

**Bioinformatic genome analysis of  
the necrotrophic wheat-  
pathogenic fungus *Phaeosphaeria  
nodorum* and related  
Dothideomycete fungi.**

**James Kyawzwar Hane**

**Bachelor of Molecular Biology (Hons.), University of Western Australia**

**This thesis is presented for the degree of Doctor of  
Philosophy**

**2011**

I declare that this thesis is my own account of my own work  
and contains as its main content work which has not been  
previously submitted for any degree at any tertiary institution.

.....

James Kyawzwar Hane

It is a truth universally acknowledged that a single Ph. D. student in possession of a whole genome sequence must be in want of a stiff drink. Of all the genomes in all the fungi in all the world, *ToxA* horizontally transfers out of mine. To lose 11 kilobases may be regarded as a misfortune; to lose 280 kilobases looks like carelessness! It was the best of times, it was the worst of times; it was the age of genomics, it was the age of bioinformatics; it was the epoch of next-generation sequencing, it was the epoch of data analysis; it was the season of transcriptomics, it was the season of proteogenomics; it was the spring of mesosyteny, it was the winter of repeat-induced point mutation; we had everything before us, we had nothing before us; we were all going directly to Curtin, but I was going the other way. All this happened, more or less.

In my younger and more vulnerable years my supervisor gave me some advice that I've been turning over in my mind ever since: "Everything's got a moral, if only you can find it". I don't believe there's an atom of meaning in it. It would be so nice if something made sense for a change. Sometimes I've believed as many as six impossible things before breakfast. Granted: I am a student of a research institution; my supervisor is watching me, he never lets me out of his sight; there's a peephole in the door, and my supervisor's expertise is the shade of biochemistry that can never see through a bioinformatics type like me.

Oh! it is absurd to have a hard and fast rule about what one should read and what one shouldn't. More than half of scientific literature depends on what one shouldn't read. The truth is rarely pure and never simple. Scientific research would be very tedious if it were either, and scientific literature a complete impossibility! I have always been of opinion that a man who desires to get a Ph. D. should know either everything or nothing. I do not approve of anything that tampers with natural ignorance. Ignorance is like a delicate exotic fruit; touch it and the bloom is gone. The whole theory of modern education is radically unsound. Fortunately in Murdoch, at any rate, education produces no effect whatsoever. If it did, it would prove a serious danger to the upper management, and probably lead to acts of violence in Bush Court.

Of all the things that drive men to research, the most common disaster, I've come to learn, is women. Here's looking at you, kid. We'll always have Perth. You must remember this - maybe not today, maybe not tomorrow, but soon and for the rest of your life.

I have never begun a thesis with more misgiving. I write this sitting in the kitchen sink. I don't write accurately - anyone can write accurately - but I write with wonderful expression. Read the directions and directly you will be directed in the right direction. When I use a word, it means just what I choose it to mean - neither more nor less. I never travel without my thesis. One should always have something sensational to read in the train. Whether I shall turn out to be the hero of my own life, or whether that station will be held by anybody else, these pages must show.

## **Abstract**

*Phaeosphaeria nodorum* (anamorph: *Stagonospora nodorum*) is the causal agent of *Stagonospora nodorum* blotch (SNB, syn. glume blotch) in wheat. *P. nodorum* is estimated to cause up to 31% wheat yield loss worldwide. Within Australia it is the primary pathogen of wheat and is estimated to cause losses of \$108 million per annum. The genome assembly of *P. nodorum* was sequenced in 2005 and was the first species in the class Dothideomycetes, a significant fungal taxon containing several major phytopathogens, to be publically released. The *P. nodorum* genome database has since evolved from basic sequence data into a powerful resource for studying the SNB host-pathogen interaction and advancing the understanding of fungal genome structure. The genes of *P. nodorum* have been annotated to a high level of accuracy and now serve as a model dataset for comparative purposes. *P. nodorum* gene annotations have been refined by a combination of several techniques including manual curation, orthology with related species, expressed sequence tag (EST) alignment, and proteogenomics. Analysis of the repetitive DNA in the *P. nodorum* genome lead to the development of software for the analysis of repeat-induced point mutation (RIP), a fungal-specific genome defence mechanism, which was a major improvement upon previous methods. Comparative genomics between *P. nodorum* and related species has highlighted a novel pattern of genome sequence conservation between filamentous fungi called 'mesosynteny' and has lead to the development of novel 'genome finishing' strategies.

## **Acknowledgements**

First and foremost, I would like to express my deep gratitude to my principal supervisor Richard Oliver, whose guidance has been invaluable. Richard provides his students with exceptional opportunities to develop and grow. Many thanks also to my supervisors Barbara Howlett and Thierry Rouxel who have been extremely helpful with constructive criticism and expert advice.

