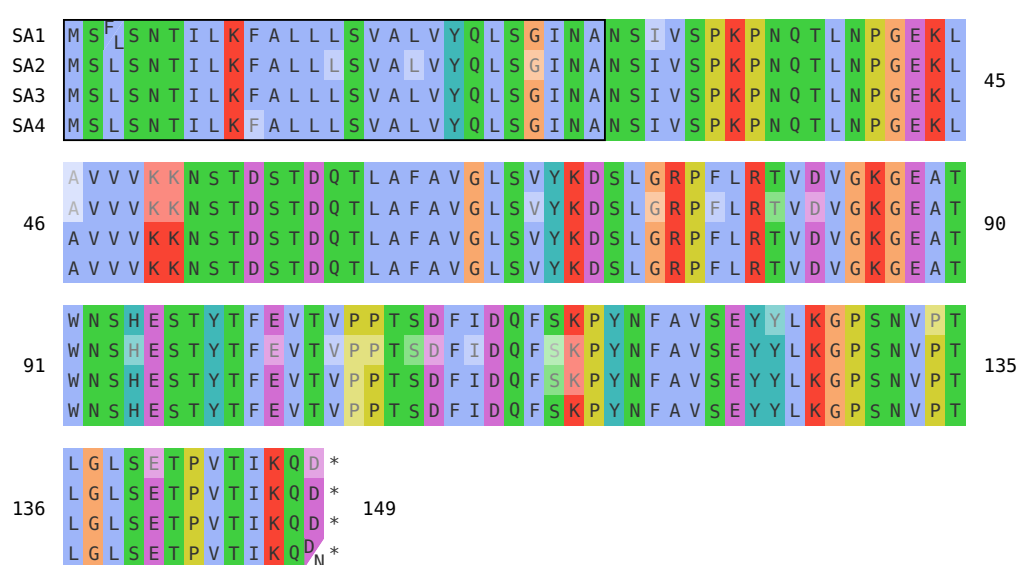


Figure B.1: Translated sequence alignment of gene PST130_02001. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

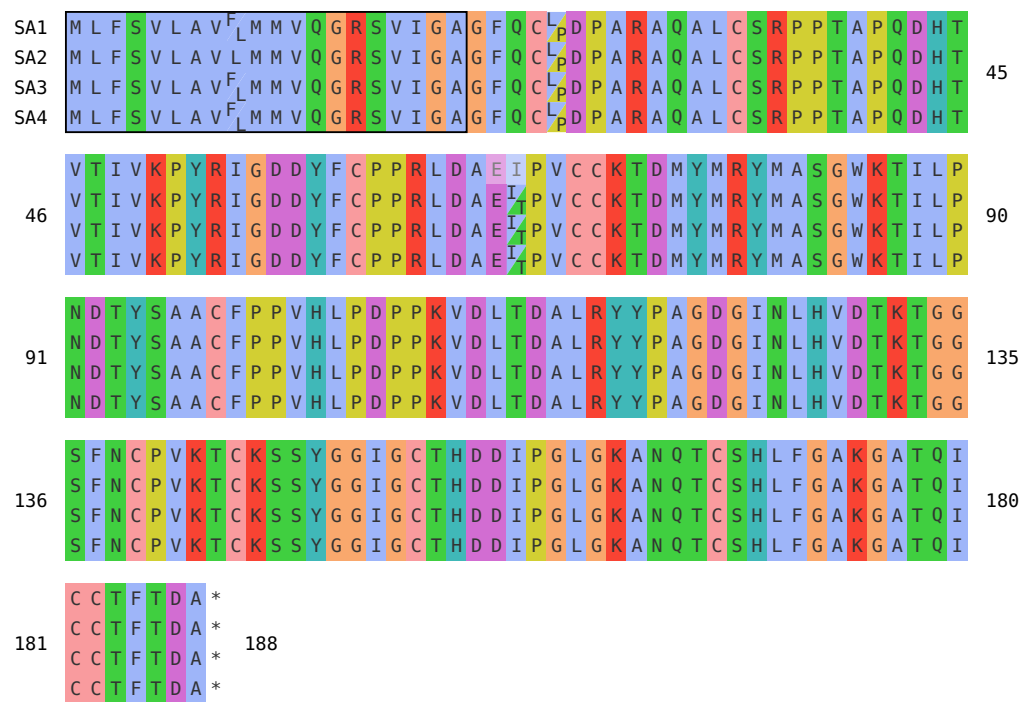


Figure B.2: Translated sequence alignment of gene PST130_02118. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

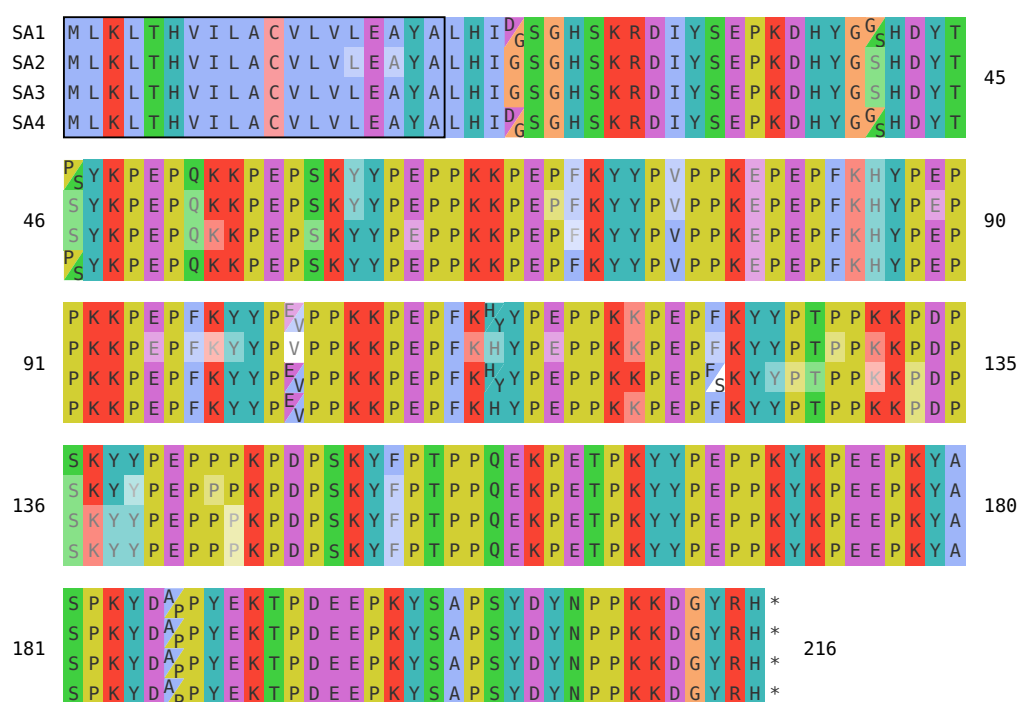


Figure B.3: Translated sequence alignment of gene PST130_02403. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

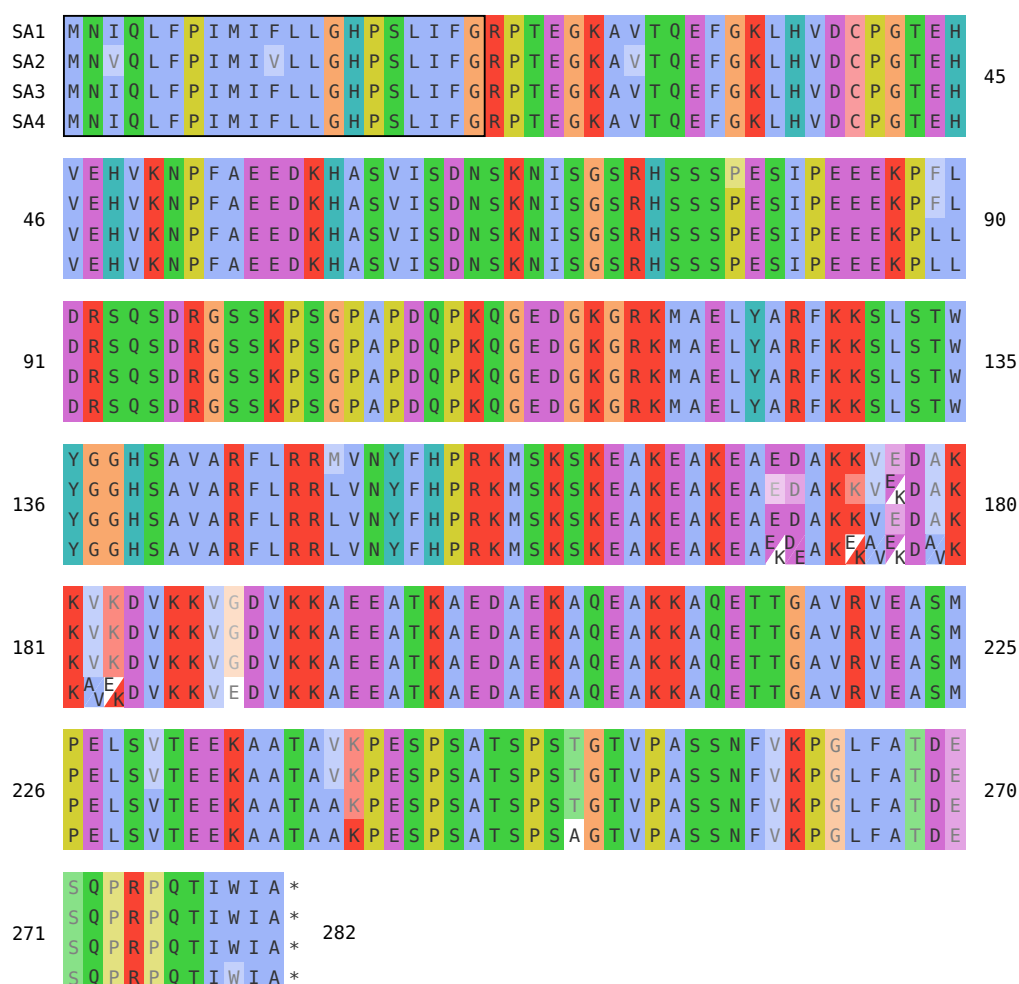


Figure B.4: Translated sequence alignment of gene PST130_05023. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.



Figure B.5: Translated sequence alignment of gene PST130_05454. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

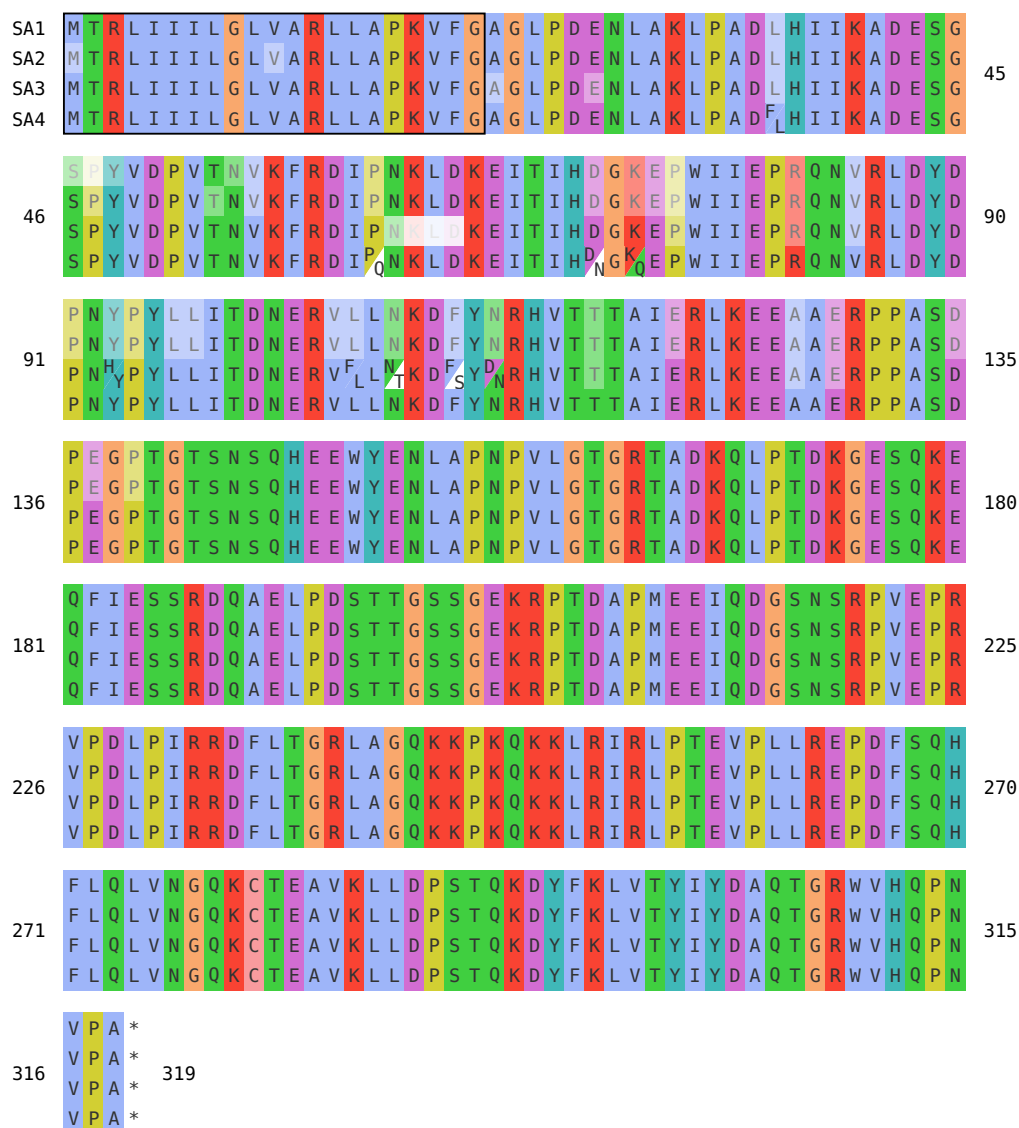
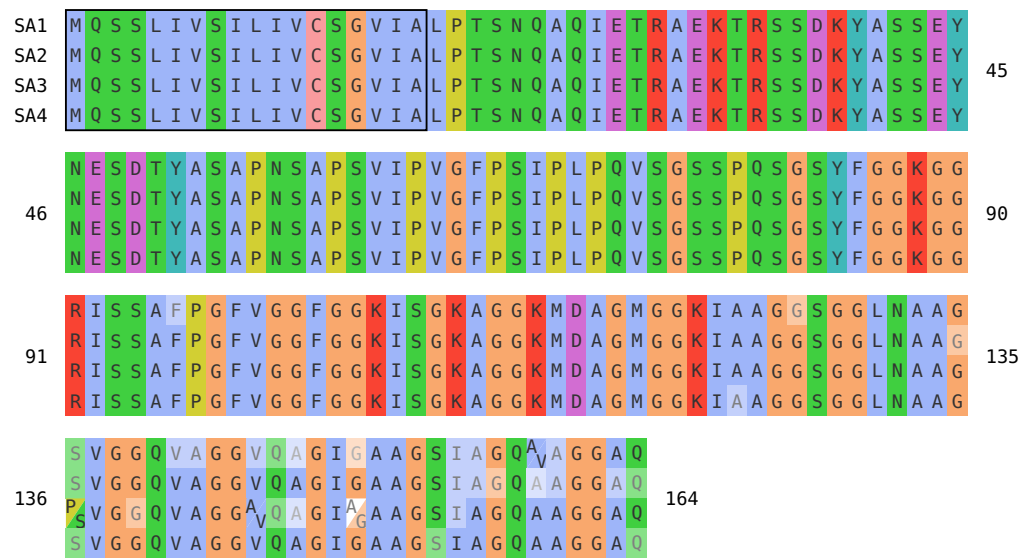


Figure B.6: Translated sequence alignment of gene PST130_05944. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.



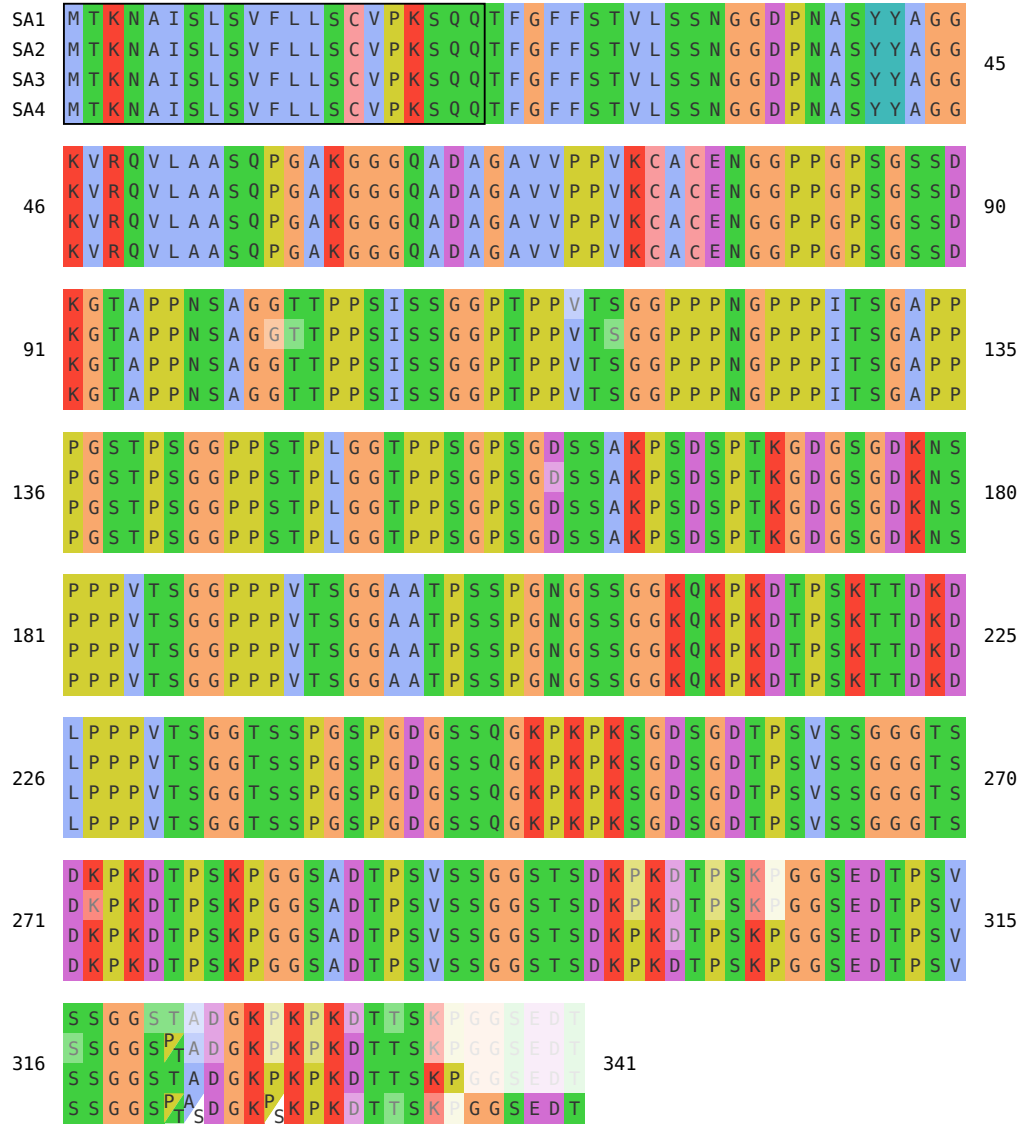


Figure B.8: Translated sequence alignment of gene PST130_06558. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

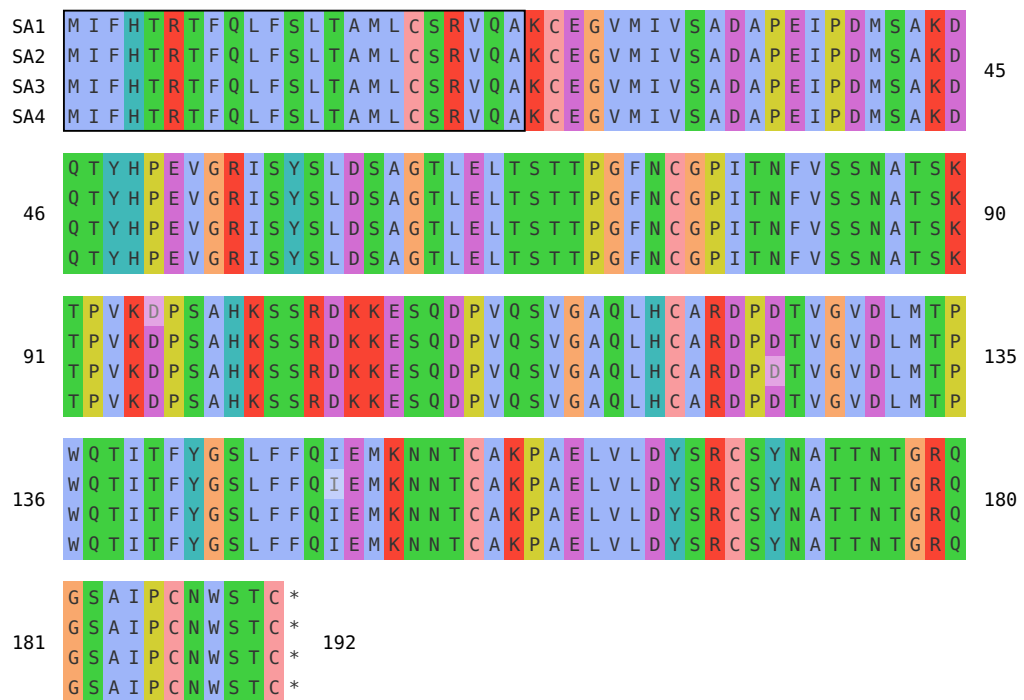
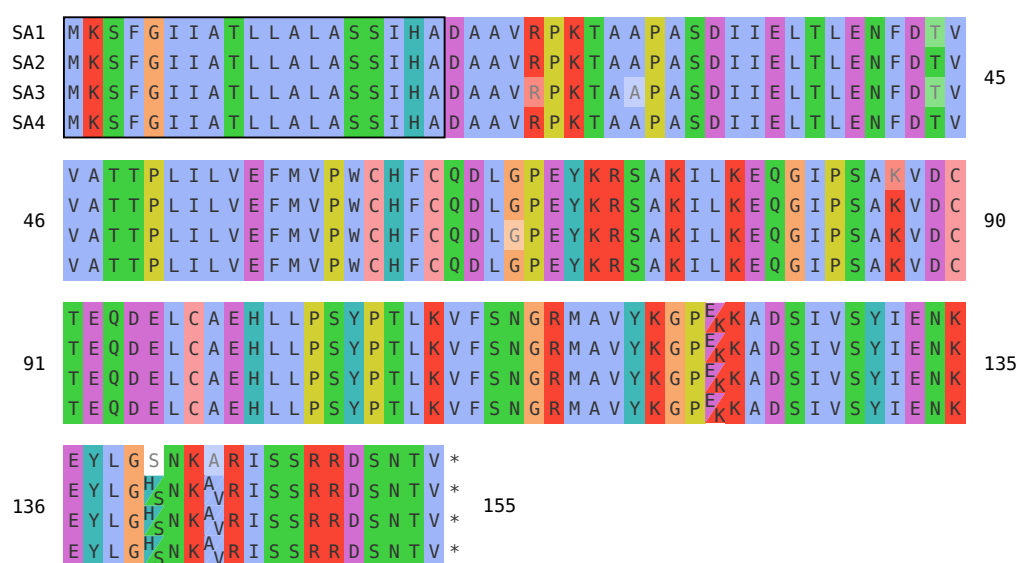


Figure B.9: Translated sequence alignment of gene PST130_07448. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.



