Title: Molecular Genetic and Genomic Characterization of a Mycotoxigenic Emerging Pathogen \textit{Fusarium proliferatum}

Name: Bandar Almiman

This is a digitised version of a dissertation submitted to the University of Bedfordshire.

It is available to view only.

This item is subject to copyright.
Molecular Genetic and Genomic Characterization of a Mycotoxigenic Emerging Pathogen *Fusarium proliferatum*

B. F. Almiman
Ph.D

2017
UNIVERSITY OF BEDFORDSHIRE
Molecular Genetic and Genomic Characterization of a Mycotoxigenic Emerging Pathogen *Fusarium proliferatum*

by

B. F. Almiman

A thesis submitted to the University of Bedfordshire
In partial fulfil of the requirements for the degree of Doctor of Philosophy

April 2017
This aim of this research was to elucidate the genotypic diversity of the mycotoxigenic species *Fusarium proliferatum* associated with diverse hosts and distributed in wide geographic locations to gain new insights into the biology of this emerging pathogen. This study developed a novel molecular genetic marker FG1056. Multilocus typing of *F. proliferatum* isolates (52) using *F. verticillioides* (2) and *F. oxysporum* (3) as references was carried out with FG1056 and a set of known genetic markers (ITS, TEF1, CAL and FUM1). This distinguished up to 10 genetic groups, 2 clusters and 23 haplotypes among the *F. proliferatum* isolates. FG1056 marker showed the highest number of SNPs (169), informative sites (89) and haplotypes (23) relative to other markers used and was comparable to the multilocus typing. Varying patterns of relationships were observed between isolates represented in the genetic groups and their host and geographic origin. Considerable biological variability was recorded among the *F. proliferatum* isolates in morphology, growth, sporulation and most notably fumonisin production (up to 140-fold differences) with reference to variable temperature, water activity and duration. *De novo* genome assemblies with the size ranging from 43.96 - 50 Mb have been developed for four diverse *F. proliferatum* isolates. *In silico* analysis led to the identification of 12,980 genes common to all isolates and up to 134 genes potentially unique to an isolate. Using these resources, FUM gene cluster (~45.3 Kb) was identified for the first time in *F. proliferatum*. Order and orientation of the 16 FUM genes and the complete flanking genes (MSF1 and ZCB1 at 5’; ANK1 and GAT1 at 3’) have been determined. This study has provided new insights into the genetic and biological diversity of *F. proliferatum* and also developed new genetic and genomic resources, which will serve as a solid platform for further research particularly to understand the regulation of fumonisins production in the laboratory and in the field.
Acknowledgments

First and foremost, I would like to express my gratitude to my director of study, Professor S. Sreenivasaprasad who has been involved in this work from the beginning to this time of completion for the guidance, support and constructive criticisms. I would also like to appreciate the efforts of my other supervisory team Dr Shaobo and Dr N. Worsfold for their continual encouragement and academic support when required. In addition to my supervisors, Dr S. Muthumeenakshi has willingly and voluntarily given this project much time and thought that led to the expansion and successful submission of the reports. I would like to also appreciate Dr Riccardo Baroncelli and Dr Steven for their good heart and time towards the data analysis.

I also wish to acknowledge Professor Lukasz Stephen at the Polish Academy, Poland, Professor Naresh Magan at the Cranfield University, UK, as well as the Institutes of science of food production (ISPA), Italy and University of Warwick HRI for the provision of isolates used in this research.

On a personal note, I will like to appreciate my friend Taiwo Shittu for his support, time and encouragement through the process and also to my other research colleagues Haroon, Guri, Farah, Ingri, Payal and Dr Elias Eze, I want to say thank you.

To all the technical staffs at the School of Life science, University of Bedfordshire, Patrick, Parbin, Micheal, Andrew, Rosie and Emma Buick, Thank you all.

My sincere appreciation goes to the Ministry of Higher Education, Saudi Arabia for the support and sponsorship of this research work.

I am so fortunate to have supportive and understanding family member for their prayer, calls, encouragement, supports and understanding throughout the period of this research.
Declaration

I, Bandar Almiman Almiman declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

Name of candidate: Signature: Date:
Dedication

This dissertation is dedicated to my parents, Mr. Fahad Almiman and Mrs Munirah Alhusainy, for their generous support during the period of this research, family, and friends, for all their prayers and good wishes.
# Table of Contents

Title Page................................................................. II

Abstract ........................................................................... III

Acknowledgements ............................................................. IV

List of Abbreviations .......................................................... V

Declaration ................................................................. IX

Dedication ................................................................. X

Table of Contents ........................................................... XI

List of Tables ...................................................................... XV

List of Figures .................................................................. XVII

1 Introduction ..................................................................... 1

1.1 Research context ......................................................... 1

1.2 Research problem ......................................................... 1

1.3 *Fusarium* genus .......................................................... 4

1.3.1 *Fusarium* species complexes ........................................ 4

1.3.2 *Fusarium fujikuroi* complex ......................................... 5

1.4 *Fusarium proliferatum*, its host range and geographic distribution ....... 6

1.4.1 Economic and health impact of *F. proliferatum* ..................... 7

1.5 Life style and pathogenicity pathway of *Fusarium* species ............ 9

1.6 Influence of environmental factors on the biology of *Fusarium* species. 11

1.6.1 Effect of temperature .................................................. 11

1.6.2 Effect of water activity ................................................ 12

1.7 Reproductive behaviour and sporulation in *Fusarium* species .......... 13

1.7.1 Sexual reproduction .................................................. 13

1.7.2 Asexual reproduction ................................................ 14

1.8 Mycotoxins and their impact ............................................. 15

1.8.1 Fumonisins .............................................................. 18
1.9 FUM gene cluster and its role in fumonisins production ........................................... 20
1.10 Genetic diversity and phylogenetic relationships in Fusarium species .................. 24
  1.10.1 DNA profiling ............................................................................................................. 26
  1.10.2 DNA sequencing (DNA barcoding) ................................................................. 26
1.11 Identification of new gene loci for Fusarium species diversity and phylogeny .......................................................... 28
1.12 Genome sequencing ....................................................................................................... 28
  1.12.1 Development of next generation sequencing ......................................................... 29
1.13 Knowledge and resource gaps in Fusarium proliferatum ........................................ 30
1.14 Aim and Objectives .................................................................................................... 31

2 Materials and Methods .................................................................................................... 32

3 Genotyping of Fusarium proliferatum isolates ................................................................. 38
  3.1 Introduction ...................................................................................................................... 38
  3.2 Materials and methods ................................................................................................. 39
    3.2.1 Liquid culture for genomic DNA extraction .......................................................... 39
    3.2.2 Genomic DNA extraction and quantification ......................................................... 39
    3.2.3 Gel electrophoresis .................................................................................................. 40
    3.2.4 Assessment of the genomic DNA .......................................................................... 40
    3.2.5 PCR Amplification, purification of products and quantification ......................... 41
    3.2.6 Preparation of 20 µl and 50 µl PCR amplification reactions ................................. 41
    3.2.7 PCR amplicon purification and quantification ....................................................... 44
    3.2.8 Nucleotide sequencing of the amplicons ............................................................... 44
    3.2.9 Bioinformatics of DNA sequences ......................................................................... 44
  3.3 Experimental approaches ............................................................................................ 45
  3.4 Results .............................................................................................................................. 46
    3.4.1 Identification of optimal PCR primers to amplify known marker regions ............ 46
    3.4.2 Identification of novel genetic markers for F. proliferatum from the FUNYBASE 51
3.4.3 Evaluation of the genetic diversity and phylogenetic relationships in *F. proliferatum* representing wide biogeographic diversity .......... 60

3.5 Discussion ................................................................................................................. 89

4 **Assessment of biological variability in *Fusarium proliferatum*** ........................................... 97

4.1 Introduction .................................................................................................................. 97

4.2 Materials and methods ............................................................................................... 98

4.2.1 Morphological and cultural studies .......................................................................... 98

4.2.2 Assessment of colony characteristics ....................................................................... 101

4.2.3 Growth rate at different temperatures ....................................................................... 101

4.2.4 Growth rate at different water activities ................................................................. 101

4.2.5 Determination of the level of sporulation ............................................................... 101

4.2.6 Determination of the level of fumonisins ............................................................... 103

4.3 Experimental approaches .......................................................................................... 105

4.4 Results ......................................................................................................................... 105

4.4.1 Morphological variability among *Fusarium* species isolates ............................... 105

4.4.2 The influence of temperature on growth ................................................................. 108

4.4.3 The influence of water activity on growth ............................................................... 110

4.4.4 The level of sporulation over time .......................................................................... 112

4.4.5 Variability in the production of fumonisins by *Fusarium proliferatum* isolates ....................................................................................................................... 114

4.5 Discussion .................................................................................................................. 115

5 **Fusarium proliferatum genome sequencing and FUM gene cluster characterisation** .................... 119

5.1 Introduction .................................................................................................................. 119

5.2 Materials and methods ............................................................................................... 120

5.2.1 DNA Extraction for Genome Sequencing ............................................................ 120

5.2.2 Genome Sequencing of four *F. proliferatum* isolates ........................................... 120
5.2.3 De novo assembly of the genome sequences of *F. proliferatum* isolates .......................................................... 121
5.2.4 Identification of the FUM gene cluster and the flanking regions in the genome of *F. proliferatum* isolates .............................................. 121

5.3 Experimental approaches ....................................................................................................................................... 122

5.4 Results ................................................................................................................................................................. 122
5.4.1 Genome sequencing, *de novo* assembly and characteristics .......................................................... 122
5.4.2 FUM gene cluster of *Fusarium proliferatum* .............................................................................................. 126
5.4.3 The FUM cluster variability among the four *F. proliferatum* isolates .............................................. 131
5.4.4 Flanking genes of the FUM cluster in *Fusarium proliferatum* ................................................................. 133
5.4.5 Variability in the protein sequences of selected genes from the FUM cluster ........................................ 135
5.4.6 Evolutionary relationship reflected by the FUM clusters ..................................................................... 145

5.5 Discussion ........................................................................................................................................................... 146

6 General Discussion, Conclusions and Perspectives .......... 152

6.1 General Discussion .............................................................................................................................................. 152
6.2 Conclusions and Future Perspectives ............................................................................................................. 157

7 References .......................................................................................................................................................... 159

8 Appendix .............................................................................................................................................................. 192
### List of Tables

Table 1.1 Examples of *Fusarium* species and their mycotoxigenic potential on diverse hosts.................................................................................................................................................. 3

Table 1.2 Selected examples of key *Fusarium* species belonging to the *F. fujikuroi* complex.................................................................................................................................................................................. 6

Table 1.3 Details of mycotoxins produced by *Fusarium* species including *F. proliferatum* and their biogeographic diversity................................................................................................................................. 17

Table 1.4 Type of fumonisins produced by some *Fusarium* species and the genomic context of the FUM gene cluster ........................................................................................................................................... 23

Table 1.5 Examples of molecular techniques used with multiple *Fusarium* species and their utility for the characterisation of pathogen biology ........................................................................................................... 25

Table 2.1 Details of the *Fusarium proliferatum* isolates along with the *F. verticillioides* and *F. oxysporum* reference isolates used in this study................................................................. 33

Table 4.1 Details of the *Fusarium proliferatum* isolates along with the *F. verticillioides* and *F. oxysporum* reference isolates used in this study................................................................. 99

Table 4.2 Quantity of glycerol added to the medium to set-up the desired levels of water activity ............................................................................................................................................................................ 101

Table 4.3 Composition of the fumonisin inducing medium ................................................................................................................................. 104

Table 5.1 Details of the next generation sequencing and assembly process and the genome characteristics of four isolates belonging to *Fusarium proliferatum* ........ 124

Table 5.2 General information about isolates and their FUM clusters size ............ 128

Table 5.3 The size range of the FUM cluster genes involved in the biosynthesis of fumonisins in *F. proliferatum*, *F. fujikuroi*, *F. verticillioides* and *F. oxysporum* .... 129

Table 5.4 Variability in the FUM cluster genes among the four *Fusarium proliferatum* isolates ......................................................................................................................................................................................... 132

Table 5.5 Variability in the genes flanking the FUM cluster among the *Fusarium proliferatum* isolates R16, KF3377, ITEM 2341 and MPVP 328.................................................. 135

Table 5.6 FUM cluster genes identified in the *Fusarium proliferatum* in this study and their putative functions with reference to previously known data in other *Fusarium* species................................................................................................................................................................................................. 149
Table 5.7 Genomic context group and the flanking regions genes identified in *Fusarium proliferatum* isolates and their putative functions based on corresponding information from other *Fusarium* species ................................................................. 150

Table 6.1 Comparative data reflecting phylogenetic information content of the novel marker FG1056 relative to known markers RPB1 and RPB2 in four *F. proliferatum* isolates and six representative *Fusarium* species ............................................. 153
List of Figures

Figure 1.1 *Fusarium* species complexes reflecting their level of diversity (species numbers shown in parentheses) and phylogenetic relationship ........................................ 5
Figure 1.2 Kernel black point symptoms caused by *F. proliferatum* isolated from different Wheat varieties in Nepal ................................................................. 8
Figure 1.3 Carnation plant infected by *F. proliferatum* ........................................ 9
Figure 1.4 Different types of disease symptoms caused by *Fusarium* pathogens (a) wilting; and (b) stem browning ................................................................. 10
Figure 1.5 The infection pathway common to various *Fusarium* species ............. 10
Figure 1.6 Sexual and asexual reproduction processes generic to *Fusarium* species 14
Figure 1.7 The morphological characters of *F. proliferatum* ............................... 15
Figure 1.8 The chemical structures of the four main types of fumonisins (A, B, C, and P) ........................................................................................................ 19
Figure 1.9 Proposed order and function of FUM genes in relation to the production of fumonisins ................................................................. 21
Figure 1.10 The order and orientation of the FUM cluster genes in *Fusarium verticillioides* (AF155773) and *F. oxysporum* (EU449979) ................................. 22
Figure 1.11 Schematic showing the five groups of the FUM gene cluster .......... 23
Figure 1.12 Structure of the ribosomal RNA gene block denoting the internal transcribed spacer region (ITS) ................................................................. 26
Figure 1.13 TEF marker region along with the location of the forward (ef1) and reverse (ef22 and ef2) primers for *Fusarium* species ........................................ 27
Figure 4.1 Haemocytometer (A) and the Grid (B) showing the 5 large squares from which the spores were counted in *Fusarium* species isolates ...................... 103
Figure 4.2 Three sets of cultures grown on PDA plates displaying morphological variability among the isolates of *Fusarium proliferatum* from a single host in different geographic locations ................................................................. 106
Figure 4.3 Three sets of cultures grown on PDA plates displaying overlapping morphologies across different *Fusarium* species host and geographical location. 107
Figure 4.4 The effect of temperature on the maximum growth rate of isolates belonging to *Fusarium proliferatum* (14), *F. verticillioides* (2) and *F. oxysporum* (3) ................................................................................................................................................109

Figure 4.5 The effect of water activity on the maximum growth rate of isolates belonging to *Fusarium proliferatum* (14), *F. verticillioides* (2) and *F. oxysporum* (3) ................................................................................................................................................111

Figure 4.6 Variability in the level of spore production over time among isolates belonging to *Fusarium proliferatum* (14), *F. verticillioides* (2) and *F. oxysporum* (3) ................................................................................................................................................113

Figure 4.7 Production of fumonisins FB1 and FB2 by four isolates of *Fusarium proliferatum* ................................................................................................................................................114

Figure 5.1 Gene distribution pattern among the *Fusarium proliferatum* isolates R16 (green), KF3377 (blue), ITEM 2341 (red) and MPVP 328 (yellow) based on OrthoVenn analysis of the predicted total gene sets ................................................................................................................125

Figure 5.2 Order and orientation of the 16 genes located within the FUM cluster in four *F. proliferatum* isolates R16 (onion, UK), MPVP 328 (maize, USA), KF3377 (garlic, Poland) and ITEM 2341 (date palm, Saudi Arabia) characterised in this study. ................................................................................................................................................127

Figure 5.3 The structure of the FUM 18 gene in *Fusarium proliferatum* (FP), *F. fujikuroi* (FF), *F. verticillioides* (FV) and *F. oxysporum* (FO) showing a variation in the number of exons in FV ................................................................................................................................................130

Figure 5.4 Order and orientation of the genes flanking the FUM21 at the 5’ and the FUM19 at the 3’ end of the FUM cluster in the *Fusarium proliferatum* isolates R16, KF3377, ITEM 2314 and MPVP 328 ................................................................................................................................................134

Figure 5.5 Multiple sequence alignment of proteins encoded by FUM21 and FUM19 flanking region genes in the four *F. proliferatum* isolates ........................................................................................................139

Figure 8.1 Multiple sequence alignment for the new genetic marker developed in this study FG1056 data from 26 haplotypes identified among *F. proliferatum* (23), *F. verticillioides* (1) and *F. oxysporum* (2) ................................................................................................................................................195
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Aspergillus</td>
</tr>
<tr>
<td>ACT</td>
<td>Actin</td>
</tr>
<tr>
<td>AFLP</td>
<td>Amplified Fragment Length Polymorphism</td>
</tr>
<tr>
<td>AFs</td>
<td>Aflatoxins</td>
</tr>
<tr>
<td>As</td>
<td>Asparagus</td>
</tr>
<tr>
<td>BEA</td>
<td>Beauvericin</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSC</td>
<td>Biological species concept</td>
</tr>
<tr>
<td>BSV</td>
<td>Bootstrap values</td>
</tr>
<tr>
<td>CAL</td>
<td>Calmodulin</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CDS</td>
<td>Coding sequence</td>
</tr>
<tr>
<td>CuSO(_4) \cdot 5H(_2)O</td>
<td>Cupric sulfate pentahydrate</td>
</tr>
<tr>
<td>D</td>
<td>Date palm</td>
</tr>
<tr>
<td>DAS</td>
<td>Diacetoxyscirpenol</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotidetriphosphate</td>
</tr>
<tr>
<td>DON</td>
<td>Deoxynivalenol</td>
</tr>
<tr>
<td>ENs</td>
<td>Enniatins</td>
</tr>
<tr>
<td>F</td>
<td>Fusarium</td>
</tr>
<tr>
<td>f. sp.</td>
<td>Formae speciale</td>
</tr>
<tr>
<td>F.o (FO)</td>
<td><em>Fusarium oxysporum</em></td>
</tr>
<tr>
<td>F.p (FP)</td>
<td><em>Fusarium proliferatum</em></td>
</tr>
<tr>
<td>F.v (FV)</td>
<td><em>Fusarium verticillioides</em></td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization</td>
</tr>
<tr>
<td>FB</td>
<td>Fumonisin type B</td>
</tr>
<tr>
<td>FBs</td>
<td>Fumonisins B series</td>
</tr>
<tr>
<td>FC</td>
<td>Fumonisin type C</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FHB</td>
<td><em>Fusarium</em> head blight</td>
</tr>
<tr>
<td>FUM</td>
<td>Fumonisin biosynthetic pathway cluster</td>
</tr>
<tr>
<td>FUM1</td>
<td>Polyketide synthase</td>
</tr>
<tr>
<td>G</td>
<td><em>Gibberella</em></td>
</tr>
<tr>
<td>GC</td>
<td>Genomic context</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
</tr>
<tr>
<td>Gr</td>
<td>Garlic</td>
</tr>
<tr>
<td>GTR</td>
<td>General-Time-Reversible</td>
</tr>
<tr>
<td>H3</td>
<td>Histone 3</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>IARC</td>
<td>International Agency for Research on Cancer</td>
</tr>
<tr>
<td>IGS</td>
<td>Intergenic spacer</td>
</tr>
<tr>
<td>ISs</td>
<td>Informative sites</td>
</tr>
<tr>
<td>It</td>
<td>Italy</td>
</tr>
<tr>
<td>ITS</td>
<td>Internal Transcribed Spacer</td>
</tr>
<tr>
<td>Kbp</td>
<td>Kilo base pair</td>
</tr>
<tr>
<td>KCl</td>
<td>Potassium chloride</td>
</tr>
<tr>
<td>Kg</td>
<td>Kilogram</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>Monopotassium phosphate</td>
</tr>
<tr>
<td>L</td>
<td>Litre</td>
</tr>
<tr>
<td>MAT</td>
<td>Mating type gene</td>
</tr>
<tr>
<td>Mbp</td>
<td>Million base pair</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>Mg</td>
<td>Mango</td>
</tr>
<tr>
<td>mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>MgCl$_2$</td>
<td>Magnesium chloride</td>
</tr>
<tr>
<td>MgSO$_4$·7H$_2$O</td>
<td>Magnesium sulfate heptahydrate</td>
</tr>
<tr>
<td>MI</td>
<td>Malaysia</td>
</tr>
</tbody>
</table>
ML  Maximum likelihood
MI  Millilitre
mm²  Square millimeters
MON  Moniliformin
mRNA  Messenger RNA
MSC  Morphological species concept
Mz  Maize
NCBI  National Centre for Biotechnological Information
NIV  Nivalenol
On  Onion
OPA  o-phthalaldehyde
OTA  Ochratoxin A
P  *Penicillium*
PCR  Polymerase chain reaction
PDA  Potato dextrose agar
PDB  Potato dextrose broth
pH  Potential of Hydrogen
PL  Poland
ppb  Parts per billion
PPV  Posterior probability values
PSC  Phylogenetic species concept
RAPD  Randomly Amplified Polymorphic DNA
RFLP  Restriction Fragment Length Polymorphism
RH  Relative humidity
RNA  Ribonucleic acid
RPB1  Subunit of RNA polymerase I
RPB2  Largest subunit of RNA polymerase II
rRNA  Ribosomal RNA
Sa  Saudi Arabia
SCAR  Sequence Characterised Amplified Regions
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SNPs</td>
<td>Single nucleotide polymorphisms</td>
</tr>
<tr>
<td>Sp</td>
<td>Spain</td>
</tr>
<tr>
<td>spp</td>
<td>Species</td>
</tr>
<tr>
<td>SRAP</td>
<td>Sequence Related Amplified Polymorphism</td>
</tr>
<tr>
<td>SSCP</td>
<td>Single Strand Conformational Polymorphism</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple Sequence Repeat</td>
</tr>
<tr>
<td>TCTs</td>
<td>Trichothecenes</td>
</tr>
<tr>
<td>TEF</td>
<td>Translation elongation factor</td>
</tr>
<tr>
<td>Tm</td>
<td>Tomato</td>
</tr>
<tr>
<td>TUB</td>
<td>β-tubulin gene</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>US</td>
<td>United State of America</td>
</tr>
<tr>
<td>USD</td>
<td>United State Dollars</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VNTR</td>
<td>Variable Number of Tandem Repeat</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>Yg</td>
<td>Yugoslavia</td>
</tr>
<tr>
<td>ZEA</td>
<td>Zearalenone</td>
</tr>
<tr>
<td>ZnSO$_4$ · 7H$_2$O</td>
<td>Zinc sulfate heptahydrate</td>
</tr>
</tbody>
</table>

**Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>µg</td>
<td>Microgram</td>
</tr>
<tr>
<td>µl</td>
<td>Microliter</td>
</tr>
<tr>
<td>ng</td>
<td>Nanogram</td>
</tr>
<tr>
<td>aw</td>
<td>Water activity</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Research context
There is a growing concern about the impact of environmental change on the health of plants, animals, humans and the ecosystems. Compelling evidence of the threat posed by environmental change to global food and nutritional security is emerging including changes in the distribution and increase in pests, pathogens and invasive species. Reduction of the potential global yield of crops caused by pests and pathogens is estimated to be around 40% (Bebber et al., 2013; Bebber et al., 2014). Adverse effects of agrochemicals on beneficial species, people and the environment are well documented (e.g. Pimentel et al., 2005; Pal and Gardener, 2006). Impending withdrawal of many pesticides in the EU and the economic burden of pesticide use, e.g. USD 4-12 Billion during 1990 to 2005 add to the challenges (Pimentel et al., 1991; Pimentel et al., 2005). In-depth knowledge of the biology, pathology and genetics of major pests and pathogens to evaluate the risk and develop effective strategies and innovative practices to controlling them is critical to achieve food and nutritional security and environmental sustainability.