I am indebted to the GRDC, without which I would not have had the financial means to pursue a scientific career, and to iVEC and the NCI which provide excellent high-performance computing facilities for research purposes at no charge. I would also like to express my gratitude to Karam Singh, who has recently hired me to work at CSIRO.

I would like to thank my friends and colleagues at the ACNFP for their professional support, expertise and friendship – I learnt a lot from all of you. KC, Zippy, Rohan, Peter, Lars, Kasia, Ormonde, Joel, Christian and Eva – my waistline will miss our squash games. Rob, Maryn, Ramisah, Francis, Judith, Julie, Madeline, Emma, Nola, Megan and Simon – I enjoyed working with you.

Many thanks to my father Zor and step-mother Teeni, my mother Kimberley, my brothers Michael and Alex and my grandmother Gwenda, who have been supportive and understanding while I was finishing my Ph. D. Last but not least, I would like to thank Angela from the bottom of my heart for her support and for not letting me give up.

# Table of contents

<b>Chapter 1: Introduction</b>	1-13
<b>Chapter 2: Dothideomycete-Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the Wheat Pathogen <i>Stagonospora nodorum</i></b>	14-38
Attribution statement	14
Abstract	17
Introduction	17
Results	18
Acquisition and analysis of the genome sequence	18
Analysis of repetitive elements	18
Mitochondrial genome	19
Gene content	19
Gene expression during infection	20
Discussion	21
Phylogenetic analysis	21
Repeated elements in the nuclear genome	21
The mitochondrial genome	21
Functional analysis of proteins	23
Gene expression during infection	29
Genome architecture	31
Methods	33
Fungal strains	33
Phylogenetic analysis	33
Repetitive elements	33
Gene content	33
EST library construction	34
Quantitative PCR	34
Accession number	34
References	34
<b>Chapter 3: RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences</b>	39-74
Attribution statement	41
Abstract	40
Background	41
Results	42
Validation of RIP detection by the alignment-based method	42
Identification of the dominant CpN to TpN di-nucleotide mutation in RIP-affected sequences	43
Di-nucleotide frequency and index analysis of RIP mutation in <i>Stagonospora nodorum</i>	43
Alignment-based analysis of RIP mutation in <i>Stagonospora nodorum</i>	43
Discussion	46
Conclusion	48
Methods	48
RIP-index analysis	48
RIP-index sequence scan	50
Alignment-based analysis	50
Validation of alignment-based RIP analysis	50

RIP analysis of <i>S. nodorum</i> <i>de novo</i> repeat families	50
Time of operation	51
References	51
Appendix 3A: RIPCAL/deRIP manual	53
Appendix 3B: Response to thesis examination comments	64
<b>Chapter 4: <i>In silico</i> reversal of repeat-induced point mutation (RIP) identifies the origins of repeat families and uncovers obscured duplicated genes</b>	<b>75-98</b>
Attribution statement	75
Abstract	77
Background	77
Results	79
Validating the deRIP process using known non-RIP affected repeats	79
Determining the roles and origin of RIP-degraded repeats in <i>S. nodorum</i>	82
Discussion	88
Conclusions	90
Methods	90
Analysis of RIP-mutation of <i>S. nodorum</i> repetitive DNA	90
Predicting the original repeat sequence prior to RIP-degradation	90
Validating the deRIP technique	91
Predicting the origin of RIP-degraded repeats	91
References	91
Appendix 4A: Response to thesis examination comments	93
<b>Chapter 5: Evolution of Linked Avirulence Effectors in <i>Leptosphaeria maculans</i> Is Affected by Genomic Environment and Exposure to Resistance Genes in Host Plants</b>	<b>99-116</b>
Attribution statement	99
Abstract	101
Introduction	101
Results	102
Virulence of Australian isolates of <i>L. maculans</i> and analysis of mutations at <i>AvrLm1</i> and <i>AvrLm6</i>	102
Characteristics of the genomic environment of <i>AvrLm1</i> and <i>AvrLm6</i>	105
Distribution and extent of RIP mutations	106
Transcription of <i>AvrLm1</i> , <i>AvrLm6</i> , <i>LmCys1</i> and <i>LmCys2</i> alleles	106
Rate of mutations of genes	107
Evolution and phylogenetic relationships of RIP and deletion alleles	107
Discussion	108
Selection pressure on frequencies of <i>AvrLm1</i> and <i>AvrLm6</i> alleles is imposed by extensive sowing of cultivars with <i>Rlm1</i> resistance	108
Repeat-induced point mutations may be ‘leaking’ from adjacent inactivated transposable elements into single copy regions	110
Deletion alleles of <i>AvrLm1</i> and <i>AvrLm6</i> had multiple origins whilst RIP alleles had a single origin	112
Deletions, non-synonymous and RIP mutations can confer virulence	112
Methods	112
<i>Brassica</i> cultivars and <i>Leptosphaeria maculans</i> isolates	112
Virulence testing	112
Gene identification and primer design	113
Allele genotyping	113