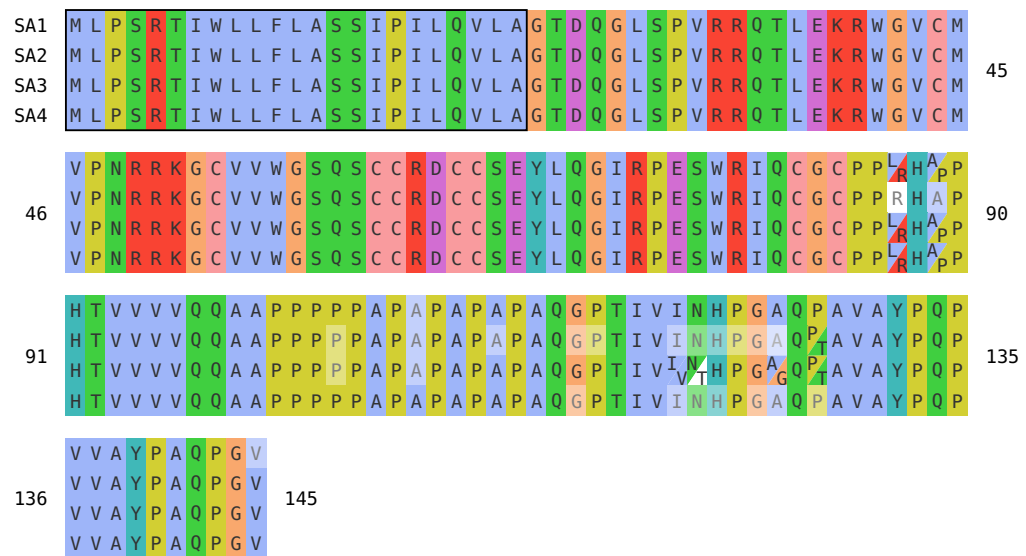


Figure B.11: Translated sequence alignment of gene PST130_07564. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

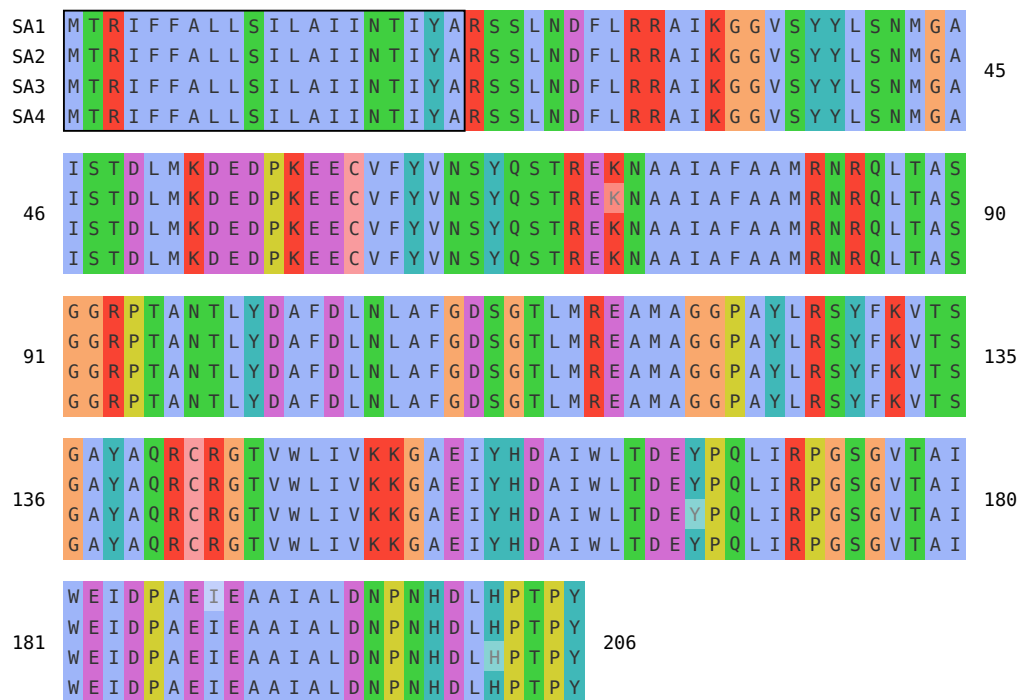


Figure B.12: Translated sequence alignment of gene PST130_08031. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

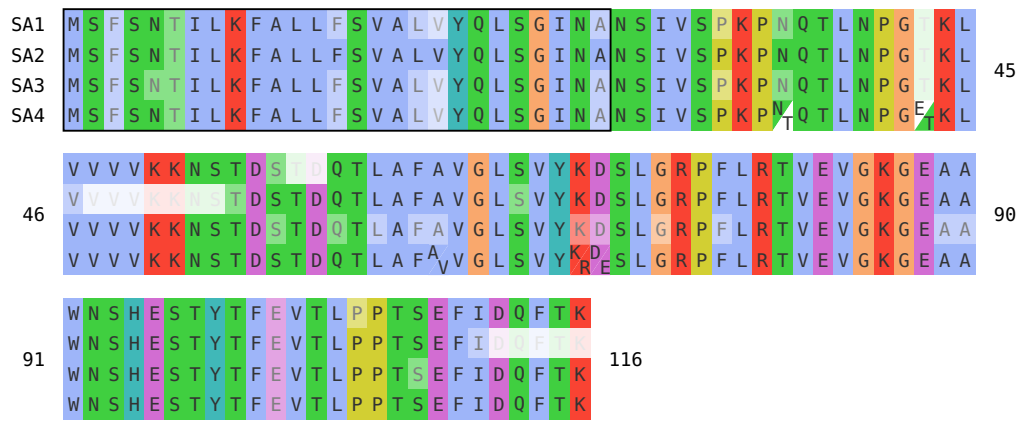


Figure B.13: Translated sequence alignment of gene PST130_08984. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

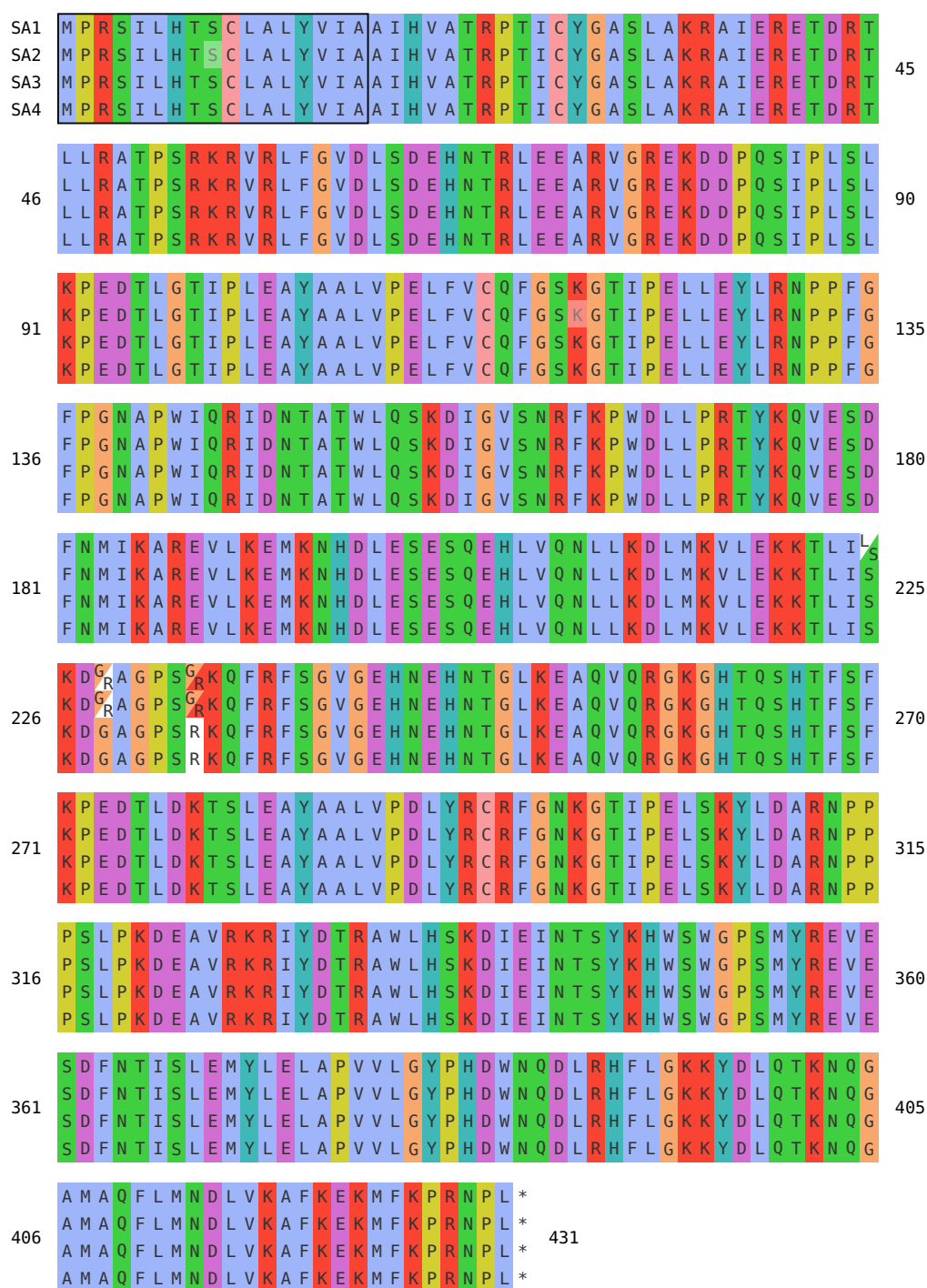


Figure B.14: Translated sequence alignment of gene PST130_09018. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

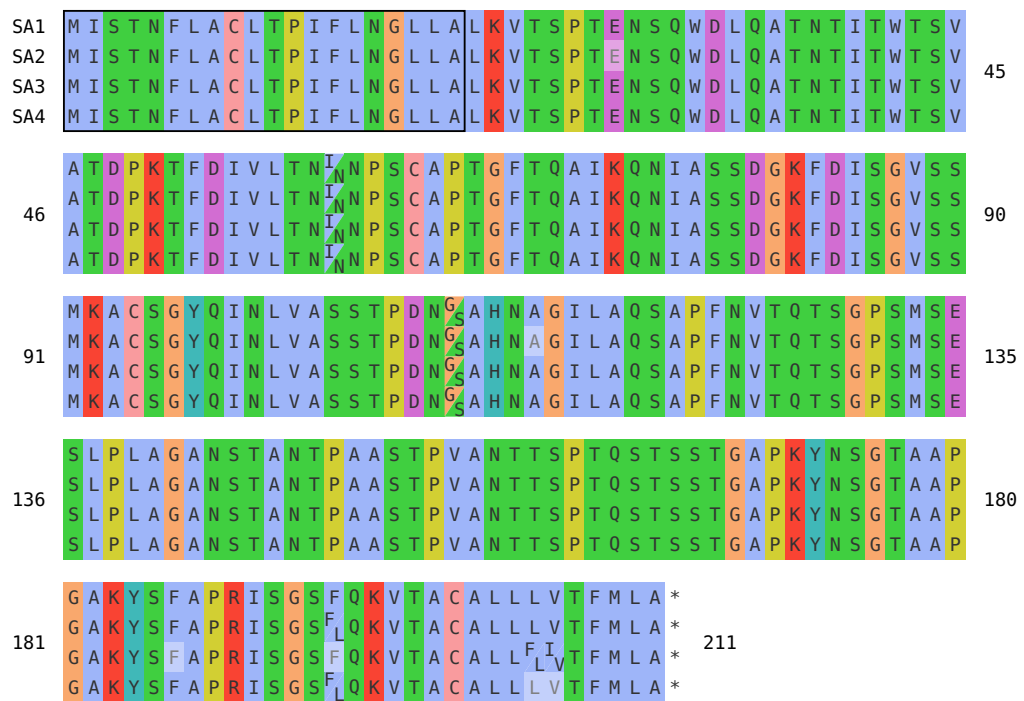


Figure B.15: Translated sequence alignment of gene PST130_09275. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

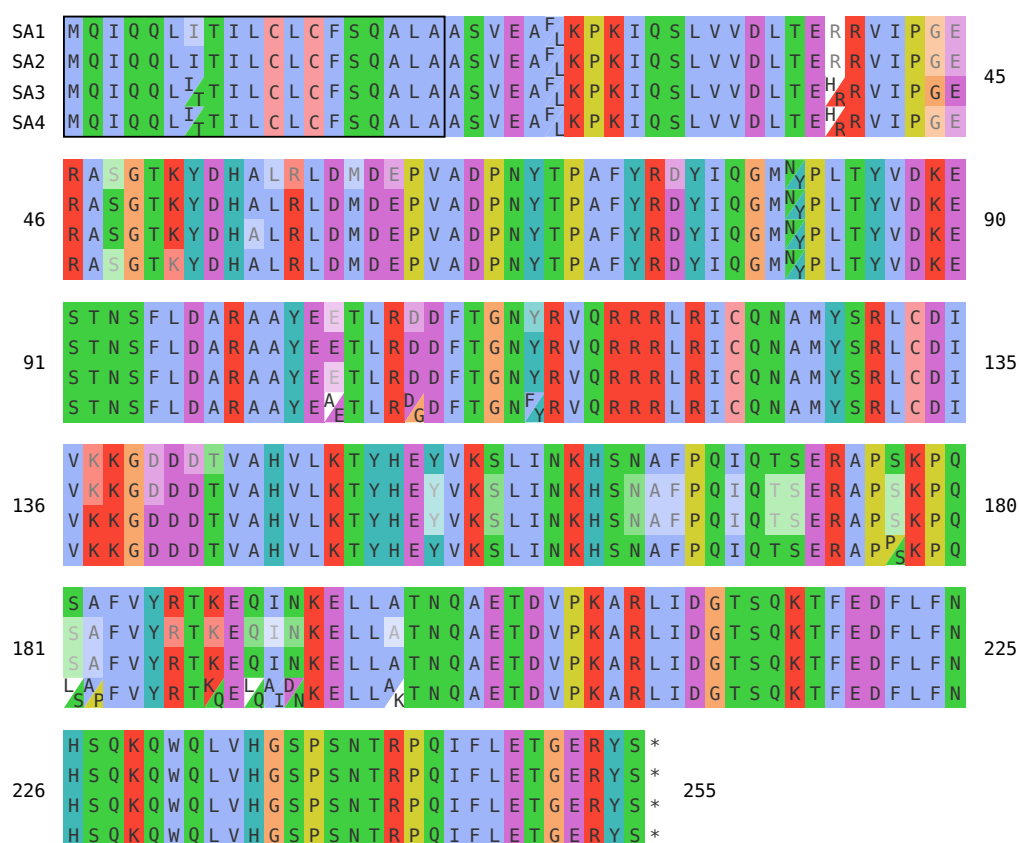


Figure B.16: Translated sequence alignment of gene PST130_10286. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

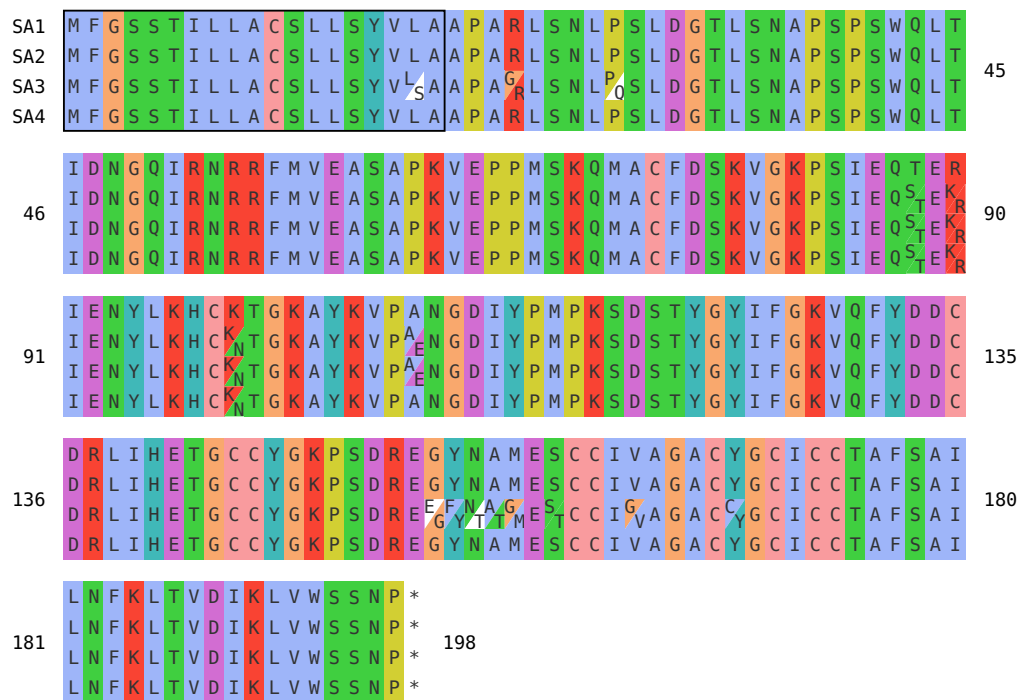


Figure B.17: Translated sequence alignment of gene PST130_12487. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.



Figure B.18: Translated sequence alignment of gene PST130_12491. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

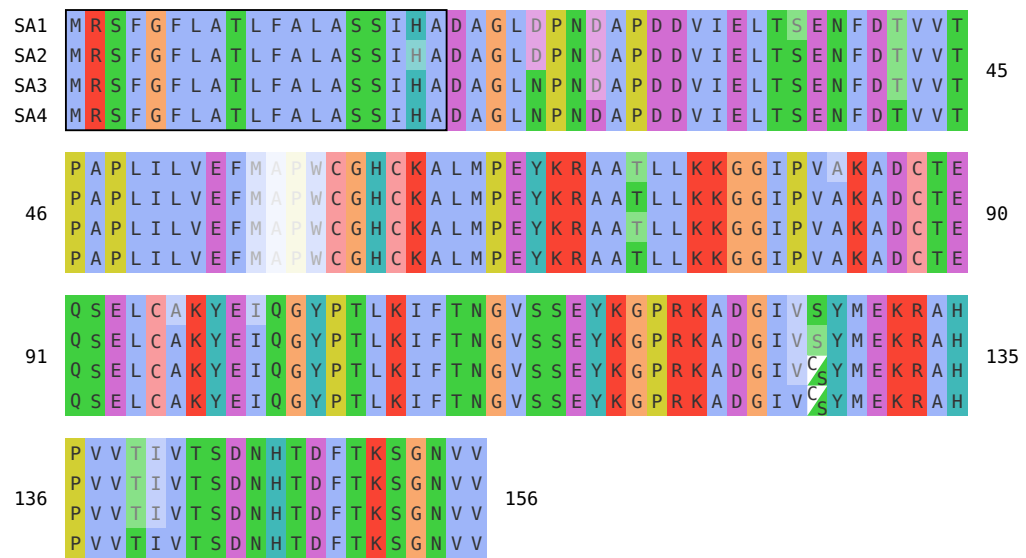


Figure B.19: Translated sequence alignment of gene PST130_12956. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

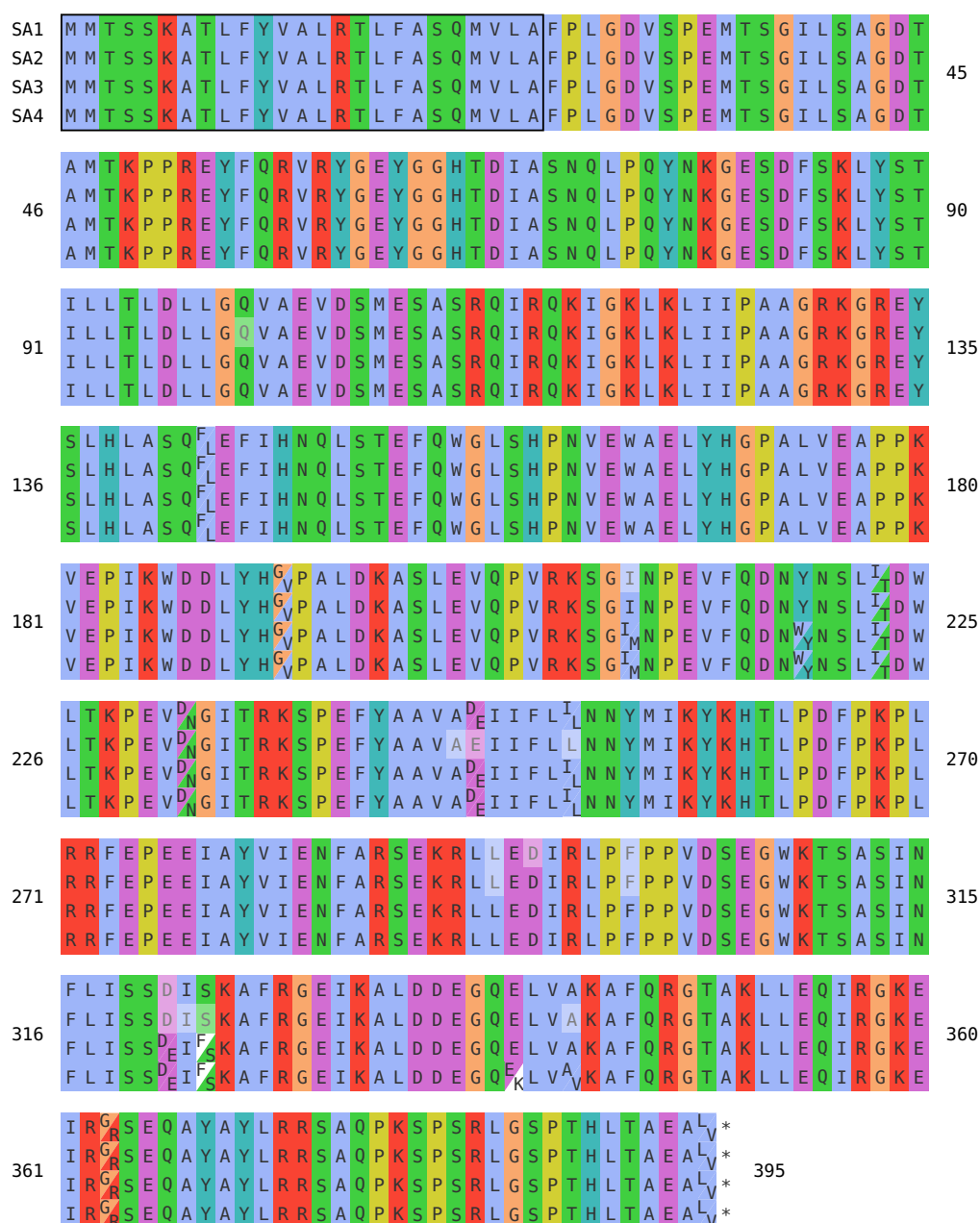


Figure B.20: Translated sequence alignment of gene PST130_13969. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

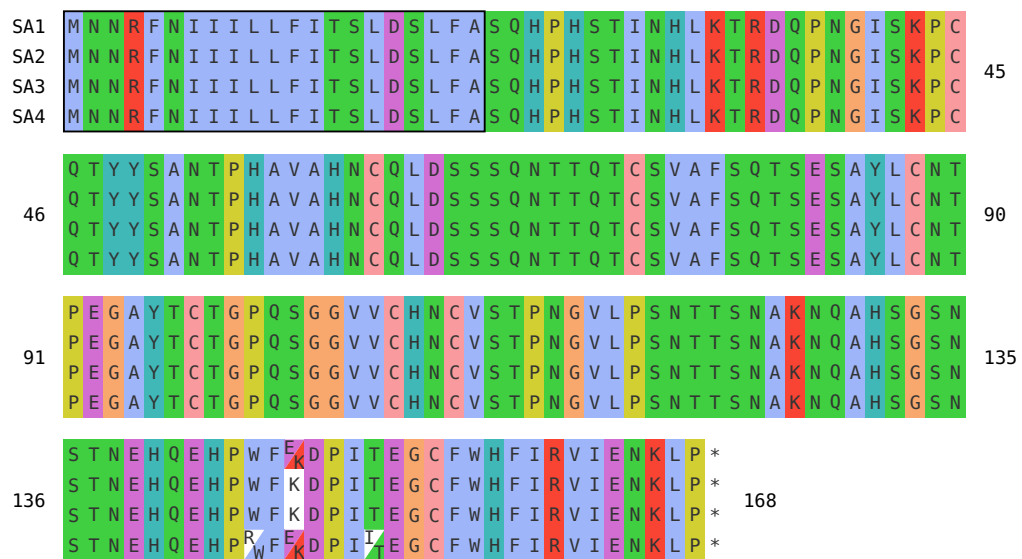


Figure B.21: Translated sequence alignment of gene PST130_14091. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

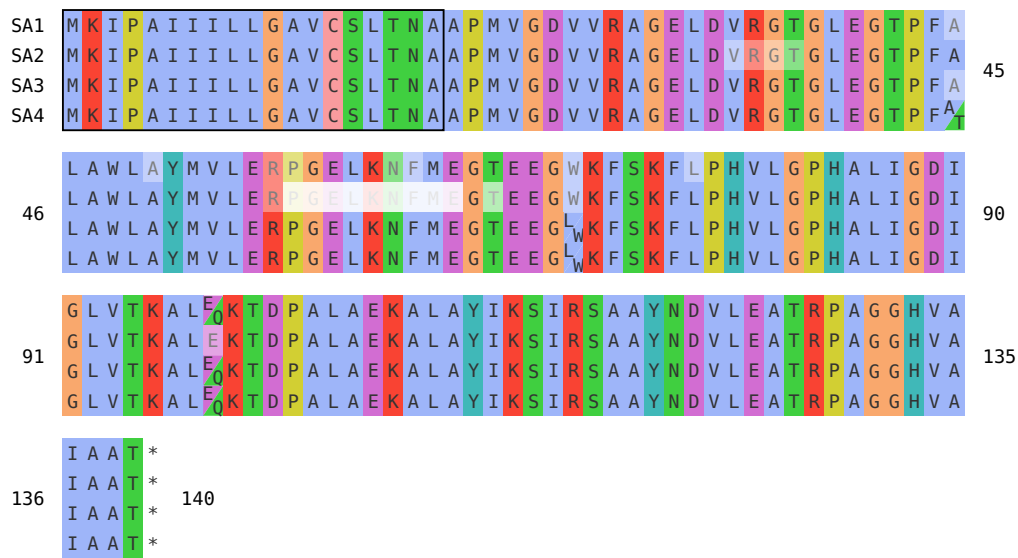


Figure B.22: Translated sequence alignment of gene PST130_14831. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

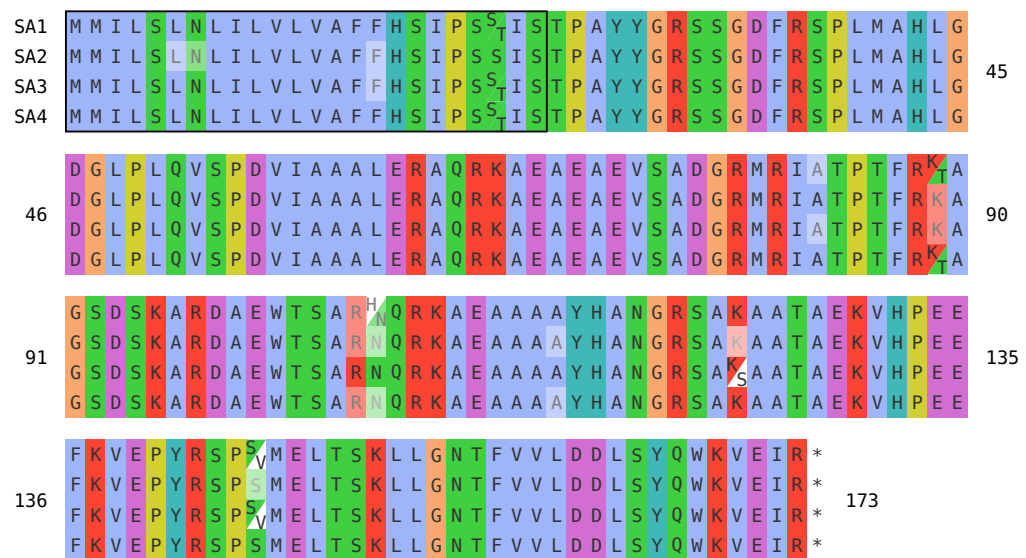


Figure B.23: Translated sequence alignment of gene PST130_16778. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