1.2 Research problem
Fusarium genus includes some of the most important crop pathogens. For example, F. graminearum and F. oxysporum have been identified among the top 10 fungal pathogens of crops based on economic and scientific importance. Major losses in various crops such as maize, wheat, onion, tomato and cotton have been reported and head blight of barley and wheat caused by F. graminearum was estimated to cost one billion dollars in USA (Llorens et al., 2004; Burgess and Bryden, 2012; Dean
et al., 2012). The number of recognised or fully described species of *Fusarium* remains uncertain due to the cryptic characteristics including overlapping or similar morphologies, host range, and life cycle. In order to reliably identify and classify fungal species such as *Fusarium*, three species concepts are used singly or in combination. This includes morphological species concept (MSC), biological species concept (BSC) and phylogenetic species concept (PSC). The MSC is essentially based on observable morphological characters such as growth and sporulation (Summerell et al., 2003); the BSC includes observable traits and interbreeding ability. Both these concepts have limitations in accurate diagnosis of *Fusarium* species importantly in defining population-species boundaries and distinguishing evolving species. In contrast, the PSC mainly relies on genetic diversity and phylogenetic relationships. PSC principles work on distinguishing or arranging *Fusarium* species isolates based on DNA sequence of genes or intergenic regions used as genetic markers (molecular markers or biomarkers) enabling the identification of genetic groups (clades) displayed on a phylogenetic tree (Summerell et al., 2003; Tibayrenc et al., 2014). The PSC has emerged as the most reliable concept to distinguish and characterise *Fusarium* species and also led to the more recent concept of Genealogical Concordance Phylogenetic Species Recognition (GCPSR; e.g. O’Donnell et al., 2015). Therefore, the number of *Fusarium* species is beginning to increase with reports such as the potential separation of 11 distinct species within *F. graminearum* (Ward et al., 2008). Furthermore, this approach has also led to the recognition of the importance of various *Fusarium* species as major pathogens associated with diverse hosts and geographic locations (Table 1.1).
Table 1.1 Examples of *Fusarium* species and their mycotoxigenic potential on diverse hosts

<table>
<thead>
<tr>
<th>Species</th>
<th>Mycotoxin</th>
<th>Hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F. verticillioides</em> and <em>F. proliferatum</em></td>
<td>Fumonisin, Fusaric Acid and Moniliformin (Yan and Dickman, 1996)</td>
<td>Maize, rice, banana, barley, pine trees, palm trees and asparagus (Jurado <em>et al.</em>, 2010)</td>
</tr>
<tr>
<td><em>F. poae</em></td>
<td>Beauvericin (Kokkonen <em>et al.</em>, 2010)</td>
<td>Wheat, corn, barley and oats (Kulik and Pszczólkowska, 2011)</td>
</tr>
<tr>
<td><em>F.avenaceum</em> and <em>F. tricinctum</em></td>
<td>Enniatins and Moniliformin (Kokkonen <em>et al.</em>, 2010)</td>
<td>Citrus, lisianthus flower, carnation, soybean, potato, hibiscus and chrysanthemum, wheat, barley, wheat and oats (Nalimet <em>et al.</em>, 2009)</td>
</tr>
<tr>
<td><em>F. graminearum</em> and <em>F. culmorum</em></td>
<td>Zearalenone, Deoxynivalenol and 3-Acetyl Deoxynivalenol (Kokkonen <em>et al.</em>, 2010)</td>
<td>Wheat, maize, oats and rice (Yli-Mattila and Gagkaeva, 2010; Popovski and Celar, 2013)</td>
</tr>
</tbody>
</table>

In this context, the identification, phylogeny and molecular diagnostics of *Fusarium* species are key areas that require further research building on the available bioinformatic resources such as the FUSARIUM-ID and Fusarium-MLST databases (Geiser *et al.*, 2004; O’Donnell *et al.*, 2015).

*Fusarium proliferatum* is beginning to be recognised as an emerging pathogen in view of its association with a variety of economically important fruit, vegetable and cereal crops, wide geographic distribution and strong mycotoxigenic potential. Our knowledge of the biology and genetics of this pathogen as well as the risk posed to plant, animal and human health remains incomplete. The present study focuses on further advancing our understanding of the genetic diversity and phylogenetic relationships, biological variability including mycotoxin production and genome level differences in *F. proliferatum*. Emerging molecular technologies such as next generation sequencing and bioinformatic resources have been utilised to characterise a collection of isolates representing the biogeographic diversity of *F. proliferatum*. 
In the following sections 1.3 and 1.4, an overview of the *Fusarium* genus and its species complexes is presented to provide the context for the focal species of this work *F. proliferatum*, its biogeographic diversity and impact. In the subsequent sections 1.5 to 1.12, research advances in key areas such as pathogenicity life cycle, environmental effects, reproductive behaviour, mycotoxigenic potential, FUM gene cluster, phylogenetic relationships and genome sequencing in relevant *Fusarium* species have been discussed in view of the limited progress in *F. proliferatum* – the focal species of this research, to identify the gaps in the knowledge and resources (section 1.13) leading to the aim and objectives (1.14).

### 1.3 *Fusarium* genus

*Fusarium* is a large genus comprising of a large number of species (~300 currently recognised) many of which are recognised as economically important crop pathogens (Nelson *et al.*, 1994; Yan and Dickman, 1996; Summerell *et al.*, 2003; Leslie and Summerel, 2006). The genus *Fusarium* belongs to the phylum - Ascomycota, class - Sordariomycetes, order - Hypocreales and the family – Nectriaceae (Moretti, 2009). The teleomorph (sexual stage) is known as *Gibberella*, whereas the anamorph (asexual stage) is known as *Fusarium* (Proctor *et al.*, 2009). *Fusarium* species can live as endophyte or saprophyte or a pathogen in nature (Marois *et al.*, 1981; Karim *et al.*, 2016), examples include *F. graminearum*, *F. oxysporum*, *F. fujikuroi* and *F. proliferatum*. All *Fusarium* species have one common taxonomic feature by the production of distinctly shaped macroconidia.

#### 1.3.1 *Fusarium* species complexes

*Fusarium* genus encompasses 20 species complexes each of which include a varying number of species (Figure 1.1; O’Donnell *et al.*, 2015). These species complexes have been grouped into 18 sections. For example, *F. fujikuroi* and *F. oxysporum* species complexes belong to the *Liseola* and *Elegans* sections, respectively (Hajek and St. Leger, 1994; O'Donnell and Cigelnik, 1997; Watanabe *et al.*, 2011; O’Donnell *et al.*, 2015). The most common mycotoxigenic *Fusarium* species are *F. graminearum*, *F. proliferatum* and *F. verticillioides* (previously known as *F. moniliforme*).
1.3.2 *Fusarium fujikuroi* complex

The first species belonging to *F. fujikuroi* complex was discovered in 1931 causing bakanae disease on rice and cotton boll rot (Ito and Kimura, 1931; Jeon *et al.*, 2013; Gupta *et al.*, 2015). This is one of largest complexes containing ~ 50 species (Figure 1.1; Table 1.2) some of which can coexist on the same host. Some of these species such as *F. verticillioides*, *F. proliferatum* and *F. fujikuroi* have been reported to show...
very similar morphological characteristics (Nirenberg et al., 1998; Steenkamp et al., 2000; Jurjevic et al., 2005; Hsuan et al., 2011; O’Donnell et al., 2015; Herron et al., 2015). This complex has a major impact on global food security as it affects a wide range of the most economically important crops. There is a lack of in-depth information about the genetic diversity, population structure and fumonisin gene cluster in *F. proliferatum*, which belongs to the *F. fujikuroi* complex (Jurado et al., 2010).

<table>
<thead>
<tr>
<th>Key Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F. circinatum</em></td>
<td>Fourie et al., 2013</td>
</tr>
<tr>
<td><em>F. fujikuroi</em></td>
<td>Nirenberg and O'Donnell, 1998</td>
</tr>
<tr>
<td><em>F. globosum</em></td>
<td>Nirenberg and O'Donnell, 1998</td>
</tr>
<tr>
<td><em>F. mangifera</em></td>
<td>Kvas et al., 2009</td>
</tr>
<tr>
<td><em>F. nygamai</em></td>
<td>Nirenberg and O'Donnell, 1998</td>
</tr>
<tr>
<td><em>F. verticillioides</em></td>
<td>Jurjevic et al., 2005</td>
</tr>
</tbody>
</table>

1.4 *Fusarium proliferatum*, its host range and geographic distribution

*F. proliferatum* was first discovered in 1971 and has emerged as a fungal pathogen that has the potential ability to adapt to different environmental conditions (Nirenberg, 1976; Leslie and Summerell, 2008). This pathogen has been reported in various countries affecting more than 26 crops that are important to food security such as maize, rice and onion (Górna et al., 2016). *F. proliferatum* isolates are very similar to some species within the *Fujikuroi* complex in terms of the colony morphology, conidia shape, pigmentation, toxin production and host range (Mulè et al., 2004a; Proctor et al., 2009).

Many studies have confirmed the distribution of *F. proliferatum* around the world affecting many economically important crops. *F. proliferatum* has been reported to occur in all continents - Asia, New Zealand, Australia, Middle East, Africa, Europe, North and South America. In Asia, *F. proliferatum* has been reported to affect a wide range of host plants such as rice, tomato, asparagus, maize, grape, mango, sunflower, chive, Chinese or Indian date and carnation (Elmer et al., 1997;
Maheshwar and Janardhana, 2010; Zhan et al., 2010; Zainuddin et al., 2011; Zhang et al., 2013a; Zhang et al., 2013b; Quazi et al., 2013; Ren et al., 2015; Wang et al., 2015; Gao et al., 2016). In the Middle East, *F. proliferatum* has been reported in Iran, Iraq and Saudi Arabia affecting sesame, maize, onion, sugarcane, rice and date palm (Abdalla et al., 2000; Alizadeh et al., 2010; Khudhair et al., 2014; Torabi et al., 2014). In Africa, *F. proliferatum* has been isolated from a number of infected plants such as cloves, garlic, onion, sorghum, maize in Egypt, Ethiopia, Nigeria and South Africa (Geleta et al., 2005; Adejumo et al., 2007; Moharam et al., 2013). The occurrence of *F. proliferatum* in many parts of Europe has been confirmed, for example on crops such as maize, clove and garlic in Spain, Switzerland and Italy (Logrieco et al., 1995; Dorn et al., 2009; Palmero et al., 2012). *F. proliferatum* has been detected in the United States and Canada affecting the yields of some important crops such as soybean, onion and garlic (Fuentes et al., 2013; Arias et al., 2015; Chang et al., 2015). Various reports have confirmed the existence of *F. proliferatum* in different South American countries such as Brazil and Argentina where they affect some important crops such as oat, blueberry, pecan, maize and wheat (Stenglein et al., 2010; Lazarotto et al., 2014; Palacios et al., 2015; Pérez et al., 2016).

1.4.1 Economic and health impact of *F. proliferatum*  
Association of *F. proliferatum* with a variety of economically important fruit and vegetable crops is well recognised (e.g. Table 1.1), but the role of this species in disease outbreak and the risk posed to plant, animal and human health is yet to be fully established. For example, *F. proliferatum* was identified in wilt disease samples of onion and garlic initially thought to have been affected by *F. oxysporum* f. sp. *cepae* in forty-one fields in Serbia during early 2000s. Both these hosts were shown to be susceptible to all *F. proliferatum* isolates obtained from these samples. Higher disease severity index was recorded on onion and six out of eleven isolates of *F. proliferatum* produced Fumonisin B1 (Stankovic et al., 2007). Similarly, rapid increases of the incidence of *F. proliferatum* have been reported during last 10 to 15 years in diverse locations such as Italy, Saudi Arabia and China on many crops such as garlic, date palm and grape (Abdalla et al., 2000, Tonti et al., 2012, Wang et al., 2015). *F. proliferatum* is able to cause different kind of diseases such as diebacks, wilt and rots at different stages of the crop. The rot and wilt infections can target different parts of plant such as seeds, roots, bulbs, fruits, crowns, stems and shoots (Figure 1.2 and 1.3). However, host-specific forms have not yet been reported within *F. proliferatum*, unlike in other *Fusarium* species where these have been clearly defined such as *F. oxysporum* f. sp. *lycopersici* and *F. solani* f. sp. *pisi*, which are
specific to tomato and pea, respectively (Ma et al., 2013). *F. proliferatum* can also colonise a range of hosts including wheat, maize and orchids without displaying visible symptoms (Munkvold et al., 1997; Sobek and Munkvold, 1999; Tsavkelova et al., 2008). In addition to its direct impact on plants, *F. proliferatum* is also able to produce mycotoxins such as fumonisins, which are known to cause different types of cancer (Fandohan et al., 2003; Scott, 2012).

*Figure 1.2 Kernel black point symptoms caused by *F. proliferatum* isolated from different Wheat varieties in Nepal*

Images A, C and E in the upper panel are uninfected or control samples, while B, D and F in the lower panel show fungal infections (Source: Desjardins et al., 2007).
1.5 **Life style and pathogenicity pathway of *Fusarium* species**

*Fusarium* species can colonise their plants by producing hyphae and/or spores that penetrate host tissues through primary or secondary infection or both (Senthilkumar et al., 2011; Burgess and Bryden, 2012). The primary or horizontal inocula mainly affect plant roots and the secondary or vertical infection can be initiated by spores spreading on and/or through the plant (Jurado et al., 2008; De Kuppler et al., 2011). Many *Fusarium* species can also live endophytically with no clear evidence or traces of any symptoms, which is a major challenge in pathogen/disease diagnosis during different types of interactions (e.g. Figure 1.4; Dar et al., 1997; Marin et al., 1998; Bacon et al., 2001; Newton et al., 2010; De Kuppler et al., 2011).
The various pathways of infection by *Fusarium* species such as *F. verticillioides* include 1) grain infection via systemic movement through stalk, 2) water splashed conidia or airborne infection towards silk and subsequently to grain and 3) through wounds caused by insects, which can also work as vectors of inoculum (Figure 1.5; Munkvold and Desjardins, 1997). It has been reported that *F. proliferatum* shares some of these infection pathways (e.g. Stankovic et al., 2007).
Fusarium species can cause direct and indirect impact in agriculture and health sectors globally. The direct effect is caused by reduction in crop productivity leading to nutritional deficiency. Contamination of food and feed with different types of toxins produced by Fusarium species affects the quality of the produce indirectly impacting on the health of animals and people (Table 1.1; e.g. Fandohan et al., 2003; IARC 2015).

1.6 Influence of environmental factors on the biology of Fusarium species

Key environmental factors such as temperature and water are known to influence diverse biological attributes such as growth, sporulation and toxin production of mycotoxigenic fungal species such as Fusarium. This includes the survival of conidia present in the air and soil in relation to their dispersal and plant invasion (e.g. Takahashi, 1997; Tang, 2009). Growth is the primary process that leads to sporulation followed by toxin production in various biological species under specific biotic and abiotic conditions including temperature (Wong et al., 1998; Deduke et al., 2012).

1.6.1 Effect of temperature

Research with isolates of F. graminearum and F. culmorum has shown that two isolates of F. culmorum from maize from different geographic locations showed different optimal growth temperature of 15°C and 25°C. This indicates that isolates belonging to the same species and host can react differently under different temperature (Hope et al., 2005). With F. proliferatum and F. verticillioides isolates from maize, 20°C to 25°C was the temperature range for optimum growth (Samapundo et al., 2005). Thirty-one isolates of Fusarium species tested at 25°C showed that isolates of F. incarnatum and F. concentricum were the fastest growing compared to isolates of F. graminearum and F. solani from guava showed optimum growth at 28°C but the isolates varied in the response (Gupta et al., 2010).

It has been reported that the level of fungal spores was much higher in the autumn and summer seasons compared to winter and spring as in the case of F. graminearum on maize (Garrett et al., 1998; Velluti et al., 2004; Lee and Jo, 2006). Similarly, effects of temperature and duration on the level of sporulation in fungi including Fusarium species have been reported (Subbarao et al., 1995; Rossi et al.,
Soil isolates of *Fusarium* species from various crops in India showed variation in the level of sporulation (Islam, 2015). *F. verticillioides* isolates from maize showed maximum sporulation at 27°C but isolates originating from different geographic locations showed different levels of spore production (Rossi *et al*., 2009).

Similarly, the level of mycotoxin production is known to vary at different temperatures. For example, the optimum temperature for mycotoxin production in *F. proliferatum* and *F. verticillioides* ranged between 20°C and 25°C, when tested up to 35°C (Marin *et al*., 1999a; Marin *et al*., 2013a). Among thirty-one isolates belonging to nine *Fusarium* species tested, *F. proliferatum* isolates showed the highest level of fumonisins production at 25°C (Stepien *et al*., 2013).

### 1.6.2 Effect of water activity

Based on the differences in responding to the degree of water present in their environment, fungi were distinguished as hydrophilic, mesophilic and xerophilic (Nielsen, 2003). Under different water activities, different groups of fungal species can be present in an environment (Nielsen *et al*., 2004; Park *et al*., 2008). Hydrophilic fungi including *Fusarium* species favour wettest conditions above 0.90 aw (Pieckova and Jesenska, 1999; Nielsen, 2003). Water activity (aw) in the host plant and the environment has been reported to influence diverse attributes such as growth, sporulation, mycotoxin production and virulence in *Fusarium* species such as *F. graminearum* (Ramirez *et al*., 2004; Velluti *et al*., 2000, 2004; Barbosa-Cánovas *et al*., 2008).

Crops can be contaminated with mycotoxins at high wet conditions (e.g. above 0.95 aw) by fungi such as *Fusarium* species (Llorens *et al*., 2004; Ramirez *et al*., 2006). For example, in 1997 in the Kangwon province in East Korea, the level of fumonisins produced by *F. verticillioides* in maize crop increased from 1.0 µg/g to 169 µg/g following a heavy rainy season (Sohn, 1999). *F. proliferatum* and *F. verticillioides* showed optimum growth and fumonisin production at the wettest condition of 0.995 aw (Marin *et al*., 1999b). Furthermore, in strains of *F. graminearum* and *F. culmorum*, the growth rate and production of DON increased when the water content increased from 0.85 to 0.995 aw (Hope *et al*., 2005).

However, the patterns of biological variability in terms of growth, sporulation and fumonisins production in response to temperature and water activity among *F. proliferatum* isolates originating from diverse hosts and geographic locations have not yet been fully investigated, which is addressed in this study.
1.7 Reproductive behaviour and sporulation in *Fusarium* species

*Fusarium* species can reproduce sexually and asexually. Sexual reproduction can occur through heterothallic (self-sterile) or homothallic (self-fertile) mating (Hajek and St. Leger, 1994; Dean *et al.*, 2012). Both sexual and asexual forms produce spores with the sexual form producing ascospores while the asexual form producing conidiospores. Spores play an important role in pathogen dissemination and the disease cycle and are used as a key feature in species identification (Ohara and Tsuge, 2004; Li *et al.*, 2006; Moretti, 2009).

1.7.1 Sexual reproduction

The sexual form has been mainly reported under laboratory-based *in vitro* conditions. The heterothallic sexual form relies on the crosses of two genetically distinct individual strains that are determined by different forms of the mating type (MAT) locus, while homothallic sexual form relies on a single isolate containing the both the mating types to form ascospores (Irzykowska and Kosiada, 2011; Ma *et al.*, 2013). The perithecium is the fruiting body produced as a result of the sexual reproduction. The spores produced are contained in structures known as asci inside perithecia (Figure 1.6). Fungal sexual reproduction has the potential to play a significant role in the diversity and evolution of pathogens including new genetically distinct strains that could lead to the breakdown of host resistance and/or fungicide resistance. However, there are over 15,000 fungal species including strongly pathogenic and toxigenic species for which no sexual stage has been identified (Irzykowska and Kosiada, 2011). For example, sexual cycle has been recognised in less than 21% of the *Fusarium* species (Ma *et al.*, 2013) and the asexual form is the predominant form reported in nature (Figure 1.6).
1.7.2 Asexual reproduction
The conidiospores produced through asexual form can be divided into three different forms namely microconidia, macroconidia and chlamydospores (Ohara and Tsuge, 2004). The microconidia are formed on the aerial mycelium as chains or clumps (Rodrigues and Menezes, 2005). The macroconidia have a distinctive septate banana shape. The microconidia and macroconidia are formed on infected plant tissue and commonly serve as the source of inoculum for disease development/spread. The chlamydospores are known to have a thick-walled structure and are not formed or observed in all species. They are soil borne and mostly affect the plant root under favourable conditions (Snyder and Hansen, 1940; Ohara et al., 2004; Ohara and Tsuge, 2004; Rodrigues and Menezes, 2005; Nicholson, 2007). Chlamydospores can survive in the soil for long periods until they find suitable host. *F. proliferatum* regularly forms microconidia and macroconidia but chlamydospores are very rare (e.g. Figure 1.7; Nicholson, 2007; Fuentes et al., 2013; Quazi et al., 2013; Chang et al., 2015).
1.8 Mycotoxins and their impact

Mycotoxins are secondary metabolites produced by some fungi and they contaminate 25 to 50% of the world food crops annually (e.g. Rice and Ross, 1994; Fandohan et al., 2003). Mycotoxins are harmful to animals and humans are considered a threat to food security as they affect agricultural economy and the health of the consumers (Logrieco et al., 2002). Exposure of ~500 million people from resource poor communities in Asia, Africa and Latin America to mycotoxins and consequent health impacts including cancer and child stunting have been highlighted in a recent report by the International Agency for Research on Cancer (IARC; Wild et al., 2015). Fumonisins mainly produced by Fusarium species are among the most prolific mycotoxins (Table 1.3). The build-up of mycotoxins in the tissue and organs of the consumers leads to cell disruption and disorder (Bennett and Klich, 2003). There are no visible signs of mycotoxin contamination of crop produce and/or food products, which makes it difficult to recognise and prevent their threat (Gelderblom et al. 1988; Marin et al., 2013b). Mycotoxin contamination of agricultural produce can occur during various tages of food production–supply...

Figure 1.7 The morphological characters of *F. proliferatum*
A) hyphae structure; B) aerial mycelium; C) phialides with spores; D) microconidia; and E) macroconidia (Source: Chang et al., 2015)
chain including improper handling, drying, transport, packaging and storage (Marin et al., 2013b). Mycotoxin production is a complex process regulated by genetic mechanisms that can also be influenced by different environmental stimuli, which is not yet fully understood. For example, different Fusarium species and isolates within a species can vary in their mycotoxin production levels in the laboratory or field (e.g. Kokkonen et al., 2010) linked to the biosynthetic mechanisms and the genetic basis.
Table 1.3 Details of mycotoxins produced by *Fusarium* species including *F. proliferatum* and their biogeographic diversity

<table>
<thead>
<tr>
<th>Mycotoxin*</th>
<th>Fungal species</th>
<th>Potential host/products</th>
<th>Geographical distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUM (FB1, FB2, FB3)</td>
<td><em>Fusarium</em> (e.g. <em>verticillioides</em>, <em>proliferatum</em>, <em>fujikuroi</em>, <em>oxysporum</em>, <em>nygamai</em> and <em>globosum</em>)</td>
<td>Sorghum, millet, corn, rice and their derivatives</td>
<td>Hot- Moderate climate regions in Africa and Europe (Nelson et al., 1994)</td>
</tr>
<tr>
<td>ZEA</td>
<td><em>Fusarium</em> (e.g. <em>verticillioides</em>, <em>culmorum</em>, <em>crookwellense</em>, <em>graminearum</em>, <em>semitectum</em>, <em>incarnatum</em>, <em>equiseti</em>, <em>pseudograminearum</em> and <em>sporotrichioides</em>)</td>
<td>All cereals and cereal-based products and banana</td>
<td>Northern regions in America, Europe and Asia (Doohan et al., 2003)</td>
</tr>
<tr>
<td>TCTs (DON, NIV, T-2, HT-2, DAS)</td>
<td><em>Fusarium</em> (e.g. <em>poae</em>, <em>crookwellense</em>, <em>culmorum</em>, <em>graminearum</em>, <em>sporotrichioides</em> and <em>sambucinum</em>)</td>
<td>Cereals and cereal based products</td>
<td>The Northern moderate climate regions America, Asia and Europe (Doohan et al., 2003; Desjardins et al., 1993)</td>
</tr>
<tr>
<td>MON</td>
<td><em>Fusarium</em> (<em>verticillioides</em>, <em>proliferatum</em>, <em>oxysporum</em>, <em>avenaceum</em>, <em>equiseti</em>, <em>acuminatum</em> and <em>subglutinans</em>)</td>
<td>Cereals and cereal based products</td>
<td>Moderate climate regions in Europe (Marasas et al., 1986; Rabie et al., 1982)</td>
</tr>
<tr>
<td>ENs (A, A1, B, B1)</td>
<td><em>Fusarium</em> (<em>proliferatum</em>, <em>tricinctum</em>, <em>sambucinum</em>, sporotrichioides, <em>langsethiae</em>, <em>acuminatum</em>, <em>avenaceum</em>, <em>lateritium</em> and <em>poae</em>)</td>
<td>Cereals and cereal based products</td>
<td>Moderate climate regions in Europe (Burmeister et al., 1987; Plattner and Nelson, 1994)</td>
</tr>
<tr>
<td>BEA</td>
<td><em>Fusarium</em> (e.g. <em>proliferatum</em>, <em>pseudoanthophilum</em>, <em>subglutinans</em>, <em>poae</em> and <em>avenaceum</em>)</td>
<td>Cereals and cereal based products</td>
<td>Moderate climate regions in Europe (Plattner and Nelson, 1994; Logrieco et al., 2002)</td>
</tr>
</tbody>
</table>

* FUM, fumonisins; ZEA, zearalenone; TCTs, trichothecenes; DON, deoxynivalenol; NIV, nivalenol; T-2, T-2 toxin; HT2, HT-2 toxin; DAS, diacetoxyscirpenol; MON, moniliformin; ENs, enniatins and BEA, beauvericin.
1.8.1 Fumonisins

Fumonisins were first discovered in maize affected by *F. verticillioides* in early 1980s in South Africa (Gelderblom et al., 1988). Some *Fusarium* pathogens such as *F. solani* do not produce fumonisins, whilst others including *F. proliferatum*, *F. fujikuroi* and *F. verticillioides* are known as prominent producers (Table 1.3). Among these species, *F. proliferatum* is the highest producer compared to highly toxigenic species such as *F. verticillioides*. The type of fumonisins produced by different fungal species including *F. verticillioides* has been elucidated but less information is available from *F. proliferatum* (Chen et al., 1992; Rheeder et al., 2002; Frisvad et al., 2007; Khaldi and Wolfe, 2011; Proctor et al., 2013).