Expression analysis	113
Phylogenetic analysis	113
References	114
Appendix 5A: Response to thesis examination comments	116
<b>Chapter 6: Effector diversification within compartments of the <i>Leptosphaeria maculans</i> genome affected by Repeat-Induced Point Mutations</b>	<b>117-128</b>
Attribution statement	117
Abstract	119
Introduction	120
Results	120
General features of the <i>L. maculans</i> genome	120
The TEs are RIP-affected	120
The compartmentalized genome of <i>L. maculans</i>	120
The ribosomal DNA is extensively affected by RIP	122
AT-blocks as niches for effectors	122
History of genome invasion by transposable elements	123
Discussion	124
Methods	126
Phylogenetic analysis	126
Sequencing and assembly	126
<i>L. maculans</i> genome annotation	126
Annotation and analysis of repeated elements	126
RIP and deRIP analysis	126
Analysis of AT-blocks	127
Identification of small secreted proteins (SSPs)	127
References	127
<b>Chapter 7: Deep proteogenomics; high throughput gene validation by multidimensional liquid chromatography and mass spectrometry of proteins from the fungal wheat pathogen <i>Stagonospora nodorum</i></b>	<b>129-139</b>
Attribution statement	129
Abstract	131
Background	132
Methods	132
Growth and maintenance of <i>Stagonospora nodorum</i>	132
Protein extraction	133
Sample preparation	133
Strong cation exchange chromatography	133
Reverse phase nano LC MALDI-MS/MS	133
Data analysis	133
Characterisation of peptide-supported genes	133
<i>De novo</i> proteogenomics	134
Results and Discussion	134
Gene models confirmed by Mascot analysis of the existing gene model database	134
Gene models confirmed by Mascot analysis of the 6-frame translated genome sequence	136
Gene models identified by Mascot analysis of the 6 frame translated un-assembled read database	137
Sifting new and modified gene models by homology criteria	137
Conclusion	138



References	139
<b>Chapter 8: A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi</b>	140-160
Attribution statement	140
Abstract	142
Background	142
Results	145
Taxonomic distribution of mesosyteny across the fungal kingdom	147
Conclusions	152
Materials & Methods	154
Whole-genome comparisons	154
Determination of significant sequence conservation	154
Analysis of syntenic regions between conserved sequences	154
Classification of syteny type	154
References	155
Appendix 8A: Response to thesis examination comments	158
<b>Chapter 9: Finished Genome of the Fungal Wheat Pathogen <i>Mycosphaerella graminicola</i> Reveals Dispensome Structure, Chromosome Plasticity and Stealth Pathogenesis</b>	161-180
Attribution statement	161
Abstract	163
Introduction	164
Results	164
Features of the finished genome	164
Sexual activation of chromosome plasticity and repeat-induced point mutation	164
Core and dispensome chromosomes are highly different	165
A new type of syteny	167
A mechanism of stealth pathogenesis	167
Discussion	169
Materials and Methods	174
Biological material	174
Initial sequencing and assembly	175
Gap closure and finishing	175
Genome annotation	175
Microarray analysis	175
Principal component analyses of core and dispensable chromosomes	175
Mesosyteny	176
Cazy annotation and growth profiling	176
Genome structure analyses	176
References	178
Appendix 9A: Response to thesis examination comments	180
<b>Chapter 10: Genomic and comparative analysis of the class Dothideomycetes</b>	181-207
Attribution statement	181
Introduction	183
Significant phytopathogenic species	183
Order Pleosporales	183
Order Capnodiales	185
Dothideomycetidae of uncertain phylogenetic placement	185

Genome sequencing in the Dothideomycetes	185
<i>Phaeosphaeria nodorum</i>	185
<i>Mycosphaerella graminicola</i>	191
<i>Leptosphaeria maculans</i>	191
Comparative Genomics	192
Synteny	192
Completion of draft genome assemblies using mesosyntenic predictions	195
Mitochondrial genomes	197
Repetitive DNA	199
Repeat-induced point mutation	200
Pathogenicity	201
Mechanisms of pathogenicity	201
Lateral gene transfer	202
Generation of diversity via RIP	203
Conclusions	203
References	204
<b>Chapter 11: Conclusion</b>	<b>208-217</b>
<b>Supplementary data</b>	<b>CD-ROM</b>