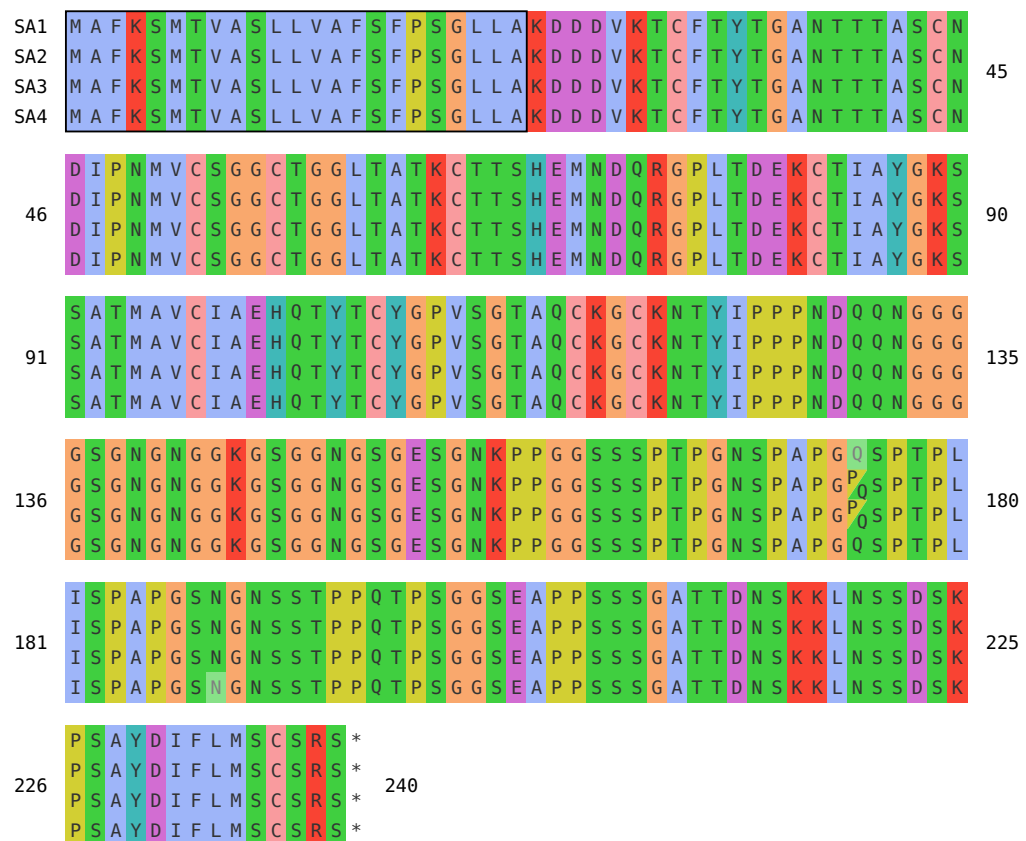


Figure B.24: Translated sequence alignment of gene PST130_17605. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

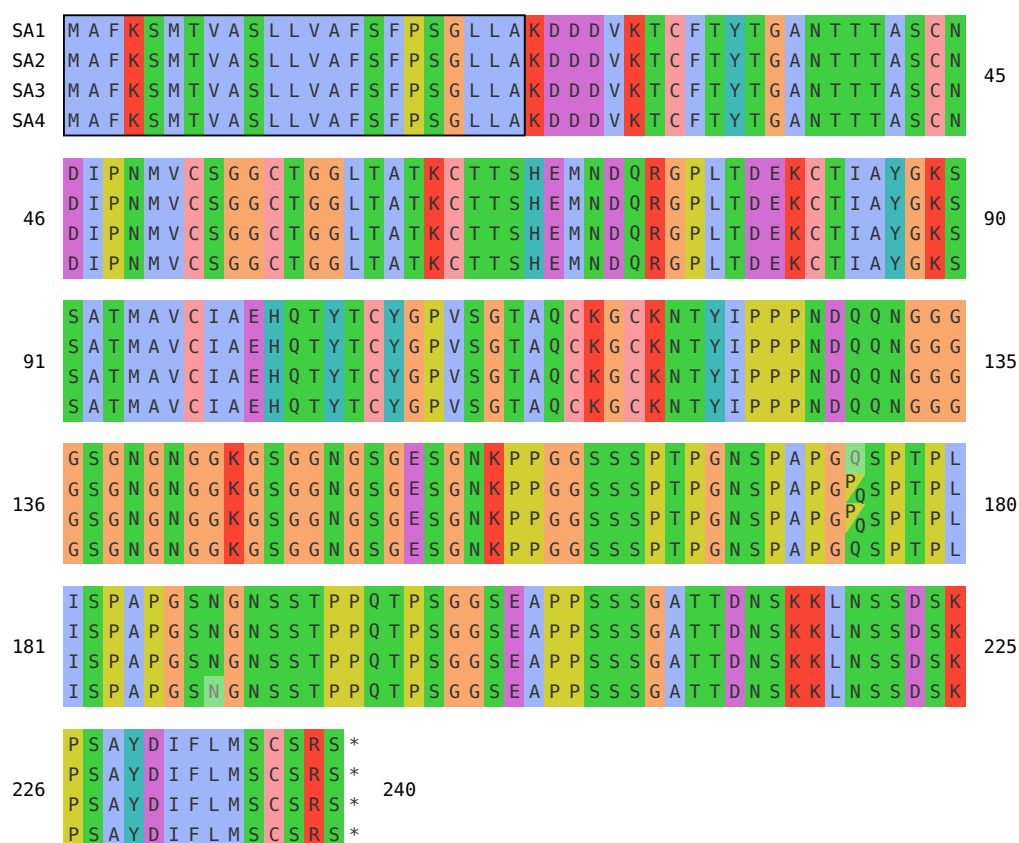


Figure B.25: Translated sequence alignment of gene PST130_17605. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

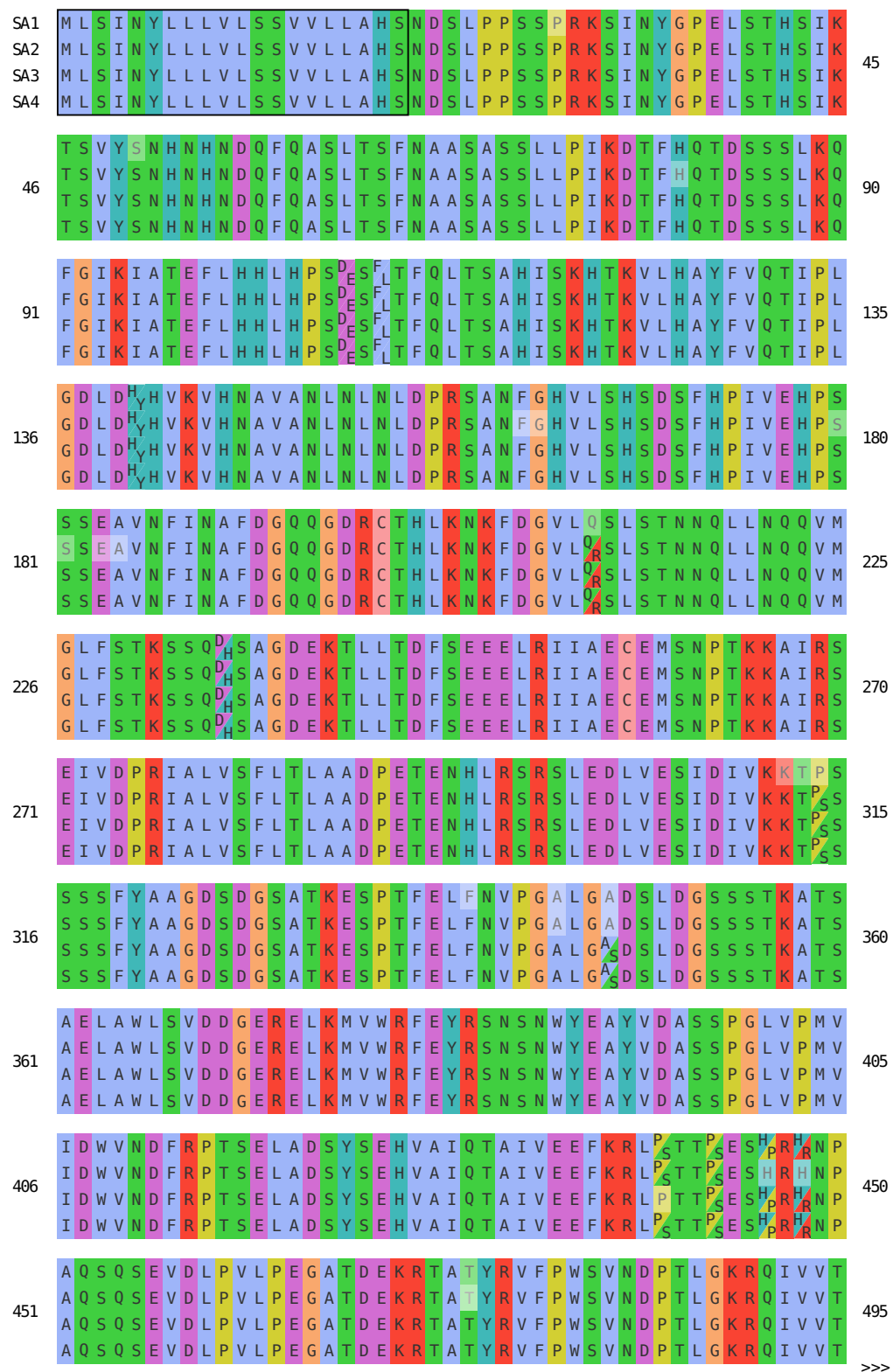


Figure B.26: See continuation on next page.

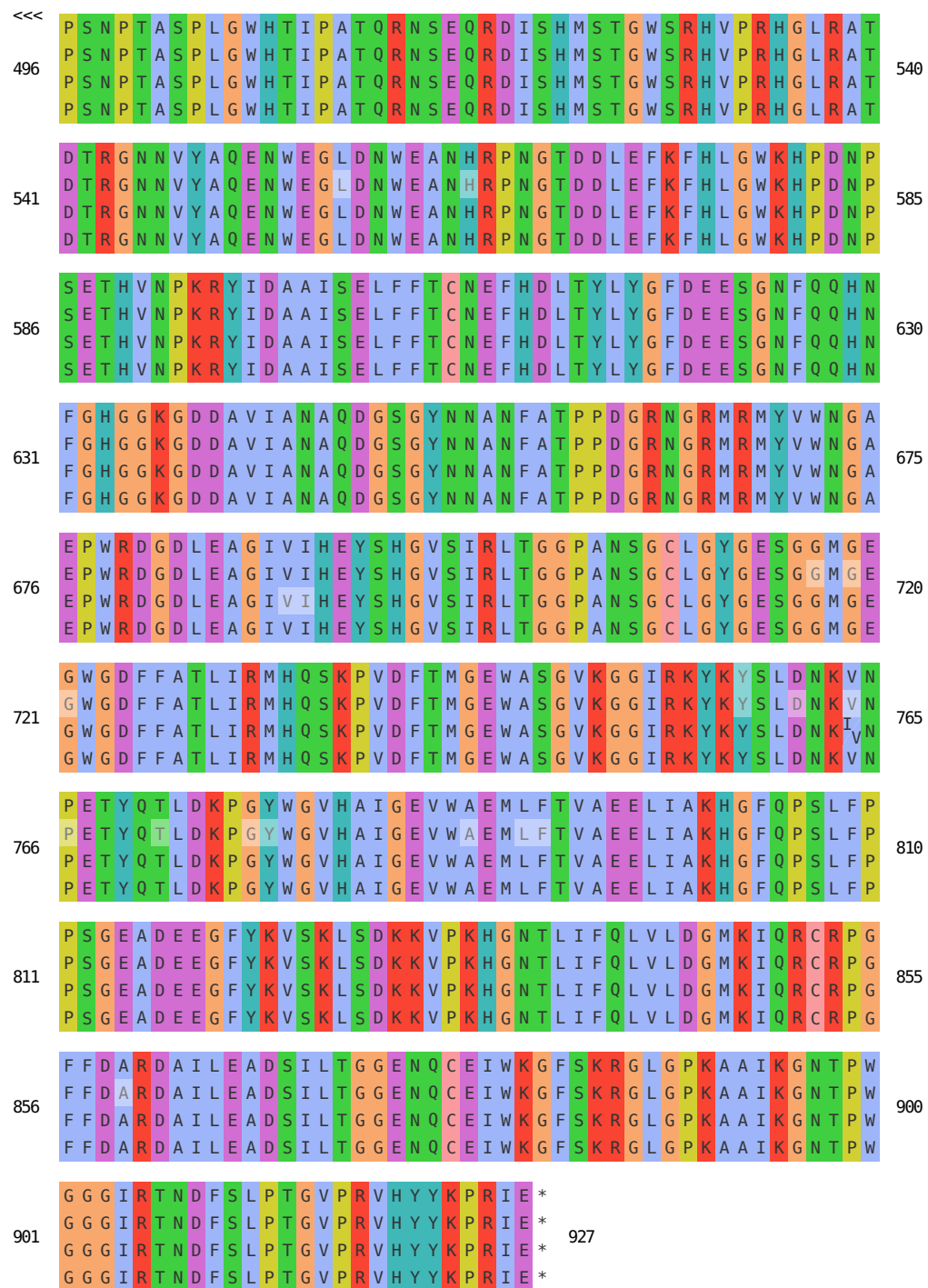


Figure B.27: Translated sequence alignment of gene PST130_07579. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

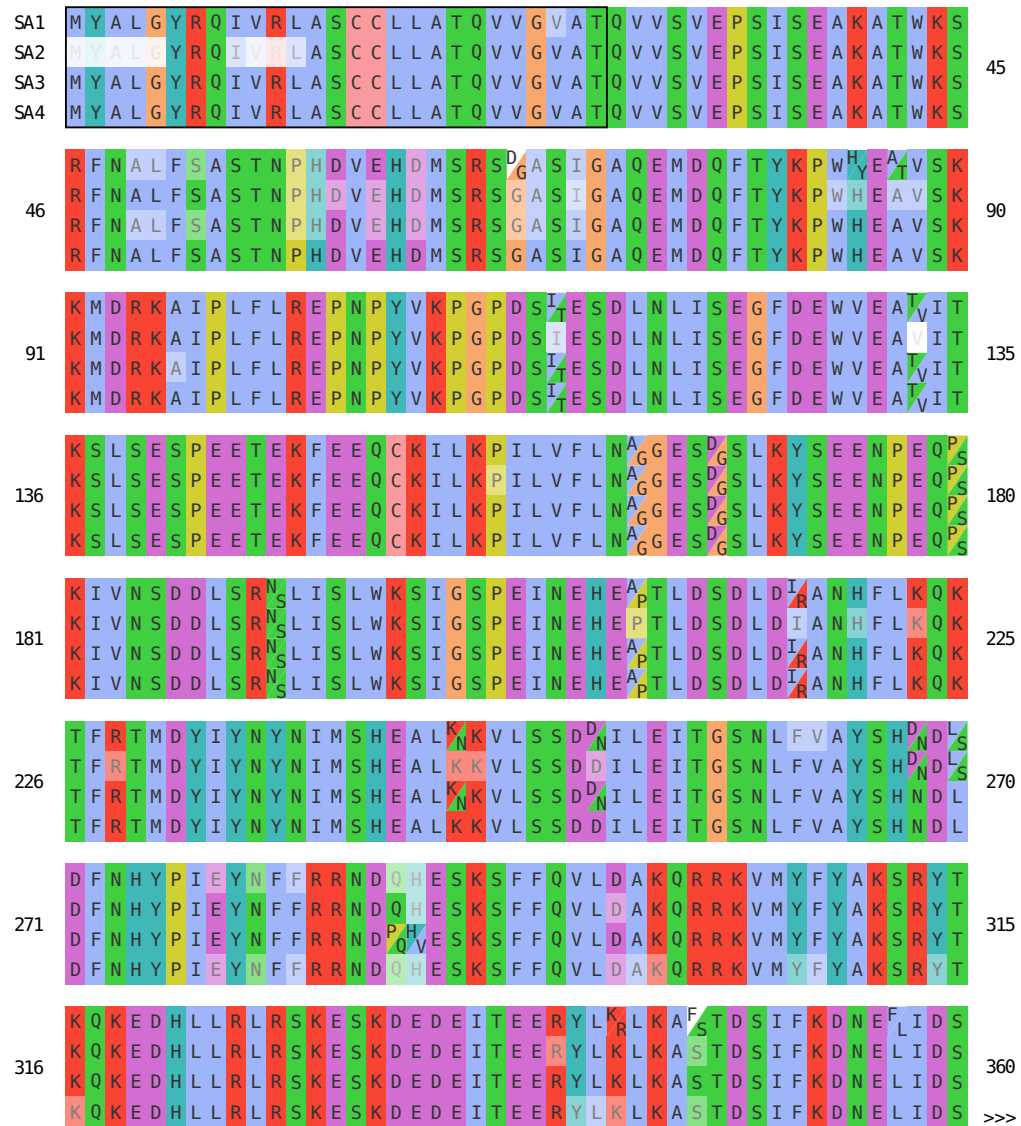


Figure B.28: See continuation on next page.

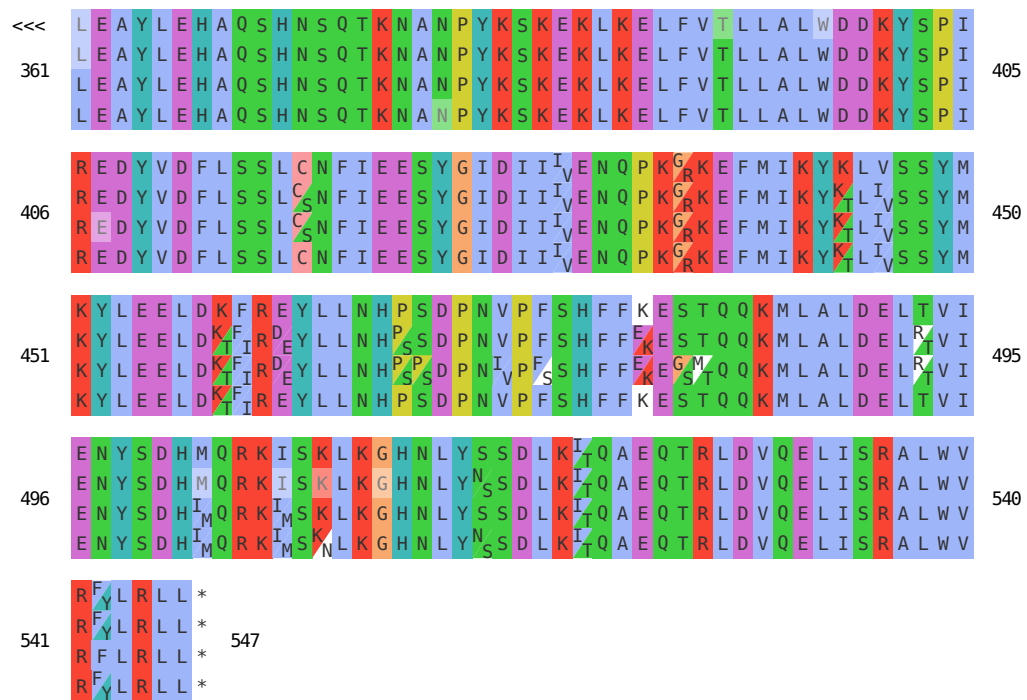


Figure B.29: Translated sequence alignment of gene PST130_15131. This gene has been identified to encode a putative effector protein (Cantu et al., 2013). The signal peptide, predicted using SignalP (version 2; Emanuelsson et al., 2007) is indicated by the black box. Alternative amino acids resulting from nonsynonymous SNPs at biallelic sites are indicated in the below diagonal triangles. Colours were assigned according to the “Clustal X Colour Scheme” used in Jalview (Waterhouse et al., 2009), categorising amino acid profiles.

Appendix C

Gene Expression Analysis of Candidate Effectors Identified in South African *Pst* Isolates

C.1 Candidate gene inspection



Figure C.1: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_02001 in SA1 and SA4.

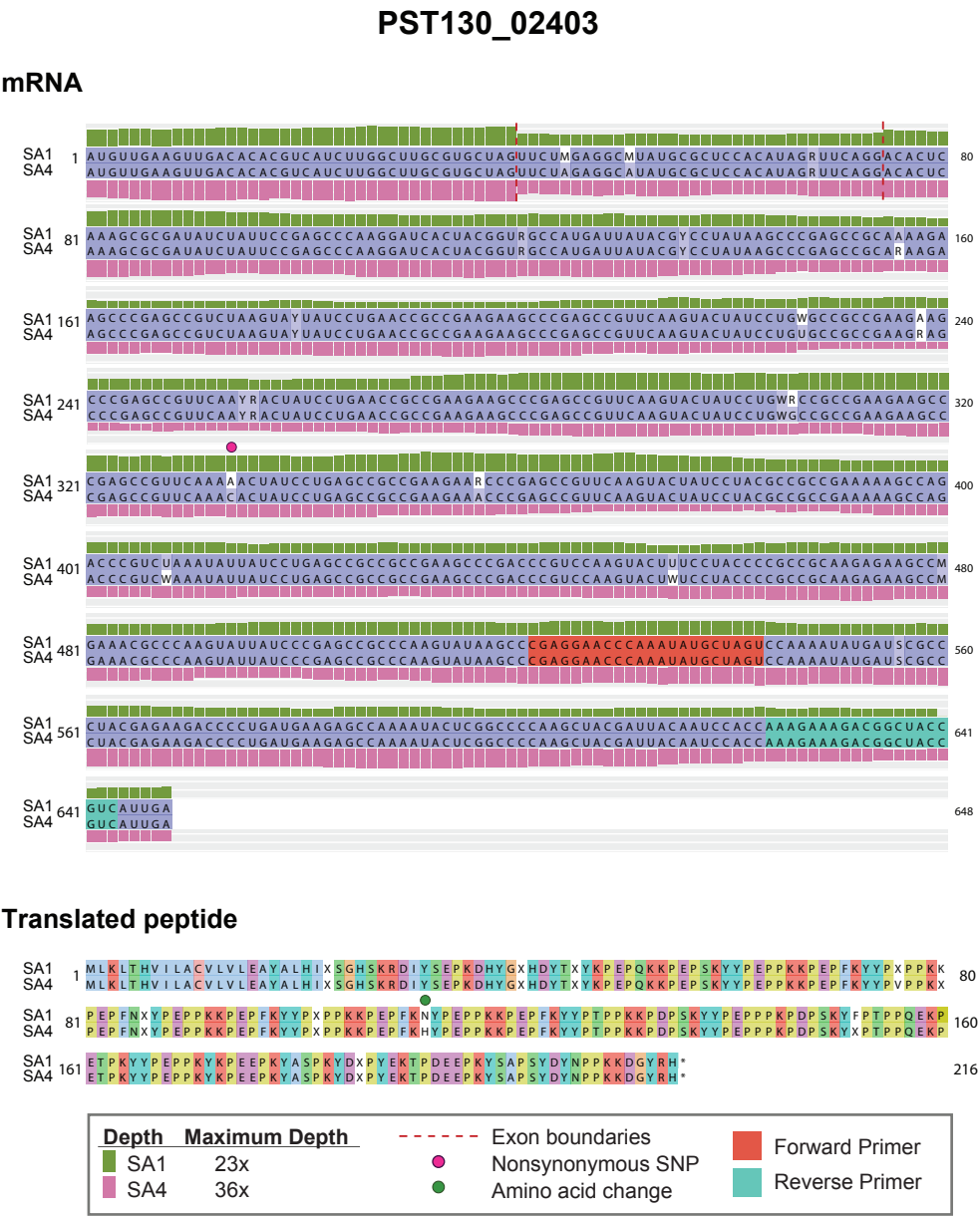


Figure C.2: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_02403 in SA1 and SA4.

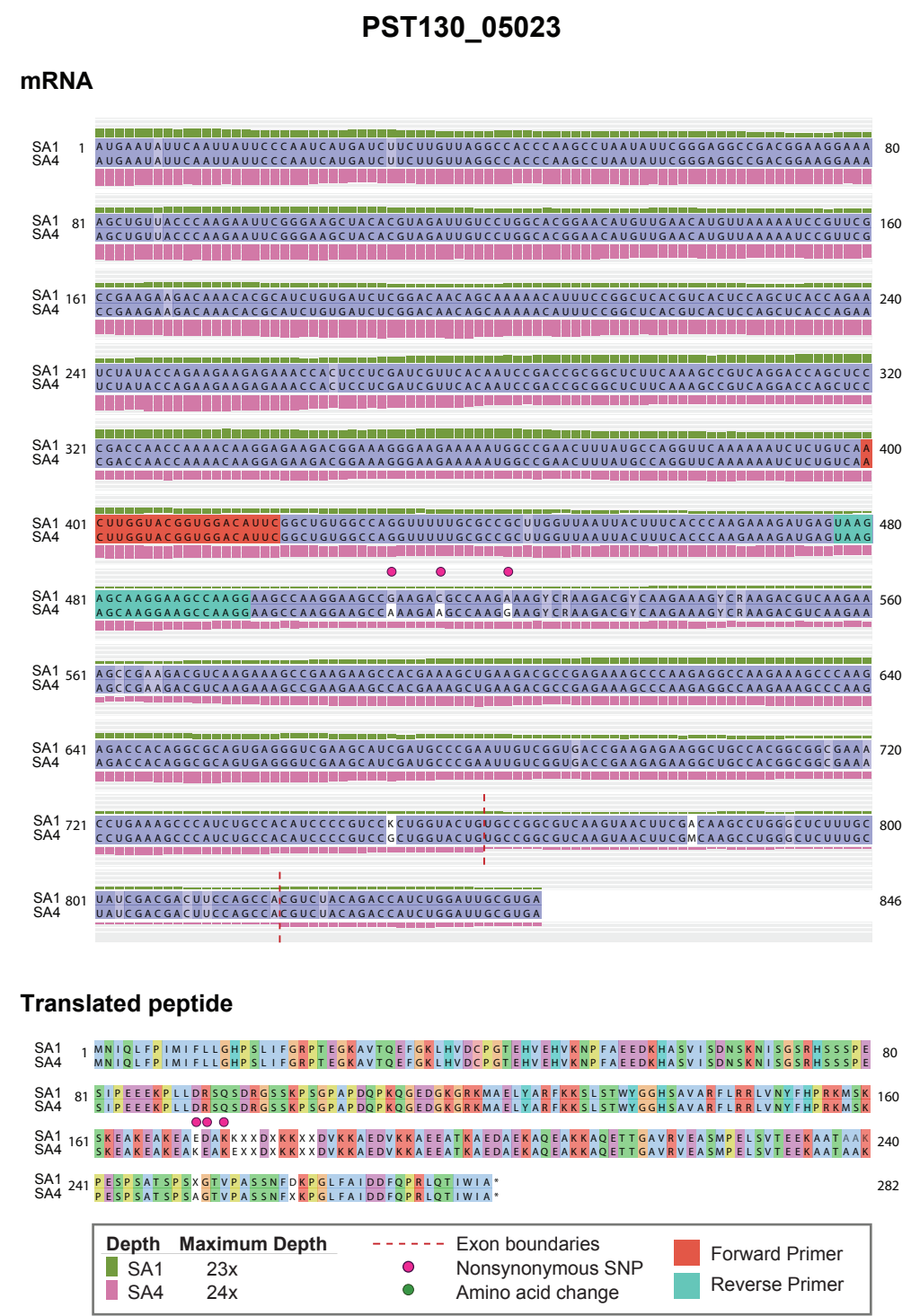


Figure C.3: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_05023 in SA1 and SA4.