Fumonisin producing *Fusarium* species such as *F. verticillioides* cause ear rot of maize leading to mycotoxin contamination of grains. Infection is increased in damaged kernals. And improper storage (e.g. above 19 % moisture content) of the grains could lead to continued growth of the fungus and the production of fumonisins (e.g. Munkvold and Desjardins, 1997; Jurado et al., 2010). The genetic basis for the levels of contamination of maize kernals by fumonisins was attributed to specific mating populations A-G of *G. fujikuroi*. Population A common on maize is recognised as prolific producers of fumonisins, whilst population F common on sorghum produce little or no fumonisins (Munkvold and Desjardins, 1997; Jurado et al., 2010). *G. fujikuroi* Population A is now recognised as *F. verticillioides* (e.g. Jurado et al., 2010). Regulation of the biosynthesis of fumonisins in infected maize kernals by *F. verticillioides* has been attributed to ZFR1 gene, which belongs to a gene family known to regulate fungal metabolism (Flaherty and Woloshuk, 2004).

Fumonisins belong to the group of mycotoxins derived from polyketides and these have been categorised in four main groups (A, B, C and P) based on their structural variability specifically the carbon backbone length and the amino group (Figure 1.8). Both fumonisin B (FB) and C (FC) have relatively similar structure compared to other fumonisin types. In terms of importance, however, FB has been the main focus in several studies due to wide occurrence and the carcinogenic property (Branham and Plattner, 1993; Musser and Plattner, 1997; Proctor et al., 2003; Sewram et al., 2005; Glenn et al., 2008; Proctor et al., 2008; Jurado et al., 2010; Visentin et al., 2012; Proctor et al., 2013; Rocha et al., 2016).
The chemical structures of the four main types of fumonisins (A, B, C, and P)

Each type consists of 19 – 20 linear carbon backbone but with different hydroxyl, methyl and tricarballylic acid groups at the chains (Deshmukh et al., 2016).

The FBs were divided into classes FB1, FB2, FB3 and FB4. FB1 occurrence is predominant in nature at about 70%, whereas FB2, FB3 and FB4 at around 30% based on a number of publications (Desjardins et al., 1996; Lukacs et al., 1996; Rheeder et al., 2002; Proctor et al., 2003; Rocha et al., 2016). Based on in vitro experiments using maize-based culture F. verticillioides was shown to produce ~17900 µg/g of FB1, whilst F. proliferatum produced ~31000 µg/g of FB1 (Rheeder et al., 2002).

Mycotoxins produced by F. verticillioides have been recognised as carcinogenic to animals and humans (Loprieno, 1975; Kellerman et al., 1990; Gelderblom et al., 1991; IARC, 1993; FDA, 2000; Fandohan et al., 2003; Scott, 2012). It has been reported that fumonisin contaminated feeds caused damage to brain function in horses and pulmonary disease in swine (Nelson et al., 1994) constituting a major threat to animal health (Nelson et al., 1993). In Brazil, higher rate of human oesophageal cancer was found in the South where maize consumption is higher and the level of fumonisin contamination was between 1.5 and 3.4 µg/g in human food products (Van der Westhuizen et al., 2003). An added complication is the co-occurrence of fumonisins with other toxins such as aflatoxins and DON in many important crops. For example in many African countries such Tanzania, Cameroon and Nigeria, both
Fumonisin and aflatoxins were found to be over the tolerable level in 20 to 80% of tested crops and in over 40% of children’s food in Tanzania (Shirima et al., 2015; Wild et al., 2015). Health impacts due to concurrent exposure of people from Africa, Latin America and Asia to fumonisins and aflatoxins have recently been highlighted by the IARC (Wild et al., 2015). In parts of Africa, it’s been reported that children are at risk of exposure to multiple mycotoxins via maize-based food products (e.g. Kimanya et al., 2014). Variability in the levels of mycotoxin contamination of maize has been linked to agro-ecological zones and seasons (Geary et al., 2016).

In parallel, efforts are underway to develop practical ways to reduce fumonisin levels in maize products (van der Westhuizen et al., 2011). Organisations such as FAO and WHO are developing guidelines of the tolerable level of fumonisins in human foods (provisional maximum tolerable daily intake of 2μg/kg body weight/day (JECFA, 2001; Fandohan et al., 2003; Palencia et al., 2003; El-Imam et al., 2012)). Novel sensor-based approaches integrating previously available analytical and immunological assays are being explored to improve throughput, automation and sensitivity (Lin and Guo, 2016).

1.9 FUM gene cluster and its role in fumonisin production

The FUM gene cluster is known to be responsible for the biosynthesis of fumonisins with the key enzyme polyketide synthase encoded by the FUM1 gene (Glenn et al., 2008). The FUM cluster in various Fusarium species is estimated to be 45 - 47 Kb consisting of 16 to 17 co-regulated genes involved in synthesis and regulatory activities (Table 1.4). Despite containing the FUM cluster, not all species or isolates of Fusarium have the same ability to produce fumonisin potentially due to changes in the gene sequence as well as differences in responding to environmental stimuli and the plant host. Comparison of FUM genes of F. verticillioides from maize and banana showed differences in the cluster size mainly suggested to be due to deletions and/or mutations in FUM1 gene affecting the production of FB (Brown et al., 2005; Glenn et al., 2008; Stepien et al., 2011a; Stepien et al., 2011b; Medina et al., 2013; Proctor et al., 2013). FUM 1 and FUM 8 are the most important genes responsible for key enzymatic activities. FUM 1 gene encodes the key enzyme polyketide synthase required for the synthesis of the fumonisin backbone, which can be enzymatically modified to form different types or levels of fumonisins. On the other hand FUM 8 is involved in the addition of distinctive amino groups associated with different groups of fumonisins such as FB or FC. While the remaining FUM genes have functions in fumonisin biosynthesis and self-protection (Proctor et al., 2003; Brown et al., 2005; Proctor et al., 2013). As shown in the schematic below (Figure 1.9; Proctor et al., 2003), FUM1 plays a key role in the
linear polyketide formation (A); Fatty acyl-coA activation is mediated by FUM10 or FUM16, alanine condensation by FUM8 and carbonyl reduction by FUM13 (B); FUM10 or FUM16 enhance the coA activation of tricarboxylic acids (TCA) via FUM11, whereas FUM6, 9, 12 or 15 have a role in the hydroxylation and esterification of TCA backbone at C14 and C15 (C); FUM6, 9, 12 or 15 have a role in the hydroxylation of the backbone at C10 (D); and FUM6, 9, 12 and 15 have a role in the hydroxylation of the backbone at C5 (E).

Figure 1.9 Proposed order and function of FUM genes in relation to the production of fumonisins
(Source: Proctor et al., 2003)

The overall order and orientation of the FUM genes are similar in the Fusarium species such as F. oxysporum and F. verticillioides with FUM21 and FUM19 as the terminal genes (Figure 1.10). It is not yet fully clear whether FUM 20 gene is present
in all species (Brown et al., 2005). Furthermore, the flanking regions in some *Fusarium* species have been reported to be different potentially due to differences in the location of the gene cluster in the genome (Table 1.4) (Proctor et al., 2003; Waalwijk et al., 2004; Glenn et al., 2008, Proctor et al., 2008; Proctor et al., 2013). Based on the genomic context, FUM gene clusters have been categorised into five groups namely GC1, GC2, GC3a, GC3b and GC4 (Figure 1.11). In the GC1 group, ZNF1 and ZBD1 genes flank the FUM21 gene present at the 5’ end of the cluster and ORF20 and ORF21 flank the FUM 19 gene at the 3’ as in the case of *F. verticillioides*, *F. ramigenum*, *F. nygamai* and *F. phyllophilum*. In the GC2 group, MFS1 and ZCB1 genes flank FUM21 and ANK1 and GAT1 flank FUM19 as in the case of *F. fujikuroi* and *F. globosum*. Whereas the flanking genes in the GC3 (e.g. in *F. anthophilum* and *F. bulbicola*) and GC4 (e.g. *F. oxysporum*) groups are less well defined (Proctor et al., 2003; Proctor et al., 2013). Overall, the genomic regions of the FUM cluster or some of the flanking genes and their function have not yet been fully defined in various *Fusarium* species including *F. proliferatum*, e.g. no sequence data or accession number available in public databases.

Figure 1.10 The order and orientation of the FUM cluster genes in *Fusarium verticillioides* (AF155773) and *F. oxysporum* (EU449979)

(Source: Proctor et al., 2008)
Figure 1.11 Schematic showing the five groups of the FUM gene cluster

Represented by the terminal genes shown in blue based on potential flanking regions genes in varied colours at the 5’ and 3’ ends of the cluster (Source: Proctor et al., 2013)

Table 1.4 Type of fumonisins produced by some *Fusarium* species and the genomic context of the FUM gene cluster

<table>
<thead>
<tr>
<th>Species</th>
<th>Fumonisin type</th>
<th>Genomic context of FUM cluster (Proctor et al., 2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F. verticillioides</em></td>
<td>FB (Desjardins et al., 2007)</td>
<td>GC1</td>
</tr>
<tr>
<td><em>F. remigenum</em></td>
<td>FB (Proctor et al., 2013)</td>
<td>GC1</td>
</tr>
<tr>
<td><em>F. fujikuroi</em></td>
<td>FB (Desjardins et al., 2007)</td>
<td>GC2</td>
</tr>
<tr>
<td><em>F. globosum</em></td>
<td>FB (Rheeder et al., 2002)</td>
<td>GC2</td>
</tr>
<tr>
<td><em>F. anthophilum</em></td>
<td>FC (Rheeder et al., 2002)</td>
<td>GC3a</td>
</tr>
<tr>
<td><em>F. bulbicola</em></td>
<td>FC (Proctor et al., 2013)</td>
<td>GC3b</td>
</tr>
<tr>
<td><em>F. oxysporum</em></td>
<td>FC (Proctor et al., 2008)</td>
<td>GC4</td>
</tr>
</tbody>
</table>
1.10 Genetic diversity and phylogenetic relationships in *Fusarium* species

*Fusarium* species poses an increasing threat to food and nutritional security as well as animal and human health. Research over the last two decades has focused on understanding the pathogen genetic diversity, population structure and phylogenetic relationships to gain new insights into their biology and evolution and to develop effective disease interventions such as resistant varieties. (Leung *et al.*, 1993; McDonald and McDermott, 1993; Chen *et al.*, 1995; McDonald, 1997; Hajibabaei *et al.*, 2007; O’Donnell *et al.*, 2015). Advances in molecular technologies have led to the identification and development of molecular markers (also referred to as genetic markers or biomarkers) that unravel DNA sequence diversity embedded within the genome (Tsui *et al.*, 2011; Singh *et al.*, 2013). Various molecular techniques that could be broadly categorised as DNA profiling (e.g. RFLP and RAPD) and DNA barcoding (markers such as TEF1 and ITS) have been used in the population study of various *Fusarium* species (Table 1.5; e.g. O’Donnell *et al.*, 2015). In some other fungal pathogens such as the globally important cereal blast pathogen *Magnaporthe oryzae*, repeat elements have been used as a probe for DNA profiling to develop new knowledge of the population diversity and structure in sub-Saharan Africa (e.g. Takan *et al.*, 2012).
### Table 1.5 Examples of molecular techniques used with multiple *Fusarium* species and their utility for the characterisation of pathogen biology

<table>
<thead>
<tr>
<th>Technique/Marker</th>
<th>Species/ origins</th>
<th>Host</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSR (simple sequence repeat)</td>
<td><em>F. proliferatum</em>, <em>F. fujikuori</em>, <em>F. globosum</em>, <em>F. oxysporum</em>, <em>F. verticilloides</em>, <em>F. thapsinum</em>, <em>F. subglutinans</em> and <em>F. andyazi</em>, Israel, Germany, USA, Austria and Taiwan</td>
<td>Onion, Garlic, Maize, Orchid, Rice, Cucumber and Asparagus</td>
<td>Limited correlation to species or host or location</td>
<td>Moncrief <em>et al.</em>, 2016</td>
</tr>
<tr>
<td>DNA sequencing / FUM1 gene marker</td>
<td><em>F. proliferatum</em>, <em>F. commune</em>, <em>F. verticillioides</em>, <em>F. subglutinans</em>, <em>F. concentricum</em> and <em>F. fujkori</em> complex, Korea</td>
<td>Rice and Corn</td>
<td><em>F. fujikuroi</em> species complex isolates were distinctive, but no correlation to host or location</td>
<td>Kim <em>et al.</em>, 2012</td>
</tr>
<tr>
<td>DNA sequencing / Calmodulin gene and ITS region</td>
<td><em>F. verticillioides</em>, <em>F. proliferatum</em> and <em>F. subglutinans</em></td>
<td>Maize, Sugar cane, Wheat, Rice, Ananas, Asparagus, Garlic and Date palm</td>
<td>Three distinct clades based on species with Calmodulin; lack of resolution with ITS</td>
<td>Mulè <em>et al.</em>, 2004a</td>
</tr>
<tr>
<td>DNA sequencing/ H3 (Histone gene)</td>
<td><em>Fusarium</em> species</td>
<td>Wheat and barley</td>
<td>2 clades with 6 groups and 18 sub-groups</td>
<td>O'Donnell <em>et al</em>, 2004</td>
</tr>
<tr>
<td>DNA sequencing/ Beta tubulin gene</td>
<td><em>F. avenaceum</em> and <em>F. acuminatum</em></td>
<td>Barley, Carnation, Turf, Wheat, Soil, Soybean, common osier, Citrus, Lisianthus, Potato, Lavendula, Hibiscus, Morning glory and Chrysanthemum</td>
<td>2 clades with 5 groups- 10 sub-groups, majority of the isolates in each group were of common origin</td>
<td>Nalim <em>et al.</em>, 2009</td>
</tr>
</tbody>
</table>
1.10.1 DNA profiling
A range of molecular techniques such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), single strand conformation polymorphism (SSCP), amplified fragment length polymorphism (AFLP), variable number of tandem repeats (VNTR) and simple sequence repeats (SSR) also known as microsatellites have been extensively used to study the genetic differences of many *Fusarium* species such as *graminearum, fujikori* and *oxysporum* complexes (e.g. Table 1.5, Geistlinger et al., 1997; McDonald, 1997; Chandra, 2011; Wang et al., 2011). The DNA profiles from these methods are scored based on the presence and absence of similarly positioned bands (McDonald, 1997). These techniques operate on different principles, for example PCR-based methods such as RAPD and AFLP do not require prior DNA information to design primers (Fearon et al., 1987; Raeder et al., 1989; McDonald, 1997). AFLP analysis showed positive discrimination among isolates of *F. graminearum* (Leissner et al., 1997).

1.10.2 DNA sequencing (DNA barcoding)
With the advances in DNA sequencing technologies, direct sequencing of PCR products has been used extensively to characterise inter and intra species differences in diverse fungi including *Fusarium* species (Vos et al., 1995; Xu, 2006; Sucher et al., 2012; O’Donnell et al., 2015). Availability of a large number of DNA sequences of various marker genes for diverse organisms in public databases, requirement of lesser amount of DNA, decreasing costs have led to use of this approach to investigate genetic diversity and phylogenetic relationships in a number of species (Hajibabaei et al., 2007; Arif et al., 2010; Cowan and Fay, 2012).

The first and most commonly used genes and/or regions to study fungal population relationships are the internal transcribed spacer (ITS), intergenic spacer (IGS), 18S rRNA (small subunit) and the 28S rRNA (large subunit) all embedded in the ribosomal RNA (rRNA) gene block (Buscot et al., 2000; Scupham et al., 2006; Schoch et al., 2009; Chandra, 2011; Schoch et al., 2012). The ITS region covers the 3’ end of the small subunit to the 5’ end of the large subunit (Figure 1.12).

![Figure 1.12 Structure of the ribosomal RNA gene block denoting the internal transcribed spacer region (ITS)](Source: Buscot et al., 2000)
The ITS region has been used extensively in the population study of pathogenic and non-pathogenic fungi to understand their evolution and to develop diagnostic tests (e.g. Sreenivasaprasad et al., 1996). To date over 170,000 ITS sequences covering more than 14,900 species have been deposited in the gene banks (Schoch et al., 2012). This region has been proposed as the standard DNA barcode for fungal species (White et al., 1990; Schoch et al., 2012). Presence of non-orthologous copies of ITS and consequent need for careful interpretation of the data has been reported in some genera such as Fusarium and Penicillium (O'Donnell and Cigelnik, 1997; O'Donnell et al., 1998a; Skouboe et al., 1999; O'Donnell et al., 2015). The 18S gene has poor species-level resolution in fungi, while the 28S has good resolution at certain taxonomic levels. Nevertheless, researchers have highlighted the potential limitations of ribosomal gene block markers compared to protein coding gene markers (Geiser et al., 2004; Barik and Tayung, 2012).

Protein coding genes have been used extensively for the identification and phylogenetic analysis of fungal species. Most of these are housekeeping genes such as actin (ACT), translation elongation factor 1-α (TEF1), β-tubulin (TUB), histone (HIS3), calmodulin (CAL) and the large subunit of RNA polymerase II (RPB1). Some are functional genes such as fumonisin (FUM) genes and mating (MAT) genes (Geiser et al., 2004; Vaquero et al., 2004; Barik and Tayung, 2012). For example, TEF1 and TUB have been used for the identification of Fusarium species (Geiser et al., 2004; Matheny et al., 2007; O'Donnell et al., 2010; Herron et al., 2015). The TEF gene marker was first developed for F. oxysporum complex (Cho et al., 1995; Mitchell et al., 1997; O'Donnell et al., 1998c). Wide utility of the TEF1 in phylogenetic analysis of Fusarium genus and other related fungi has been attributed to its high informative nature as well as the absence of non-orthologous copies (Geiser et al., 2004). The first pair of primers ef1 and ef2 for the TEF1 region was designed to amplify around 700 bp covering the region in exon 1 and exon 4 in all Fusarium species (Figure 1.13; Geiser et al., 2004).

![Figure 1.13 TEF marker region along with the location of the forward (ef1) and reverse (ef22 and ef2) primers for Fusarium species](source: Geiser et al., 2004)
Other genetic markers such as CAL (Mulè et al., 2004a), TEF1 (Stepien et al., 2011a), FUM 1 (Kim et al., 2012), FUM 8 (Bargen et al., 2009) and TUB (Nalim et al., 2009) have shown potential in assessing the diversity of Fusarium species associated with a range of hosts and originating from various geographic locations (Table 1.5). These results have also suggested the distinctive nature of the isolates derived from garlic using the FUM1 marker (Stepien et al., 2011a). Among the common genetic markers used in phylogenetic studies of Fusarium species such as ITS, CAL, TEF1, TUB and HIS3, it’s been suggested that the TEF1 showed better resolution to discriminate species within the F. fujikuroi complex (Geiser et al., 2004; Kristensen et al., 2005; Herron et al., 2015).

1.11 Identification of new gene loci for Fusarium species diversity and phylogeny

A primary objective of ongoing genome studies in Fusarium species is to identify new loci that are useful in assessing the genetic diversity and phylogenetic relationships (O’Donnell et al., 2015). A database containing 246 single copy orthologs that were extracted from 21 fungal genomes known as fungal phylogenomic database (FUNYBASE) has been developed (Marthey et al., 2008). The key pathogens in FUNYBASE include Fusarium graminearum, Aspergillus fumigatus, Trichoderma reesei and Magnaporthe grisea. FUNYBASE is a distinctive resource to identify new maker loci showing high variability useful for phylogenetic analysis of some fungal pathogens belonging to ascomycetes (Walker et al., 2012; Armitage et al., 2015). This resource and the available genome sequences of Fusarium species have been used in the present study to identify new markers for assessing the genetic diversity and phylogenetic relationships in F. proliferatum.

1.12 Genome sequencing

DNA sequencing developed by Frederick Sanger around 1977 combined with capillary electrophoresis enabled automated high-throughput sequencing (Schuster, 2007; Swerdlow et al., 1990; Hunkapiller et al., 1991). This led to the sequencing of the first eukaryotic species Saccharomyces cerevisiae in 1996 (Goffeau et al., 1996). Mycologists have built on this landmark breakthrough to sequence more fungal species such as Neurospora crassa (Galagan et al., 2003), Magnaporthe grisea (Dean et al., 2012), which was the first plant fungal pathogen genome to be sequenced, Aspergillus fumigatus (Nierman et al., 2005), Trichoderma (Kubicek et al., 2011) and Colletotrichum (O’connell et
Development of next generation sequencing technologies during this period has revolutionised the whole field of genomics (e.g. Schuster, 2007; Reis-Filho, 2009).

1.12.1 Development of next generation sequencing

Next generation sequencing (NGS) platforms use sequencing by synthesis method leading to very high throughput, faster and cost-effective technologies that and can be performed by small laboratories (Schuster, 2007; Reis-Filho, 2009). NGS includes different platforms such as the Roche 454, Illumina (GA/HiSeq), Sequencing by Oligo Ligation Detection (SOLiD), PacBio, Oxford-Nanopore (Shendure and Ji, 2008; Liu et al., 2012). NGS-based genome sequencing includes key steps such as DNA extraction, fragment library preparation, generation of sequence reads, assembly of the reads, gene prediction and functional annotation (Shendure and Ji, 2008). A number of programmes are becoming available to enable rapid and efficient assembly of fungal genomes. This includes programs such as Geneious, BWA and SAMtools used for reference genome assembly and Velvet, Spades, SOAPdenovo for de novo assembly (Zhang et al., 2011; Abbas et al., 2014). The genome assembly quality is assessed using programmes such as QUAST (Gurevich et al., 2013) and BUSCO (Baroncelli et al., 2016) and to verify various parameters including the genome coverage, number of contigs and the N50 (Feldmeyer et al., 2011, Lin et al., 2011). Gene prediction based on features such as polyadenylation site, splice sites, branch and coding sites, start and stop codons sites (Picardi and Pesole, 2010) has been achieved through programs such as AUGUSTUS (Stanke and Waack, 2003). Blast2GO and/or a pipeline approach are used for functional annotation of the genome (e.g. Conesa et al., 2005; Gotz et al., 2008; Baroncelli et al., 2016).

These developments have led to a large number of de novo and resequencing projects in many fungal genera including Fusarium species (Cuomo et al., 2007). In parallel, web-accessible resources integrating strain identification, phlogenetics and comparative genomics of Fusarium species are also being developed (Park et al., 2011). Genome sequences of F. oxysporum, F. verticillioides, F. fujikuroi and F. pseudograminearum have become available (Chiara et al., 2015; King et al., 2015; Niehaus et al., 2016). Comparative analysis of various Fusarium species genomes has revealed gene content differences, species- and isolate-specific genes and arrangement of gene clusters. Genome sequence of two isolates of F. proliferatum has become available very recently, but the full FUM gene cluster responsible for the production of key mycotoxins such as fumonisins is yet to be identified (Niehaus et al., 2016). These resources provide an exciting opportunity to develop and apply integrated bimolecular methodologies to further characterize F. proliferatum.
1.13 Knowledge and resource gaps in *Fusarium proliferatum*
Following its initial discovery in the early 1970s, *F. proliferatum* has emerged as a key pathogen affecting more than 26 crops in diverse geographic locations across continents. However, the research progress with *F. proliferatum* has been relatively limited compared to other major pathogens such as *F. oxysporum*, *F. graminearum* and *F. verticilliodes*, as discussed in the previous sections of this chapter. Research carried out over the last 10 – 15 years, has highlighted the biological and genetic diversity of *F. proliferatum* isolates mainly associated with specific crop hosts/geographic locations. However, our knowledge and understanding of the biological and genetic variability patterns and phylogenetic relationships in *F. proliferatum* representing its global biogeographic diversity remains incomplete. There is also a clear recognition by the *Fusarium* research community of the need to identify and develop new genetic markers in this context. In addition, although the strong mycotoxigenic potential of *F. proliferatum* is well recognised, the FUM gene cluster responsible for the biosynthesis of fumonisins has not yet been fully delineated in this species. Consequently, there is very limited information available of the overall size of the fumonisin gene cluster, the number of genes, their order and orientation including the flanking genes, and the mutations potentially related to the regulation of the level and type of fumonisins produced in *F. proliferatum*. Furthermore, from the start of this work in June 2012 till close to the submission of this thesis, there have been no reports focused on the genome level characterisation of *F. proliferatum* aligned to its biogeographic diversity (Niehaus *et al.*, 2016 mainly focuses on the “Comparative ‘Oomics’ of the *Fusarium fujikuroi* Species Complex”). Thus, the knowledge and resource gaps identified above, based on the information presented in previous sections, provide the rationale for the aim and objectives of the present research study.
1.14 Aim and Objectives

The overall aim of this research is to elucidate the genotypic diversity of the mycotoxigenic species *Fusarium proliferatum* associated with diverse hosts and distributed in wide geographic locations to gain new insights into the biology of this emerging pathogen. This will be achieved through the following objectives:

1) To determine the genetic diversity and phylogenetic relationships of *F. proliferatum* isolates representing the biogeographic variability of the pathogen utilising a set of existing and new molecular markers.

2) To assess the biological variability particularly growth, sporulation and fumonisin production in *F. proliferatum* isolates with reference to key environmental parameters such as temperature and water activity and to decipher any links to the genetic relationships.

3) To develop the genome sequence data of selected *F. proliferatum* isolates using NGS technology and to carry out comparative analysis including the characterization of the FUM gene cluster responsible for the production of fumonisins.

The overall purpose of this research is to generate new knowledge and resources that will contribute to the development of strategies and practices for effective control of crop diseases caused by *F. proliferatum*. 
2 Materials and Methods

2.1 *Fusarium proliferatum, F. verticillioides and F. oxysporum* isolates used

Fifty-two *F. proliferatum* isolates originating from diverse hosts and geographic locations were used in this research (Table 2.1). Five isolates belonging to *F. verticillioides* (2) and *F. oxysporum* (3) were used as references. This collection was established by securing the isolates from research collaborators in the UK and other European Countries as well as the Agro-Food Microbial Culture Collection (ITEM) based at the Institute of Sciences of Food Production, Bari, Italy. Isolates were stored under sterile conditions at room temperature in Universal vials containing water as agar plugs carrying the mycelium and spores prepared from potato dextrose agar cultures. [Chapters 3, 4 and 5]
Table 2.1 Details of the *Fusarium proliferatum* isolates along with the *F. verticillioides* and *F. oxysporum* reference isolates used in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>Isolate code*</th>
<th>Species</th>
<th>Host /Plant part</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2/2</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>2</td>
<td>A6/1</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>3</td>
<td>R16</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>4</td>
<td>A40</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>5</td>
<td>A8</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>6</td>
<td>ITEM 1950</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>Italy</td>
</tr>
<tr>
<td>7</td>
<td>ITEM 1951</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>Italy</td>
</tr>
<tr>
<td>8</td>
<td>ITEM 1952</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>Italy</td>
</tr>
<tr>
<td>9</td>
<td>ITEM 1953</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>Italy</td>
</tr>
<tr>
<td>10</td>
<td>ITEM 4418</td>
<td><em>F. proliferatum</em></td>
<td>Allium cepa root</td>
<td>Yugoslavia</td>
</tr>
<tr>
<td>11</td>
<td>ITEM 4419</td>
<td><em>F. proliferatum</em></td>
<td>Allium cepa root</td>
<td>Yugoslavia</td>
</tr>
<tr>
<td>12</td>
<td>ITEM 4358</td>
<td><em>F. proliferatum</em></td>
<td>Allium sativum root</td>
<td>Yugoslavia</td>
</tr>
<tr>
<td>13</td>
<td>ITEM 4359</td>
<td><em>F. proliferatum</em></td>
<td>Allium sativum root</td>
<td>Yugoslavia</td>
</tr>
<tr>
<td>14</td>
<td>KF3363</td>
<td><em>F. proliferatum</em></td>
<td>Allium sativum</td>
<td>Poland</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scientific Name</td>
<td></td>
<td>Country</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>------------------</td>
<td>---</td>
<td>-----------</td>
</tr>
<tr>
<td>15</td>
<td>KF3366</td>
<td><em>F. proliferatum</em></td>
<td><em>Allium sativum</em></td>
<td>Poland</td>
</tr>
<tr>
<td>16</td>
<td>KF3377</td>
<td><em>F. proliferatum</em></td>
<td><em>Allium sativum</em></td>
<td>Poland</td>
</tr>
<tr>
<td>17</td>
<td>ITEM 1449</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>18</td>
<td>ITEM 1450</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>19</td>
<td>ITEM 1452</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>20</td>
<td>ITEM 1453</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>21</td>
<td>ITEM 1454</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>22</td>
<td>ITEM 1455</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>23</td>
<td>ITEM 1451</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>24</td>
<td>ITEM 1477</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>25</td>
<td>ITEM 1478</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>26</td>
<td>ITEM 1480</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Italy</td>
</tr>
<tr>
<td>27</td>
<td>KF3341</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Poland</td>
</tr>
<tr>
<td>28</td>
<td>KF3360</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Poland</td>
</tr>
<tr>
<td>29</td>
<td>KF3362</td>
<td><em>F. proliferatum</em></td>
<td><em>Asparagus officinalis</em></td>
<td>Poland</td>
</tr>
<tr>
<td>30</td>
<td>18 F.P</td>
<td><em>F. proliferatum</em></td>
<td><em>Zea mays</em></td>
<td>Spain</td>
</tr>
<tr>
<td>31</td>
<td>19 F.P</td>
<td><em>F. proliferatum</em></td>
<td><em>Zea mays</em></td>
<td>Spain</td>
</tr>
<tr>
<td></td>
<td>ITEM</td>
<td>Organism</td>
<td>Species</td>
<td>Country</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>32</td>
<td>7595</td>
<td>F. proliferatum</td>
<td>Zea mays</td>
<td>USA</td>
</tr>
<tr>
<td>33</td>
<td>MPVP 328</td>
<td>F. proliferatum</td>
<td>Zea mays</td>
<td>USA</td>
</tr>
<tr>
<td>34</td>
<td>3270</td>
<td>F. proliferatum</td>
<td>Solanum lycopersicum crown</td>
<td>Italy</td>
</tr>
<tr>
<td>35</td>
<td>3272</td>
<td>F. proliferatum</td>
<td>Solanum lycopersicum crown</td>
<td>Italy</td>
</tr>
<tr>
<td>36</td>
<td>3274</td>
<td>F. proliferatum</td>
<td>Solanum lycopersicum crown</td>
<td>Italy</td>
</tr>
<tr>
<td>37</td>
<td>3268</td>
<td>F. proliferatum</td>
<td>Solanum lycopersicum root</td>
<td>Italy</td>
</tr>
<tr>
<td>38</td>
<td>3273</td>
<td>F. proliferatum</td>
<td>Solanum lycopersicum root</td>
<td>Italy</td>
</tr>
<tr>
<td>39</td>
<td>3275</td>
<td>F. proliferatum</td>
<td>Solanum lycopersicum root</td>
<td>Italy</td>
</tr>
<tr>
<td>40</td>
<td>4297</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera</td>
<td>Spain</td>
</tr>
<tr>
<td>41</td>
<td>4298</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera</td>
<td>Spain</td>
</tr>
<tr>
<td>42</td>
<td>4302</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera</td>
<td>Spain</td>
</tr>
<tr>
<td>43</td>
<td>4303</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera</td>
<td>Spain</td>
</tr>
<tr>
<td>44</td>
<td>2336</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera root</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>45</td>
<td>2337</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera root</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>46</td>
<td>2341</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera root</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>47</td>
<td>2342</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera root</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>48</td>
<td>2343</td>
<td>F. proliferatum</td>
<td>Phoenix dactylifera root</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td></td>
<td>ITEM</td>
<td>Species</td>
<td>Host/Plant</td>
<td>Country</td>
</tr>
<tr>
<td>---</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>49</td>
<td>ITEM 2339</td>
<td><em>F. proliferatum</em></td>
<td><em>Phoenix dactylifera leaves</em></td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>50</td>
<td>ITEM 2979</td>
<td><em>F. proliferatum</em></td>
<td><em>Mangifera indica</em></td>
<td>Malaysia</td>
</tr>
<tr>
<td>51</td>
<td>ITEM 2982</td>
<td><em>F. proliferatum</em></td>
<td><em>Mangifera indica</em></td>
<td>Malaysia</td>
</tr>
<tr>
<td>52</td>
<td>ITEM 2984</td>
<td><em>F. proliferatum</em></td>
<td><em>Mangifera indica</em></td>
<td>Malaysia</td>
</tr>
<tr>
<td>53</td>
<td>MPVP 294</td>
<td><em>F. verticillioides</em></td>
<td><em>Zea mays</em></td>
<td>Italy</td>
</tr>
<tr>
<td>54</td>
<td>ITEM 1744</td>
<td><em>F. verticillioides</em></td>
<td><em>Zea mays</em></td>
<td>Italy</td>
</tr>
<tr>
<td>55</td>
<td>NL70-7</td>
<td><em>F. oxysporum</em></td>
<td><em>Allium Cepa</em></td>
<td>UK</td>
</tr>
<tr>
<td>56</td>
<td>A28</td>
<td><em>F. oxysporum</em></td>
<td><em>Allium Cepa</em></td>
<td>UK</td>
</tr>
<tr>
<td>57</td>
<td>A25</td>
<td><em>F. oxysporum</em></td>
<td><em>Allium Cepa</em></td>
<td>UK</td>
</tr>
</tbody>
</table>

*Isolates source:*

ITEM – Isolates from the Culture collection at the Institutes of science of food production (ISPA), Italy

KF – Isolates provided by Professor Lukasz Stephen at the Polish Academy, Poland.

MPVP and F.P – Isolates were provided by Professor Naresh Magan at the Cranfield University, UK.

A, R and NL – Isolates from joint research projects at Warwick HRI – University of Bedfordshire, UK.
2.2 Revival and sub-culture of the isolates
Fungal isolates were initially cultured on the PDA plates for 3-5 days at 20°C in an incubator (Panasonic, UK) to revive them from storage. Isolates were sub-cultured on PDA plates and maintained in an incubator set at the required temperature. [Chapters 3, 4 and 5]

2.3 Preparation and use of solid and liquid media
*Fusarium* species isolates were cultured on potato dextrose agar (PDA; Sigma, 70139-500G) or in potato dextrose broth (PDB; Sigma, P6685-250G) as required for specific experiments. The media were prepared as directed by the manufacturer and autoclaved under standard conditions of 120°C and 15 PSI for 15 minutes. For routine culturing on solid medium, approximately 20 ml of autoclaved PDA at 60°C was dispensed into 9 cm diameter Petri dishes in a laminar flow bench. The medium was allowed to cool and solidify for ~20 minutes and used for inoculations in a class 2 microbiological safety cabinet or stored at 4°C for further use. For preparing liquid cultures, autoclaved PDB was dispensed into suitable containers such as microfuge tubes or 3 cm diameter Petri dishes or tissue culture flasks as required. [Chapters 3, 4 and 5]
3 Genotyping of *Fusarium proliferatum* isolates

3.1 Introduction

*Fusarium proliferatum* (belongs to the *F. fujikuroi* complex) is an emerging fungal pathogen that affects a range of important food crops such as rice, maize, onion, tomato, sugarcane and date palm in diverse geographic locations (Yan and Dickman, 1996; Abdalla et al., 2000; Stankovic et al., 2007; Khudhair et al., 2014; Torabi et al., 2014). In addition to the threat to food security, *F. proliferatum* poses a serious hazard to animal and human health by producing mycotoxins such as fumonisins (Leslie and Summerell, 2006). *Fusarium* species isolates contained within the 20 species complexes show overlapping biological attributes including colony characters, sporulation, toxin production and host range (Leslie and Summerell, 2008). In this context, the inadequacies of the Morphological Species Concept (MSC) and Biological Species Concept (BSC) for reliable identification, characterisation and diagnosis of *Fusarium* pathogens and the increased reliance on Phylogenetic Species Concept (PSC) are well documented (Summerell et al., 2003; Tibayrenc and Ayala, 2014; O’Donnell et al., 2015).

Widely known molecular markers used for phylogenetic analysis of *Fusarium* species include the internal transcribed spacer (ITS), translation elongation factor (TEF1), Calmodulin (CAL) and β tubulin (TUB), histone (H3) and actin (ACT) and recent work has also been testing the utility of RPB1 and RPB2 (Geiser et al., 2004; Vaquero et al., 2004; Nalim et al., 2009; Schmitt et al., 2009; Stepień et al., 2011a; Stockinger et al., 2014). However, the need for the development of new genetic markers for further characterization of isolates and species in evolving *Fusarium* species complexes such as *F. fujikuroi*, *F. oxysporum* and *F. solani* is well recognized (O’Donnell et al., 2015).

The aim and objectives of the work described in this chapter are focused in genotypic characterization of *F. proliferatum* isolates (part of the *F. fujikuroi* complex) from diverse hosts and geographic locations utilizing known markers as well as new marker(s) developed in this study.
3.2 Materials and methods

3.2.1 Liquid culture for genomic DNA extraction
Small blocks of mycelial mass carrying minimal agar were cut-out from sub-cultured plates of each isolate using a sterile loop or a scalpel. The mycelial blocks were introduced into 1 ml of PDB contained in 2 ml microfuge tubes. These cultures were maintained in a stationary incubator at 20°C for 5 days to induce growth. At day 5, the mycelium was harvested from the liquid medium by centrifugation for 5 minutes at 14,000 RPM (17,000 g) by retaining the pellet and the liquid was carefully drained and discarded as per biological waste disposal procedures. The mycelial mat was rinsed with 2 ml sterile water to remove any traces of the medium and the liquid discarded as above. The mycelial mat of each of the isolates was used for genomic DNA extraction either immediately or after storage at -20°C.

3.2.2 Genomic DNA extraction and quantification
Two methods were employed for the extraction of fungal genomic DNA used in PCR-based amplification of short fragments of marker genes/regions, as described below. This included a chelex-based method, and a column-based method to standardize the optimum DNA extraction method for the PCR analysis. Between these two methods, the column-based method yielded consistently better quantity and quality of DNA and this method was used for the experiments in this chapter. In addition, a third method of genomic DNA extraction was used to achieve DNA quantity and quality suitable for genome sequencing and the details of this method are presented in Chapter 5.

3.2.2.1 Chelex-based method, DNA extraction
Equal amount of chelex® (Bio-Rad; 142-1253) and acidic sand (Sigma; 274739-500G) were added to approximately 40 mg of mycelial mat in a 2 ml Eppendorf tube and 300 µl of deionised water (MilliQ) was added. The mycelial mat was ground thoroughly for approximately 15 minutes until the mycelial clumps were broken and incubated at 70°C for 10 minutes. The mixture was centrifuged at 14,000 RPM for 5 minutes to pellet the debris. Approximately 150 µl of the supernatant containing the DNA was transferred into a fresh microfuge tube carefully avoiding the debris and the DNA samples were stored at -20°C for further use.
3.2.2.2 Column-based method, DNA extraction

For the extraction of the genomic DNA using the column-based method, the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma; G2N350-1KT) was used. Prior to the use, the wash solution was prepared as directed by the manufacturer.

To approximately 40 mg mycelial mat in a 2 ml microfuge tube, equal amount of chelex®, sand and 350 µl of lysis solution A and 50 µl of lysis solution B were added from the kit. The mycelial mat was ground thoroughly using a sterile pestle for approximately 15 minutes until the mycelial clumps were broken and a homogenous mixture was achieved. Further steps were followed as described in the manufacturer protocol (Sigma; G2N350-1KT). In the final step to elute the DNA, 100 µl of sterile water prewarmed at 65°C was added to the binding column and incubated at the room temperature for 5 minutes. The DNA was eluted by centrifugation at 14,000 RPM for a minute. The elution process was repeated by loading the first eluate onto the binding column and centrifugation to ensure maximum yield and concentration of the DNA. The binding column was discarded and the genomic DNA was stored at -20°C for further use. This method was used for the PCR-based amplification and sequencing of short fragments of marker genes/regions.

3.2.3 Gel electrophoresis

Agarose gels were used for routine electrophoresis of genomic DNA (0.8%) and PCR products (1.0%) for experiments carried out in this chapter and chapter 5. The gels were prepared by adding 0.8 g or 1.0 g of agarose powder (Sigma; A9539-250G) to 100 ml 1X TAE buffer. The gel suspension was boiled in a microwave oven for approximately 3 minutes to ensure the agarose is fully melted. The gel solution was allowed to cool to approximately 50°C, 5 µl of ethidium bromide (Sigma; E1510-10ml) was added and poured into a gel tray. The comb was placed and the gel was allowed to set for at least 30 minutes. Agarose gel electrophoresis was carried out to visualize the DNA and assess the quantity, quality (e.g. RNA free) and integrity (e.g. non-degraded) in the case of genomic DNA and to assess the size and the quantity of marker gene fragments amplified by PCR.

3.2.4 Assessment of the genomic DNA

DNA samples were subjected to gel electrophoresis and Nanodrop methods to check the quantity, quality and integrity. In the electrophoretic method, known concentrations of Lambda DNA (30 ng, 60 ng, 90 ng and 120 ng) were included in the gels. DNA concentration was estimated by comparing the intensity of the florescence in each sample with the known concentration of Lambda DNA. The Nanodrop method was also used to determine the concentration of the DNA using a known concentration (50 ng/µl) of Lambda DNA as reference and the quality of each of the sample was recorded based on the 260/230 and
260/280 ratios. An example of the genomic DNA assessment by gel electrophoresis is shown in Figure 3.1.

![Figure 3.1 Agarose gel electrophoresis of genomic DNA samples](image)

Lanes 1 - 4 are test DNA samples loaded at 1 µl, lanes 5 - 8 are the same set of DNA samples loaded at 2 µl and lanes 9 - 12 are known concentrations of Lambda DNA as reference at 30 ng, 60 ng, 90 ng and 120 ng, respectively.

### 3.2.5 PCR Amplification, purification of products and quantification

Initial optimization experiments involved 20 µl PCR amplifications followed by gel electrophoresis. For generating DNA sequence data of marker regions a series of steps were followed routinely. These included an initial 20 µl PCR followed by electrophoresis to visualise the product, a scaled-up 50 µl and electrophoresis, column-based purification of the combined products and quantification of the purified DNA to determine the template concentration for sequencing.

### 3.2.6 Preparation of 20 µl and 50 µl PCR amplification reactions

Details of the various primer pairs used in PCR amplifications are presented below (Table 3.1). For 20 µl reactions, 1 µl of DNA, 1 µl forward primer (20 µM stock), 1 µl reverse primer (20 µM) and 10 µl of Biomix containing Taq DNA polymerase, dNTPs, MgCl₂, buffer and red dye (Sigma, UK) and 7 µl of deionised water were added. For negative controls with no DNA, 8 µl of deionised water was added along with all other components. For 50 µl PCR reactions, all components were scaled-up 2.5 times. PCR amplifications were performed in a thermal cycler (Bio-rad) with 35 cycles consisting of denaturation at 94°C for one minute, one minute at the annealing temperature specified for each marker (Table 3.1) and extension at 72°C for one minute. In addition, an initial denaturation at 95°C for 3 minutes and final extension at 72°C for 5 minutes was included for each amplification reaction. Following PCR, the products (5 µl volume) were resolved on 1 % agarose gel stained with ethidium bromide for 60
minutes at 70 Volts. The DNA bands were visualised under UV and the images were recorded using a gel-doc system (Bio-rad).
Table 3.1. Details of the 11 pairs of primers used in the study for initial PCR screening and subsequent amplifications

<table>
<thead>
<tr>
<th>Gene locus and function</th>
<th>Primer*</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Primer length (bp)</th>
<th>Annealing temperature (°C)</th>
<th>Estimated amplicon size (bp)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal transcribed spacer (ITS), Intergenic region of rRNA gene block</td>
<td>ITS1 F ITS4 R</td>
<td>TCCGTAGGTGAAACCTGCGG TCCCTCGCTTATGATATGC</td>
<td>19 20</td>
<td>60</td>
<td>600</td>
<td>White et al., 1990</td>
</tr>
<tr>
<td>Translation elongation factor (TEF), Protein synthesis</td>
<td>Ef728M F Ffuj_tef R</td>
<td>CATCGAGAAAGTTCGAGAAGG ATGGTTAGTTGACACGTTGAC</td>
<td>20 21</td>
<td>60</td>
<td>230</td>
<td>Carbone and Kohn, 1990</td>
</tr>
<tr>
<td></td>
<td>Ef728M F Tef1 R</td>
<td>CATCGAGAAAGTTCGAGAAGG GCCATCCTGGAGATACCAGC</td>
<td>20 21</td>
<td>61</td>
<td>600</td>
<td>Carbone and Kohn, 1990; Kullnig-Gradinger et al., 2002</td>
</tr>
<tr>
<td>Calmodulin (CAL), Affinity for calcium</td>
<td>CLOX1 F CLOX2 R</td>
<td>CAGCAAAAGCAGACACCTATAAATCTCC TTTCTCAACGACCTGTTGACT</td>
<td>27 26</td>
<td>60</td>
<td>534</td>
<td>Mulè et al., 2004a</td>
</tr>
<tr>
<td></td>
<td>VER1 F VER2 R</td>
<td>CTTCCCTGCTATTTTCTC AATTGCCATTTATATAC</td>
<td>19 26</td>
<td>59</td>
<td>578</td>
<td>Mulè et al., 2004b</td>
</tr>
<tr>
<td></td>
<td>PRO1 F PRO2 R</td>
<td>CTTCCGCAAGTTTCTC TGTCAAACTCTGACCTTGTGG</td>
<td>19 22</td>
<td>61</td>
<td>585</td>
<td>Mulè et al., 2004b</td>
</tr>
<tr>
<td></td>
<td>CL1 F CLA2 R</td>
<td>GARTWCAAGGAGGCTTCTC TTTGTACATGAGGGCC</td>
<td>20 20</td>
<td>55</td>
<td>670</td>
<td>O'Donnell et al., 1998</td>
</tr>
<tr>
<td>FUM1 gene, Fumonisin polyketide synthase</td>
<td>STE F STE R</td>
<td>CACATCTGAGGCGATCC ATATGGCCAGCTGATA</td>
<td>18 19</td>
<td>60</td>
<td>1126</td>
<td>Stepien et al., 2011</td>
</tr>
<tr>
<td></td>
<td>GLE F GLE R</td>
<td>TGAGATAAAATAGTCTTTGCGA CGTGATACGTGATTGAC</td>
<td>25 25</td>
<td>60</td>
<td>450</td>
<td>Glenn et al., 2008</td>
</tr>
<tr>
<td></td>
<td>FUM1 F FUM1 R</td>
<td>GTCTACCAGGATACCCCACAGACACATG AATACGTGCGGCGCTGCTCATG</td>
<td>27 27</td>
<td>70</td>
<td>1000</td>
<td>Karthikeyan et al., 2010</td>
</tr>
<tr>
<td></td>
<td>FUM5 F FUM6 R</td>
<td>ACACCAAAGCTCCTACAGGGA AGGTATCGGGCCACCAGCT</td>
<td>22 17</td>
<td>55</td>
<td>450</td>
<td>Flaherty et al., 2003</td>
</tr>
</tbody>
</table>

*The F and R indicate Forward and Reverse primers, respectively
3.2.7 PCR amplicon purification and quantification
Amplified DNA from the 20 and 50 µl reactions after electrophoresis was pooled together and purified using a PCR product purification kit (QIAGEN; 28006) following the manufacturer’s protocol. The purified DNA was eluted in 30 µl deionised water pre-warmed at 65°C. For quantitative estimation of the purified DNA, a 10 µl mixtures containing 2 µl purified amplicon, 2 µl 5x DNA loading dye and 6µl sigma water were electrophoresed as described earlier. The amplicon size and concentration were estimated with reference to a DNA quantification marker Easyladder 1 (Bioline; BIO-33062).