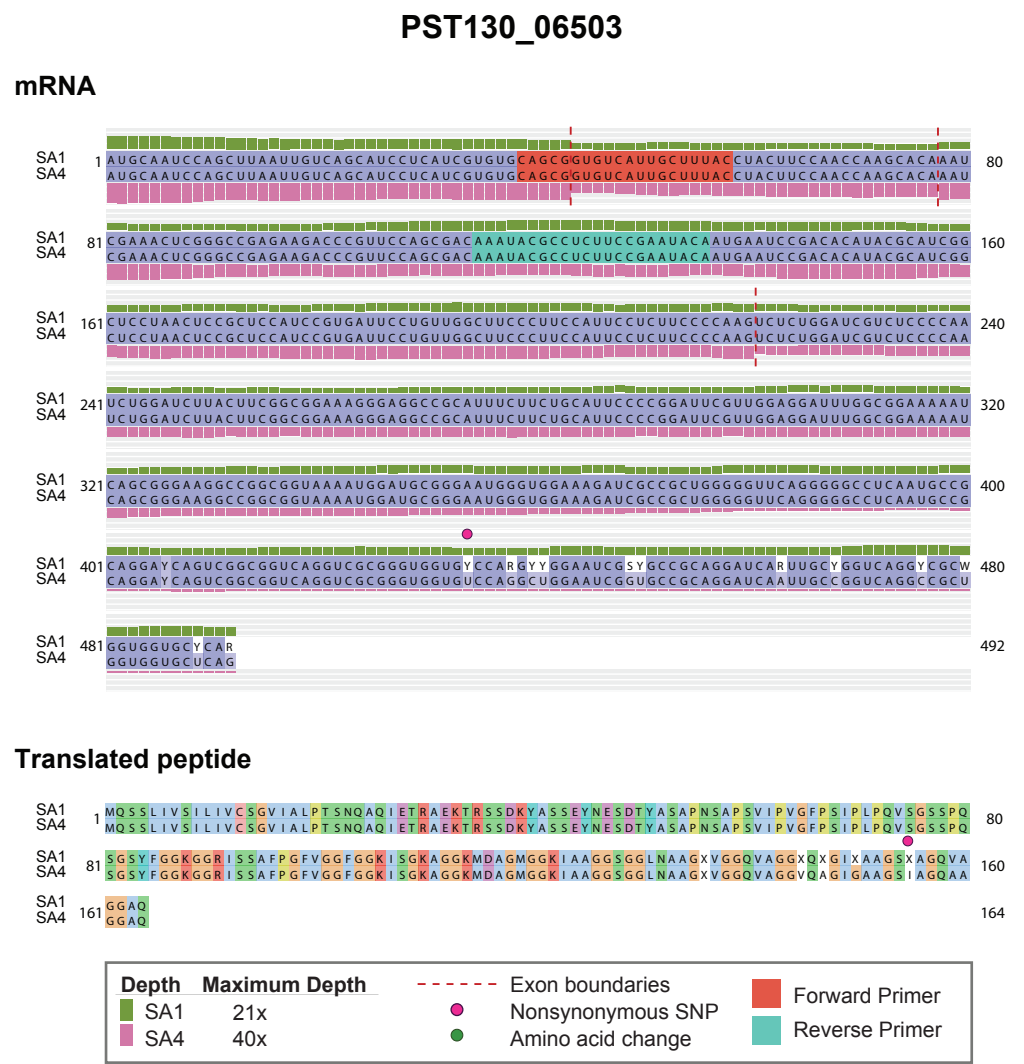


Figure C.4: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_06503 in SA1 and SA4.

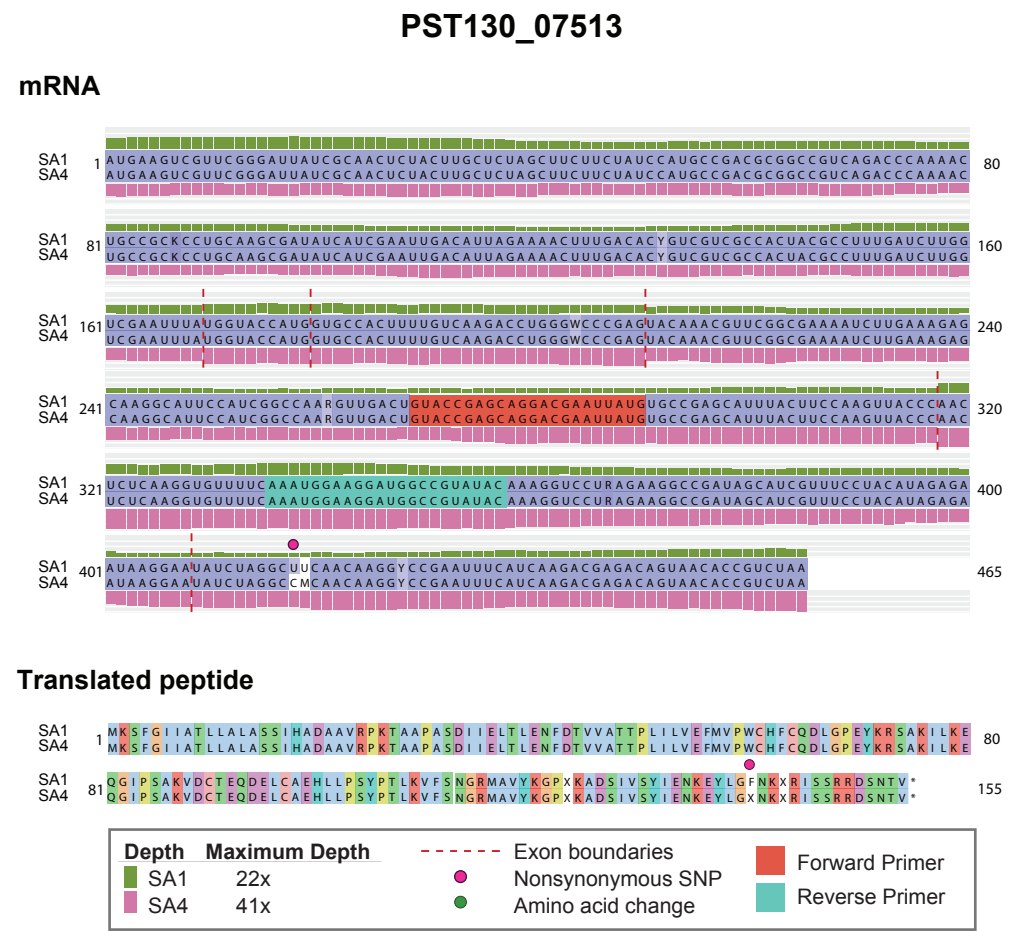


Figure C.5: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_07513 in SA1 and SA4.

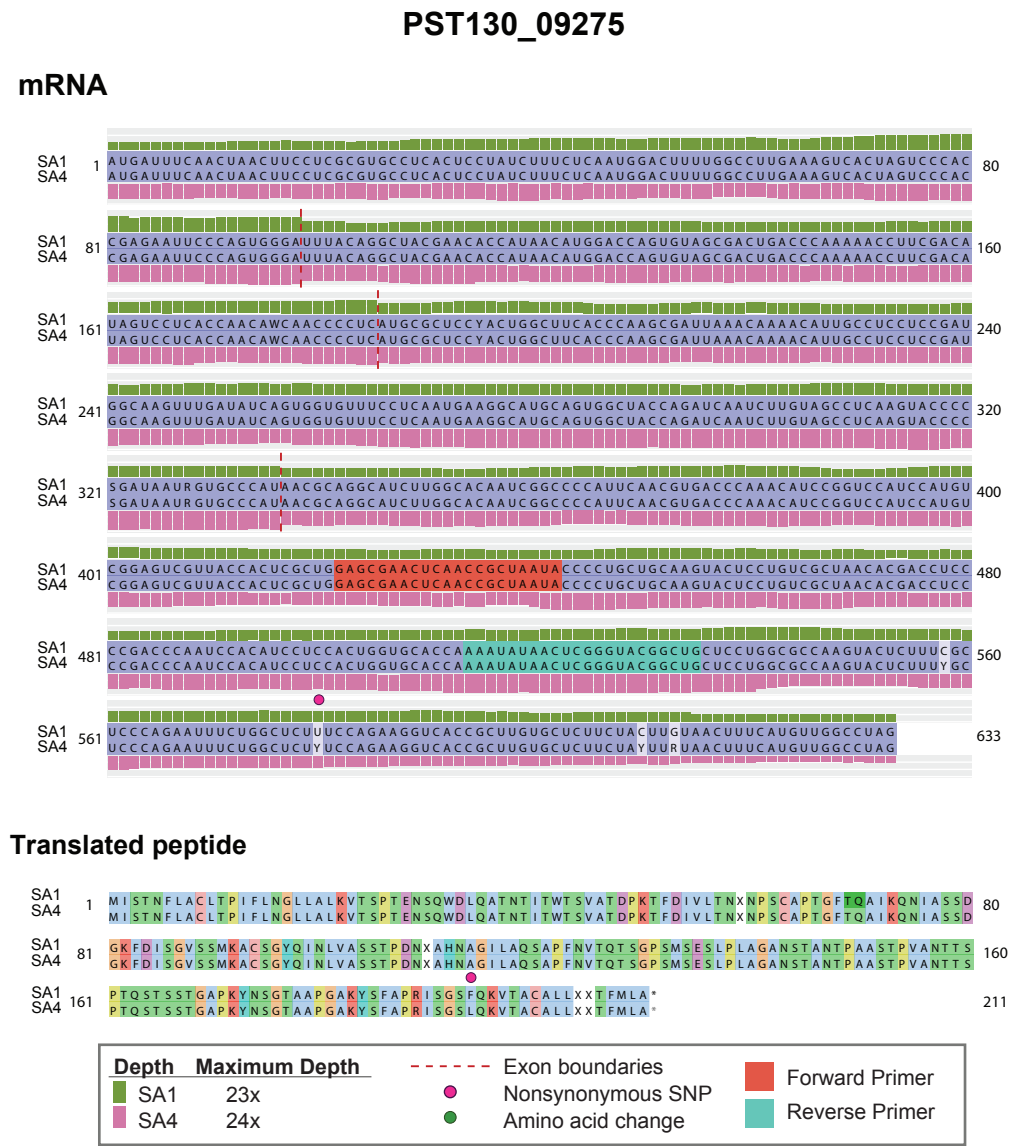


Figure C.6: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_09275 in SA1 and SA4.

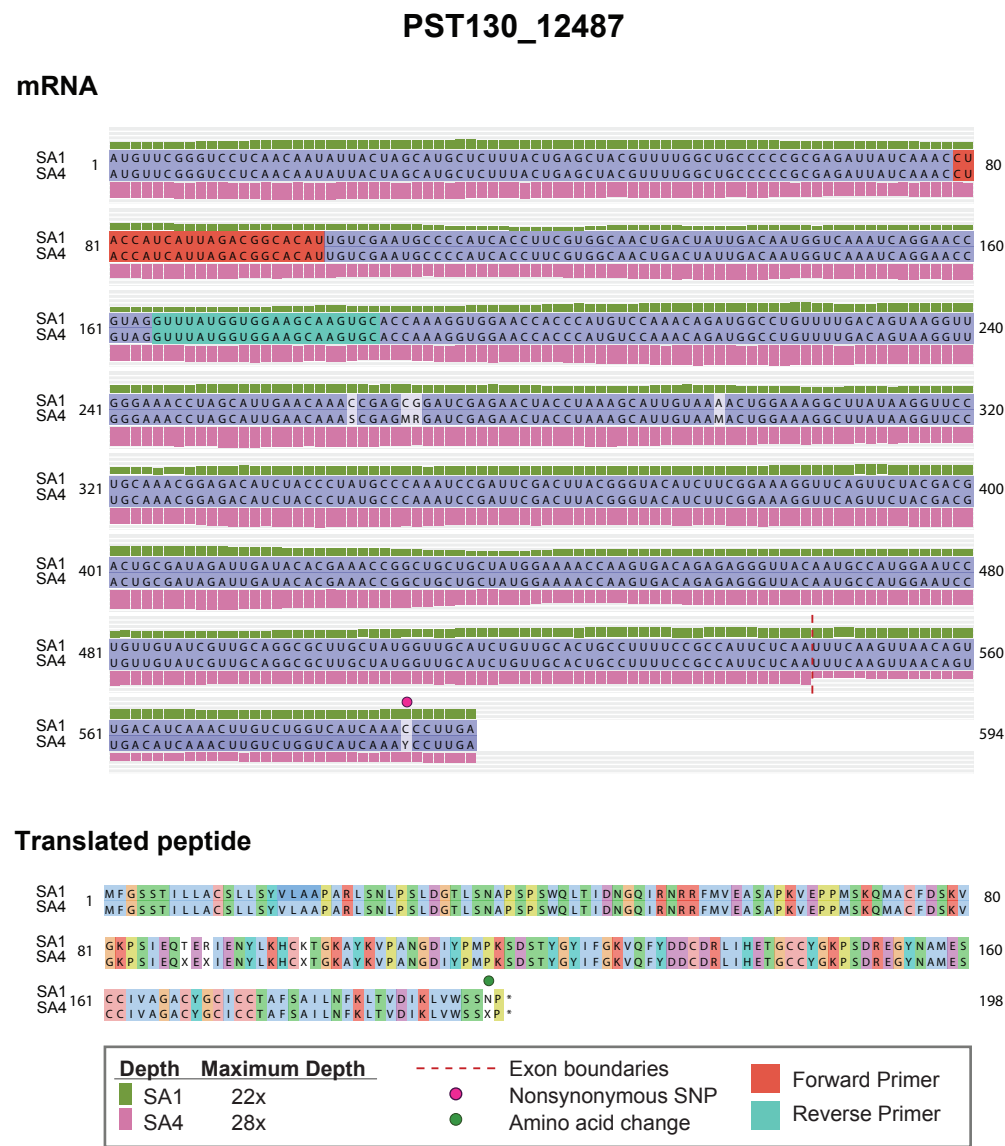


Figure C.7: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_12487 in SA1 and SA4.

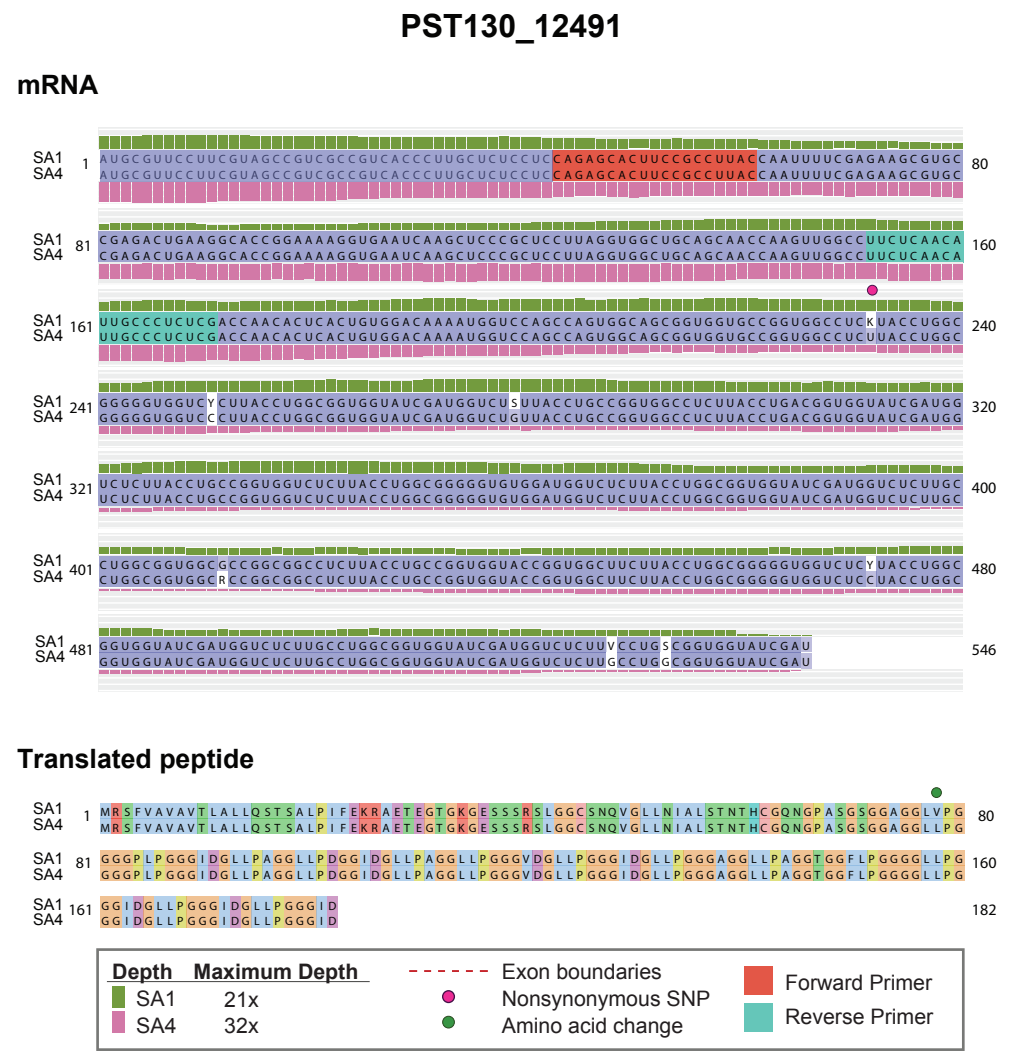


Figure C.8: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_12491 in SA1 and SA4.

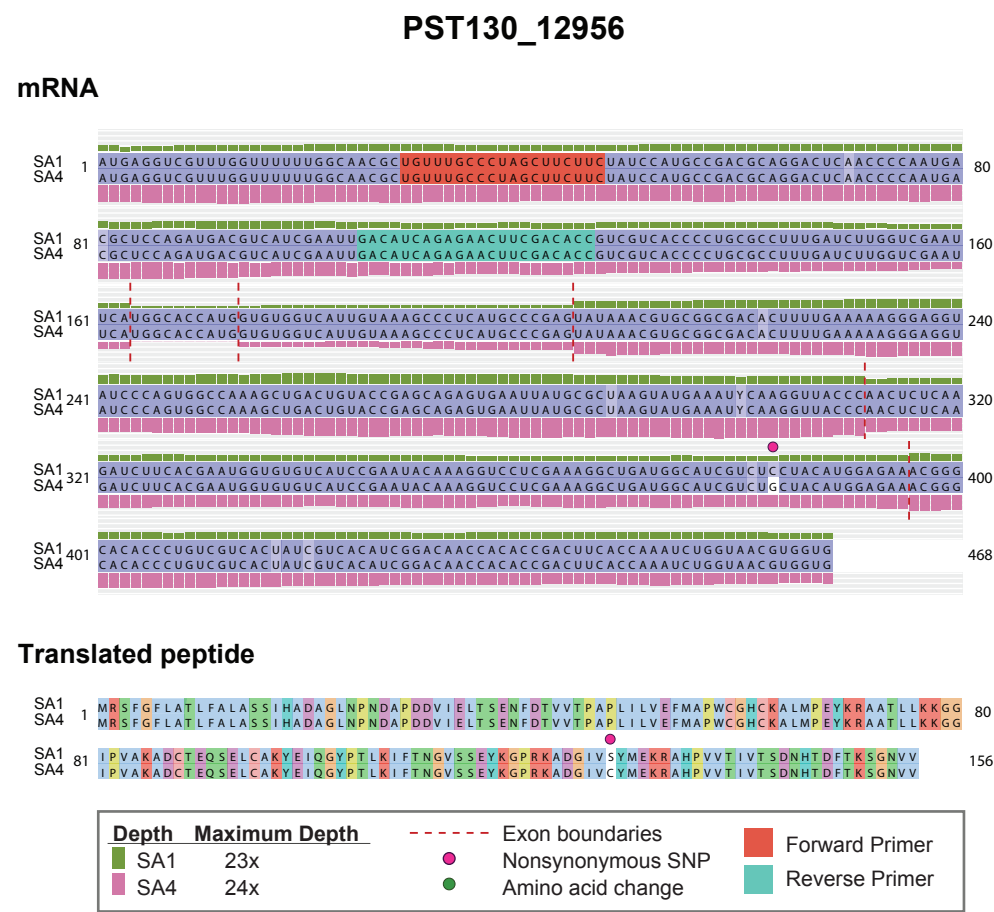


Figure C.9: Nonsynonymous polymorphisms and primer design of the candidate effector gene PST130_12956 in SA1 and SA4.

C.2 Additional figures of statistical analyses

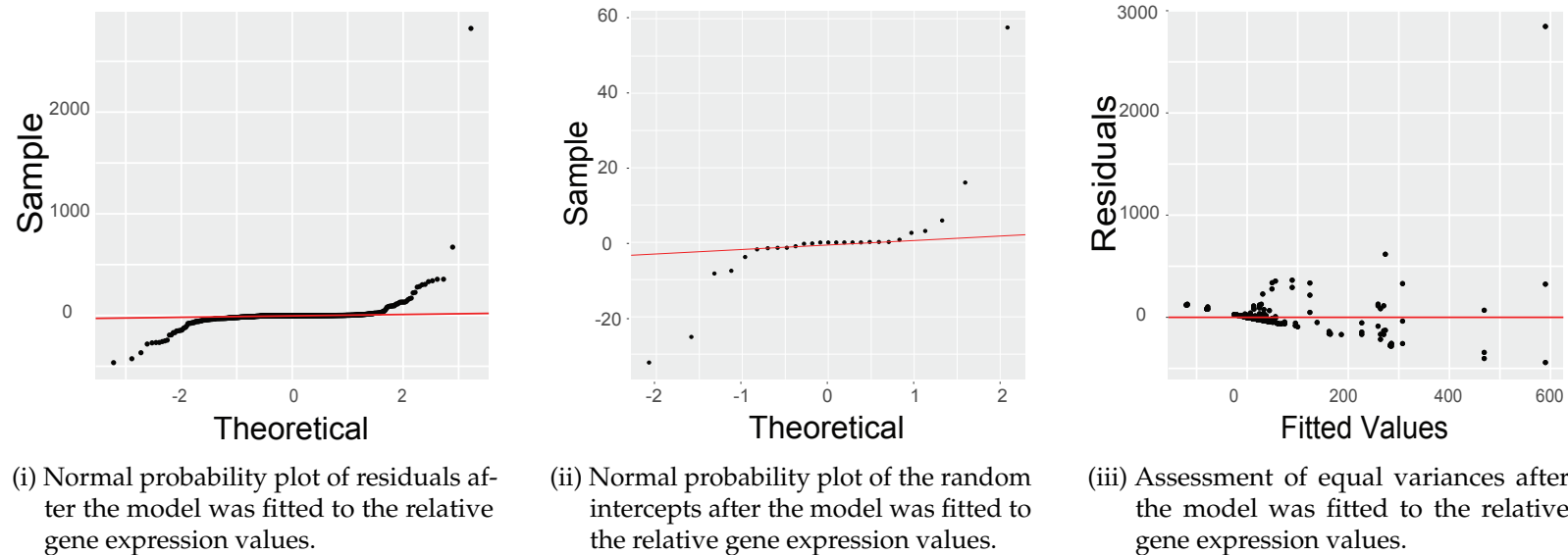


Figure C.10: Graphical tests for normality and equal variances of the residuals and random intercepts. The relative gene expression dataset was evaluated applying the assumptions that linear mixed models are based on. Normal probability plots of the random intercept dataset (i) and the residuals (ii) showed deviation from normality. The fan like pattern observed in the plot to assess equal variances (iii) revealed that variances were not equal, as is required for using a linear model. This indicated that the relative gene expression dataset was not a good fit for a linear mixed model, as it violated the assumptions of the model type.