3.2.8 Nucleotide sequencing of the amplicons
DNA sequencing was carried out utilising the service providing facility at the University of Cambridge. Sequencing reactions were carried out based on Sanger method. For each sequencing reaction the template DNA was provided in 10 ul containing 20 ng of the template DNA for each 100 bases of the target fragment and the sequencing primer was provided as a 10 µM stock. For all marker regions, sequencing reactions were performed routinely using the forward primer. Any ambiguous bases were resolved by utilising sequence data generated using the reverse primer. Bioedit programme was used to view the DNA sequence trace files, edit the data where required and to export the data to other software for further analysis.

3.2.9 Bioinformatics of DNA sequences

3.2.9.1 Multiple sequence alignment of marker data
DNA sequence data generated from each isolate for each marker were exported to the Geneious package version 9.1.5 (Biomatters) and analysed with various algorithms available as plugins both for multiple sequence alignment and phylogenetic analysis, unless otherwise stated. Multiple sequences for each marker were initially aligned using MAFFY. The alignment was visually inspected, edited and/or trimmed where required to ensure the sequence data for each isolate covers the same region (i.e. starting and ending with conserved sequences both at the 5’ and 3’ ends) and realigned using the MUSCLE to generate best-fit alignment for each gene marker. These data were used to generate sequence alignment of multilocus concatenated (combined sequences) data of ITS, CAL, FUM1, TEF and FG2 markers. Two sets of alignments of the multilocus concatenated data were generated either with or without the FUM1 marker based on the set of isolates that were included in the analysis.
3.2.9.2 Phylogenetic analysis
Phylogenetic trees were generated both from the single locus and the multilocus alignment data using two different methods. Bayesian phylogenies were generated based on Markov Chain Monte Carlo (MCMC) analysis using the MrBayes plugin within Geneious (Kearse et al., 2012). The MCMC analysis was performed using the General-Time-Reversible (GTR) evolutionary model with 1,100,000 cycles set at 4 heated chains and burn-in length of 100,000 were run and sampled every 200 generations. Geneious was used to estimate 70% Consensus phylogenies at 70% were estimated from each MrBayes run where the first 25% trees were discarded as burn-in. Maximum likelihood (ML) analysis was performed using the GTR evolution model with 500 bootstrap replication analysis and the trees selected for 70% consensus phylogenies. Phylogenetic groups or clades were identified by visual inspection of the tree topology.

3.3 Experimental approaches
Fifty-two F. proliferatum isolates reflecting the global biogeographic diversity of this species have been characterized along with F. verticillioides (2 isolates) and F. oxysporum (3 isolates) used as references/out-groups (Table 2.1). Fungal universal barcode marker ITS was used as a baseline reference. Fungal housekeeping gene markers Translation elongation factor (TEF1) and Calmodulin (CAL) were used utilizing optimal primer pairs identified through an initial screening with a selected set of six isolates belonging to F. proliferatum, F. verticillioides and F. oxysporum. The fumonisin polyketide synthase gene marker FUM1 was similarly used with an optimal primer pair. FUNYBASE and the Fusarium species genomic resources available via the JGI-NCBI interface were used to carry out extensive bioinformatic analysis first to identify potentially suitable new markers for which primers were subsequently designed and initially PCR screened as above. This led to the identification and use of a new marker FG1056, which has been used to characterize all the isolates. Nucleotide sequence data generated was subjected to a range of analysis to assess the genetic diversity (e.g. number of SNPs, informative sites and haplotypes) and phylogenetic relationships (based on Bayesian posterior probabilities and Maximum likelihood bootstrap values) allied to the host and geographic diversity.
3.4 Results

3.4.1 Identification of optimal PCR primers to amplify known marker regions

3.4.1.1 Translation Elongation Factor (TEF1) gene marker
With the Translation Elongation Factor (TEF1) gene marker, two primer pairs were selected from published data in *Fusarium* species and initially screened with six isolates representing *F. proliferatum* (3), *F. verticillioides* (1) and *F. oxysporum* (2). Primer pair A (EF728M and Tef1R) yielded a ~600 bp fragment with comparable level of strong amplification in all isolates. Primer pair B (EF728M and Ffuj_tefR) yielded a much smaller fragment (~230 bp) at a lower level of amplification (Figure 3.2). Primer pair A was selected for generating the TEF1 marker nucleotide sequence data for all isolates namely *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3).

3.4.1.2 Calmodulin (CAL) gene marker
With the Calmodulin (CAL) gene marker, four primer pairs were selected from published data in *Fusarium* species and initially screened with the six isolates. These primer pairs showed varying patterns with the primer pair A (PRO 1 & 2) yielding an amplicon only with the *F. proliferatum* isolates. Primer pair B (VER 1 & 2) showed strong amplification of the *F. verticillioides* isolates and a weak product with only one of the *F. proliferatum* isolates. Primer pair C (CLOX 1 & 2) showed amplification with only the two *F. oxysporum* isolates. However, primer pair D (CL1 & CL2A) yielded consistently strong uniform sized fragment of ~670 bp with all isolates (Figure 3.3). Primer pair D was selected for generating the CAL marker nucleotide sequence data for all isolates namely *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3).
Figure 3.2 Amplification patterns of the translation elongation factor (TEF1) gene marker regions in three *Fusarium* species

Isolates using two sets of primers Isolates used were *F. verticillioides* (F.v): 1) MPVP 294; *F. proliferatum* (F.p): 2) MPVP 328, 3) ITEM 7595 and 4) A2/2; *F. oxysporum* (F.o): 5) NL70-7 and 6) A28. N, negative control without DNA; L, DNA ladder. Primer pairs used were A (EF728M and Tef1 R) and B (EF728M and Ffuj_tefR) yielding ~600 bp and ~230 bp amplicons, respectively.
Figure 3.3 Amplification pattern of the Calmodulin (CAL) gene marker regions in three *Fusarium* species isolates using four sets of primers

Isolates used were *F. verticillioides* (F.v): 1) MPVP 294; *F. proliferatum* (F.p): 2) MPVP 328, 3) ITEM 7595 and 4) A2/2; *F. oxysporum* (F.o): 5) NL70-7 and 6) A28. N, negative control without DNA; L, DNA ladder. The primer pairs used were A (PRO 1 & 2), B (VER 1 & 2), C (CLOX 1 & 2) and D (CL1 & CLA2) yielding ~585 bp, ~570 bp, ~530 bp and ~670 bp, respectively.
3.4.1.3 Fumonisin polyketide synthase (FUM1) gene marker

With the fumonisin polyketide synthase (FUM1) gene marker, four pairs were selected from published data in *Fusarium* species and initially screened with the six isolates. None of the primer pairs showed consistent amplification patterns across all isolates tested. The primer pair A (STE F & R) showed a ~1120 bp fragment with strong to very low levels of amplification in *F. proliferatum* and *F. verticillioides* isolates and a much smaller fragment (~300 bp) in *F. oxysporum*. Primer pair B (GLE F & R) amplified a ~450 bp product with only the *F. verticillioides* isolate; whereas the primer pair C (FUM1 F & R) showed poor levels and variable patterns of amplification across the isolates. Primer pair D (FUM 5 & 6) yielded a ~450 bp product with consistently strong levels of amplification across *F. proliferatum* and *F. verticillioides* isolates though no product was detected with *F. oxysporum* (Figure 3.4). As *F. proliferatum* is the focal species for this study and also in view of the uncertainty of the FUM gene cluster sequence in *F. oxysporum*, Primer pair D was selected for generating the FUM1 marker nucleotide sequence data for the *F. proliferatum* isolates and the two *F. verticillioides* reference isolates.
Figure 3.4 Amplification pattern of the FUM 1 gene marker regions in three *Fusarium* species isolates using four sets of primers

Isolates used were *F. proliferatum* (F.p): 1) MPVP 328, 2) ITEM 7595, 3) A2/2 and 4) R16; *F. oxysporum* (F.o): 5) NL70-7; *F. verticillioides* (F.v): 6) MPVP 294. N, negative control without DNA; L, DNA ladder. The primer pairs used were A (STE F & R), B (GLE F & R), C (FUM1 F & R) and D (FUM 5 & 6) yielding ~1120 bp, ~570 bp, ~1000 bp and ~450 bp, respectively.
3.4.2 Identification of novel genetic markers for *F. proliferatum* from the FUNYBASE

From the 246 single copy orthologs contained in the fungal phylogenomic database (FUNYBASE), 60 protein sequences were selected based on their mean amino acid identity values. These included 20 from the low (23.7 to 34.5%), 20 from the medium (50.2 to 53.0%) and 20 from the high (71.7 to 83.7%) mean identity values. In addition, six further sequences were selected based on a recent report of their utility in the intra- and interspecies phylogenies of Alternaria (Armitage *et al.*, 2015). So, these 66 protein sequences were downloaded from *Fusarium graminearum* as this is the only *Fusarium* species represented within the FUNYBASE. These protein sequences were used to perform a series of bioinformatic analysis utilising the available genome sequences of *Fusarium* species to evaluate the potential of the corresponding genes as novel genetic markers for *F. proliferatum*, the focal species of this study (Figure 3.5). For example, the genome sequences of four *F. fujikuroi* isolates were utilised to initially assess the intra-species variability based on the number of SNPs and Informative sites for each of the 66 genes (Table 3.2). This led to the identification 20 genes which showed higher levels of variability compared to the data for ITS, TEF1 and CAL markers in *F. fujikuroi* isolates. From these data sets, three gene loci MS550, FG1056 and FG644 showing the highest levels of variability were identified (Table 3.3) as potential new gene markers. Subsequently, genome sequences of five different *Fusarium* species were analysed to identify conserved sequences flanking the variable regions for each of the three selected marker genes. And these conserved sequences were used to design new primer pairs (Table 3.4; Table 3.5).

The newly designed primer pairs for the novel loci were tested in PCR and the amplification patterns for the gene loci MS550 and FG644 were inconsistent across the *F. proliferatum* (4), *F. verticillioides* (1) and *F. oxysporum* (1) isolates screened. In contrast, with the gene locus FG1056, consistent amplification pattern was recorded for all six isolates (e.g. C-FG1056, PS1, Figure 3.6) and the amplicons yielded up to 800 base nucleotide sequence reads (data not shown). To enable resource efficient and robust PCR – DNA sequencing of the full set of isolates in this study, two primer sets targeting amplicon size of smaller than 500 bp were designed from these sequences. Primer set 2 (PS2, FG1056-2 F & R) as well as the primer set 3 (PS3, FG1056-3 F & R) yielded consistent amplifications across all six isolates (Figure 3.5) and the primer set 2 was used for generating the nucleotide sequence data of the new genetic marker FG1056 for all isolates namely *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3) in the study.
Figure 3.5 Flow diagram summarizing the key steps used in the identification of a novel marker gene for genotyping 52 Fusarium proliferatum isolates representing wide biogeographic diversity along with F. verticillioides (2) and F. oxysporum (3) reference.
Table 3.2 Bioinformatics analysis results for the 66 protein/gene loci selected from the FUNYBASE reflecting the number of SNPs and their location in different regions of the genes

<table>
<thead>
<tr>
<th>No.</th>
<th>FUNYBASE* Protein/Gene Id.</th>
<th>Protein** sequence identity</th>
<th>Gene size (bases)</th>
<th>SNPs/Informative sites</th>
<th>SNPs in gene regions+ (bases)</th>
<th>Similarity§ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MS393</td>
<td>23.7</td>
<td>1937</td>
<td>1/0</td>
<td>1 in F 100</td>
<td>93-96%</td>
</tr>
<tr>
<td>2</td>
<td>MS401</td>
<td>26.7</td>
<td>3576</td>
<td>3/0</td>
<td>2 in M 800</td>
<td>92-97%</td>
</tr>
<tr>
<td>3</td>
<td>MS547</td>
<td>26.8</td>
<td>3357</td>
<td>9/3</td>
<td>5 in L 900</td>
<td>90-96%</td>
</tr>
<tr>
<td>4</td>
<td>FG819</td>
<td>27.1</td>
<td>1913</td>
<td>2/1</td>
<td>2 in L 1200</td>
<td>93-98%</td>
</tr>
<tr>
<td>5</td>
<td>MS487</td>
<td>28</td>
<td>1140</td>
<td>1/0</td>
<td>1 in L 200</td>
<td>93-96%</td>
</tr>
<tr>
<td>6</td>
<td>MS578</td>
<td>28.1</td>
<td>5444</td>
<td>11/0</td>
<td>8 in L 1800</td>
<td>93-97%</td>
</tr>
<tr>
<td>7</td>
<td>FG1000</td>
<td>28.3</td>
<td>2944</td>
<td>10/1</td>
<td>4 in M 900</td>
<td>92-95%</td>
</tr>
<tr>
<td>8</td>
<td>FG1020</td>
<td>31</td>
<td>5254</td>
<td>6/0</td>
<td>4 in L 2000</td>
<td>94-96%</td>
</tr>
<tr>
<td>9</td>
<td>MS431</td>
<td>32.1</td>
<td>3562</td>
<td>2/0</td>
<td>1 in F 500 &amp; 1 in L 500</td>
<td>94-97%</td>
</tr>
<tr>
<td>10</td>
<td>MS447</td>
<td>32.1</td>
<td>3044</td>
<td>1/1</td>
<td>1 in F 360</td>
<td>91-96%</td>
</tr>
<tr>
<td>11</td>
<td>MS378</td>
<td>32.4</td>
<td>5362</td>
<td>11/0</td>
<td>7 in L 2300</td>
<td>93-97%</td>
</tr>
<tr>
<td>12</td>
<td>MS541</td>
<td>32.5</td>
<td>3703</td>
<td>7/2</td>
<td>5 in M 840</td>
<td>88-96%</td>
</tr>
<tr>
<td>13</td>
<td>FG535</td>
<td>32.9</td>
<td>2633</td>
<td>4/0</td>
<td>4 in 1000</td>
<td>93-96%</td>
</tr>
<tr>
<td>14</td>
<td>MS517</td>
<td>33</td>
<td>666</td>
<td>2/0</td>
<td>2 in M 120</td>
<td>94-96%</td>
</tr>
<tr>
<td>15</td>
<td>MS377</td>
<td>33.1</td>
<td>2458</td>
<td>8/0</td>
<td>6 in M 660</td>
<td>94-98%</td>
</tr>
<tr>
<td>16</td>
<td>FG746</td>
<td>33.2</td>
<td>848</td>
<td>2/1</td>
<td>2 in L 200</td>
<td>92-96%</td>
</tr>
<tr>
<td>17</td>
<td>MS497</td>
<td>33.2</td>
<td>3239</td>
<td>5/1</td>
<td>4 in L 860</td>
<td>94-97%</td>
</tr>
<tr>
<td>18</td>
<td>MS400</td>
<td>34.2</td>
<td>1759</td>
<td>10/4</td>
<td>5 in L 600</td>
<td>92-93%</td>
</tr>
<tr>
<td>19</td>
<td>FG1056</td>
<td>34.2</td>
<td>2463</td>
<td>23/5</td>
<td>16 in L 900</td>
<td>89-93%</td>
</tr>
<tr>
<td>20</td>
<td>MS345</td>
<td>34.5</td>
<td>6099</td>
<td>4/2</td>
<td>2 in L 240</td>
<td>93-96%</td>
</tr>
<tr>
<td>21</td>
<td>FG893</td>
<td>50.2</td>
<td>1929</td>
<td>11/1</td>
<td>10 in L 960</td>
<td>95-96%</td>
</tr>
<tr>
<td>22</td>
<td>FG579</td>
<td>50.5</td>
<td>1060</td>
<td>1/1</td>
<td>1 in F 60</td>
<td>94-98 %</td>
</tr>
<tr>
<td>23</td>
<td>FG756</td>
<td>50.5</td>
<td>1784</td>
<td>0/0</td>
<td>NA</td>
<td>94-96 %</td>
</tr>
<tr>
<td>24</td>
<td>FG723</td>
<td>50.6</td>
<td>1971</td>
<td>2/1</td>
<td>2 in M 900</td>
<td>87-98%</td>
</tr>
<tr>
<td>25</td>
<td>FG635</td>
<td>50.7</td>
<td>1415</td>
<td>3/3</td>
<td>3 in L 960</td>
<td>93-97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>MS384</td>
<td>50.7</td>
<td>3370</td>
<td>7/1</td>
<td>5 in M 900</td>
<td>94-96 %</td>
</tr>
<tr>
<td>27</td>
<td>MS333</td>
<td>51.2</td>
<td>1460</td>
<td>1/0</td>
<td>1 in F 360</td>
<td>90-98%</td>
</tr>
<tr>
<td>28</td>
<td>FG832</td>
<td>51.5</td>
<td>2700</td>
<td>5/2</td>
<td>4 in F 300</td>
<td>90-98 %</td>
</tr>
<tr>
<td>29</td>
<td>MS353</td>
<td>51.9</td>
<td>2300</td>
<td>10/2</td>
<td>7 in L 540</td>
<td>92-93%</td>
</tr>
<tr>
<td>30</td>
<td>FG708</td>
<td>51.9</td>
<td>3300</td>
<td>5/0</td>
<td>3 in M 400</td>
<td>95-97%</td>
</tr>
<tr>
<td>31</td>
<td>FG901</td>
<td>52.1</td>
<td>990</td>
<td>7/0</td>
<td>6 in L 700</td>
<td>92-94%</td>
</tr>
<tr>
<td>32</td>
<td>MS430</td>
<td>52.2</td>
<td>1540</td>
<td>14/4</td>
<td>7 in M 780</td>
<td>91-94%</td>
</tr>
<tr>
<td>33</td>
<td>FG529</td>
<td>52.3</td>
<td>1720</td>
<td>8/3</td>
<td>7 in L 780</td>
<td>93-94%</td>
</tr>
<tr>
<td>34</td>
<td>FG716</td>
<td>52.4</td>
<td>3050</td>
<td>6/0</td>
<td>4 in F 400</td>
<td>92-97%</td>
</tr>
<tr>
<td>35</td>
<td>FG948</td>
<td>52.4</td>
<td>1500</td>
<td>1/1</td>
<td>1 in F 500</td>
<td>91-95%</td>
</tr>
<tr>
<td>36</td>
<td>MS411</td>
<td>52.5</td>
<td>1220</td>
<td>0</td>
<td>NA</td>
<td>97-99%</td>
</tr>
<tr>
<td>37</td>
<td>MS375</td>
<td>52.8</td>
<td>1960</td>
<td>2/0</td>
<td>2 in F 400</td>
<td>95-98%</td>
</tr>
<tr>
<td>38</td>
<td>FG897</td>
<td>52.8</td>
<td>1680</td>
<td>4/4</td>
<td>4 in L 900</td>
<td>91-95%</td>
</tr>
<tr>
<td>39</td>
<td>FG834</td>
<td>52.9</td>
<td>2750</td>
<td>4/2</td>
<td>3 in M 600</td>
<td>94-97%</td>
</tr>
<tr>
<td>40</td>
<td>FG964</td>
<td>52.9</td>
<td>2040</td>
<td>5/1</td>
<td>3 in L 600</td>
<td>92-96%</td>
</tr>
<tr>
<td>41</td>
<td>FG598</td>
<td>53</td>
<td>2200</td>
<td>0</td>
<td>NA</td>
<td>96-98%</td>
</tr>
<tr>
<td>42</td>
<td>FG507</td>
<td>71.7</td>
<td>2841</td>
<td>5/1</td>
<td>4 in L 900</td>
<td>95-97%</td>
</tr>
<tr>
<td>43</td>
<td>FG465</td>
<td>72.2</td>
<td>1445</td>
<td>6/0</td>
<td>5 in L 600</td>
<td>96-98%</td>
</tr>
<tr>
<td>44</td>
<td>FG551</td>
<td>73.2</td>
<td>2037</td>
<td>4/0</td>
<td>3 in L 900</td>
<td>89-96%</td>
</tr>
<tr>
<td>45</td>
<td>FG651</td>
<td>73.2</td>
<td>2313</td>
<td>5/1</td>
<td>3 in M 800</td>
<td>95-98%</td>
</tr>
<tr>
<td>46</td>
<td>MS305</td>
<td>73.4</td>
<td>1141</td>
<td>4/0</td>
<td>3 in M 700</td>
<td>89-94%</td>
</tr>
<tr>
<td>47</td>
<td>FG691</td>
<td>73.6</td>
<td>2437</td>
<td>1/0</td>
<td>In L 600</td>
<td>96-98%</td>
</tr>
<tr>
<td>48</td>
<td>FG696</td>
<td>74</td>
<td>3261</td>
<td>1/0</td>
<td>1 in L 600</td>
<td>93-95%</td>
</tr>
<tr>
<td>49</td>
<td>MS422</td>
<td>74.1</td>
<td>2138</td>
<td>4/0</td>
<td>3 in L 800</td>
<td>95-99%</td>
</tr>
<tr>
<td>50</td>
<td>FG459</td>
<td>74.2</td>
<td>2595</td>
<td>3/2</td>
<td>2 in F 600</td>
<td>93-98%</td>
</tr>
<tr>
<td>51</td>
<td>MS407</td>
<td>74.4</td>
<td>2380</td>
<td>2/1</td>
<td>2 in M 1100</td>
<td>93-96%</td>
</tr>
<tr>
<td>52</td>
<td>FG542</td>
<td>76.1</td>
<td>1791</td>
<td>3/2</td>
<td>3 in M 900</td>
<td>93-98%</td>
</tr>
<tr>
<td>53</td>
<td>FG612</td>
<td>76.5</td>
<td>1360</td>
<td>0/0</td>
<td>NA</td>
<td>93-97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>FG861</td>
<td>77</td>
<td>3548</td>
<td>11/2</td>
<td>6 in L 960</td>
<td>94-97%</td>
</tr>
<tr>
<td>55</td>
<td>FG644</td>
<td>77.5</td>
<td>3089</td>
<td>40/3</td>
<td>30 in M 860</td>
<td>94-97%</td>
</tr>
<tr>
<td>56</td>
<td>FG533</td>
<td>77.9</td>
<td>2016</td>
<td>6/3</td>
<td>3 in L 600</td>
<td>93-94%</td>
</tr>
<tr>
<td>57</td>
<td>FG850</td>
<td>78.2</td>
<td>2149</td>
<td>1/0</td>
<td>1 in L 700</td>
<td>95-96%</td>
</tr>
<tr>
<td>58</td>
<td>FG503</td>
<td>78.5</td>
<td>3604</td>
<td>6/1</td>
<td>4 in M 1440</td>
<td>94-97%</td>
</tr>
<tr>
<td>59</td>
<td>FG699</td>
<td>81.9</td>
<td>1870</td>
<td>0</td>
<td>NA</td>
<td>94-98%</td>
</tr>
<tr>
<td>60</td>
<td>MS418</td>
<td>83.4</td>
<td>1016</td>
<td>0</td>
<td>NA</td>
<td>91-97%</td>
</tr>
<tr>
<td>61</td>
<td>FG556</td>
<td>83.7</td>
<td>2345</td>
<td>4/2</td>
<td>2 in F 180</td>
<td>96-97%</td>
</tr>
<tr>
<td>62</td>
<td>MS550</td>
<td>35</td>
<td>1543</td>
<td>6/2</td>
<td>6 in L 840</td>
<td>92-95%</td>
</tr>
<tr>
<td>63</td>
<td>MS320</td>
<td>40</td>
<td>2120</td>
<td>12/2</td>
<td>6 in F 720</td>
<td>92-95%</td>
</tr>
<tr>
<td>64</td>
<td>MS294</td>
<td>44.2</td>
<td>4391</td>
<td>8/1</td>
<td>6 in L 840</td>
<td>89-94%</td>
</tr>
<tr>
<td>65</td>
<td>FG864</td>
<td>48.1</td>
<td>1655</td>
<td>22/9</td>
<td>16 in L 600</td>
<td>91-93%</td>
</tr>
<tr>
<td>66</td>
<td>MS432</td>
<td>48.6</td>
<td>1402</td>
<td>3/3</td>
<td>2 in M 700</td>
<td>91-95%</td>
</tr>
</tbody>
</table>