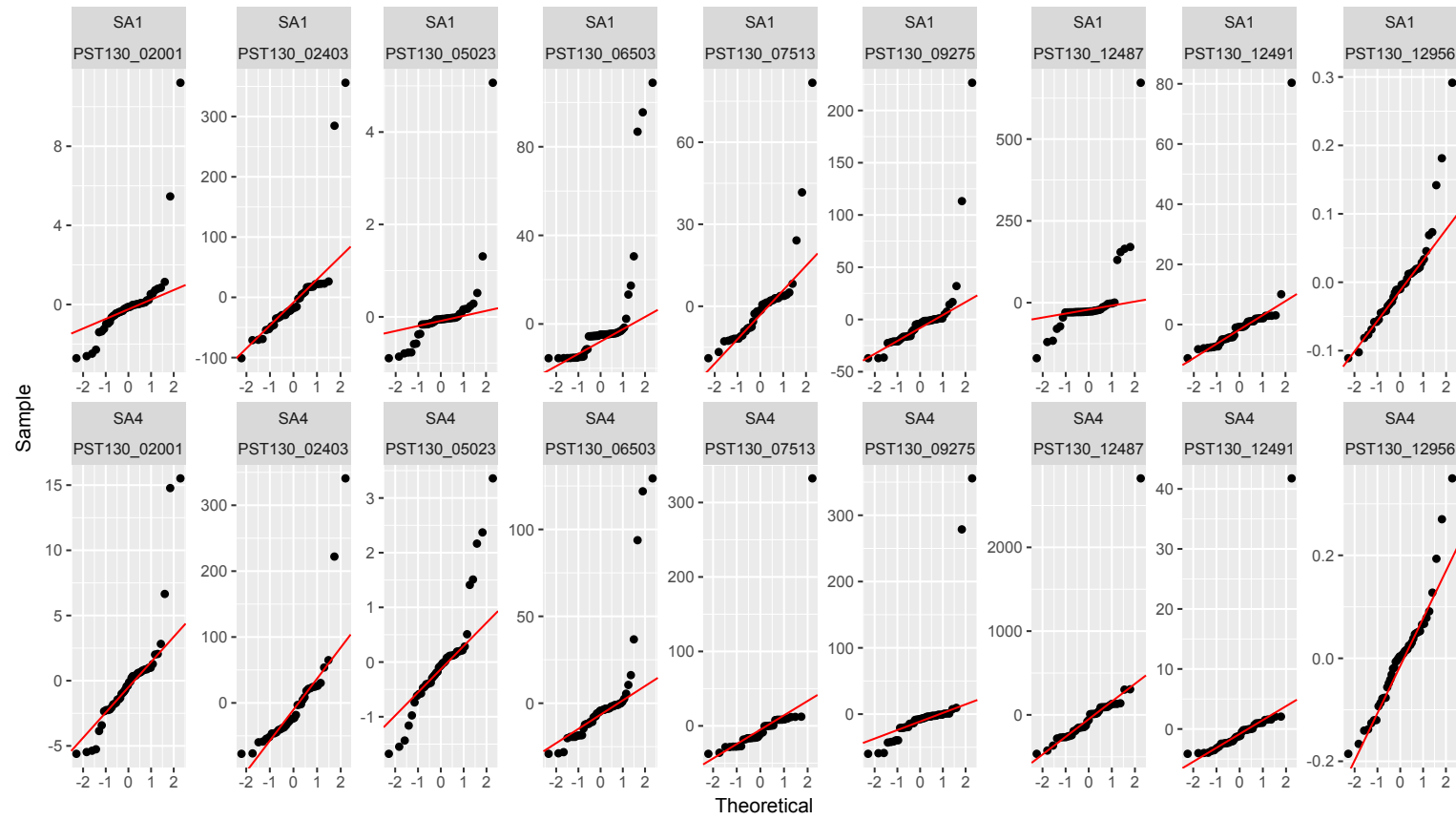


Figure C.11: Gene and isolate specific tests for equal variances after the model was fitted to the relative gene expression values.

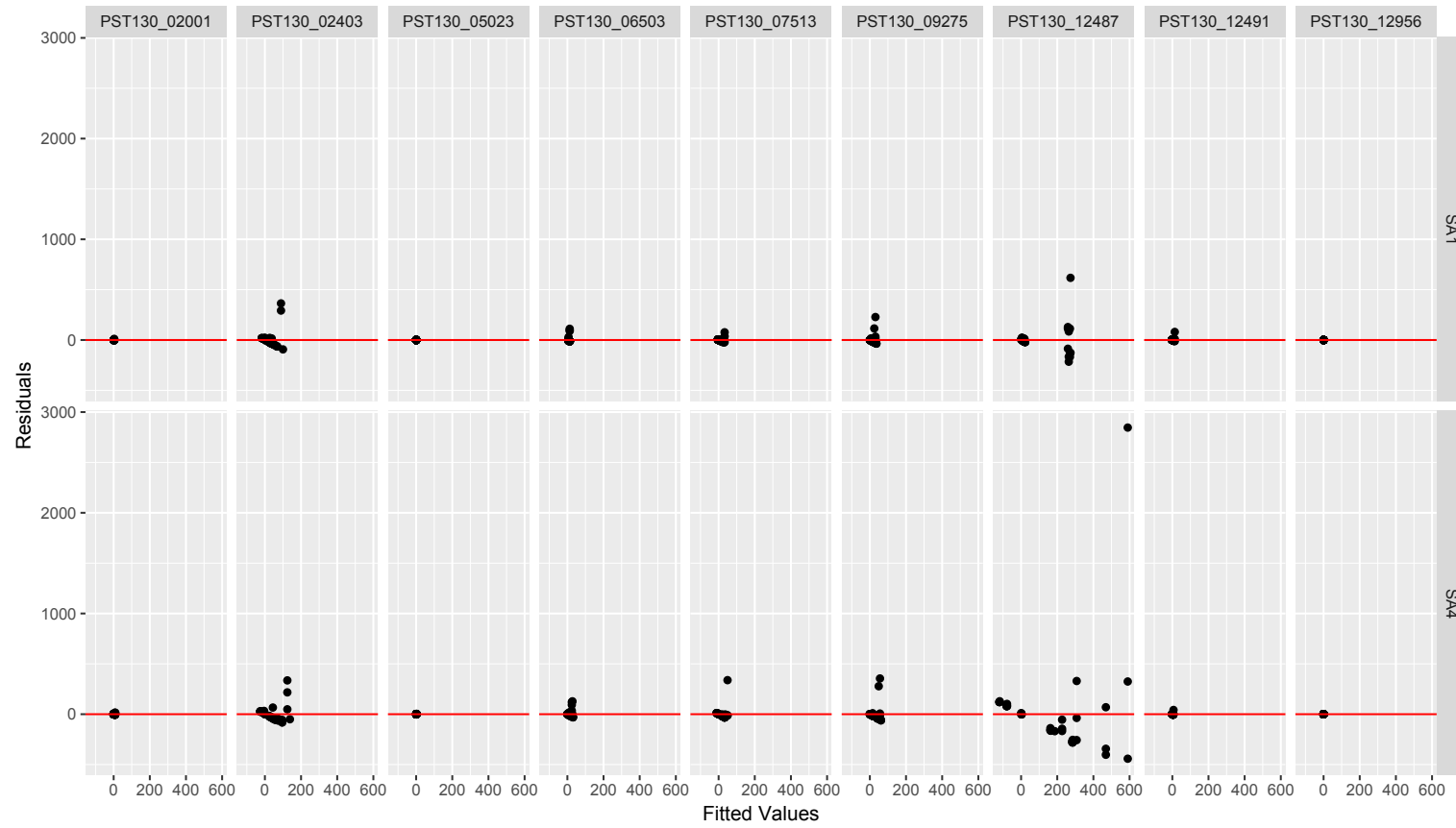
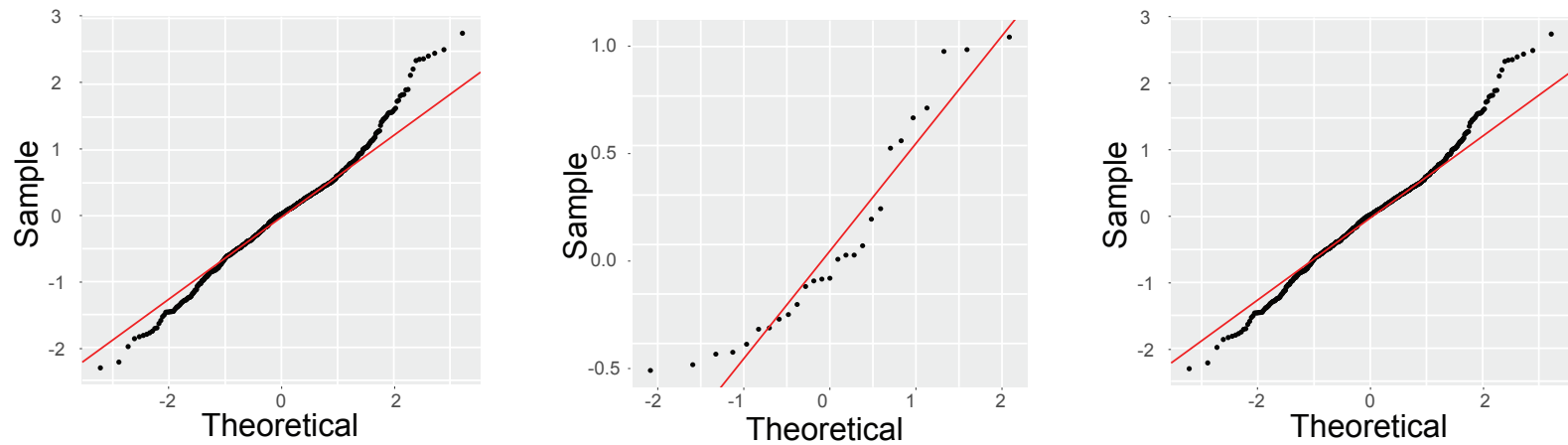


Figure C.12: Gene and isolate specific tests for equal variances after the model was fitted to the relative gene expression values.



(i) Normal probability plot of residuals after the model was fitted to the $\log_{10}$ transformed relative gene expression values.

(ii) Normal probability plot of the random intercepts after the model was fitted to the $\log_{10}$ transformed relative gene expression values.

(iii) Assessment of equal variances after the model was fitted to the $\log_{10}$ transformed relative gene expression values.

Figure C.13: Graphical tests for normality and equal variances of the residuals and random intercepts following a $\log_{10}$ transformation. The relative gene expression dataset was $\log_{10}$ transformed and reevaluated for the assumptions that linear mixed models are based on. (i) Normal probability plot of the residuals and (ii) random intercepts of the $\log_{10}$ transformed relative expression values. A much closer relation was observed between the data and the curve indicating normality (in red), compared to the untransformed data (Figure C.12), (iii) residuals randomly scattered around the horizontal axis were as expected in a normally distributed dataset.

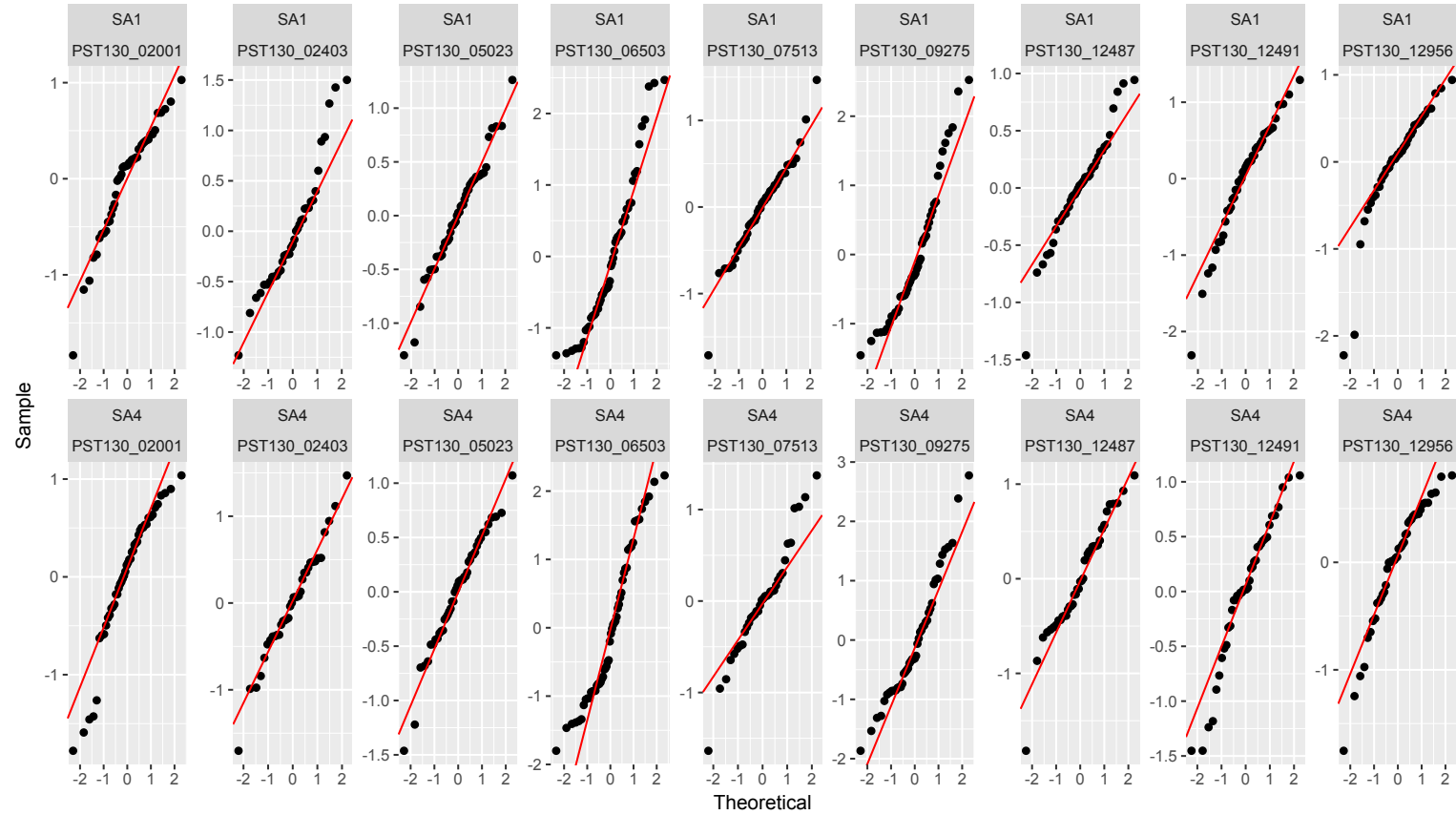


Figure C.14: Gene and isolate specific normal probability plots of the residuals after the model was fitted to the $\log_{10}$ transformed relative gene expression values.

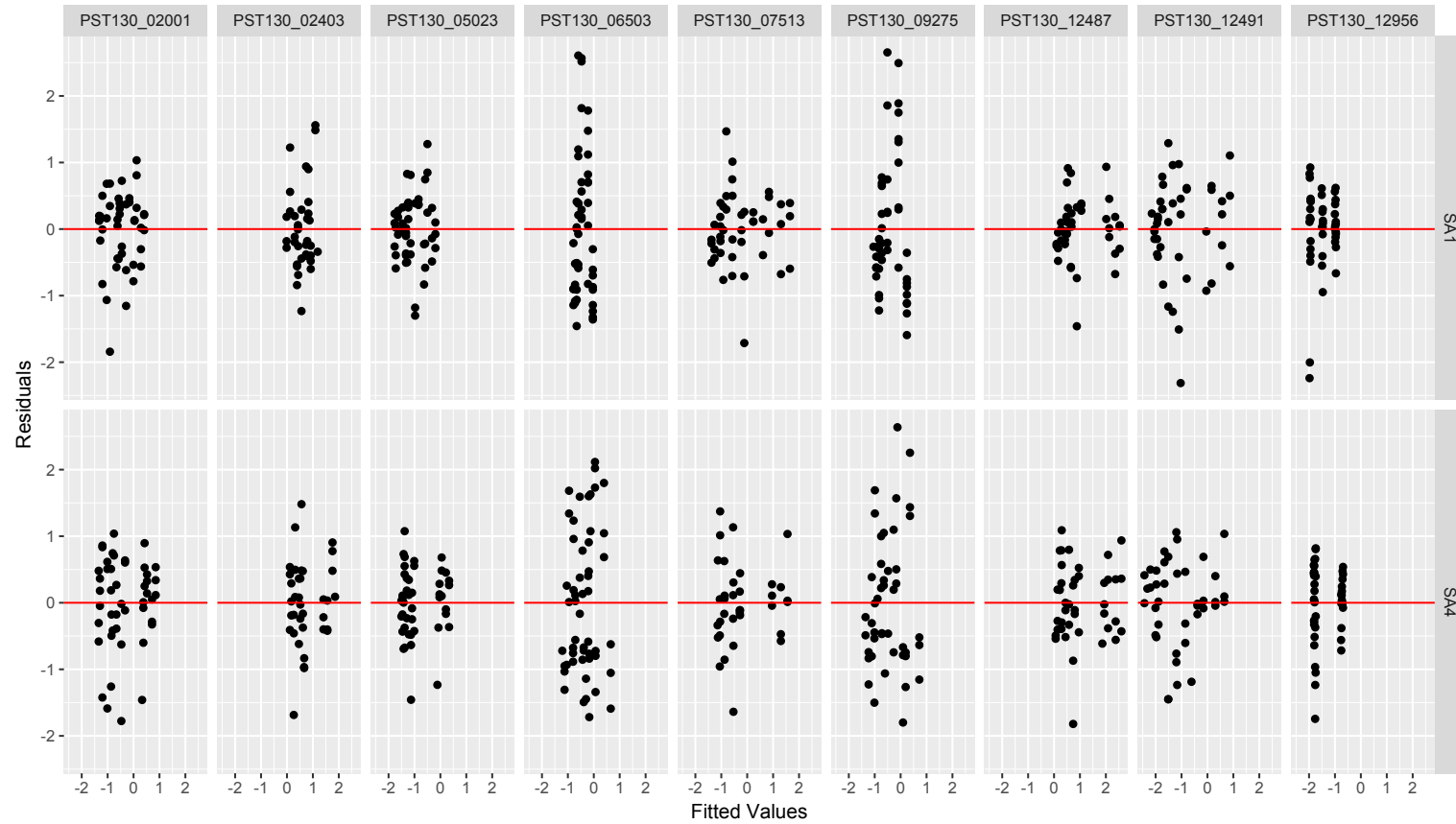


Figure C.15: Gene and isolate specific tests for equal variances after the model was fitted to the $\log_{10}$ transformed relative gene expression values.

C.3 Variability in RT-qPCR

RT-qPCR is a sensitive, multi-step method where technical variation can easily be introduced, yielding variable data. Ling et al. (2012) recommended the use of three biological replicates with two to three RT-PCR repeats. It is a lengthy process, where every step needs to be skillfully performed with high precision, using calibrated instruments and keeping consumables constant. High quality template DNA is needed for biologically useful results, as limited template material can reduce the sensitivity of qPCR (Derveaux et al., 2010).

Relative quantification of target gene expression complicates the process because only one reference gene can be used in the Pfaffl gene expression quantification model (Pfaffl, 2001). The use of a single reference gene is not advised, as more accurate results, with higher statistical significance are obtained when multiple reference genes are used. This requires the use of more complicated quantification models with built-in calibration schemes, efficiency calculators and methods to determine confidence measures, all adding to the total cost of qPCR experiments (Derveaux et al., 2010).

The technical sensitivity of the RT-qPCR assay would ideally require every step in this five-step process to be replicated, for example, replicated RNA extractions of the same tissue sample and replicate cDNA synthesis. This is often impractical, both in terms of time and cost. In this study, the amount of sample tissue that could be harvested per time point was an additional limitation. To include two more reference genes in the study would have increased the number of PCR reactions by three times, significantly increasing required time and resources. Even if this was not a limiting factor, raw sample material would not have yielded enough RNA to be used in all reactions, and a whole different experimental approach would have been needed. The experimental setup performed was the most ambitious approach that could be accommodated.

In the rest of this section best practices to keep technical noise to a minimum

will be discussed and suggestions of how to further improve this aspect in future studies will be made.

C.3.1 Variation in the application of treatments to biological replicates

To keep technical variation to a minimum, special care must be taken to apply inoculum evenly to all plants. In this study, variation could have been introduced by the specific location of trays in the glasshouse, as some plants could have received more sunlight than others. In future, a mock inoculation can be considered as an alternative negative control.

Run-to-run variation is not considered as true biological variation, but should not be regarded as negligible. The sample “maximisation assay setup”, described in Section 6.2.6, was applied with the aim to avoid this variation. Due to all the possible introduced variation, solid conclusions rely on the inclusion of multiple, independent biological replicates (Derveaux et al., 2010). However, due to physical capacity limitations, different biological replicates were assessed in different PCR runs. Assays would be spread out over even more runs if mock inoculation samples were added. Derveaux et al. (2010), offers a word of caution about inter-run variation, proposing inclusion of an identical sample across all plates as an inter-run calibrator. The limitation in the present study was that the positive control sample was not identical between all plates due to quantity constraints.

C.3.2 Variation introduced by the RNA extraction process

One must take the utmost care during the RNA extraction process as this is the most vulnerable part of the RT-qPCR experiment, due to the unstable nature of RNA. Fleige and Pfaffl (2006) also emphasise that it is important to use intact RNA for RT-qPCR and states that the Bioanalyzer 2100 measurement is a stable and reliable method for the quantification and quality assessment of RNA (Fleige and Pfaffl, 2006). Bustin and Nolan (2004) argue that RNA purification is the

critical determinant of reproducibility and biological relevance of the subsequent result.

Approximately half of the inoculated leaf sample for each time point was used in the first attempt of RNA extraction. The quantity of sampling material is therefore a limiting factor for RNA extraction replicates. With the current method, a maximum of two RNA extraction replicates would be possible, but that would leave no material to repeat an extraction that was unsuccessful. Sample processing, storing and transportation were controlled and kept consistent to minimise variability.

C.3.3 Variation introduced by the reverse transcription process

In this study, random hexamers were used as primers for the RT process. Using this nonspecific random oligonucleotides allowed the assessment of multiple targets in each sample and is known to yield the highest quantity and least bias cDNA. The alternative to prime RT reactions with, are thymine oligonucleotides (oligo-dTs). The advantages of this type of primers are that it is more specific to mRNA. RNA needs to be intact and of very high quality for this method as it will not prime RNAs without an adenine tail consisting of multiple adenine nucleotides (polyA tail). Judging the RNA integrity number (RIN) range, the methodology of random hexamers rendered it more suitable for the samples used in this study. An alternative priming method could be tested in future experiments, where a mixture of random hexamers and oligo-dTs is used, as suggested by Taylor and Mrkusich (2014).

Some may argue that the one-step method of RT-qPCR reduces technical noise to some extent. The two-step method of RT-qPCR was chosen in this work as expression of multiple genes were assessed. Interest in more than one part of the transcriptome also meant that qPCR experiments were done over a considerable time frame of about three months and therefore storing a more stable form of nucleic acid was beneficial.

C.3.4 Variation introduced by RT-qPCR

Two objectives drive experimental layout (Derveaux et al., 2010). Firstly, in the “gene maximisation method”, the expression profiles of different genes in a sample are compared. Multiple genes are assessed together on the same sample DNA per plate and samples are spread across different plates. In the present study, the alternative layout was used. Nine biological replicates were assessed for each gene and time point, allowing one biological replicate per 96-well PCR plate. This type of layout is also known as the “sample maximisation method” where samples that will be compared are preferably run together. The standardised values of the two treatments were compared with one another in each plate at each time point. In this way, gene expression changes can be investigated over the course of the infection process, to assess differences in the expression pattern between the two treatments. Because the number of wells per run is a limitation, the nine genes were assessed in different PCR runs. To reduce technical variability, three repeats of each sample were evaluated (Ling et al., 2012). Although isolates could be compared across time points for each gene assay within a run, inter-run variability could not be accounted for, as the nine biological replicates were assessed on nine different plates and the internal control had to be taken from different samples due to too little quantities. Standardisation as explained by Willems et al. (2008) could have been considered if a third treatment or mock treatment was present.

C.3.5 Variation introduced by primers

Primer design

Primers that were specific to target sequences were designed with the help of online tools. Primers were designed with suitable characteristics regarding GC content, with low probability to form secondary structures and with the desired

annealing temperature, using gene sequences.

Empirical proof of primer dimer absence can be found by assessment of the melt curves. Primers-dimers are usually shorter than target amplicons and would therefore form a peak at low melting temperatures which are clearly visible on the melt curve (Kubista et al., 2006). In such a case the melt curve will have two peaks, one for the secondary structures and one for the amplicons.

Designing primers for long gene sequences that include introns can be more complex. The possibility of alternative splicing in short genes of fungal pathogens has been indicated before (Grützmann et al., 2014), which sets the stage for future investigations on alternative splicing in *Pst* effector coding genes. Effectors are often short peptides (Saunders et al., 2012), which are consequently not rich in intronic regions. However, when genes have alternatively spliced isoforms, target identification can be tricky, complicating the primer design in turn (Derveaux et al., 2010). Some of the candidates assessed in this work have been shown to be alternatively spliced from RNA-Seq datasets.

By annealing primers at various points along the mRNA, information about the expression of specific exons or the full transcript length can be gathered. To avoid amplification of gDNA, primers can further be designed to span exon-exon boundaries (Thellin et al., 1999). In future, more attention could be given to incorporate this step in the primer design protocol to improve primer specificity. This would not be possible in all cases however, as for example where short exons exist with high sequence variability between compared entities, as other criteria such as the absence of SNPs in primer sequences and identical amplicons need to be met. This will furthermore highly increase the cost and time needed per gene assay. End point analysis by gel electrophoresis remains a good indication that a PCR product of the intended size is obtained (Wittwer et al., 1997). Gel purification and sequencing of the product can be performed for a more specific confirmation. In this study, after optimisation, one primer pair was used for each gene assayed.

Efficiency

Primer efficiencies were determined by implementing dilution series assays. Care should be taken that the starting concentration of the series is concentrated enough to allow six or seven serial dilutions that still contain sufficient template to yield an accurate result. Quantifications at low template concentrations were either not successful in the programmed 40 cycle PCRs, or less reproducible.

C.3.6 Choice of reference genes

The *Pst* β -tubulin was used as reference gene. It has been used widely across many species, but there are controversial reports in the literature about the stability of many of the genes traditionally considered to have stable expression profiles and specifically using them as reference genes in qPCR (Murphy and Polak, 2002; Jain et al., 2006; Schmidt and Delaney, 2010). If the reference gene has not been tested before in the same experimental conditions, it is recommended that more than one reference gene should be included when the relative quantification method is used (Thellin et al., 1999). Thellin et al. (1999) suggest using rRNA 18S and 28S as internal standards. The use of three to five reference genes is proposed for accurate normalisation (Derveaux et al., 2010). It is advised that more reference genes should be used to evaluate relative gene expression in future studies.

RT-qPCR remains to be a process full of grey areas, but it has become a prime method in various biological research fields (Schmittgen and Livak, 2008). Although the variability in data quality and reporting has been addressed by setting up the MIQE guidelines (Bustin et al., 2009), vague reporting of methodology and statistical analysis continues to mislead newcomers to the field.

The sensitivity of RT-qPCR makes it a powerful tool, but necessitates that every step is performed with great accuracy to keep variability, which is inevitably introduced into every part of the multi-step process, to a minimum. Technical

repeats aim to identify outliers that are not caused by true biological variation, but rather due to accumulation of technical inconsistency. Even when no obvious outliers can be identified, the measured C_T values are always a combination of true biological variance and technical noise introduced during the process (Ling et al., 2012).

C.3.7 Results of efficiency corrected relative gene expression

Sufficient amounts of biological replicates are needed to make sound conclusions. The trouble with RT-qPCR is that inter-plate variability can jeopardise conclusions. Appendix C, Figure C.16, displays the relative expression data that was standardised to the reference gene by the efficiency corrected method (Schmittgen and Livak, 2008) and $\log_{10}$ transformed. From this data, it is difficult to conclude a true biological result. Patterns seen across plates can indicate an experimental error. For example, high relative expressions were seen in all genes in SA1, plate four at time point 9. Similar behaviour was observed at time points 3 and 5, although to a lesser extent.

Appendix C, Figure C.17, indicates relative gene expression profiles as determined by the efficiency corrected relative quantification method. It is clear from Figures C.16 and C.17 that the Pfaffl method is not suitable for data with so much variability. To achieve success in RT-qPCR assays in future, it is advised to use multiple reference genes and deploy the developed software available, as reviewed (Ruijter et al., 2013). Improvement of both throughput and accuracy can be achieved using a 384-well PCR platform.

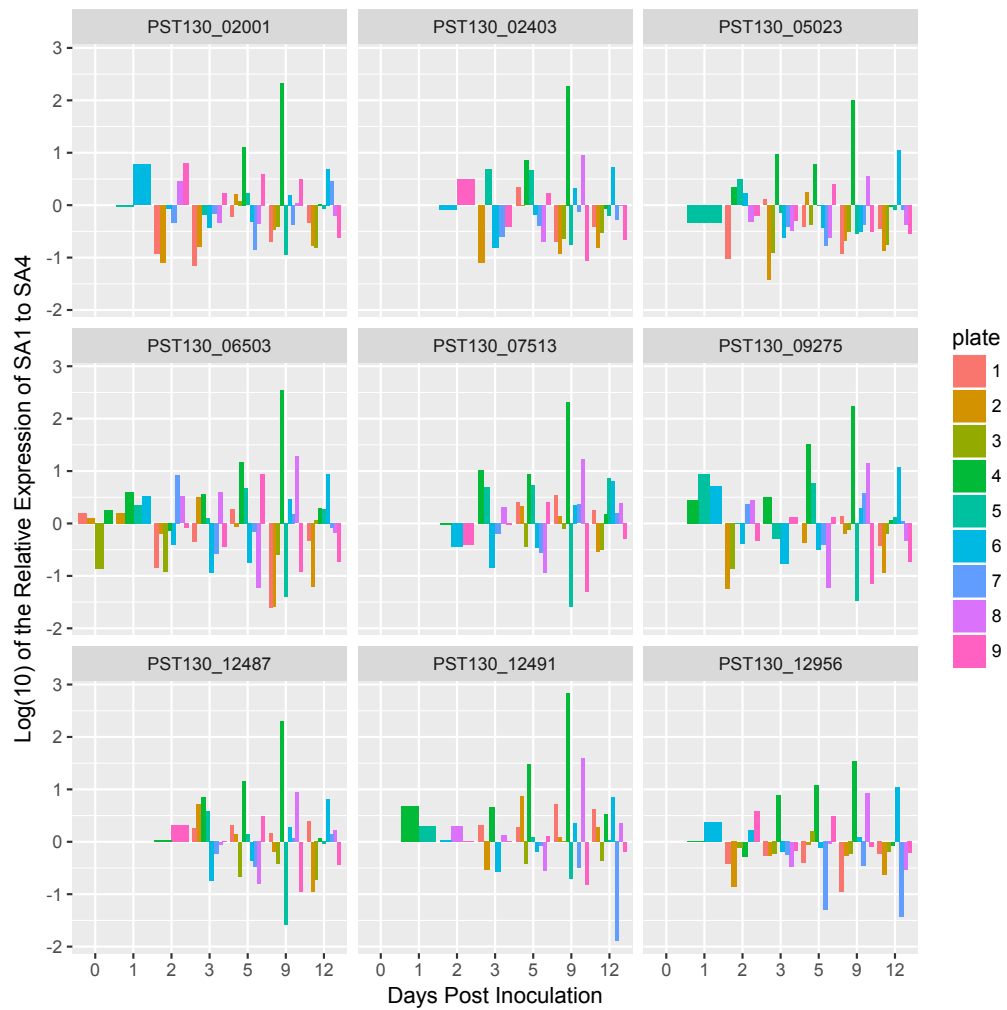


Figure C.16: High inter-run variability in relative expression patterns is due to the sum of the effects of inter-assay variability and the variability between different biological replicates. It is difficult to distinguish between the two sources of variability. This highlights the need for a calibration method when experiments include more than one qPCR that need to be compared.

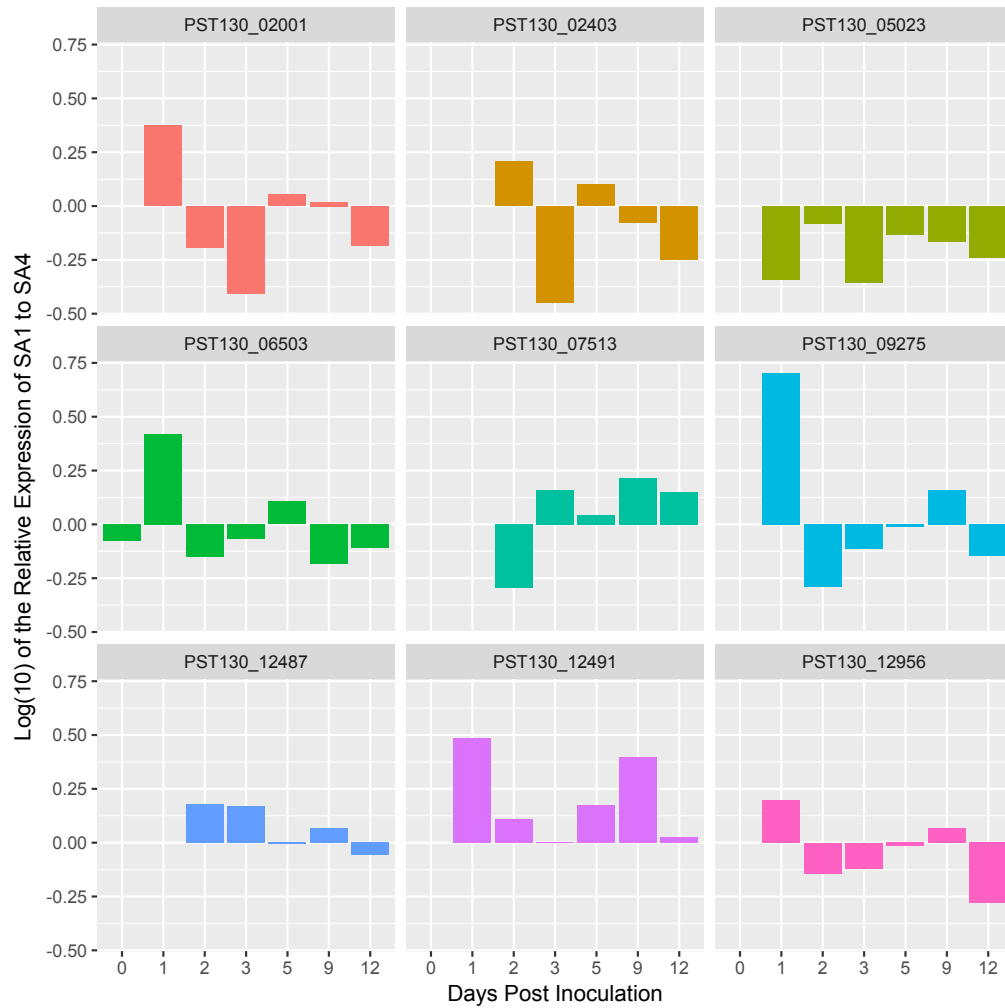


Figure C.17: The Pfaffl method of relative gene expression shows the relative gene expression of SA1 to SA4. A positive value indicates a higher expression in SA1, while a negative value indicates a higher expression in SA4. This method does not correct for inter-run variability and risks making false conclusions.

Appendix D

Analysis of the Current Stripe Rust Threat in South Africa

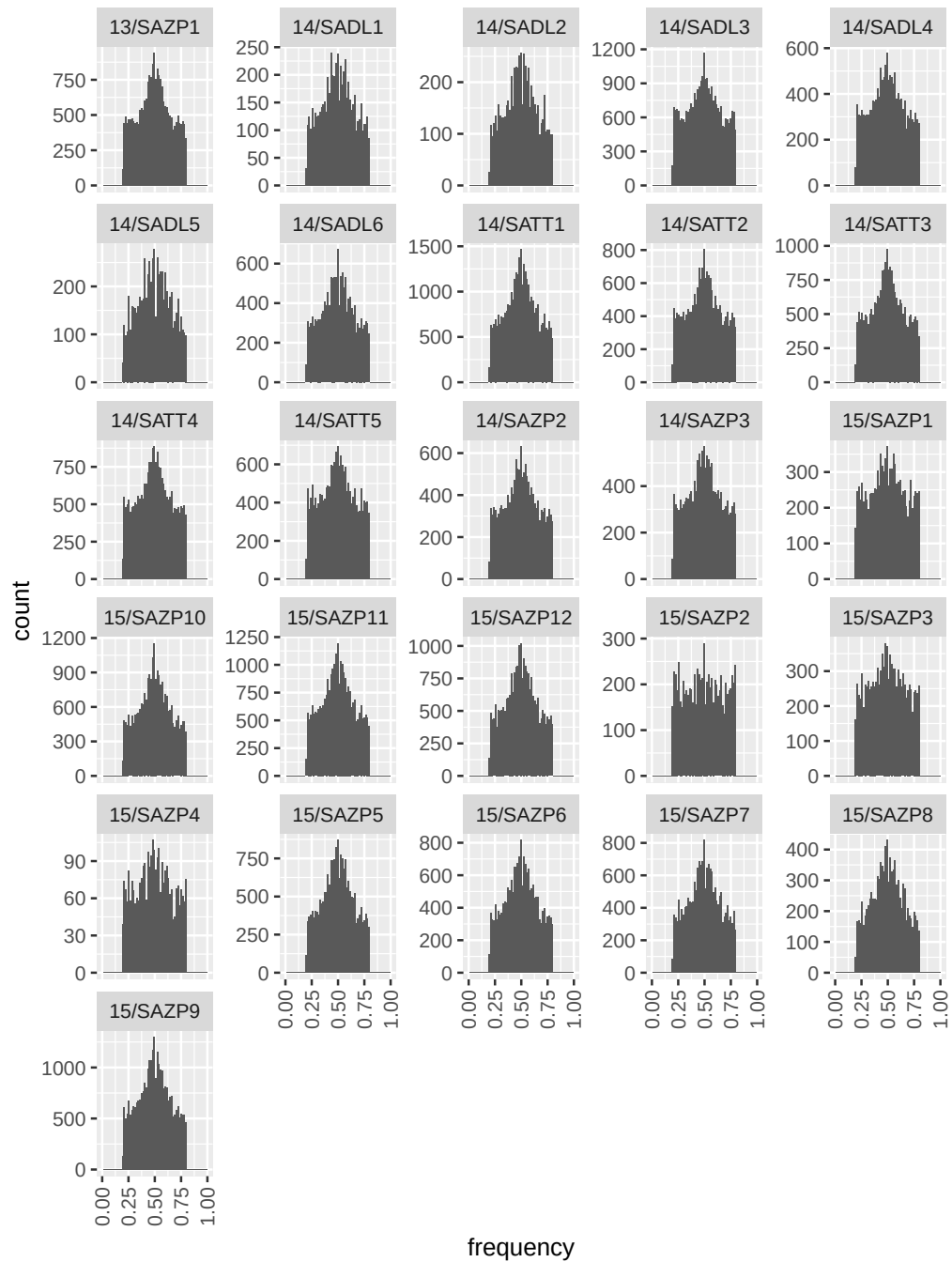


Figure D.1: Read frequency graphs from heterokaryotic SNP sites for the South African field isolates (analysed in Chapter 7) that were collected between 2013 and 2015. See Table 7.2 for further identification purposes.

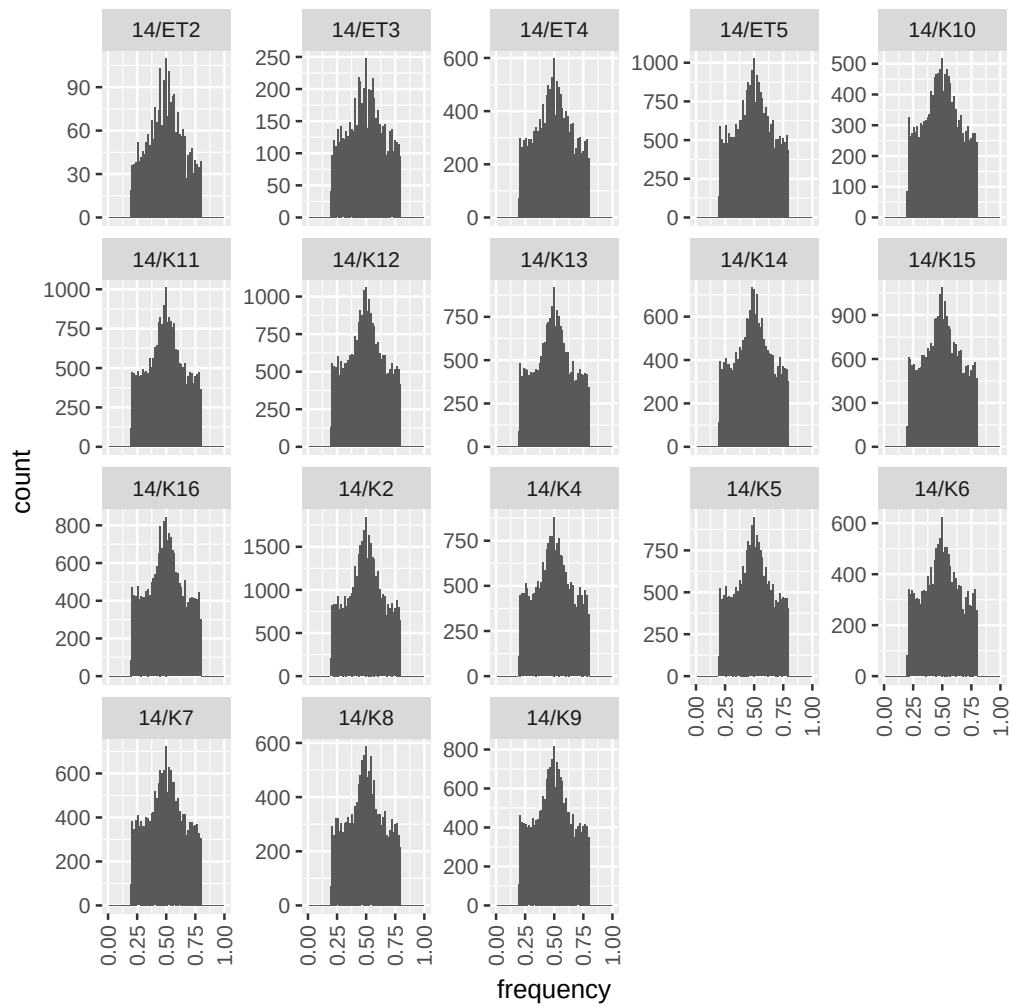


Figure D.2: Read frequency graphs from heterokaryotic SNP sites for the East African field isolates (analysed in Chapter 7) that were collected in 2014. See Table 7.2 for further identification purposes.

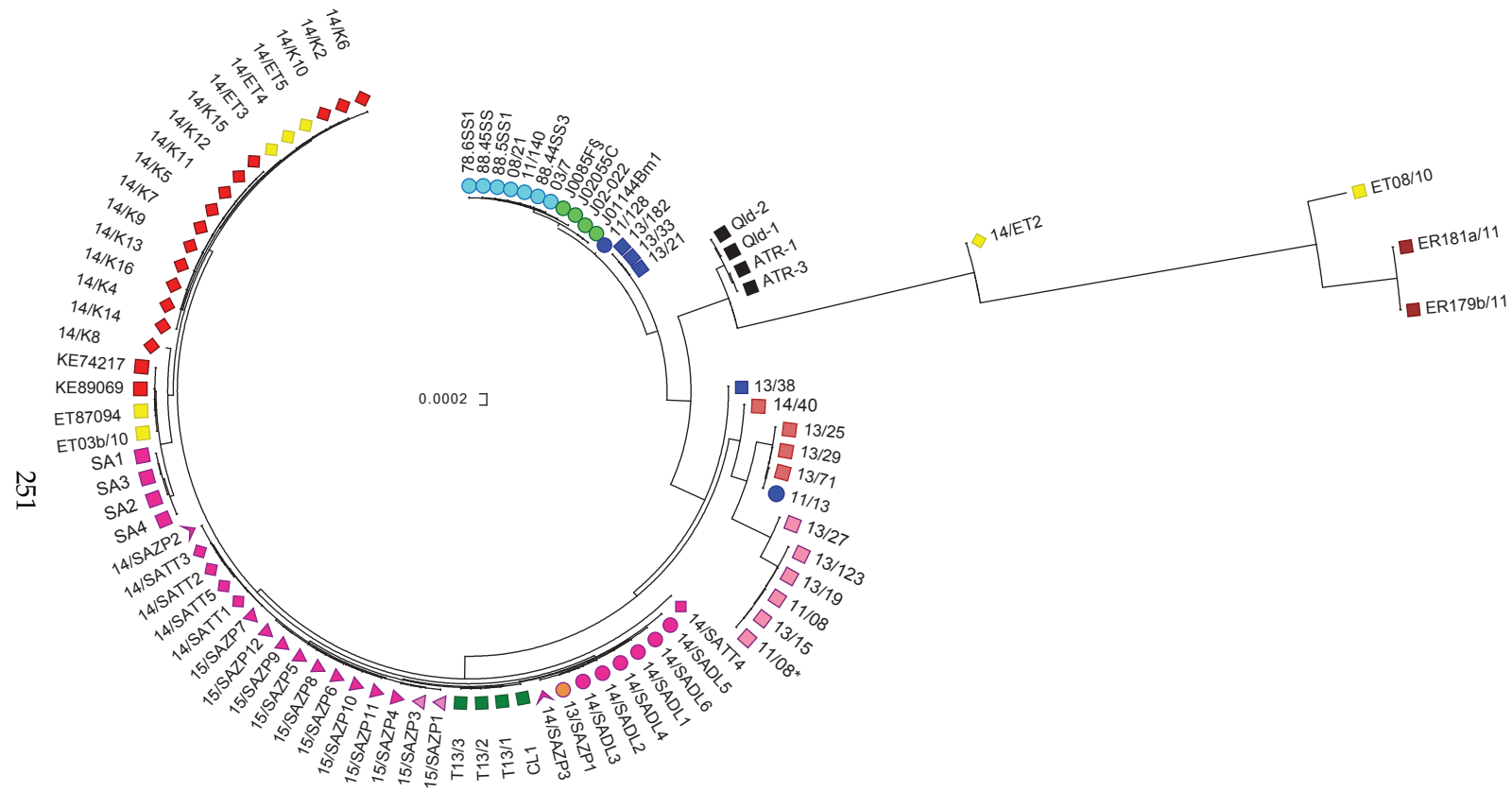


Figure D.3: Circular relative distance maximum likelihood phylogenetic tree. The relative distance maximum likelihood phylogenetic tree describes the relative relationship between isolates described in Figure 7.3 where branch lengths are ignored and only topology was considered. The group East Africa (B) isolates absent from Figure 7.4 is displayed in this dendrogram. The key in Figure 7.3 also applies here.

Table D.1: Differential testing of South African *Pst* isolates previously defined as patho-type 6E16A- on an extended set of wheat seedling testers

Isolate ID		SA1 (6E16A-)		13/SAZP1	
Wheat variety	Resistance	Rep1	Rep2	Rep1	Rep2
Suwon 92/ Omar	<i>Yr 4, Su</i>	;c	;c	;c	;c
Strubbes Dickoph	<i>Yr Sd, 25</i>	;c	;c	;c	;c
Moro	<i>Yr 10, Mor</i>	;c	;c	;cn	;cn
Vilmorin 23	<i>Yr 3a, 4a</i>	;cn	;cn	;cn	;cn
Heines Kolben	<i>Yr 2, 6</i>	3+	3+	3+	3+
Lee	<i>Yr 7</i>	4	4	4	4
Chinesees 166	<i>Yr 1</i>	;	-	-	;
Heines VII	<i>2, 25, HVII</i>	;cn	;cn	;cn	;cn
AvS/YrSp	<i>Yr Sp</i>	-	-	-	-
AvS/Yr32	<i>Yr 32 (YrCV)</i>	;1=c	-	;c	-
Compair	<i>Yr 8</i>	-	-	-	3
Nord Desprez	<i>Yr 3a, 4a</i>	;c	;c	;c	;c
Heines Peko	<i>Yr 2, 6, 25</i>	;cn	;cn	;cn	;cn
Reichersberg 42	<i>Yr 7, 25</i>	;cn	;cn	;cn	;cn
Hybrid 46	<i>Yr 4b</i>	;	;	;	;
Clement	<i>Yr 2, 9, 25, Cle</i>	;	;	;	;
T. Spelta	<i>Yr 5</i>	-	-	;	;
Kalyansona	<i>Yr 2</i>	4	4	4	4
Avocet S	Susc control	4	4	3++	4
Yr1/6* AvS	<i>Yr 1</i>	;	;	;	;
Yr5/6* AvS	<i>Yr 5</i>	;	-	;	;
Yr6/6* AvS	<i>Yr 6</i>	4	4	4	4
Yr7/6* AvS	<i>Yr 7</i>	4	4	4	4
Yr8/6* AvS	<i>Yr 8</i>	-	-	-	3+
Yr9/6* AvS	<i>Yr 9</i>	-	-	;c	-
Yr10/6* AvS	<i>Yr 10</i>	-	-	-	-
Yr15/6* AvS	<i>Yr 15</i>	;	-	;	;
Yr17/6* AvS	<i>Yr 17</i>	3=c	3=c	3=c	3=c
Yr24/6* AvS	<i>Yr 24</i>	;cn	;cn	;nc	;cn
Yr26/3* AvS	<i>Yr 26</i>	;cn	-	;cn	-
Yr27/6* AvS	<i>Yr 27</i>	;1=cn	;1=cn	;cn	;1=cn
Avocet R	<i>Yr A</i>	;c	;c	;1cn	;1c
Yr Sk	<i>Yr27</i>	1cn	1cn	;1cn	2c
Selkirk	<i>Yr27</i>	;1=cn	;1=cn	;1cn	;1cn
Ambition	<i>Unknown</i>	;cn	;cn	;cn	;cn
Apache	<i>Yr7, Yr17, Other</i>	;cn	;cn	;cn	;cn
Brigadier	<i>Yr17, Other</i>	;	;	;	;
Cadenza	<i>Yr6, Yr7, Other</i>	;cn	;cn	;cn	;cn
Carstens V	<i>Yr32, Other</i>	;	;	;	;
Claire	<i>Yr16</i>	;c	;c	;c	;c
Crusoe	<i>Unknown</i>	;	;	;	;
Delphi	<i>Unknown</i>	;cn	;cn	;cn	;cn
Kranich	<i>Unknown</i>	;cn	;cn	;	;
KWS Sterling	<i>Unknown</i>	;	;	;	;
Monterey	<i>Unknown</i>	;c	;c	;c	;c
Mosaic	<i>Unknown</i>	;	;	;	;
Rendezvous	<i>Yr17, Other</i>	;	;	;	;
Revelation	<i>Unknown</i>	;	;	;	;
Robigus	<i>Unknown</i>	;c	;c	;c	;c
Solstice	<i>Unknown</i>	;	;	;	;
Trident	<i>Yr17</i>	33	33c	33c	3c
Talon	<i>Yr32</i>	;	;	;	;
Vuka	Susc control (UK)	;cn	;cn	;cn	;cn
Warrior	<i>Unknown</i>	;	;	;	;

Rep, replicate | susc, susceptible | ; fleck | = very small pustule | 1-4 size of pustule | c chlorosis | n necrosis

Table D.2: Differential testing of South African *Pst* isolates previously defined as patho-type 6E22A+ on an extended set of wheat seedling testers