* Gene loci shaded in grey were also used by Armitage et al., 2015

** Protein sequence identity refers to the amino acid identity across all fungal species in the FUNYBASE

* F, M and L refer to the first, middle and last parts of the nucleotide sequence of the genes

NA, not applicable

§ BlastN values based on the level of nucleotide similarity for each gene from the genome sequences available for four *Fusarium fujikuroi* isolates.
Table 3.3 The nucleotide base differences reflected by the number of SNPs among the 20 genes selected from FUNYBASE and the SNP locations in the genes, along with comparable information for known genetic markers

<table>
<thead>
<tr>
<th>Gene name (code)*</th>
<th>SNPs / Informative sites in the gene / marker region size (bases)</th>
<th>SNP rich regions in the gene** (bases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FG832</td>
<td>5 /2 in 2700</td>
<td>4 in F 300</td>
</tr>
<tr>
<td>FG864</td>
<td>6 /2 in 1543</td>
<td>6 in L 840</td>
</tr>
<tr>
<td>FG901</td>
<td>7 /0 in 900</td>
<td>7 in 900</td>
</tr>
<tr>
<td>MS384</td>
<td>7 /1 in 3370</td>
<td>5 in M 900</td>
</tr>
<tr>
<td>MS541</td>
<td>7 /2 in 3703</td>
<td>5 in M 840</td>
</tr>
<tr>
<td>MS377</td>
<td>8 /0 in 2458</td>
<td>6 in M 600</td>
</tr>
<tr>
<td>MS320</td>
<td>8 /1 in 4391</td>
<td>6 in L 840</td>
</tr>
<tr>
<td>FG529</td>
<td>8 /3 in 1720</td>
<td>7 in L 780</td>
</tr>
<tr>
<td>MS547</td>
<td>9 /3 in 3357</td>
<td>5 in L 900</td>
</tr>
<tr>
<td>FG1000</td>
<td>10 /1 in 2944</td>
<td>4 in M 900</td>
</tr>
<tr>
<td>MS353</td>
<td>10 /2 in 2300</td>
<td>7 in L 540</td>
</tr>
<tr>
<td>MS400</td>
<td>10 /4 in 1759</td>
<td>5 in L 500</td>
</tr>
<tr>
<td>FG893</td>
<td>11 /1 in 1929</td>
<td>10 in M 960</td>
</tr>
<tr>
<td>FG861</td>
<td>11 /2 in 3548</td>
<td>6 in L 960 bp</td>
</tr>
<tr>
<td>MS378</td>
<td>11 /1 in 5362</td>
<td>6 in L 840</td>
</tr>
<tr>
<td>MS294</td>
<td>12 /2 in 2120</td>
<td>6 in F 720</td>
</tr>
<tr>
<td>MS430</td>
<td>14 /4 in 1540</td>
<td>7 in M 780</td>
</tr>
<tr>
<td>MS550</td>
<td>22 /9 in 1655</td>
<td>16 in L 600</td>
</tr>
<tr>
<td>FG1056</td>
<td>23 /5 in 2463</td>
<td>16 in L 900</td>
</tr>
<tr>
<td>FG644</td>
<td>40 /3 in 3089</td>
<td>30 in M 860</td>
</tr>
<tr>
<td>CAL</td>
<td>1 in 718</td>
<td>1 in 718</td>
</tr>
<tr>
<td>TEF1</td>
<td>3 in 489</td>
<td>3 in 489</td>
</tr>
<tr>
<td>ITS1</td>
<td>1 in 600</td>
<td>1 in 600</td>
</tr>
</tbody>
</table>

* The grey shades indicate known markers Calmodulin (CAL), Translation Elongation Factor (TEF1) and the Internal transcribed spacer (ITS1) widely used in *Fusarium species*; full details of genes and their putative functions are available in FUNYBASE

** F, M and L refer to the first, middle and last parts of the nucleotide sequences of the genes
Table 3.4 Details of the *Fusarium* species isolates for which the genome sequences available at NCBI were used to locate the variable regions in the new marker genes identified and to design new primers pairs

<table>
<thead>
<tr>
<th>No.</th>
<th>Species</th>
<th>Host</th>
<th>Geographic location</th>
<th>Accession number</th>
<th>Purpose in bioinformatics analysis and work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>F. fujikuroi</em></td>
<td>Rice</td>
<td>Taiwan</td>
<td>JRVF01000729.1</td>
<td>Identification of variable region and primer design</td>
</tr>
<tr>
<td>2</td>
<td><em>F. fujikuroi</em></td>
<td>Rice</td>
<td>Thailand</td>
<td>JRVH01001192.1</td>
<td>Identification of variable region and primer design</td>
</tr>
<tr>
<td>3</td>
<td><em>F. fujikuroi</em></td>
<td>Little Bluestem</td>
<td>Kansas (USA)</td>
<td>JRVG000000000.1</td>
<td>Identification of variable region and primer design</td>
</tr>
<tr>
<td>4</td>
<td><em>F. fujikuroi</em></td>
<td>Rice</td>
<td>South Korea</td>
<td>KB205953.1</td>
<td>Identification of variable region and primer design</td>
</tr>
<tr>
<td>5</td>
<td><em>F. nygamai</em></td>
<td>unknown</td>
<td>unknown</td>
<td>JF776866.1</td>
<td>Primer design based on conserved flanking sequences</td>
</tr>
<tr>
<td>6</td>
<td><em>F. oxysporum</em></td>
<td>Corm</td>
<td>China</td>
<td>AMGP01000499.1</td>
<td>Primer design based on conserved flanking sequences</td>
</tr>
<tr>
<td>7</td>
<td><em>F. oxysporum</em></td>
<td>Seeds</td>
<td>Switzerland</td>
<td>JNNQ000000000.1</td>
<td>Primer design based on conserved flanking sequences</td>
</tr>
<tr>
<td>8</td>
<td><em>F. verticilloides</em></td>
<td>unknown</td>
<td>unknown</td>
<td>AAIM2000120.1</td>
<td>Primer design based on conserved flanking sequences</td>
</tr>
<tr>
<td>9</td>
<td><em>F. circinatum</em></td>
<td>Pinus radiata</td>
<td>USA</td>
<td>CM004511.1</td>
<td>Primer design based on conserved flanking sequences</td>
</tr>
<tr>
<td>10</td>
<td><em>F. circinatum</em></td>
<td>Pine</td>
<td>USA</td>
<td>JRVE000000000.1</td>
<td>Primer design based on conserved flanking sequences</td>
</tr>
</tbody>
</table>
Table 3.5 Details of the new primers designed for the new genetic markers identified in this study utilising the FUNYBASE and the genome sequences of diverse *Fusarium* species in JGI - NCBI databases

<table>
<thead>
<tr>
<th>Gene code/ Primer name</th>
<th>Putative gene function</th>
<th>Primer sequence (5' to 3')</th>
<th>Primer length (bp)</th>
<th>Annealing temp (°C)</th>
<th>Amplicon size (approx.) bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS550 R</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG644 F</td>
<td>Actin and tubulin assembly</td>
<td>A C A T G A T G C T A G C T C C A C C C G</td>
<td>20</td>
<td>63</td>
<td>660</td>
</tr>
<tr>
<td>FG644 R</td>
<td></td>
<td>T C T T C T C A A C C T T G G C C T C</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG1056 F</td>
<td>DNA repair and telomere stability</td>
<td>G C T T G T C T T G C G A A G G A T G G G A C C A T C T G G C T A T C C T C G</td>
<td>20</td>
<td>60</td>
<td>800</td>
</tr>
<tr>
<td>FG1056 R</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG1056-2 R</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FG1056-3 F</td>
<td></td>
<td>T C G T G A C C C A G A C T A C A A G</td>
<td>20</td>
<td>62</td>
<td>400</td>
</tr>
<tr>
<td>FG1056-3 R</td>
<td></td>
<td>C A T G T C G G T T A T G A A G A G C G</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The F and R in this table indicate forward and reverse primers, respectively; the grey shade indicates the reverse prime; PS1-3 refers to three different primer pairs.
Figure 3.6 Amplification pattern of the variable regions in three Fusarium species isolates using five sets of primers for the three novel marker gene loci identified from the FUNYBASE.

Isolates used were F. proliferatum (F.p): 1) MPVP 328, 2) ITEM 7595, 3) A2/2 and 4) R16; F. verticillioides (F.v): 5) MPVP 294; F. oxysporum (F.o): 6) NL70-7. N, negative control without DNA; L, DNA ladder. Primer pairs used were A (MS550 F & R), B (FG644 F & R), C-FG1056 (PS1, PS2 and PS3). With the FG1056 marker gene, primers pairs PS1, PS2 and PS3 amplified ~ 1000 bp, ~ 480 bp and ~ 400 bp fragments, respectively.
3.4.3 Evaluation of the genetic diversity and phylogenetic relationships in *F. proliferatum* representing wide biogeographic diversity

Genotypic characterisation of 52 *F. proliferatum* isolates (part of *fujikuroi* species complex) representing diverse host and geographic origins was carried out using five genetic markers. These included the new genetic marker FG1056 (gene involved in DNA repair and telomere stability) developed in this study, fungal universal DNA barcode marker ITS (rRNA gene block internal transcribed spacer), widely used housekeeping gene markers TEF1 (translation elongation factor) and CAL (Calmodulin) and the fumonisin biosynthetic gene FUM1 (fumonisin polyketide synthase). Along with the *F. proliferatum* isolates, two isolates of *F. verticillioides* (also belonging to *fujikuroi* species complex) and three isolates of *F. oxysporum* (part of *oxysporum* species complex) were used as references or outgroups. Raw sequence data were checked for accuracy and to resolve any ambiguities using Bioedit software. Multiple sequence alignments were generated for each marker, manually inspected and edited to ensure optimal alignment. End-trimmed multiple sequence alignments with uniform start and end sequences for each isolate were used in phylogeny reconstruction. Two widely followed molecular phylogenetic methodologies namely Bayesian analysis based on Markov Chain Monte Carlo (MCMC) posterior probability (BA) and Maximum likelihood analysis based on parsimony with bootstrap values (ML) were used. In general, the overall tree topologies were comparable with each of the makers as well as the concatenated data combining multiple markers. Phylogenetic trees generated with BA, however, provided a higher level of resolution compared to the ML phylogenies. Overall, both the BA and ML phylogenetic trees exhibited three major clades with the focus species of this study represented by major Clade A comprising *F. proliferatum* isolates (52) and major Clades B and C representing isolates belonging to *F. verticillioides* (2) and *F. oxysporum* (3), respectively. Isolates genetic groups (clades) were identified based on posterior probability or bootstrap values of 70 % or above (values shown on the nodes of the trees), whilst other isolates showing close relationships but not supported at the set value of 70 % were categorised as genetic clusters. Single nucleotide polymorphisms (SNPs) reflecting the level of base substitutions and the Informative sites (ISs) reflecting substitution of a nucleotide base in at least two isolates or more were computed for each of the genetic makers among the *F. proliferatum* isolates (52) and corresponding data was collated relative to the reference species *F. verticillioides* (2) and *F. oxysporum* (3). Isolates with one or more nucleotide base substitutions were identified as distinctive haplotypes.
3.4.3.1 Internal transcribed spacer (ITS) region

The end-trimmed sequence of the internal transcribed spacer (ITS) genetic marker was ~454 bases in length. Multiple sequence alignment of the data from 57 isolates belonging to *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3) showed 54 SNPs and 52 ISs (informative sites). Bayesian phylogenetic tree, generated from the multiple sequence alignment, with 70 % posterior probability value (Figure 3.7) distinguished the species as three clades with Clade A representing *F. proliferatum* isolates (52) and Clades B and C representing isolates belonging to *F. verticillioides* (2) and *F. oxysporum* (3), respectively (Figure 3.7). The *F. proliferatum* isolates were clustered together with 78 – 100 % posterior probability values (PPV) comprising a large cluster of 47 isolates from diverse host and geographic origins. Group A1 represented 5 isolates originating from onion crops in the UK and Italy with 78 % PPV. The Maximum likelihood analysis distinguished the species into three clades as above with 92 to 100 % bootstrap values (BSV), but the 52 isolates belonging to *F. proliferatum* were all clustered together with no genetic groups apparent (Figure 3.8).
Figure 3.7 Bayesian consensus tree based on the internal transcribed spacer (ITS) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (Cluster 1 and group A1), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1.
Figure 3.8 Maximum likelihood consensus tree based on the internal transcribed spacer (ITS) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1.
3.4.3.2 Translation elongation factor (TEF1) gene

The end-trimmed sequence of the translation elongation factor (TEF1) genetic marker was \(~358\) bases in length. Multiple sequence alignment of the data from 57 isolates belonging to \(F.\ proliferatum\) (52), \(F.\ verticillioides\) (2) and \(F.\ oxysporum\) (3) showed 36 SNPs and 32 ISs. Bayesian phylogenetic tree, generated from the multiple sequence alignment, with 70 % posterior probability value distinguished the species as three clades with Clade A representing \(F.\ proliferatum\) isolates (52) and Clades B and C representing isolates belonging to \(F.\ verticillioides\) (2) and \(F.\ oxysporum\) (3), respectively (Figure 3.9). The \(F.\ proliferatum\) isolates within clade A revealed four genetic groups with 90 – 100 % PPV including 17 isolates with no clear relationships to their host but some links to their geographic origin. The other 35 isolates originating from diverse host and geographic locations were represented as two clusters (Figure 3.9). The Maximum likelihood analysis distinguished the species into three clades as above with 71 to 100 % bootstrap values (BSV), but 50 isolates belonging to \(F.\ proliferatum\) were clustered together with only one Group of two isolates apparent (Figure 3.10).
Figure 3.9 Bayesian consensus tree based on the translation elongation factor (TEF1) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1.
Figure 3.10 Maximum likelihood consensus tree based on the translation elongation factor (TEF1) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
3.4.3.3 Calmodulin (CAL) gene

The end-trimmed sequence of the Calmodulin (CAL) genetic marker was ~650 bases in length. Multiple sequence alignment of the data from 57 isolates belonging to *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3) showed 129 SNPs and 62 ISs.

Bayesian phylogenetic tree, generated from the multiple sequence alignment, with 70 % posterior probability value distinguished the species as three clades with Clade A representing *F. proliferatum* isolates (52) and Clades B and C representing isolates belonging to *F. verticillioides* (2) and *F. oxysporum* (3), respectively (Figure 3.11). The *F. proliferatum* isolates within clade A revealed six genetic Groups A1 to A6 85 – 100 % PPV. These groups showed two different patterns, with three groups each including of isolates originating from the same host and geographic location. The other three groups each included isolates from different host and geographic origins. A further 17 isolates were clustered together indicating their genetic similarity (Figure 3.11). The Maximum likelihood analysis distinguished the species into three clades as above with 98 to 100 % bootstrap values (BSV), but 48 isolates belonging to *F. proliferatum* were clustered together with only one Group of four isolates apparent (Figure 3.12).
Figure 3.11 Bayesian consensus tree based on the Calmodulin (CAL) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
Figure 3.12 Maximum likelihood consensus tree based on the Calmodulin (CAL) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
3.4.3.4 Fumonisin polyketide synthase (FUM1) gene

The end-trimmed sequence of the fumonisin polyketide synthase (FUM1) gene marker was ~287 bases in length. The FUM1 PCR-sequencing approach did not work with 4 *F. proliferatum* isolates and the 3 *F. oxysporum* isolates. Multiple sequence alignment of the data from 50 isolates belonging to *F. proliferatum* (48) and *F. verticillioides* (2) showed 69 SNPs and 58 informative regions. Bayesian phylogenetic tree, generated from the multiple sequence alignment, with 70 % posterior probability value distinguished the species as two clades with Clade A representing *F. proliferatum* isolates (48) and Clades B representing the 2 isolates belonging to *F. verticillioides* (Figure 3.13). The *F. proliferatum* isolates within clade A revealed three genetic Groups A1 to A3 with 92 – 98 % PPV, including 8 isolates among which a set of 4 isolates originating from date palm in Saudi Arabia belonged to one genetic group. The remaining 40 isolates originating from diverse hosts and geographic locations were clustered together indicting their close genetic relationships (Figure 3.13). The Maximum likelihood analysis distinguished the species into two clades as above, but among the 48 isolates belonging to *F. proliferatum*, 46 were clustered together with only one group of two isolates apparent (Figure 3.14).
Figure 3.13 Bayesian consensus tree based on the fumonisin polyketide synthase (FUM1) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among 48 isolates of *F. proliferatum* (A) and 2 of *F. verticillioides* (B). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
Figure 3.14 Maximum likelihood consensus tree based on the fumonisin polyketide synthase (FUM1) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among 48 isolates of *F. proliferatum* (A) and 2 of *F. verticillioides* (B). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
3.4.3.5 FG1056 (DNA repair and telomere stability) gene

FG1056 (gene involved in repair of DNA breaks and telomere stability) is a new genetic marker developed specifically for *F. proliferatum* in this study from the FUNYBASE, also utilising the genomic resources available for *Fusarium* species via the BROAD institute database via the NCBI interface.

The end-trimmed sequence of the genetic marker FG1056 was ~485 bases in length. Multiple sequence alignment of the data from 57 isolates belonging to *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3) showed 190 SNPs and 118 ISs. Bayesian phylogenetic tree, generated from the multiple sequence alignment, with 70 % posterior probability value distinguished the species as three clades with Clade A representing *F. proliferatum* isolates (52) and Clades B and C representing isolates belonging to *F. verticillioides* (2) and *F. oxysporum* (3), respectively (Figure 3.15). The *F. proliferatum* isolates within clade A revealed 10 genetic groups with 72 – 100 % PPV and 2 clusters. The genetic groups showed different patterns with at least 4 groups each including isolates originating from the same host and geographic location, e.g. group of isolates from garlic in Poland. In some of the other groups, although an association to host and geographic location appeared to exist, the relationships were less clear. Out of 52 *F. proliferatum* isolates, 22 belonged to two clusters comprising isolates of diverse origin (Figure 3.15). The Maximum likelihood analysis distinguished the species into three clades as above, with 95 – 100 % bootstrap values. Five genetic groups with 73 – 100 % BSV were formed from among the 52 *F. proliferatum* isolates, with at least 3 groups each including isolates from same host and geographic origin (Figure 3.16).
Figure 3.15 Bayesian consensus tree based on the FG1056 (gene involved in DNA repair and telomere stability) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
Figure 3.16 Maximum likelihood consensus tree based on the FG1056 (gene involved in DNA repair and telomere stability) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
3.4.3.6 Multilocus sequence typing excluding FUM1
The end-trimmed data from each genetic marker namely ITS, TEF1, CAL and FG1056 were concatenated and the total sequence length was ~1956 bases. Multiple sequence alignment of the data from 57 isolates belonging to *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3) showed 419 SNPs and 270 informative regions. Bayesian phylogenetic tree, generated from the multiple sequence alignment, with 70 % posterior probability value distinguished the species as three clades with Clade A representing *F. proliferatum* isolates (52) and Clades B and C representing isolates belonging to *F. verticillioides* (2) and *F. oxysporum* (3), respectively (Figure 3.17). The *F. proliferatum* isolates within Clade A revealed 10 genetic groups (A1 to A10). At least four of these groups each included isolates originating from the same host and geographic location e.g. 3 isolates from garlic in Poland. In some of the other groups, some association to host and geographic origin appeared to exist e.g. in one of the genetic groups 6 of the 7 isolates were from tomato in Italy. Another group included 6 isolates from onion from UK and Yugoslavia (Figure 3.17). The Maximum likelihood analysis distinguished the isolates into three clades as above, with 100 % BSV, but offered a lesser level of resolution with only four groups based on 72 – 100 % BSV. Two of the groups each included isolates originating from the same host and geographic location (Figure 3.18).
Figure 3.17 Bayesian consensus tree based on multilocus (ITS, TEF1, CAL and FG1056) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1.
Figure 3.18 Maximum likelihood consensus tree based on multilocus (ITS, TEF1, CAL and FG1056) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among the 52 isolates of *F. proliferatum* (A), 2 of *F. verticillioides* (B) and 3 of *F. oxysporum* (C). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1.
3.4.3.7 Multilocus sequence typing including FUM1

The end-trimmed data from each genetic marker namely ITS, TEF1, CAL, FG1056 and FUM1 were concatenated and the total sequence length was ~2238 bases. Multiple sequence alignment of the data from 50 isolates belonging to *F. proliferatum* (48) and *F. verticillioides* (2) showed 488 SNPs and 328 ISs. Bayesian phylogenetic tree generated from the multiple sequence alignment, with 70 % posterior probability value distinguished the species into two clades with Clade A representing *F. proliferatum* isolates (48) and Clade B with the two *F. verticillioides* isolates (Figure 3.19). The *F. proliferatum* isolates within Clade A revealed 10 genetic groups (A1 – A10) and one cluster of five isolates. At least five of these groups each included isolates originating from the same host and geographic location e.g. 4 isolates from Asparagus in Italy. In some of the other groups, some association to host and geographic origin appeared to exist. Other genetic groups included isolates originating from diverse hosts and geographic locations. The Maximum likelihood analysis distinguished the isolates into two clades as above, but offered a lesser level of resolution with only five groups based on 71 – 100 % BSV identified within clade A. Three of the groups each included isolates originating from the same host and geographic location (Figure 3.20).
Figure 3.19 Bayesian consensus tree based on multilocus (ITS, TEF1, CAL, FG1056 and FUM1) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among 48 isolates of *F. proliferatum* (A) and 2 of *F. verticillioides* (B). The numbers above each branch indicate the posterior probability value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
Figure 3.20 Maximum likelihood consensus tree based on multilocus (ITS, TEF1, CAL, FG1056 and FUM1) genetic marker sequences

Reflecting the genetic diversity and phylogenetic relationships among 48 isolates of *F. proliferatum* (A) and 2 of *F. verticillioides* (B). The numbers above each branch indicate the bootstrap value. Isolates represent diverse host and geographic origins and further details have been provided in Chapter 2, Table 2.1
3.4.3.8 Genetic diversity reflected by SNPs, Informative sites and Haplotypes

Genotyping of the 52 *F. proliferatum* isolates representing wide biogeographic diversity revealed varying levels of nucleotide sequence divergence among the genetic markers used at the intra-species level. Corresponding data at the inter-species level has also been generated by comparison to the reference species *F. verticillioides* and *F. oxysporum* (Table 3.6; Figures 3.21 - 3.25). Overall, the number of SNPs, Informative sites and Haplotypes varied with the different genetic markers. For example, lowest number of SNPs and ISs was recorded with the internal transcribed spacer (ITS) marker (3 and 1, respectively among *F. proliferatum* isolates), whereas the highest number of SNPs and ISs (40 and 7, respectively) was recorded with Calmodulin (CAL) among the widely used known genetic markers. FG1056, the new genetic marker developed in this study specifically targeting *F. proliferatum* revealed at least a 4-fold higher level of SNPs and ISs (169 and 89, respectively; Table 3.6). In terms of haplotype diversity, FUM1 among the known markers revealed the highest level with 18 identified among the *F. proliferatum* isolates (e.g. Figures 3.21 - 3.24). FG1056 provided the highest level of resolution with 23 haplotypes identified among the *F. proliferatum* isolates and 26 identified among the three species (e.g. Figure 3.25). The full alignment of the FG1056 marker for all identified haplotypes is shown in the Appendix (Figure 8.1).
Table 3.6 Genetic diversity reflected by SNPs, Informative sites and Haplotypes among *Fusarium proliferatum* isolates and corresponding data with reference to *F. verticillioides* and *F. oxysporum*