Isolate ID		SA4		15/SAZP4	
Wheat variety	Resistance	Rep1	Rep2	Rep1	Rep2
Suwon 92/ Omar	<i>Yr 4, Su</i>	;c	;c	;c	;c
Strubbes Dickoph	<i>Yr Sd, 25</i>	3=c	3++	mixed: ;c 2	-
Moro	<i>Yr 10, Mor</i>	;cn	;c	;c	;c
Vilmorin 23	<i>Yr 3a, 4a</i>	;1=c	;1=c	;1=c	;1=c
Heines Kolben	<i>Yr 2, 6</i>	4	4	4	4
Lee	<i>Yr 7</i>	4	4	4	4
Chineses 166	<i>Yr 1</i>	;	-	;	-
Heines VII	<i>2, 25, HVII</i>	;1+cn	;1cn	3c	3c
AvS/ <i>YrSp</i>	<i>Yr Sp</i>	-	-	-	-
AvS/ <i>Yr32</i>	<i>Yr 32 (YrCV)</i>	-	-	;1c (single)	-
Compair	<i>Yr 8</i>	-	-	-	3
Nord Desprez	<i>Yr 3a, 4a</i>	;1=c	;1=c	;c	;c1=c
Heines Peko	<i>Yr 2, 6, 25</i>	3-c	3-c	3-c	3+c
Reichersberg 42	<i>Yr 7, 25</i>	3++c	3++c	3++c	3++c
Hybrid 46	<i>Yr 4b</i>	;	;	;	;
Clement	<i>Yr 2, 9, 25, Cle</i>	;	;	;	;
T. Spelta	<i>Yr 5</i>	;	-	;	-
Kalyansona	<i>Yr 2</i>	4	4	4	4
Avocet S	Susc control	4	4	4	4
<i>Yr1/6*</i> AvS	<i>Yr 1</i>	;	;	;	;
<i>Yr5/6*</i> AvS	<i>Yr 5</i>	;	-	-	-
<i>Yr6/6*</i> AvS	<i>Yr 6</i>	4	4	4	4
<i>Yr7/6*</i> AvS	<i>Yr 7</i>	4	4	4	4
<i>Yr8/6*</i> AvS	<i>Yr 8</i>	3	-	-	-
<i>Yr9/6*</i> AvS	<i>Yr 9</i>	;	-	-	;c
<i>Yr10/6*</i> AvS	<i>Yr 10</i>	-	-	-	-
<i>Yr15/6*</i> AvS	<i>Yr 15</i>	;	-	;	-
<i>Yr17/6*</i> AvS	<i>Yr 17</i>	3=c	3=c	3++	3++
<i>Yr24/6*</i> AvS	<i>Yr 24</i>	;c	;c	;c	;c
<i>Yr26/3*</i> AvS	<i>Yr 26</i>	-	-	;c (single)	-
<i>Yr27/6*</i> AvS	<i>Yr 27</i>	;1=cn	;1=cn	;1=cn	-
Avocet R	<i>Yr A</i>	4	4	mixed: ;c 4	mixed: ;c 4
<i>Yr Sk</i>	<i>Yr27</i>	;12cn	;12cn	2+c	2+c
Selkirk	<i>Yr27</i>	2cn	2cn	3=cn	3=cn
Ambition	<i>Unknown</i>	;cn	;cn	;1=1=cn	;1=cn
Apache	<i>Yr7, Yr17, Other</i>	;cn	;cn	;cn	;cn
Brigadier	<i>Yr17, Other</i>	;	;	;	;
Cadenza	<i>Yr6, Yr7, Other</i>	3c	3c	3+c	3+c
Carstens V	<i>Yr32, Other</i>	;cn	;cn	;cn	;cn
Claire	<i>Yr16</i>	;1=cn	;cn	;cn	;cn
Crusoe	<i>Unknown</i>	;	;	;	;
Delphi	<i>Unknown</i>	;cn	;cn	;cn	;cn
Kranich	<i>Unknown</i>	;cn	;cn	;cn	;1=cn
KWS Sterling	<i>Unknown</i>	;cn	;cn	;cn	;cn
Monterey	<i>Unknown</i>	;cn	;cn	;1=c	;1=c
Mosaic	<i>Unknown</i>	;cn	;cn	;cn	;cn
Rendezvous	<i>Yr17, Other</i>	;cn	;cn	;cn	;cn
Revelation	<i>Unknown</i>	;cn	;cn	;cn	;cn
Robigus	<i>Unknown</i>	;cn	;cn	;cn	;cn
Solstice	<i>Unknown</i>	;	;	;c	;c
Trident	<i>Yr17</i>	3++	3c	4	3++c
Talon	<i>Yr32</i>	;cn	;cn	;cn	;cn
Vuka	Susc control (UK)	3=c	3=c	3++	3++
Warrior	<i>Unknown</i>	;	;	;	;

Rep, replicate | susc, susceptible | ; fleck | = very small pustule | 1-4 size of pustule | c chlorosis | n necrosis
 mixed, more than one infection type among individual plants in inoculation replicate | single, only a single plant scored

Bibliography

- AgriOrbit, 2017. Uncertainty over Western Cape wheat cultivation conditions. URL <https://agriorbit.com/uncertainty-western-cape-wheat-cultivation-conditions/>. [Online; accessed 20/01/2018].
- Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., Enjalbert, J., and de Vallavieille-Pope, C. 2014. Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen *Puccinia striiformis* f.sp. *tritici*. *PLoS Pathogens*, 10:e1003903.
- Ali, S., Rodriguez-Algaba, J., Thach, T., Sørensen, C. K., Hansen, J. G., Lassen, P., Nazari, K., Hodson, D. P., Justesen, A. F., and Hovmøller, M. S. 2017. Yellow rust epidemics worldwide were caused by pathogen races from divergent genetic lineages. *Frontiers in Plant Science*, 8:1057.
- Allison, O. C. and Isenbeck, K. 1930. Biological specialization of *Puccinia glumarum tritici* Eriksson and Henning. *Phytopathologische Zeitschrift*, 2.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25:3389–3402.
- Ames, B. N. 1979. Identifying environmental chemicals causing mutations and cancer. *Science*, 204:587–593.
- Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19: 535–544.
- Andrews, S., 2010. FastQC A Quality Control tool for High Throughput Sequence Data. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. [Online; accessed 20/01/2018].
- Anikster, Y. 1984. The formae speciales. In Bushnell, W. R. and Roelfs, A. P., editors, *The Cereal Rusts*. Orlando.

- Badebo, A., Stubbs, R. W., van Ginkel, M., and Gebeyehu, G. 1990. Identification of resistance genes to *Puccinia striiformis* in seedlings of Ethiopian and CIM-MYT bread wheat varieties and lines. *Netherlands Journal of Plant Pathology*, 96: 199–210.
- Badebo, A., Assefa, S., and Fehrman, H. 2008. Yellow rust resistance in advanced lines and commercial cultivars of bread wheat from Ethiopia. *East African Journal of Sciences*, 2:29–34.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Beddow, J. M., Pardey, P. G., Chai, Y., Hurley, T. M., Kriticos, D. J., Braun, H.-J., Park, R. F., Cuddy, W. S., and Yonow, T. 2015. Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants*, 1:15132.
- Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannonouche, P., Wider, G., Peter, M., Lehmann, A. R., Hofmann, K., and Dikic, I. 2005. Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science*, 310:1821–1824.
- Bockus, W. W. and Wiese, M. V., editors. 2010. *Compendium of wheat diseases and pests*. St. Paul, Minn, 3rd ed edition.
- Bofkin, L. and Goldman, N. 2006. Variation in evolutionary processes at different codon positions. *Molecular Biology and Evolution*, 24:513–521.
- Bolton, M. D., Kolmer, J. A., and Garvin, D. F. 2008. Wheat leaf rust caused by *Puccinia triticina*. *Molecular Plant Pathology*, 9:563–575.
- Boshoff, W. H. P. and Pretorius, Z. A. 1999. A new pathotype of *Puccinia striiformis* f. sp. *tritici* on wheat in South Africa. *Plant Disease*, 83:591–591.
- Boshoff, W. H. P., Pretorius, Z. A., and Van Niekerk, B. D. 2002. Establishment, distribution, and pathogenicity of *Puccinia striiformis* f. sp. *tritici* in South Africa. *Plant Disease*, 86:485–492.
- Boshoff, W. H. P., Pretorius, Z. A., and Van Niekerk, B. D. 2003. Fungicide efficacy and the impact of stripe rust on spring and winter wheat in South Africa. *South African Journal of Plant and Soil*, 20:11–17.
- Bozkurt, T. O., Schornack, S., Banfield, M. J., and Kamoun, S. 2012. Oomycetes, effectors, and all that jazz. *Current opinion in plant biology*, 15:483–492.
- Brown, J. K. M. 2003. Little else but parasites. *Science*, 299:1680–1681.
- Brown, J. K. and Hovmöller, M. S. 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297: 537–541.

- Bubić, I., Wagner, M., Krmpotić, A., Saulig, T., Kim, S., Yokoyama, W. M., Jonjić, S., and Koszinowski, U. H. 2004. Gain of virulence caused by loss of a gene in murine cytomegalovirus. *Journal of Virology*, 78:7536–7544.
- Bueno-Sancho, V., Persoons, A., Hubbard, A., Cabrera-Quio, L. E., Lewis, C. M., Corredor-Moreno, P., Bunting, D. C. E., Ali, S., Chng, S., Hodson, D. P., Madariaga Burrows, R., Bryson, R., Thomas, J., Holdgate, S., and Saunders, D. G. O. 2017. Pathogenomic analysis of wheat yellow rust lineages detects seasonal variation and host specificity. *Genome Biology and Evolution*, 9:3282–3296.
- Burns, M. J., Nixon, G. J., Foy, C. A., and Harris, N. 2005. Standardisation of data from real-time quantitative PCR methods—evaluation of outliers and comparison of calibration curves. *BMC Biotechnology*, 5:31.
- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., and Wittwer, C. T. 2009. The MIQE Guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55:611–622.
- Bustin, S. A. and Nolan, T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques: JBT*, 15:155.
- Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., Daelen, R. v., Lee, T. v. d., Diergaarde, P., Groenendijk, J., Töpsch, S., Vos, P., Salamini, F., and Schulze-Lefert, P. 1997. The barley Mlo gene: A novel control element of plant pathogen resistance. *Cell*, 88:695–705.
- Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., Jurka, J., Michelmore, R. W., and Dubcovsky, J. 2011. Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE*, 6:e24230.
- Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., Dubcovsky, J., Saunders, D. G., and Uauy, C. 2013. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics*, 14:270.
- Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., Stajich, J. E., and Ramírez, L. 2016. Transposable elements versus the fungal genome: Impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics*, 12: e1006108.
- Chen, J., Upadhyaya, N. M., Ortiz, D., Sperschneider, J., Li, F., Bouton, C., Breen, S., Dong, C., Xu, B., Zhang, X., Mago, R., Newell, K., Xia, X., Bernoux, M., Taylor, J. M., Steffenson, B., Jin, Y., Zhang, P., Kanyuka, K., Figueroa, M., Ellis,

- J. G., Park, R. F., and Dodds, P. N. 2017. Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50* resistance in wheat. *Science*, 358:1607–1610.
- Chen, W., Wellings, C., Chen, X., Kang, Z., and Liu, T. 2014. Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*: *Puccinia striiformis*, yellow rust. *Molecular Plant Pathology*, 15:433–446.
- Chen, X. M., Line, R. F., and Leung, H. 1993. Relationship between virulence variation and DNA polymorphism in *Puccinia striiformis*. *Phytopathology*, 83: 1489–1497.
- Chen, X., Penman, L., Wan, A., and Cheng, P. 2010. Virulence races of *Puccinia striiformis* f. sp. *tritici* in 2006 and 2007 and development of wheat stripe rust and distributions, dynamics, and evolutionary relationships of races from 2000 to 2007 in the United States. *Canadian Journal of Plant Pathology*, 32:315–333.
- Chen, X. 2005. Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. *Canadian Journal of Plant Pathology*, 27:314–337.
- Chen, Y.-E., Cui, J.-M., Su, Y.-Q., Yuan, S., Yuan, M., and Zhang, H.-Y. 2015. Influence of stripe rust infection on the photosynthetic characteristics and antioxidant system of susceptible and resistant wheat cultivars at the adult plant stage. *Frontiers in Plant Science*, 6:779.
- Cheng, P., Ma, Z., Wang, X., Wang, C., Li, Y., Wang, S., and Wang, H. 2014. Impact of UV-B radiation on aspects of germination and epidemiological components of three major physiological races of *Puccinia striiformis* f. sp. *tritici*. *Crop Protection*, 65:6–14.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6:80–92.
- Coram, T. E., Wang, M., and Chen, X. 2008. Transcriptome analysis of the wheat—*Puccinia striiformis* f. sp. *tritici* interaction. *Molecular Plant Pathology*, 9: 157–169.
- Cuomo, C. A., Bakkeren, G., Khalil, H. B., Panwar, V., Joly, D., Linning, R., Sakthikumar, S., Song, X., Adiconis, X., and Fan, L. 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3: Genes, Genomes, Genetics*, 7:361–376.
- DAFF, 2015. A profile of the South African wheat market value chain. URL <http://www.nda.agric.za/doaDev/sideMenu/Marketing/AnnualPublications/CommodityProfiles/fieldcrops/WheatMarketValueChainProfile2015.pdf>. [Online; accessed 20/01/2018].
- DAFF, 2016. A profile of the South African wheat market value chain. URL <http://www.nda.agric.za/doaDev/sideMenu/Marketing/AnnualPublications/>

- CommodityProfiles/fieldcrops/WheatMarketValueChainProfile2016.pdf. [Online; accessed 20/01/2018].
- Dangl, J. L. and Jones, J. D. 2001. Plant pathogens and integrated defence responses to infection. *Nature*, 411:826.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12:499–510.
- de Vallavieille-Pope, C., Huber, L., Leconte, M., and Bethenod, O. 2002. Preinoculation effects of light quantity on infection efficiency of *Puccinia striiformis* and *P. tritici* on wheat seedlings. *Phytopathology*, 92:1308–1314.
- de Vallavieille-Pope, C., Ali, S., Leconte, M., Enjalbert, J., Delos, M., and Rouzet, J. 2012. Virulence dynamics and regional structuring of *Puccinia striiformis* f. sp. *tritici* in France between 1984 and 2009. *Plant Disease*, 96:131–140.
- Dean, R., Van Kan, J. a. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd, J. J., Dickman, M., Kahmann, R., Ellis, J., and Foster, G. D. 2012. The top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13:414–430.
- Denbel, W. 2014. Epidemics of *Puccinia striiformis* f. sp. *tritici* in Arsi and West Arsi zones of Ethiopia in 2010 and identification of effective resistance genes. *Journal of Natural Sciences Research*, 4:33–39.
- Derveaux, S., Vandesompele, J., and Hellemans, J. 2010. How to do successful gene expression analysis using real-time PCR. *Methods*, 50:227–230.
- Dobon, A., Bunting, D. C. E., Cabrera-Quio, L. E., Uauy, C., and Saunders, D. G. O. 2016. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics*, 17: 380.
- Dodds, P. N. and Rathjen, J. P. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 11:539.
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Teh, T., Wang, C.-I. A., Ayliffe, M. A., Kobe, B., and Ellis, J. G. 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, 103:8888–8893.
- Dodds, P. N., Rafiqi, M., Gan, P. H. P., Hardham, A. R., Jones, D. A., and Ellis, J. G. 2009. Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytologist*, 183:993–1000.

- Dong, S., Raffaele, S., and Kamoun, S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, 35:57–65.
- Dou, D. and Zhou, J.-M. 2012. Phytopathogen effectors subverting host immunity: different foes, similar battleground. *Cell Host & Microbe*, 12:484–495.
- Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F. 1998. Rates of spontaneous mutation. *Genetics*, 148:1667–1686.
- Du Plessis, A. 1933. The history of small-grains culture in South Africa. *Annals of the University of Stellenbosch*, 8:1652–1752.
- Duan, X., Tellier, A., Wan, A., Leconte, M., Vallavieille-Pope, C. d., and Enjalbert, J. 2010. *Puccinia striiformis* f.sp. *tritici* presents high diversity and recombination in the over-summering zone of Gansu, China. *Mycologia*, 102:44–53.
- Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kües, U., Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Peer, Y. V. d., Rouzé, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J., and Martin, F. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, 108:9166–9171.
- Edgerton, M. D. 2009. Increasing crop productivity to meet global needs for feed, food, and fuel. *Plant Physiology*, 149:7–13.
- Egorov, T. A., Odintsova, T. I., Pukhalsky, V. A., and Grishin, E. V. 2005. Diversity of wheat anti-microbial peptides. *Peptides*, 26:2064–2073.
- El Gueddari, N. E., Rauchhaus, U., Moerschbacher, B. M., and Deising, H. B. 2002. Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. *New Phytologist*, 156:103–112.
- Elyasi-Gomari, S. and Petrenkova, V. P. 2011. Virulence of *Puccinia striiformis* f. sp. *tritici* in Khuzestan province of Iran. *American Journal of Experimental Agriculture*, 1:281.
- Emanuelsson, O., Brunak, S., Heijne, G. v., and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2: 953.
- Enjalbert, J., Duan, X., Leconte, M., Hovmøller, M. S., and De Vallavieille-Pope, C. 2005. Genetic evidence of local adaptation of wheat yellow rust (*Puccinia*

- striiformis* f. sp. *tritici*) within France: Geographic structure of yellow rust in France. *Molecular Ecology*, 14:2065–2073.
- Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14:2611–2620.
- FAS USDA, 2016. Grain and feed annual report—Republic of South Africa. URL https://gain.fas.usda.gov/Recent%20GAIN%20Publications/Grain%20and%20Feed%20Annual_Pretoria_South%20Africa%20-%20Republic%20of_3-24-2016.pdf. [Online; accessed 20/01/2018].
- FAS USDA, 2017. United states department of agriculture—foreign agricultural service: Production, supply and distribution report. URL <https://apps.fas.usda.gov/psdonline/app/index.html#/app/home/statsByCountry>. [Online; accessed 20/01/2018].
- Fernández-Ortuño, D., Torés, J. A., Vicente, A. d., and Pérez-García, A. 2007. Multiple displacement amplification, a powerful tool for molecular genetic analysis of powdery mildew fungi. *Current Genetics*, 51:209–219.
- Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. 2008. *Longitudinal data analysis*.
- Fleige, S. and Pfaffl, M. W. 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*, 27:126–139.
- Flood, J. 2010. The importance of plant health to food security. *Food Security*, 2: 215–231.
- Flor, H. 1956. The complementary genic systems in flax and flax rust. *Advances in Genetics*, 8:29–54.
- Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M. J., Pennington, H. G., Kamoun, S., and Banfield, M. J. 2017. Effectors of filamentous plant pathogens: Commonalities amid diversity. *Microbiology and Molecular Biology Reviews*, 81: e00066–16.
- Garnica, D. P., Nemri, A., Upadhyaya, N. M., Rathjen, J. P., and Dodds, P. N. 2014. The ins and outs of rust haustoria. *PLoS Pathogens*, 10:e1004329.
- Gilroy, E. M., Breen, S., Whisson, S. C., Squires, J., Hein, I., Kaczmarek, M., Turnbull, D., Boevink, P. C., Lokossou, A., Cano, L. M., Morales, J., Avrova, A. O., Pritchard, L., Randall, E., Lees, A., Govers, F., van West, P., Kamoun, S., Vleeshouwers, V. G. A. A., Cooke, D. E. L., and Birch, P. R. J. 2011. Presence/absence, differential expression and sequence polymorphisms between PiAVR2 and PiAVR2-like in phytophthora infestans determine virulence on R2 plants. *New Phytologist*, 191:763–776.

- Glen, H. F. 2002. *Cultivated plants of Southern Africa: Botanical names, common names, origins, literature*.
- Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J., and Thordal-Christensen, H. 2010. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics*, 11:317.
- GRAIN SA, 2017. CEC Wheat per province: Production Info—Area Grown, Yields and Estimates. URL <http://www.grainsa.co.za/report-documents?cat=14>. [Online; accessed 20/01/2018].
- Griffiths, A. J. F., Wessler, S. R., Carroll, S. B., and Doebley, J. F. 2015. *Introduction to Genetic Analysis*. New York, NY, eleventh edition edition.
- Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics*, 11:1–21.
- Grützmann, K., Szafranski, K., Pohl, M., Voigt, K., Petzold, A., and Schuster, S. 2014. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA Research*, 21:27–39.
- Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., and Duplessis, S. 2011. The Poplar-Poplar rust interaction: Insights from genomics and transcriptomics. *Journal of Pathogens*, pages 1–11.
- Hane, J. K. and Oliver, R. P. 2010. In silico reversal of repeat-induced point mutation (RIP) identifies the origins of repeat families and uncovers obscured duplicated genes. *BMC Genomics*, 11:655.
- Harris, M. O., Friesen, T. L., Xu, S. S., Chen, M. S., Giron, D., and Stuart, J. J. 2015. Pivoting from arabidopsis to wheat to understand how agricultural plants integrate responses to biotic stress. *Journal of Experimental Botany*, 66:513–531.
- Hartl, D. L. and Clark, A. G. 1998. *Principles of population genetics*.
- Hawksworth, D., Kirk, P., Sutton, B., and Pegler, D., editors. 1995. *Ainsworth & Bisby's Dictionary of the Fungi*. 8th ed edition.
- Henikoff, S., Till, B. J., and Comai, L. 2004. TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiology*, 135:630–636.
- Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, 11:1026–1030.
- Hogenhout, S. A., Van der Hoorn, R. A., Terauchi, R., and Kamoun, S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions*, 22:115–122.

- Holland, N. T., Smith, M. T., Eskenazi, B., and Bastaki, M. 2003. Biological sample collection and processing for molecular epidemiological studies. *Mutation Research/Reviews in Mutation Research*, 543:217–234.
- Hovmøller, M. S. and Justesen, A. F. 2007a. Rates of evolution of avirulence phenotypes and DNA markers in a northwest European population of *Puccinia striiformis* f. sp. *tritici*: Clonal evolution of virulence. *Molecular Ecology*, 16: 4637–4647.
- Hovmøller, M. S., Justesen, A. F., and Brown, J. K. M. 2002. Clonality and long-distance migration of *Puccinia striiformis* f.sp. *tritici* in north-west Europe. *Plant Pathology*, 51:24–32.
- Hovmøller, M. S., Yahyaoui, A. H., Milus, E. A., and Justesen, A. F. 2008. Rapid global spread of two aggressive strains of a wheat rust fungus. *Molecular Ecology*, 17:3818–3826.
- Hovmøller, M. S., Walter, S., and Justesen, A. F. 2010. Escalating threat of wheat rusts. *Science*, 329:369–369.
- Hovmøller, M. S., Walter, S., Bayles, R. A., Hubbard, A., Flath, K., Sommerfeldt, N., Leconte, M., Czembor, P., Rodriguez-Algaba, J., Thach, T., Hansen, J. G., Lassen, P., Justesen, A. F., Ali, S., and de Vallavieille-Pope, C. 2016. Replacement of the European wheat yellow rust population by new races from the centre of diversity in the near-Himalayan region. *Plant Pathology*, 65:402–411.
- Hovmøller, M. S. and Justesen, A. F. 2007b. Appearance of atypical *Puccinia striiformis* f. sp. *tritici* phenotypes in north-western Europe. *Australian Journal of Agricultural Research*, 58:518–524.
- Huang, X., Feng, H., and Kang, Z. 2012. Selection of reference genes for quantitative real-time PCR normalization in *Puccinia striiformis* f. sp. *tritici*. *Journal of Agricultural Biotechnology*, 20:181–187.
- Hubbard, A., Pritchard, L., E, C., and S, H., 2014. United Kingdom Cereal Pathogen Virulence Survey. Annual Report. URL <https://cereals.ahdb.org.uk/media/1131354/Annual-Report-UKCPVS-2014.pdf>. [Online; accessed 20/01/2018].
- Hubbard, A., Lewis, C. M., Yoshida, K., Ramirez-Gonzalez, R. H., de Vallavieille-Pope, C., Thomas, J., Kamoun, S., Bayles, R., Uauy, C., and Saunders, D. 2015. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biology*, 16:23.
- Huerta-Espino, J., Singh, R. P., Germán, S., McCallum, B. D., Park, R. F., Chen, W. Q., Bhardwaj, S. C., and Goyeau, H. 2011. Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica*, 179:143–160.

- Huggett, J. F., Foy, C. A., Benes, V., Emslie, K., Garson, J. A., Haynes, R., Hellemans, J., Kubista, M., Mueller, R. D., and Nolan, T. 2013. The digital MIQE guidelines: minimum information for publication of quantitative digital PCR experiments. *Clinical Chemistry*, 59:892–902.
- Hussein, S. and Pretorius, Z. A. 2005. Leaf and stripe rust resistance among Ethiopian grown wheat varieties and lines. *SINET: Ethiopian Journal of Science*, 28:23–32.
- IndexMundi, 2017. South Africa wheat imports by year. URL <https://www.indexmundi.com/agriculture/?country=za&commodity=wheat&graph=imports>. [Online; accessed 20/01/2018].
- ITA USDC, 2017. South Africa - agricultural sector. URL <https://www.export.gov/article?id=South-Africa-agricultural-equipment>. [Online; accessed 20/01/2018].
- Jain, M., Nijhawan, A., Tyagi, A. K., and Khurana, J. P. 2006. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, 345:646–651.
- Jia, F., Lo, N., and Ho, S. Y. W. 2014. The impact of modelling rate heterogeneity among sites on phylogenetic estimates of intraspecific evolutionary rates and timescales. *PLoS ONE*, 9:e95722.
- Jiao, M., Tan, C., Wang, L., Guo, J., Zhang, H., Kang, Z., and Guo, J. 2017. Basidiospores of *Puccinia striiformis* f. sp. *tritici* succeed to infect barberry, while urediniospores are blocked by non-host resistance. *Protoplasma*, 254:2237–2246.
- Jin, Y. 2011. Role of *Berberis* spp. as alternate hosts in generating new races of *Puccinia graminis* and *P. striiformis*. *Euphytica*, 179:105–108.
- Jin, Y., Szabo, L. J., and Carson, M. 2010. Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* as an alternate host. *Phytopathology*, 100:432–435.
- Johnson, R. 1978. Induced resistance to fungal diseases with special reference to yellow rust of wheat. *Annals of Applied Biology*, 89:107–110.
- Johnson, R., Stubbs, R., Fuchs, E., and Chamberlain, N. 1972. Nomenclature for physiologic races of *Puccinia striiformis* infecting wheat. *Transactions of the British Mycological Society*, 58:475–480.
- Joly, D. L., Feau, N., Tanguay, P., and Hamelin, R. C. 2010. Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). *BMC Genomics*, 11:422.