<table>
<thead>
<tr>
<th>Genetic marker</th>
<th>End-trimmed sequence length (bases)</th>
<th>Similarity % <em>F. proliferatum</em></th>
<th>Similarity % 3 Fusarium spp.</th>
<th>*SNPs/**Informative sites in <em>F. proliferatum</em></th>
<th>SNPs/Informative sites in the 3 Fusarium spp.</th>
<th>Number of haplotypes in <em>F. proliferatum</em></th>
<th>Number of haplotypes in 3 Fusarium spp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal transcribed spacer (ITS)</td>
<td>~454</td>
<td>99.6</td>
<td>90.3</td>
<td>3/1</td>
<td>54/52</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Translation elongation factor (TEF1)</td>
<td>~358</td>
<td>96.3</td>
<td>92.2</td>
<td>16/11</td>
<td>36/32</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Calmodulin (CAL)</td>
<td>~650</td>
<td>95.8</td>
<td>86.3</td>
<td>40/7</td>
<td>129/62</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Fumonisin biosynthetic gene (FUM1)</td>
<td>~287</td>
<td>94.4</td>
<td>78</td>
<td>22/6</td>
<td>69/58</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>FG1056 (New marker, DNA repair gene)</td>
<td>~485</td>
<td>74</td>
<td>76.3</td>
<td>169/89</td>
<td>190/118</td>
<td>23</td>
<td>26</td>
</tr>
</tbody>
</table>

* SNPs, Single nucleotide polymorphisms; ** Informative sites, nucleotide positions at which at least two isolates show a base change
Figure 3.21 The internal transcribed spacer (ITS) gene marker data from six haplotypes

Multiple sequence alignment reflecting the most variable region in the internal transcribed spacer (ITS) gene marker data from six haplotypes identified among *F. proliferatum* (3), *F. verticillioides* (1) and *F. oxysporum* (2). These 6 haplotypes represent the 57 isolates from diverse host and geographic origins. Species designation of the isolates is as follows: *F. proliferatum* (ITEM 2984, R16 and ITEM 4304), *F. verticillioides* (ITEM 1744) and *F. oxysporum* (A25 and NL70-7). Asterisk denotes conserved/identical base. Lack of asterisk denotes an insertion/deletion, or single nucleotide polymorphism (SNP) or informative site (ISs). The four nucleotide bases have been labelled with different colour: A (red), C (blue), G (yellow) and T (green). Base substitutions identified as Informative sites only are fully coloured.
Figure 3.22 The translation elongation factor (TEF1) gene marker data from 18 haplotypes

Multiple sequence alignment reflecting the most variable region in the translation elongation factor (TEF1) gene marker data from 18 haplotypes identified among *F. proliferatum* (16), *F. verticillioides* (1) and *F. oxysporum* (1). These 18 haplotypes represent the 57 isolates from diverse host and geographic origins. Species designation of the isolates is as follows: *F. proliferatum* (16 isolates), *F. verticillioides* (1) and *F. oxysporum* (NL70-7). Asterisk denotes conserved/identical base. Lack of an asterisk denotes an insertion/deletion, or single nucleotide polymorphism (SNP) or informative site (ISs). The four nucleotide bases have been labelled with different colour: A (red), C (blue), G (yellow) and T (green). Base substitutions identified as Informative sites only are fully coloured.
Figure 3.23 The Calmodulin (CAL) gene marker data from 16 haplotypes

Multiple sequence alignment reflecting the most variable region in the Calmodulin (CAL) gene marker data from 16 haplotypes identified among *F. proliferatum* (11), *F. verticillioides* (2) and *F. oxysporum* (3). These 16 haplotypes represent the 57 isolates from diverse host and geographic origins. Species designation of the isolates is as follows: *F. proliferatum* (ITEM 1744 and MPVP 294), *F. verticillioides* (ITEM 7595, (Ma-Ue), KF3377, (Gr-Eo)) and *F. oxysporum* (A28, NL70-7 and A25). Asterisk denotes conserved/identical base. Lack of asterisk denotes an insertion/deletion, or single nucleotide polymorphism (SNP) or informative site (ISs). The four nucleotide bases have been labelled with different colour: A (red), C (blue), G (yellow) and T (green). Base substitutions identified as Informative sites only are fully coloured.
Figure 3.24 The FUM1 (fumonisin polyketide synthase) gene marker data from 19 haplotypes

Multiple sequence alignment reflecting the most variable region in the FUM1 (fumonisin polyketide synthase) gene marker data from 19 haplotypes from *F. proliferatum* (18) and *F. verticillioides* (1). These 19 haplotypes represent the 52 isolates from diverse host and geographic origins. Species designation of the isolates is as follows: *F. proliferatum* (18 isolates) and *F. verticillioides* (ITEM 1744). Asterisk denotes conserved/identical base. Lack of asterisk denotes an insertion/deletion, or single nucleotide polymorphism (SNP) or informative site (ISs). The four nucleotide bases have been labelled with different colour: A (red), C (blue), G (yellow) and T (green). Base substitutions identified as Informative sites only are fully coloured.
Figure 3.25 The FG1056 data for the identification of 26 haplotypes

Multiple sequence alignment reflecting the most variable region in the new genetic marker developed in this study FG1056 data from identified 26 haplotypes among *F. proliferatum* (23), *F. verticillioides* (1) and *F. oxysporum* (2). These 26 haplotypes represent the 57 isolates from diverse host and geographic origins. Species
designation of the isolates is as follows: \textit{F. proliferatum} (11 isolates), \textit{F. verticillioides} (ITEM 1744) and \textit{F. oxysporum} (NL70-7 and A28). Asterisk denotes conserved/identical base. Lack of asterisk denotes an insertion/deletion, or single nucleotide polymorphism (SNP) or informative site (ISs). The four nucleotide bases have been labelled with different colour: A (red), C (blue), G (yellow) and T (green). Base substitutions identified as Informative sites only are fully coloured.

3.5 Discussion

The aim of this chapter was to explore the genetic diversity and understand the phylogenetic relationships in \textit{F. proliferatum} utilising a collection of 52 isolates originating from different hosts and diverse geographic location, using isolates of \textit{F. verticillioides} (2) and \textit{F. oxysporum} (3) as references and/or outgroups. To achieve this, the universal fungal DNA barcoding marker ITS (Schoch \textit{et al}., 2012; Stern \textit{et al}., 2012; Das and Deb, 2015) and housekeeping genes TEF1 (Carbone and Kohn, 1990; Kullnig-Gradinger \textit{et al}., 2002; Geiser \textit{et al}., 2004) and CAL (O'Donnell \textit{et al}., 1998; Mule \textit{et al}., 2004a) widely used as fungal genetic diversity markers were selected. The fumonisin polyketide synthase gene FUM1 used with various \textit{Fusarium} species (Flaherty \textit{et al}., 2003; Glenn \textit{et al}., 2008; Kathikeyan \textit{et al}., 2010; Stepien \textit{et al}., 2011a) has also been selected. Furthermore, the potential to develop novel genetic markers for the focal pathogen in this study \textit{F. proliferatum} was explored utilising the FUNYBASE (Marthey \textit{et al}., 2008; Armitage \textit{et al}., 2015) and the current genomic resources available for various \textit{Fusarium} species (Chiara \textit{et al}., 2015; King \textit{et al}., 2015; Niehaus \textit{et al}., 2016).

Optimal primers and PCR conditions were identified (Table 3.1) for the TEF1, CAL and FUM1 markers by screening various primers available in the literature with six isolates representing \textit{F. proliferatum}, \textit{F. verticillioides} and \textit{F. oxysporum} in PCR (Figures 3.2 – 3.4). To test the utility of novel genetic markers, a combination of bioinformatic and comparative genomic analysis was carried out utilising the single copy gene orthologs universal to fungal species available in the FUNYBASE. This led to the identification of three potential novel marker loci MS550, FG644 and FG1056 on the basis of their higher level of SNPs and ISs in \textit{Fusarium} species, identified by bioinformatics, relative to widely used markers such as ITS, TEF1 and CAL (Figure 3.3). Following initial primer design, PCR screening and/or DNA sequencing of MS550, FG644 and FG1056 loci was carried out with the six selected isolates. This led to the development of optimal primers to enable high throughput generation of DNA sequence data for a highly variable phylogenetically informative region of FG1056 (Table 3.5; Figure 3.6). FG1056 is a novel genetic marker developed in this study specifically targeting \textit{F.}
proliferatum along with the reference species *F. verticillioides* and *F. oxysporum*. Nucleotide sequence data generated from each of these markers was utilised for phylogenetic analysis either as individual data sets for each marker or as concatenated data sets for multiple markers. Multiple sequence alignment followed by Bayesian as well as Maximum likelihood analysis with the genetic group identification set at 70% or above enabled the generation of single locus as well as multilocus phylogenies for the 52 isolates of the *F. proliferatum* representing its biogeographic diversity, along with *F. verticillioides* (2 isolates) and *F. oxysporum* (3 isolates) used as references or outgroups.

The internal transcribed spacer (ITS) region is widely recognised and used as an universal DNA barcode marker for the identification and resolution of a diverse array of fungal species (Stern *et al.*, 2012; Schoch *et al.*, 2012; Das and Deb, 2015). In this study, ITS was effective at the interspecies level with the phylogenetic tree displaying three distinct clades A, B and C (Figure 3.8), representing *F. proliferatum*, *F. verticillioides* and *F. oxysporum* isolates, respectively. However, ITS as a genetic marker was much less effective at the intraspecies level among the *F. proliferatum* isolates in view of the high sequence similarity (99.6%) with low SNPs (3) and ISs (1). And only one group was formed within the *F. proliferatum* clade, this group A1 consisted of five isolates originating from onion in the UK and Italy (Table 3.6; Figure 3.8). This pattern is similar to the results reported among isolates of *F. subglutinans* and *F. verticillioides* (Mule *et al.*, 2004). The utility of the ITS marker in fungal genetic diversity and phylogenetic studies remains uncertain (Pryor and Gilbertson, 2000; Schoch *et al.*, 2012; Clarkson *et al.*, 2013). Some studies in *Fusarium* species and other fungal pathogens such as *Penicillium* have reported limitations (O'Donnell and Cigelnik, 1997; O'Donnell *et al.*, 1998a; Skouboe *et al.*, 1999; O'Donnell *et al.*, 2015). In other genera such as *Alternaria*, ITS was useful at inter- and intra- species levels (Woudenberg *et al.*, 2013; Armitage *et al.*, 2015).

The TEF1 markers developed from different parts of the translation elongation factor gene have been used extensively in the population study of many fungal pathogens including *Fusarium* species (e.g. Carbone and Kohn, 1990; Kullnig-Gradinger *et al.*, 2002; Geiser *et al.*, 2004; O'Donnell *et al.*, 2015). In this study, the TEF1 marker was useful both at the inter- and intra- species levels with the clear distinction of clades A, B and C representing the 57 isolates belonging to the three *Fusarium* species (Figure 3.9). The *F. proliferatum* isolates within clade A showed 96.3% similarity, 16 SNPs and 11 ISs along with the differentiation of at least 4 groups (A1, A2, A3 and A4, Figure 3.9). The group A2 showed strong relationship with 94% posterior probability value (PPV) among isolates from garlic in Poland, whilst the other groups (A1, A3 and A4 with 90-100% PPV were not specifically linked to a particular host or geographical location. These results are consistent with previous relationships reported among *F. proliferatum* isolates (Stepien *et al.*, 2011a). Similarly, the utility of TEF1 marker to
characterise isolates belonging to *F. verticilodes* and *F. subglutinans* (Geiser *et al*., 2004; Barik and Tayung, 2012; Nugroho *et al*., 2013), as well as other fungi such as *Aspergillus niger* and *Penicillium* (Peterson, 2004; Perrone *et al*., 2011) has been reported.

In this study, previously designed primers spanning variable regions of the Calmodulin gene (CAL) were tested initially (O'Donnell *et al*., 1998; Mule *et al*., 2004a; Mule *et al*., 2004b). The selected CAL marker region showed a higher level of variability among the *F. proliferatum* isolates (95.8% similarity and 40 SNPs) compared to the ITS and TEF1 (e.g. 96.3% similarity and 16 SNPs) markers. The CAL marker clearly distinguished the three *Fusarium* species into Clades A, B and C representing the isolates belonging to *F. proliferatum*, *F. verticilloides* and *F. oxysporum*, respectively. Among the *F. proliferatum* isolates in Clade A 6 genetic groups A1 – A6 were distinguished with strong PPVs (Figure 3.11). Groups A2, A3, A4 and A5 showed strong relationship among isolates originating from the same host and geographical location, e.g. group A4 with 97% PPV consisted of *F. proliferatum* isolates from garlic in Poland. On the other hand, groups A1 and A6 with 85% and 100% PPV consisted of isolates originating from different crops and geographical locations. This study has shown the utility of CAL gene marker in characterising the genetic diversity and phylogenetic relationships among isolates representing diverse *F. proliferatum* populations. However, the efficacy of this marker particularly at the intraspecies level varies in other *Fusarium* species as nearly 100% similarity has been reported within pathogen population of *F. andiyazi*, *F. subglutinans*, *F. verticilloides* and *F. oxysporum*, (Mule *et al*., 2004a; Mule *et al*., 2004b; Leyva *et al*., 2015). Nonetheless, CAL has been considered as a highly reliable genetic marker for phylogenetic analysis in various fungal pathogens such as *Magnaporthe oryzae* (Couch and Kohn, 2002; Couch *et al*., 2005), *Colletotrichum* species (Liu *et al*., 2016).

The fumonisin polyketide synthase (FUM1) gene present within the FUM gene cluster is known to play a key functional role in the biosynthesis of fumonisin group of mycotoxins (Proctor *et al*., 1999; Proctor *et al*., 2013). In this study, utility of the previously available primers spanning variable regions of the FUM1 gene was initially tested by PCR and/or sequencing and a potential primer pair/marker region was selected (Table 3.1; Figure 3.4) to characterise the 57 isolates belonging *Fusarium* species *F. proliferatum* (52), *F. verticilloides* (2) and *F. oxysporum* (3). DNA sequence data for the marker region was successfully generated from 48 *F. proliferatum* and two *F. verticilloides* isolates. Four *F. proliferatum* isolates (KF3341, KF3360 and KF3362 originating from Asparagus in Poland and 19 F.P from maize in Spain) as well as the three *F. oxysporum* isolates (Table 2.1, Chapter 2) did not yield a product despite repeating the PCR amplifications. It is pertinent to note that the same genomic DNA samples of each of these isolates have been used to successfully generate amplicons for other genetic markers ITS, TEF1, CAL and FG1056. Taken together these
observations suggest mutations and/or deletions in the FUM1 gene in these set of isolates, although at this stage the moot possibility of the entire FUM cluster being not present cannot be ruled out and requires further investigation.

The FUM1 marker formed two clear clades A and B representing the *F. proliferatum* isolates (48) and *F. verticillioides* isolates (2). FUM1 revealed 94.4% similarity with 22 SNPs among the *F. proliferatum* isolates forming three genetic groups A1, A2 and A3 with 92 - 98% PPVs including the isolates from Date palm in Saudi Arabia which formed a distinct group (Figure 3.13). The FUM1 marker was not most effective with the *F. proliferatum* isolates in this study both in terms of the SNPs displayed (22 compared to 40 with CAL) and the number of genetic groups distinguished. This locus has been suggested as a useful marker to identify mycotoxigenic strains of *Fusarium* species and to assess phylogenetic relationships (Glenn 2008; von Bargen 2009; Karthikeyan et al., 2010; Stepien et al., 2011a). More recent research has reported the utility of FUM1 along with FUM8 and FUM21 genes in distinguishing mycotoxigenic isolates of various *Fusarium* species including *F. verticillioides* and *F. thapsina* (Divakara et al., 2014). Furthermore, FUM1 marker region could not be amplified in at least four of the *F. proliferatum* isolates used in this study. Whether this is due to the potential deletion of the entire FUM1 gene or even the whole FUM gene cluster or due to more localised mutations cannot be resolved at this stage. However, the data generated from the FUM gene clusters have shown a high level of nucleotide substitutions in the FUM1 gene among the four *F. proliferatum* isolates for which the genome sequencing and assembly has been completed in this study. It is pertinent to note the horizontal transfer of the FUM gene cluster into the species belonging to *fujikuroi* species complex from unknown donors. This has also led to the recognition of different genomic contexts (GCs) where the FUM gene cluster has been integrated in the genomes of different *Fusarium* species (Proctor et al., 2013).

The novel genetic marker FG1056 was initially identified through bioinformatic analysis of 66 loci selected on the basis of their level of identity at the intergeneric level (Table 3.2) out of the 246 single copy orthologs in FUNYBASE (Marthey et al., 2008; Armitage et al., 2015) for initial bioinformatic screening. This gene encodes a protein that is widely conserved in many fungi and functions in the repair of DNA double-strand breaks and in the stability of telomeres (Johzuka and Ogawa, 1995; Marthey et al., 2008). Among the various markers used in this study, FG1056 showed the highest level of variability among the 52 isolates representing the biogeographic diversity of *F. proliferatum* with a 74% similarity, 169 SNPs and 89 ISs and distinguished the three species into clades A, B and C representing *F. proliferatum*, *F. verticillioides* and *F. oxysporum* isolates, respectively. FG1056 revealed 10 genetic groups (A1 to A10) among the *F. proliferatum* isolates with high PPVs (Figure 3.15).
Some of these groups showed strong relationship among isolates originating from the same hosts and geographical locations, e.g. isolates from garlic in Poland. Other genetic groups were made up of isolates from different hosts and geographic locations indicating the lack of specialisation in some contexts of this emerging pathogen.

This is the first report of the identification and use of the FG1056 genetic marker to characterise diverse isolates of *Fusarium* species. The novel genetic marker FG1056 developed specifically for *F. proliferatum* provided a high level of resolution as reflected by the number of SNPs, ISs and Haplotypes among the 57 *Fusarium* species isolates characterised (Table 3.6). The level of resolution offered by FG1056 was much higher compared to other widely used markers including ITS, TEF1 and CAL (e.g. Geiser et al., 2004; Mule et al., 2004a; Stepien et al., 2011a; Schoch et al., 2012) utilised in this study with the same set of isolates (Table 3.6).

Development of FG1056 as a novel marker addresses the need of additional loci for the identification, molecular diagnostics and robust phylogenetic information as highlighted in a recent review on *Fusarium* (O’Donnell et al., 2015). This is particularly relevant in view of the existence of a number of *Fusarium* species complexes such as *F. oxysporum, F. solani* and *F. fujikuroi*, which includes *F. proliferatum*. In this study, the single locus phylogenies developed reflected the varying rates of evolution in nuclear encoded genes and intergenic regions such as ITS, but only offered a limited level of resolution. The new marker FG1056 developed was a clear exception to this trend with 10 genetic groups and 2 cluster shown comparable to the multilocus sequencing typing (MLST) phylogeny based on ITS, CAL, TEF1 and FG1056. The primers developed for this marker need to be further tested for the successful amplification and sequencing of isolates belonging to other key species and assess whether similar levels of high resolution can be generated to decipher intra- and inter-species diversity and relationships.

MLST of the 57 isolates belonging to *F. proliferatum* (52), *F. verticillioides* (2) and *F. oxysporum* (3) was performed based on the concatenated sequences of ITS, TEF1, CAL and FG1056. MLST data of 1947 bases provided a clearly higher level of resolution compared to the phylogenies based on single locus typing with cladess A, B and C representing *F. proliferatum, F. verticillioides* and *F. oxysporum*, respectively (Figure 3.17). And also revealed several genetic groups with strong correlation to the host and geographic origin of the *F. proliferatum* isolates. MLST based on the addition of FUM1 to ITS, TEF1, CAL and FG1056 was also carried out with the 50 isolates belonging to *F. proliferatum* (48) and *F. verticillioides* (2). The combined data of 2234 bases distinguished clades A and B representing *F. proliferatum* and *F. verticillioides* isolates, respectively. And also revealed 10 genetic groups several of which showed strong relationship to the host and geographic origin of the *F. proliferatum*
isolates (Figure 3.19). These results clearly highlight that the MLST phylogenies are more informative than the single locus phylogenies in the context of *F. proliferatum*, which is consistent with observations in a wide range of *Fusarium* species (e.g. O’Donnell et al., 2015). However, non-amplification of the FUM1 marker in some of the isolates needs to be resolved either based on the genome sequences of four isolates of *F. proliferatum* in this study or by utilising other emerging FUM gene markers such as FUM8 and FUM21 (Divakara et al., 2014).

The genetic marker data developed in this study has enabled the identification of distinct haplotypes among the 52 *F. proliferatum* isolates originating from diverse hosts and geographic locations. The number of haplotypes varied from 3 to 23 depending on how informative the markers are. For example, the novel genetic marker developed in this study FG1056 identified 23 haplotypes, whilst ITS distinguished only 3 haplotypes (Table 3.6). This provides a platform for future research to monitor the haplotype diversity and distribution pattern among populations of *F. proliferatum* associated with various hosts in different geographic locations. This would answer questions such as how the haplotypes change over time and space, as well as in relation to diverse host systems.

The biogeographically diverse *F. proliferatum* isolates characterised revealed a high degree of genetic diversity and complex phylogenetic relationships. The prevalent genetic diversity was partitioned into genetic groups that were supported by posterior probability value (PPV) and bootstrap support value (BSV) from at least 70% up to 100%, as well as clusters of genetically similar isolates that were not supported by a minimum of 70% PPV and/or BSV. The *F. proliferatum* genetic groups identified by the multilocus typing with molecular markers as well as the novel marker FG1056 revealed varying patterns in terms of the relationship to the key biological attribute of host association and the geographic origin of the isolates. These patterns include: 1) genetic groups that comprised of isolates originating from the same host and geographic location with notable examples such as *F. proliferatum* isolates from Garlic in Poland, isolates from Date palm in Saudi Arabia and isolates from Asparagus in Italy; 2) genetic groups which included isolates originating from the same host but from different geographic locations with a notable example such as *F. proliferatum* isolates from onion in the UK and Italy; 3) genetic groups which included isolates originating from different host and geographic locations. A further pattern observed was genetically similar isolates that formed clusters with no clear link to either the host and/or geographic origin. The above patterns of genetic partitioning observed in *F. proliferatum* have been reported in other fungal species similarly associated with a wide range of hosts and geographic locations. A notable example of these patterns is the anthracnose pathogen *Collettrichum* species (Baroncelli et al., 2015, 2016).
The pattern 1 referred to above in *F. proliferatum* is suggestive of the evolution of host and environment adapted populations within this species/pathogen. The caveat to this key observation is that the present study focused on covering the biogeographic diversity of the species, which did not permit the testing of a larger number of isolates. However, this pattern of adaptive divergence has been reported in *F. graminearum* (O’Donnell *et al.*, 2000; O’Donnell *et al.*, 2004) as well as in other fungal pathogens such as *Colletotrichum acutatum* (Guerber *et al.*, 2003; Sreenivasaprasad and Talhinhas, 2005). Similarly, *F. oxysporum* isolates are well recognised to exist as host-specific forms, for e.g. *F. oxysporum* f. sp. *cepa* isolates infecting onion (Taylor *et al.*, 2016). Similarly, infraspecific genetic groups initially identified within species complexes have later been designated as separate species allied to biological attributes in various fungal systems (Starkey *et al.*, 2007; Ward *et al.*, 2008; Shivas and Tan, 2009; Faedda *et al.*, 2011; Damm *et al.*, 2012).

The pattern 2 found in *F. proliferatum*, where the genetic groups include isolates from a single host but different geographic locations is suggestive of pathogen introductions due to trade including movement of germplasm, food products, agricultural equipment and/or people. Multiple introductions of the anthracnose pathogen *Colletotrichum* species into the UK through strawberry trade have recently been reported (Baroncelli *et al.*, 2015). In the current context, for example, there is a lot of movement of onion within the UK industry (V Vagany, unpublished).