- Jombart, T., Devillard, S., and Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11:94.
- Jones, J. D. and Dangl, J. L. 2006. The plant immune system. *Nature*, 444:323–329.
- Justesen, A. F., Ridout, C. J., and Hovmøller, M. S. 2002. The recent history of *Puccinia striiformis* f.sp. *tritici* in Denmark as revealed by disease incidence and AFLP markers. *Plant Pathology*, 51:13–23.
- Kamoun, S. 2007. Groovy times: filamentous pathogen effectors revealed. *Current Opinion in Plant Biology*, 10:358–365.
- Kang, Z. 2017. *Stripe rust*. New York, NY.
- Karlen, Y., McNair, A., Perseguers, S., Mazza, C., and Mermod, N. 2007. Statistical significance of quantitative PCR. *BMC Bioinformatics*, 8:131.
- Keet, J.-H., 2015. *The invasion potential of selected Berberis species in South Africa*. PhD thesis, University of the Free State.
- Kim, D., Alptekin, B., and Budak, H. 2018. CRISPR/Cas9 genome editing in wheat. *Functional & Integrative Genomics*, 18:31–41.
- Kimura, M. and Ohta, T. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61:763.
- Kiran, K., Rawal, H. C., Dubey, H., Jaswal, R., Devanna, B., Gupta, D. K., Bhardwaj, S. C., Prasad, P., Pal, D., Chhuneja, P., Balasubramanian, P., Kumar, J., Swami, M., Solanke, A. U., Gaikwad, K., Singh, N. K., and Sharma, T. R. 2016. Draft genome of the wheat rust pathogen (*Puccinia triticina*) unravels genome-wide structural variations during evolution. *Genome Biology and Evolution*, 8: 2702–2721.
- Kiran, K., Rawal, H. C., Dubey, H., Jaswal, R., Bhardwaj, S. C., Prasad, P., Pal, D., Devanna, B. N., and Sharma, T. R. 2017. Dissection of genomic features and variations of three pathotypes of *Puccinia striiformis* through whole genome sequencing. *Scientific Reports*, 7:42419.
- Kirk, P., Cannon, P., Minter, D., and Stalpers, J., editors. 2008. *Dictionary of the Fungi*. 10th ed edition.
- Klug, W. S., editor. 2012. *Concepts of Genetics*. San Francisco, 10th ed edition.
- Knott, D. 1989. *Introduction. The Wheat Rusts-Breeding for Resistance (Monograph on Theoretical and Applied Genetics, volume 12*.
- Kolmer, J. A. 2005. Tracking wheat rust on a continental scale. *Current Opinion in Plant Biology*, 8:441–449.

- Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., Ståhlberg, A., and Zoric, N. 2006. The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27:95–125.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10:R25.
- Lee, W.-S., Hammond-Kosack, K. E., and Kanyuka, K. 2012. Barley stripe mosaic virus-mediated tools for investigating gene function in cereal plants and their pathogens: virus-induced gene silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiology*, 160: 582–590.
- Lei, Y., Wang, M., Wan, A., Xia, C., See, D. R., Zhang, M., and Chen, X. 2017. Virulence and molecular characterization of experimental isolates of the stripe rust pathogen (*Puccinia striiformis*) indicate somatic recombination. *Phytopathology*, 107:329–344.
- Leonard, K. J. and Szabo, L. J. 2005. Stem rust of small grains and grasses caused by *Puccinia graminis*. *Molecular Plant Pathology*, 6:99–111.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25: 2078–2079.
- Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25.
- Li, W. H., Wu, C. I., and Luo, C. C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, 2:150–174.
- Librado, P. and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25:1451–1452.
- Ling, D., Pike, C. J., and Salvaterra, P. M. 2012. Deconvolution of the confounding variations for reverse transcription quantitative real-time polymerase chain reaction by separate analysis of biological replicate data. *Analytical Biochemistry*, 427:21–25.
- Ling, P., Wang, M., Chen, X., and Campbell, K. 2007. Construction and characterization of a full-length cDNA library for the wheat stripe rust pathogen (*Puccinia striiformis* f. sp. *tritici*). *BMC Genomics*, 8:145.

- Little, R. and Manners, J. G. 1969. Somatic recombination in yellow rust of wheat (*Puccinia striiformis*): II. Germ tube fusions, nuclear number and nuclear size. *Transactions of the British Mycological Society*, 53:251–258.
- Liu, C., Pedersen, C., Schultz-Larsen, T., Aguilar, G. B., Madriz-Ordeñana, K., Hovmøller, M. S., and Thordal-Christensen, H. 2016. The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. *New Phytologist*, pages 1–13.
- Livak, K. J. and Schmittgen, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*, 25:402–408.
- Lorrain, C., Hecker, A., and Duplessis, S. 2015. Effector-mining in the poplar rust fungus *Melampsora larici-populina* secretome. *Frontiers in Plant Science*, 6:1051.
- Lowe, I., Cantu, D., and Dubcovsky, J. 2011. Durable resistance to the wheat rusts: Integrating systems biology and traditional phenotype-based research methods to guide the deployment of resistance genes. *Euphytica*, 179:69–79.
- Ma, J., Huang, X., Wang, X., Chen, X., Qu, Z., Huang, L., and Kang, Z. 2009. Identification of expressed genes during compatible interaction between stripe rust (*Puccinia striiformis*) and wheat using a cDNA library. *BMC Genomics*, 10: 586.
- Maddison, A. C. and Manners, J. G. 1972. Sunlight and viability of cereal rust uredospores. *Transactions of the British Mycological Society*, 59:429–443.
- Malinovskiy, F. G., Fangel, J. U., and Willats, W. G. 2014. The role of the cell wall in plant immunity. *Frontiers in Plant Science*, 5:178.
- Mallard, S., Gaudet, D., Aldeia, A., Abelard, C., Besnard, A. L., Sourdille, P., and Dedryver, F. 2005. Genetic analysis of durable resistance to yellow rust in bread wheat. *Theoretical and Applied Genetics*, 110:1401–1409.
- Mandiyan, V., Andreev, J., Schlessinger, J., and Hubbard, S. R. 1999. Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-associated protein β . *The EMBO Journal*, 18:6890–6898.
- Mao, F., Leung, W.-Y., and Xin, X. 2007. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnology*, 7:76.
- Markell, S. and Milus, E. 2008. Emergence of a novel population of *Puccinia striiformis* f. sp. *tritici* in eastern United States. *Phytopathology*, 98:632–639.
- Mboup, M., Leconte, M., Gautier, A., Wan, A., Chen, W., de Vallavieille-Pope, C., and Enjalbert, J. 2009. Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genetics and Biology*, 46: 299–307.

- McDonald, B. A. 2004. Population genetics of plant pathogens. *The Plant Health Instructor*. URL <http://www.apsnet.org/edcenter/advanced/topics/PopGenetics/Pages/default.aspx>. [Online; accessed 20/01/2018].
- McDonald, B. A. and Linde, C. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40:349–379.
- McDonald, J. H. and Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351:652.
- McIntosh, R. A. A catalogue of gene symbols for wheat. In *Proceedings of the 6th International Wheat Genetics Symposium, Kyoto, Japan, 1983*.
- McIntosh, R. A., Wellings, C. R., and Park, R. F. 1995. *Wheat rusts: an atlas of resistance genes*. Melbourne.
- Mehta, D., Menke, A., and Binder, E. B. 2010. Gene expression studies in major depression. *Current Psychiatry Reports*, 12:135–144.
- Mendgen, K., Struck, C., Voegelé, R. T., and Hahn, M. 2000. Biotrophy and rust haustoria. *Physiological and Molecular Plant Pathology*, 56:141–145.
- Milus, E., Seyran, E., and McNew, R. 2006. Aggressiveness of *Puccinia striiformis* f. sp. *tritici* isolates in the south-central United States. *Plant Disease*, 90:847–852.
- Milus, E. A., Kristensen, K., and Hovmøller, M. S. 2009. Evidence for increased aggressiveness in a recent widespread strain of *Puccinia striiformis* f. sp. *tritici* causing stripe rust of wheat. *Phytopathology*, 99:89–94.
- Miyata, T. and Yasunaga, T. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution*, 16:23–36.
- Moldenhauer, J., Moerschbacher, B. M., and van der Westhuizen, A. J. 2006. Histological investigation of stripe rust (*Puccinia striiformis* f.sp. *tritici*) development in resistant and susceptible wheat cultivars. *Plant Pathology*, 55:469–474.
- Murphy, C. L. and Polak, J. M. 2002. Differentiating embryonic stem cells: GAPDH, but neither HPRT nor β -tubulin is suitable as an internal standard for measuring RNA levels. *Tissue Engineering*, 8:551–559.
- Naccache, S. N., Federman, S., Veeraraghavan, N., Zaharia, M., Lee, D., Samayoa, E., Bouquet, J., Greninger, A. L., Luk, K.-C., Enge, B., Wadford, D. A., Messenger, S. L., Genrich, G. L., Pellegrino, K., Grard, G., Leroy, E., Schneider, B. S., Fair, J. N., Martinez, M. A., Isa, P., Crump, J. A., DeRisi, J. L., Sittler, T., Hackett, J., Miller, S., and Chiu, C. Y. 2014. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Research*, 24:1180–1192.

- Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3:418–426.
- Niks, R. E. 1989. Morphology of infection structures of *Puccinia striiformis* var. *dactylidis*. *European Journal of Plant Pathology*, 95:171–175.
- Oerke, E.-C. and Dehne, H.-W. 2004. Safeguarding production—losses in major crops and the role of crop protection. *Crop Protection*, 23:275–285.
- Olsen, O., Wang, X., and von Wettstein, D. 1993. Sodium azide mutagenesis: preferential generation of AT→GC transitions in the barley Ant18 gene. *Proceedings of the National Academy of Sciences*, 90:8043–8047.
- Panstruga, R. and Dodds, P. N. 2009. Terrific protein traffic: The mystery of effector protein delivery by filamentous plant pathogens. *Science*, 324:748–750.
- Panwar, V. and Bakkeren, G. 2017. Investigating gene function in cereal rust fungi by plant-mediated virus-induced gene silencing. In *Wheat Rust Diseases*, volume 1659, pages 115–124. New York, NY.
- Parker, I. M. and Gilbert, G. S. 2004. The evolutionary ecology of novel plant-pathogen interactions. *Annual Review of Ecology, Evolution, and Systematics*, 35: 675–700.
- Parlevliet, J. E. 2002. Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica*, 124:147–156.
- Persoons, A., Morin, E., Delaruelle, C., Payen, T., Halkett, F., Frey, P., De Mita, S., and Duplessis, S. 2014. Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in Plant Science*, 5:450.
- Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., and Kamoun, S. 2015. Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Molecular Plant-Microbe Interactions*, 28:689–700.
- Petre, B., Lorrain, C., Saunders, D. G., Win, J., Sklenar, J., Duplessis, S., and Kamoun, S. 2016a. Rust fungal effectors mimic host transit peptides to translocate into chloroplasts: Effectors use molecular mimicry to target chloroplasts. *Cellular Microbiology*, 18:453–465.
- Petre, B., Saunders, D. G. O., Sklenar, J., Lorrain, C., Krasileva, K. V., Win, J., Duplessis, S., and Kamoun, S. 2016b. Heterologous expression screens in *Nicotiana benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLoS ONE*, 11:e0149035.
- Pfaffl, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29:e45–e45.

- Pfeifer, G., You, Y., and Besaratinia, A. 2005. Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571:19–31.
- Pretorius, Z. A., Boshoff, W. H. P., and Kema, G. H. J. 1997. First report of *Puccinia striiformis* f. sp. *tritici* on wheat in South Africa. *Plant Disease*, 81:424–424.
- Pretorius, Z. A., Pakendorf, K. W., Marais, G. F., Prins, R., and Komen, J. S. 2007. Challenges for sustainable cereal rust control in South Africa. *Australian Journal of Agricultural Research*, 58:593.
- Pretorius, Z., Bender, C., and Visser, B. 2015. The rusts of wild rye in South Africa. *South African Journal of Botany*, 96:94–98.
- Prins, R. and Agenbag, G., 2013. The establishment of a molecular service laboratory for wheat breeding in south africa. Poster presentation: 12th International Wheat Genetics Symposium, Yokohama, Japan.
- Prins, R., Pretorius, Z. A., Bender, C. M., and Lehmensiek, A. 2011. QTL mapping of stripe, leaf and stem rust resistance genes in a Kariëga × Avocet S doubled haploid wheat population. *Molecular Breeding*, 27:259–270.
- Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959.
- Pryce-Jones, E., Carver, T. I. M., and Gurr, S. J. 1999. The roles of cellulase enzymes and mechanical force in host penetration by *Erysiphe graminis* f. sp. *hordei*. *Physiological and Molecular Plant Pathology*, 55:175–182.
- Quinlan, A. R. and Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26:841–842.
- Rambaut, A. and Grass, N. C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*, 13:235–238.
- Ramburan, V. P., Pretorius, Z. A., Louw, J. H., Boyd, L. A., Smith, P. H., Boshoff, W. H. P., and Prins, R. 2004. A genetic analysis of adult plant resistance to stripe rust in the wheat cultivar Kariëga. *Theoretical and Applied Genetics*, 108: 1426–1433.
- Rao, H. S. and Sears, E. 1964. Chemical mutagenesis in *Triticum aestivum*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1:387–399.
- Rapilly, F. 1979. Yellow rust epidemiology. *Annual Review of Phytopathology*, 17: 59–73.
- Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. 2013. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, 8:e66428.

- Rodriguez-Algaba, J., Walter, S., Sørensen, C. K., Hovmøller, M. S., and Justesen, A. F. 2014. Sexual structures and recombination of the wheat rust fungus *Puccinia striiformis* on *Berberis vulgaris*. *Fungal Genetics and Biology*, 70:77–85.
- Roelfs, A. P., Singh, R. P., and Saari, E. E. 1992. *Rust Diseases of Wheat: Concepts and Methods of Disease Management*.
- Roelfs, A. P. and Hettel, G. 1992. *Rust diseases of wheat*.
- Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8:103–106.
- Rovenich, H., Boshoven, J. C., and Thomma, B. P. 2014. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Current Opinion in Plant Biology*, 20:96–103.
- Ruijter, J. M., Pfaffl, M. W., Zhao, S., Spiess, A. N., Boggy, G., Blom, J., Rutledge, R. G., Sisti, D., Lievens, A., and De Preter, K. 2013. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods*, 59:32–46.
- Rutledge, R. G. and Cote, C. 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Research*, 31:e93–e93.
- SAGL, 2012. The Southern African Grain Laboratory NPC: South African Winter Cereal Production. URL <http://www.sagl.co.za/Portals/0/Wheat%20crop%202011%202012/Average%20yield%20per%20province.pdf>. [Online; accessed 20/01/2018].
- Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., Anderson, N., Soto, M. F. D., Rouse, M., Szabo, L., Bowden, R. L., Dubcovsky, J., and Akhunov, E. 2017. Variation in the *AvrSr35* gene determines *Sr35* resistance against wheat stem rust race Ug99. *Science*, 358:1604–1606.
- Salemi, M., Vandamme, A.-M., and Lemey, P. 2009. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*.
- Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., and Raffaele, S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE*, 7:e29847.
- Scally, A. 2016. The mutation rate in human evolution and demographic inference. *Current Opinion in Genetics & Development*, 41:36–43.
- Schlötterer, C. 2004. The evolution of molecular markers—just a matter of fashion? *Nature Reviews Genetics*, 5:63–69.
- Schmidt, G. W. and Delaney, S. K. 2010. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Molecular Genetics and Genomics*, 283:233–241.

- Schmittgen, T. D. and Livak, K. J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3:1101–1108.
- Schumann, G. L. and Leonard, K. 2000. Stem rust of wheat (black rust). *The Plant Health Instructor*. URL <https://www.apsnet.org/edcenter/intropp/lessons/fungi/Basidiomycetes/Pages/StemRust.aspx>.
- Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., Dodds, P. N., Figueroa, M., Park, R. F., and Rathjen, J. P. 2018. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. tritici reveals high interhaplotype diversity. *mBio*, 9:e02275–17.
- Selitrennikoff, C. P. 2001. Antifungal proteins. *Applied and Environmental Microbiology*, 67:2883–2894.
- Sharma, I. 2012. *Disease resistance in wheat*, volume 1.
- Sharma-Poudyal, D., Chen, X. M., Wan, A. M., Zhan, G. M., Kang, Z. S., Cao, S. Q., Jin, S. L., Morgounov, A., Akin, B., and Mert, Z. 2013. Virulence characterization of international collections of the wheat stripe rust pathogen, *Puccinia striiformis* f. sp. tritici. *Plant Disease*, 97:379–386.
- Sharp, E. L. 1967. Atmospheric ions and germination of uredospores of *Puccinia striiformis*. *Science*, 156:1359–1360.
- Shaw, M. W. and Osborne, T. M. 2011. Geographic distribution of plant pathogens in response to climate change. *Plant Pathology*, 60:31–43.
- Shiferaw, B., Kassie, M., Jaleta, M., and Yirga, C. 2014. Adoption of improved wheat varieties and impacts on household food security in Ethiopia. *Food Policy*, 44:272–284.
- Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., Lawlor, R. T., and Scarpa, A. 2013. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS ONE*, 8:1–8.
- Simmonds, N. W. 1991. Genetics of horizontal resistance to diseases of crops. *Biological Reviews*, 66:189–241.
- Smit, H., Tolmay, V., Barnard, A., Jordaan, J., Koekemoer, F., Otto, W., Pretorius, Z., Purchase, J., and Tolmay, J. 2010. An overview of the context and scope of wheat (*Triticum aestivum*) research in South Africa from 1983 to 2008. *South African Journal of Plant and Soil*, 27:81–96.
- Speed, T. 2004. Statistics and gene expression analysis. *Biostatistical Genetics and Genetic Epidemiology*, pages 1–13.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30:1312–1313.

- Steele, K. A., Humphreys, E., Wellings, C. R., and Dickinson, M. J. 2001. Support for a stepwise mutation model for pathogen evolution in Australasian *Puccinia striiformis* f.sp. *tritici* by use of molecular markers. *Plant Pathology*, 50:174–180.
- Stergiopoulos, I. and de Wit, P. J. 2009. Fungal effector proteins. *Annual Review of Phytopathology*, 47:233–263.
- Stotz, H. U., Mitrousia, G. K., de Wit, P. J., and Fitt, B. D. 2014. Effector-triggered defence against apoplastic fungal pathogens. *Trends in Plant Science*, 19:491–500.
- Stubbs, R. W. 1988. *Pathogenicity analysis of yellow (stripe) rust of wheat and its significance in a global context*.
- Stubbs, R. 1985. Stripe rust. In *Diseases, Distribution, Epidemiology, and Control*, pages 61–101.
- Szabo, L. J. and Bushnell, W. R. 2001. Hidden robbers: the role of fungal haustoria in parasitism of plants. *Proceedings of the National Academy of Sciences of the United States of America*, 98:7654–7655.
- Sørensen, C. K., Justesen, A. F., and Hovmøller, M. S. 2012. 3-D imaging of temporal and spatial development of *Puccinia striiformis* haustoria in wheat. *Mycologia*, 104:1381–1389.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30:2725–2729.
- Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., and Nguyen, M. 2010. A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines. *Methods*, 50:S1–S5.
- Taylor, S. C. and Mrkusich, E. M. 2014. The state of RT-quantitative PCR: firsthand observations of implementation of minimum information for the publication of quantitative real-time PCR experiments (MIQE). *Journal of Molecular Microbiology and Biotechnology*, 24:46–52.
- Thach, T., Ali, S., Justesen, A., Rodriguez-Algaba, J., and Hovmøller, M. 2015. Recovery and virulence phenotyping of the historic ‘Stubbs collection’ of the yellow rust fungus *Puccinia striiformis* from wheat: Long-term storage of rust fungi. *Annals of Applied Biology*, 167:314–326.
- Thach, T., Ali, S., de Vallavieille-Pope, C., Justesen, A., and Hovmøller, M. 2016. Worldwide population structure of the wheat rust fungus *Puccinia striiformis* in the past. *Fungal Genetics and Biology*, 87:1–8.
- Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. 1999. Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology*, 75:291–295.

- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14:178–192.
- Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E., and Rubin, G. M. 2007. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biology*, 8:R145.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7:562–578.
- United Nations. World population prospects: The 2017 revision, key findings and advance tables. Technical report, United Nations, Department of Economic and Social Affairs, Population Division, 2017.
- Upadhyaya, N. M., Mago, R., Staskawicz, B. J., Ayliffe, M. A., Ellis, J. G., and Dodds, P. N. 2013. A bacterial type III secretion assay for delivery of fungal effector proteins into wheat. *Molecular Plant-Microbe Interactions*, 27:255–264.
- van der Hoorn, R. A. and Kamoun, S. 2008. From guard to decoy: A new model for perception of plant pathogen effectors. *The Plant Cell Online*, 20:2009–2017.
- Van der Plank, J. 1968. *Disease resistance in plants*.
- Van Niekerk, H. 2001. Southern Africa wheat pool. In *The World Wheat Book: The History of Wheat Breeding*.
- VanGuilder, H. D., Vrana, K. E., and Freeman, W. M. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, 44:619.
- Vieira, M. L. C., Santini, L., Diniz, A. L., and Munhoz, C. d. F. 2016. Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology*, 39:312–328.
- Visser, B., Herselman, L., and Pretorius, Z. A. 2016. Microsatellite characterisation of South African *Puccinia striiformis* races. *South African Journal of Plant and Soil*, 33:161–166.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J., and Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23:4407–4414.
- Wahl, I., Anikster, Y., Manisterski, J., and Segal, A. 1984. *Evolution at the center of origin*, volume 1.
- Walter, S., Ali, S., Kemen, E., Nazari, K., Bahri, B. A., Enjalbert, J., Hansen, J. G., Brown, J. K., Sicheritz-Pontén, T., Jones, J., de Vallavieille-Pope, C., Hovmøller, M. S., and Justesen, A. F. 2016. Molecular markers for tracking the origin and

- worldwide distribution of invasive strains of *Puccinia striiformis*. *Ecology and Evolution*, 6:2790–2804.
- Wang, B., Sun, Y., Song, N., Zhao, M., Liu, R., Feng, H., Wang, X., and Kang, Z. 2017. *Puccinia striiformis* f. sp. *tritici* microRNA-like RNA 1 (*Pst*-milR1), an important pathogenicity factor of *Pst*, impairs wheat resistance to *Pst* by suppressing the wheat pathogenesis-related 2 gene. *New Phytologist*, 215: 338–350.
- Wang, C.-F., Huang, L.-L., Buchenauer, H., Han, Q.-M., Zhang, H.-C., and Kang, Z.-S. 2007. Histochemical studies on the accumulation of reactive oxygen species (O_2^- and H_2O_2) in the incompatible and compatible interaction of wheat—*Puccinia striiformis* f. sp. *tritici*. *Physiological and Molecular Plant Pathology*, 71:230–239.
- Wang, M. and Chen, X. 2013. First report of oregon grape (*Mahonia aquifolium*) as an alternate host for the wheat stripe rust pathogen (*Puccinia striiformis* f. sp. *tritici*) under artificial inoculation. *Plant Disease*, 97:839–839.
- Wang, X., Tang, C., Zhang, G., Li, Y., Wang, C., Liu, B., Qu, Z., Zhao, J., Han, Q., Huang, L., Chen, X., and Kang, Z. 2009. cDNA-AFLP analysis reveals differential gene expression in compatible interaction of wheat challenged with *Puccinia striiformis* f. sp. *tritici*. *BMC Genomics*, 10:289.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25:1189–1191.
- Wellings, C. R. 2007. *Puccinia striiformis* in Australia: a review of the incursion, evolution, and adaptation of stripe rust in the period 1979–2006. *Australian Journal of Agricultural Research*, 58:567.
- Wellings, C. R., McIntosh, R. A., and Walker, J. 1987. *Puccinia striiformis* f. sp. *tritici* in Eastern Australia possible means of entry and implications for plant quarantine. *Plant Pathology*, 36:239–241.
- Wellings, C. R., McIntosh, R. A., and Hussain, M. 1988. A new source of resistance to *Puccinia striiformis* f. sp. *tritici* in spring wheats (*Triticum aestivum*). *Plant Breeding*, 100:88–96.
- Wellings, C. R. 2011. Global status of stripe rust: a review of historical and current threats. *Euphytica*, 179:129–141.
- Willems, E., Leyns, L., and Vandesompele, J. 2008. Standardization of real-time PCR gene expression data from independent biological replicates. *Analytical Biochemistry*, 379:127–129.
- Winter, B. 2013. Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.

- Wittwer, C. T., Herrmann, M. G., Moss, A. A., and Rasmussen, R. P. 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, 22:130–139.
- Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24:1586–1591.
- Yang, Z. and Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, 17:32–43.
- Yin, C. and Hulbert, S. 2015. Host induced gene silencing (HIGS), a promising strategy for developing disease resistant crops. *Gene Technology*, 04:130.
- Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., Kamoun, S., and Terauchi, R. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *magnaporthe oryzae*. *The Plant Cell*, 21:1573–1591.
- Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., and Burbano, H. A. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife*, 2.
- Yuan, J. S., Reed, A., Chen, F., and Stewart, C. N. 2006. Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7:85.
- Zadoks, J. C. 1961. Yellow rust on wheat studies in epidemiology and physiologic specialization. *European Journal of Plant Pathology*, 67:69–256.
- Zadoks, J., Chang, T., and Konzak, C. 1974. A decimal code for the growth stages of cereals. *Weed research*, 14:415–421.
- Zhang, Y., Qu, Z., Zheng, W., Liu, B., Wang, X., Xue, X., Xu, L., Huang, L., Han, Q., Zhao, J., and Kang, Z. 2008. Stage-specific gene expression during urediniospore germination in *Puccinia striiformis* f. sp. *tritici*. *BMC Genomics*, 9: 203.
- Zhao, J., Zhang, H., Yao, J., Huang, L., and Kang, Z. 2011. Confirmation of *Berberis* spp. as alternate hosts of *Puccinia striiformis* f. sp. *tritici* on wheat in China. *Mycosystema*, 30:895–900.
- Zhao, J., Wang, L., Wang, Z., Chen, X., Zhang, H., Yao, J., Zhan, G., Chen, W., Huang, L., and Kang, Z. 2013. Identification of eighteen *Berberis* species as alternate hosts of *Puccinia striiformis* f. sp. *tritici* and virulence variation in the pathogen isolates from natural infection of barberry plants in China. *Phytopathology*, 103:927–934.

Zheng, W., Huang, L., Huang, J., Wang, X., Chen, X., Zhao, J., Guo, J., Zhuang, H., Qiu, C., Liu, J., Liu, H., Huang, X., Pei, G., Zhan, G., Tang, C., Cheng, Y., Liu, M., Zhang, J., Zhao, Z., Zhang, S., Han, Q., Han, D., Zhang, H., Zhao, J., Gao, X., Wang, J., Ni, P., Dong, W., Yang, L., Yang, H., Xu, J.-R., Zhang, G., and Kang, Z. 2013. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications*, 4:2673.

Zillinsky, F. J. 1983. *Common Diseases of Small Grain Cereals. A Guide to Identification*.