The pattern 3 recorded in *F. proliferatum* was the inclusion of isolates associated with diverse hosts and geographic locations in some of the genetic groups with strong PPV support. The genetic clusters found also represent similar biogeographic diversity. This would suggest that these isolates are less specialised and potentially capable of infecting different hosts under suitable conditions. Further pathology assays under controlled conditions and/or natural conditions, where feasible are essential to test this hypothesis (e.g. Baroncelli *et al.*, 2015). It is often the case that patterns 2 and 3 can be found interlinked due to biological, ecological and other related factors and these patterns can also be viewed together as one complex pattern of the prevailing populations of a pathogen in the context of crop production and food supply systems across different countries and/or continents.

In conclusion, the work carried out in this chapter has generated extensive multilocus sequence typing data delineating the genetic diversity and phylogenetic relationships in *F. proliferatum* utilising a collection of isolates representative of the biogeographic diversity in this species. This provides a framework that would enable further characterisation of *F. proliferatum* populations associated with various crops and present in different geographic locations. This would pave the way for the development of additional knowledge and
strategies to effectively control and manage this emerging pathogen, which is well known for its mycotoxigenic potential.
4 Assessment of biological variability in *Fusarium proliferatum*

4.1 Introduction

*Fusarium* species can exist as saprophytes, endophytes and pathogens and tolerate diverse environmental conditions (e.g. Barik and Tayung, 2012). There are increasing reports of multiple *Fusarium* species co-occurring on a single host in the same geographic location, e.g. four species on maize in Mexico and eight species on pineapple in Costa Rica (e.g. Stepien et al., 2013). Furthermore, *Fusarium* species also display a considerable level of morphological variability making species identification and disease diagnosis based on the phenotype difficult and unreliable (Barik and Tayung, 2012). Isolates within a species are also known to differ considerably in their mycotoxicogenic potential. The currently described *Fusarium* species (~300) have been further grouped into 20 species complexes or major clades (O’Donnell et al., 2015). A vast number of crops are affected by *Fusarium* pathogens, with the added threat of the mycotoxins such as fumonisins produced by some species contaminating food and feed (e.g. Leyva-Madrigal et al., 2015). Some of the *Fusarium* species such as *F. proliferatum* have a wide geographic distribution affecting fruits, vegetables and cereals indifferent climatic regions (e.g. UK, Italy and Saudi Arabia). In this context, the aim and objectives of the work described in this chapter are focused on assessing the biological variability among *F. proliferatum* isolates originating from diverse hosts and geographic locations along with reference isolates of *F. verticillioides* (2) and *F. oxysporum* (3). The biological attributes assessed include growth, sporulation and fumonisin production with reference to key environmental parameters such as temperature and water activity.
4.2 Materials and methods

4.2.1 Morphological and cultural studies
Nineteen isolates were used in various experiments including the assessment of growth and sporulation as well as observation of colony characteristics and determination of the production of fumonisins (Table 4.1).
### Table 4.1 Details of the *Fusarium proliferatum* isolates along with the *F. verticillioides* and *F. oxysporum* reference isolates used in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>Isolate code*</th>
<th>Species</th>
<th>Host /Plant part</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2/2</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>2</td>
<td>A6/1</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>3</td>
<td>R16</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>4</td>
<td>A40</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>5</td>
<td>A8</td>
<td><em>F. proliferatum</em></td>
<td>Allium Cepa</td>
<td>UK</td>
</tr>
<tr>
<td>6</td>
<td>KF3377</td>
<td><em>F. proliferatum</em></td>
<td>Allium sativum</td>
<td>Poland</td>
</tr>
<tr>
<td>7</td>
<td>ITEM 1453</td>
<td><em>F. proliferatum</em></td>
<td>Asparagus officinalis</td>
<td>Italy</td>
</tr>
<tr>
<td>8</td>
<td>18 F.P</td>
<td><em>F. proliferatum</em></td>
<td>Zea mays</td>
<td>Spain</td>
</tr>
<tr>
<td>9</td>
<td>19 F.P</td>
<td><em>F. proliferatum</em></td>
<td>Zea mays</td>
<td>Spain</td>
</tr>
<tr>
<td>10</td>
<td>ITEM 7595</td>
<td><em>F. proliferatum</em></td>
<td>Zea mays</td>
<td>USA</td>
</tr>
<tr>
<td>11</td>
<td>MPVP 328</td>
<td><em>F. proliferatum</em></td>
<td>Zea mays</td>
<td>USA</td>
</tr>
<tr>
<td>12</td>
<td>ITEM 3275</td>
<td><em>F. proliferatum</em></td>
<td>Solanum lycopersicum root</td>
<td>Italy</td>
</tr>
<tr>
<td>13</td>
<td>ITEM 2341</td>
<td><em>F. proliferatum</em></td>
<td>Phoenix dactylifera root</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>14</td>
<td>ITEM 2984</td>
<td><em>F. proliferatum</em></td>
<td>Mangifera indica</td>
<td>Malaysia</td>
</tr>
<tr>
<td>Item</td>
<td>Isolate Code</td>
<td>Species</td>
<td>Host</td>
<td>Country</td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>---------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>15</td>
<td>MPVP 294</td>
<td><em>F. verticillioides</em></td>
<td><em>Zea mays</em></td>
<td>Italy</td>
</tr>
<tr>
<td>16</td>
<td>ITEM 1744</td>
<td><em>F. verticillioides</em></td>
<td><em>Zea mays</em></td>
<td>Italy</td>
</tr>
<tr>
<td>17</td>
<td>NL70-7</td>
<td><em>F. oxysporum</em></td>
<td><em>Allium Cepa</em></td>
<td>UK</td>
</tr>
<tr>
<td>18</td>
<td>A28</td>
<td><em>F. oxysporum</em></td>
<td><em>Allium Cepa</em></td>
<td>UK</td>
</tr>
<tr>
<td>19</td>
<td>A25</td>
<td><em>F. oxysporum</em></td>
<td><em>Allium Cepa</em></td>
<td>UK</td>
</tr>
</tbody>
</table>

*Isolates sources:*

ITEM – Isolates from the Culture collection at the Institutes of science of food production (ISPA), Italy

KF – Isolates provided by Professor Lukasz Stephen at the Polish Academy, Poland.

MPVP and F.P – Isolates were provided by Professor Naresh Magan at the Cranfield University, UK.

A, R and NL – Isolates from joint research projects at Warwick HRI – University of Bedfordshire, UK.
4.2.2 Assessment of colony characteristics
Fungal isolates were grown on PDA plates for 10 days at 20°C in an incubator. Culture images were captured from the upper and lower sides of the plates to record the colony morphology as well as any pigmentation.

4.2.3 Growth rate at different temperatures
Fungal cultures were prepared on 9 cm PDA plates by inoculating 0.5 cm agar plugs from actively growing areas of 5-day-old cultures. Five replicates were prepared for each isolate and the cultures were incubated for 6 days at 20, 25 and 30°C. Radial growth was measured on 2nd, 4th and 6th days to determine the growth rate.

4.2.4 Growth rate at different water activities
The water activity level of the PDA medium was adjusted by adding required amounts of glycerol (Table 4.2) prior to autoclaving. The autoclaved medium was dispensed into 9 cm Petri dishes, which were inoculated with plugs as above. Five replicates were maintained for each isolate and incubated at 20°C for 6 days to determine the growth rate as described above.

<table>
<thead>
<tr>
<th>Water activity (aw)</th>
<th>0.995</th>
<th>0.98</th>
<th>0.97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycerol (g per 100 ml)</td>
<td>2.76</td>
<td>9.2</td>
<td>14.72</td>
</tr>
</tbody>
</table>

* Source: Professor N Magan and Dr A Madina, Cranfield University (Ramirez et al., 2004; Formenti et al., 2012).

4.2.5 Determination of the level of sporulation
Fungal isolates replicated 5 times were grown on PDA plates for 10, 15, 20 and 25 days at 20°C in an incubator (Rossi et al., 2009; Islam, 2015). To prepare the spore suspension, 4 agar plugs (equivalent to 78.53 mm² surface area) were cut-out from the mature areas of the culture using a sterile 5 mm cork borer and transferred to a 10 ml universal tube containing 3
ml of sterile water. The suspension was vortexed for approximately 30 seconds and mixed well by inversion (approximately 30 times) (Gupta et al., 2010). The tubes were left standing at room temperature for 15 minutes to allow the spores to settle at the bottom uniformly. Then the upper 2.5 ml layer was discarded and the left over 0.5 ml, which contains the spores was vortexed for approximately 30 seconds (Li et al., 2006). To count the spores, using a consistent volume, 7 µl of the spore suspension was pipetted onto the hemocytometer chamber and 2 µl lactophenol cotton blue was added to aid visualisation. The hemocytometer chamber contains 25 large squares and the standard volume of each large square is 0.02 µl. Spores were counted from 5 large squares marked in red in Figure 4.1 and the average number of spores per large square was calculated. First, the number of spores per µl was calculated by multiplying the average number of spores per one large square by 50 (conversion factor of 0.02 µl to 1 µl). Next, this value was multiplied by 500 to calculate the total number of spores in the 500 µl suspension (78.53 mm² culture surface area) prepared for each sample. Finally, this value was divided by 78.53 to calculate the number of spores per mm² surface area of the fungal culture on PDA under the experimental conditions used.
4.2.6 Determination of the level of fumonisins

Four *F. proliferatum* isolates originating from diverse hosts and geographic locations, selected for genome sequencing and FUM cluster characterisation, were also used for the fumonisin production analysis to enable investigations of the link between genetic basis and fumonisin production also with reference to the host and geographic diversity. The isolates used include ITEM 2341 (date palm, Saudi Arabia), KF3377 (garlic, Poland) and MPVP 328 (maize, USA) for which previous data on the production of fumonisins was available from different research groups (Abdalla et al., 2000; Stepień et al., 2011c; Lazzaro et al., 2013) along with a new isolate R16 from onion in the UK. For each isolate, four 5 mm agar plugs cut-out from actively growing mycelium were inoculated in tissue culture flasks containing 50 ml fumonisin inducing medium (Table 4.3). The cultures were incubated at 25°C in a stationary incubator for 27 days (Jurado et al., 2008). Culture filtrates collected were extracted by the addition of equal volume of an acetonitrile, methanol, and water solution, gentle mixing and filtration through 2 layers of sterile Whatman no. 1 paper. Buffer solution was added to the filtrate and passed through an immunoaffinity column, which retains fumonisins. The column was washed and the fumonisins were eluted using methanol. The eluate was analysed by HPLC using pre-column derivatisation with o-phthaldialdehyde (OPA)
to facilitate fluorescence detection. The fumonisin concentration for each isolate was determined by comparison with FB1 and FB2 external standards. This fumonisin estimation analysis was done using an external service provider Sciantec Analytical Services, Stockbridge Technology Centre, Selby YO8 3SD, UK. The analytical methods followed for the various steps are based on standard procedures available in the literature (e.g. Jiménez et al., 2003; López-Errasquín et al., 2007; Medina et al., 2013). Further details of the proportion of the solvent mixtures and the columns used, however, are not available in view of the ‘Commercial in Confidence’ nature of the external service provider’s business operation.

Table 4.3 Composition of the fumonisin inducing medium*

<table>
<thead>
<tr>
<th>No.</th>
<th>Ingredients</th>
<th>Quantities (g/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Malt extract</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>Yeast extract</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>Peptone</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>KH$_2$PO$_4$</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>MgSO$_4$ · 7H$_2$O</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>KCl</td>
<td>0.3</td>
</tr>
<tr>
<td>7</td>
<td>ZnSO$_4$ · 7H$_2$O</td>
<td>0.05</td>
</tr>
<tr>
<td>8</td>
<td>CuSO$_4$ · 5H$_2$O</td>
<td>0.01</td>
</tr>
<tr>
<td>9</td>
<td>Fructose</td>
<td>20</td>
</tr>
</tbody>
</table>

* Jurado et al., 2008
4.3 Experimental approaches

In various assays carried out, a selected set of isolates belonging to *F. proliferatum* (14), *F. verticillioides* (2) and *F. oxysporum* (3) were used (Table 4.1). Effect of different temperatures (20, 25 and 30°C) on the growth of the 19 isolates was assessed over a six-day period. Effect of water activity (aw) ranging from 0.995, 0.99, 0.98, to 0.97 on growth was assessed over six days utilising fungal cultures grown on modified PDA medium and incubated at the general optimum temperature (20°C). Radial growth data from five replicates for each isolate was used to calculate the maximum growth rate. The level of sporulation was assessed at different temporal points (10, 15, 20 and 25 days) utilising fungal cultures grown on PDA at general optimal temperature and water activity levels (20°C and 0.99 aw, respectively). Data of the spore counts generated using a haemocytometer with spore suspensions prepared from culture plugs were converted to calculate the level of sporulation per mm² for each isolate. The level of production of fumonisins (FBs) was determined for four isolates representing diverse host and geographic origins. Filtrates from the cultures grown (at 25°C for 27 days) in the fumonisin inducing medium were used to quantify FB1 and FB2 following a two steps immunoaffinity and HPLC process utilising standard references.

4.4 Results

4.4.1 Morphological variability among *Fusarium* species isolates

The extent of morphological variability among the isolates belonging to *F. proliferatum* along with *F. verticillioides* and *F. oxysporum* was clearly evident from the colony characteristics observed using cultures grown on PDA plates (e.g. Figure 4.2 and 4.3). Different isolates of *F. proliferatum* from the same host, irrespective of their geographic origin showed clearly distinctive morphologies reflecting the extent of biological variability encountered within this species. For example, in isolate A40 from onion in the UK (A), the upper side of the colony showed white to yellowish cottony mycelium, while the lower side showed mainly yellowish pigmentation. In isolate ITEM 1950 from onion in Italy (B), the colony was mainly pinkish with limited white aerial mycelium around the inoculum on the upper side and in the lower side, the pigmentation was mostly pinkish. Isolate ITEM 1951 also from onion in Italy (C) showed very distinctive features with the upper side of the colony showing predominantly white aerial mycelium and the lower side showing dark to light purple pigmentation (Figure 4.2).

Conversely, some isolates belonging to *F. proliferatum*, *F. verticillioides* and *F. oxysporum* from various hosts and geographic locations showed considerable overlap in colony characteristics. This is well reflected by the three isolates shown each belonging to a different *Fusarium* species, a different host and at least from two different geographic locations. All the isolates showed pinkish red mycelium and pigmentation in the upper and lower sides of the colony, respectively (Figure 4.3). These observations highlight the highly variable morphologies within *F. proliferatum* as well as the
overlap with *F. verticillioides* and *F. oxysporum*, clearly evidencing the fact that morphological aspects are not very reliable for species identification and/or disease diagnosis in these systems.

Figure 4.2 Three sets of cultures grown on PDA plates displaying morphological variability among the isolates of *Fusarium proliferatum* from a single host in different geographic locations
A, B and C, represent isolates A40 from onion in UK and ITEM 1950 as well as ITEM 1951 from onion in Italy, respectively.
Figure 4.3 Three sets of cultures grown on PDA plates displaying overlapping morphologies across different *Fusarium* species host and geographical location

A, B and C, represent isolates ITEM 3274, *F. proliferatum* from tomato in Italy; NL70-7, *F. oxysporum* from onion in UK; and MPVP 294, *F. verticillioides* from maize in Italy, respectively.
4.4.2 The influence of temperature on growth
Among the 14 *F. proliferatum* isolates assessed, nine showed negative effects in growth rate when the temperature increased from 20°C to 30°C, four were not affected (ITEM 1453, ITEM 7595, MPVP 328, KF 3377) and only ITEM 2341 (Date palm, Saudi Arabia) showed a positive effect. Both *F. verticillioides* isolates (MPVP 294 and ITEM 1744) were not affected by temperature; and all three *F. oxysporum* isolates (A25, A28 and NL70-7) showed a negative effect in growth with the increasing temperature. In general, the growth rate was higher at 20°C than 30°C for most isolates. The highest growth rate was 15 mm per day such as for A6/1 and R16 isolates at 20°C, whereas for most isolates the growth rate averaged between ~10-14 mm per day (Figure 4.4).
*Fusarium proliferatum* (FP), *F. verticillioides* (FV) and *F. oxysporum* (FO)

Figure 4.4 The effect of temperature on the maximum growth rate of isolates belonging to *Fusarium proliferatum* (14), *F. verticillioides* (2) and *F. oxysporum* (3)

Points represent raw data, orange lines represent ordinary least squares (OLS) regression data and dark grey shading represents bootstrapped 95% confidence intervals around the fitted regressions. All isolates are *F. proliferatum* except ITEM 1744 and MPVP 294 (*F. verticillioides*) and A25, A28 and NL7-07 (*F. oxysporum*).
4.4.3 The influence of water activity on growth

Increasing water activity in general had a positive effect on the maximum growth rate in most isolates and none showed a negative effect. Overall, 13 out of 19 isolates demonstrated a very strong increase in growth rate from ~5 - 10 mm per day when the water activities increased. For example, NL70-7 and A25 showed growth rate ~15-20 mm per day at 0.995 aw. Whereas four isolates demonstrated increased growth rate at lower than ~5 mm per day over water activities (e.g. ITEM 2984 and ITEM 1453). In contrast, two isolates MPVP 328 and ITEM 7595 did not show significant variability across different water activities. Isolate KF 3377 scored the lowest growth rate ~6 mm per day at 0.97 aw (Figure 4.5). The 14 \textit{F. proliferatum} isolates showed varying response patterns to increasing water activity including a strong positive response (e.g. A8 and ITEM 3275), moderate positive response (e.g. 18 F.P. and ITE M2984), as well as no response (MPVP 328 and ITEM 7595). Among the reference species, both \textit{F. verticillioides} isolates MPVP 294 and ITEM 1744 showed a moderate response; \textit{F. oxysporum} isolates A25, A28 and NL70-7 all showed a strong positive response (Figure 4.5).
*Fusarium proliferatum* (FP), *F. verticillioides* (FV) and *F. oxysporum* (FO)

Figure 4.5 The effect of water activity on the maximum growth rate of isolates belonging to *Fusarium proliferatum* (14), *F. verticillioides* (2) and *F. oxysporum* (3)

Points represent raw data, blue lines represent ordinary least squares (OLS) regression data and dark grey shading represents bootstrapped 95% confidence intervals around the fitted regressions. All isolates are *F. proliferatum* except ITEM 1744 and MPVP 294 (*F. verticillioides*) and A25, A28 and NL7-07 (*F. oxysporum*).
4.4.4 The level of sporulation over time
In general, there was a large variability in the level of spore production across isolates assessed with a ten-fold difference observed between the highest and lowest producers (e.g. *F. proliferatum* isolates ITEM 7595 and ITEM 1453 with ~3000 and ~35000 spores per mm$^2$; Figure 4.6). The 14 isolates belonging to *F. proliferatum* showed three distinct patterns of sporulation based on spores per mm$^2$ categorised as high (30,000 or more), medium (10,000 to 20,000) and low (less than 10,000). At least five isolates categorised as high showed a rate of over 30,000 spores per mm$^2$ (e.g. 19F.P, A8, ITEM 1453, ITEM 3275 and ITEM 2984); three isolates (A40, A6/1 and MPVP 328) showed medium level sporulation and six isolates (e.g. 18F.P, A2/2 and R16) belonged to the low category (Figure 4.6). Among the reference species, both *F. verticillioides* isolates MPVP 294 and ITEM 1744 showed a high level of sporulation (~40,000); *F. oxysporum* isolates A25, A28 and NL70-7 all showed a low level of sporulation over the duration tested (Figure 4.6).

Among the *F. proliferatum* isolates, some showed considerable difference in the time taken for maximum sporulation. For example, isolates ITEM 3275 and 19F.P showed peak levels at 10 and 15 days, respectively; whereas, isolates A8 and ITEM 1453 showed peak level at 20 days. On the other hand, some isolates did not show much difference in the level of sporulation over the tested duration, e.g. ITEM 2341, ITEM 7595 and KF3377 (Figure 4.6).
Fusarium proliferatum (FP), F. verticillioides (FV) and F. oxysporum (FO)

Figure 4.6 Variability in the level of spore production over time among isolates belonging to Fusarium proliferatum (14), F. verticillioides (2) and F. oxysporum (3)
Points represent raw data, green lines represent non-linear regression data fitted by loess and dark grey shading represents bootstrapped 95% confidence intervals around the fitted regressions. All isolates are *F. proliferatum* except ITEM 1744 and MPVP 294 (*F. verticillioides*) and A25, A28 and NL7-07 (*F. oxysporum*). Note that replicates were sampled destructively, so points represent individual replicates, not repeated measurements on the same replicates.

4.4.5 Variability in the production of fumonisins by *Fusarium proliferatum* isolates

The production of fumonisins FB1 and FB2 varied considerably among the four *F. proliferatum* isolates tested. However, isolates producing high quantities of FB1 also produced large quantities of FB2 (Figure 4.7 A and B). The isolate ITEM 2341 (Date palm, Saudi Arabia) produced more than twice as much FB1 and FB2 as the next best isolate KF3377 (Garlic, Poland). MPVP 328 (Maize, USA) produced a low level of FB1 and almost undetectable level of FB2, whilst R16 (Onion, UK) showed the reverse pattern (Figure 4.7 A and B).

![Figure 4.7 Production of fumonisins FB1 and FB2 by four isolates of *Fusarium proliferatum*](image)

Isolates originated from: ITEM 2341, Date palm in Saudi Arabia; KF3377, Garlic in Poland; MPVP 328, Maize in USA; and R16, Onion in UK. Note different scale used on Y-axis for FB1 and FB2.
4.5 Discussion

There was no correlation between the morphological characteristics reflected by the colony appearance and pigmentation of the 14 F. proliferatum isolates and their host and geographic origin. For example, isolates originating from the same host and geographic location showed distinctive morphologies as in the case of ITEM 1950 and ITEM 1951 from onion in Italy (Figure 4.2). This clearly highlights the morphological diversity existing within F. proliferatum. Similar morphological variability has been reported among a number of isolates of F. oxysporum from guava in Bangladesh (Hussain et al., 2012). Furthermore, some of the F. proliferatum isolates also exhibited overlapping morphological characteristics with isolates of the reference species F. verticillioides and F. oxysporum as in the case of isolates ITEM 3274, MPVP 294 and NL70-7, respectively (Figure 4.3). Variability shown by related Fusarium species in their biological attributes such as the colony morphology and pigmentation has previously been described (e.g. Summerell et al., 2003; Leslie and Summerell, 2008). The overall findings presented in this study clearly emphasise the challenges faced by traditional mycologists and pathologists in reliably identifying isolates of F. proliferatum and/or diagnosing the pathogen in crop diseases (e.g. Barik and Tayung, 2012; O'Donnell et al., 2015). This issue is exacerbated by the fact that in an increasing number of crops and food production systems, co-occurrence of multiple Fusarium species on a single host in the same geographic location is being reported (e.g. Stepien et al., 2013; Leyva-Madrigal et al., 2015). An added complexity is the fact that Fusarium species can exist as saprophytes, endophytes and pathogens and tolerate diverse environmental conditions (e.g. Barik and Tayung, 2012).

The series of experiments presented in this chapter aimed to characterise the growth and sporulation responses of a selected set of isolates of F. proliferatum (14), along with reference isolates of F. verticillioides (2) and F. oxysporum (3) to key environmental change parameters such as temperature and water activity. The results clearly demonstrate that the F. proliferatum isolates, also with reference to the F. verticillioides and F. oxysporum isolates, show considerable biological variability reflected by greatly varied growth rates and sporulation and that these rates are affected differently by changes to temperature and water activity.

In terms of maximum growth rate and response to increasing temperature from 20 to 30°C, the F. proliferatum isolates showed three patterns: 1) Four isolates that were not affected by the increasing temperature; 2) Nine isolates that showed a negative effect and 3) One isolate ITEM 2341 from Date palm in Saudi Arabia which showed a positive effect reaching its maximum growth rate at 30°C, the highest temperature tested (Figure 4.4). It is pertinent to note that this particular isolate ITEM 2341 was the highest producer of fumonisins FB1 and FB2 - more than twice
compared to the next high producing isolate KF3377. Whereas the maximum growth rate under different temperatures ranged between 10 and 15 mm per day and the higher growth rate for most isolates was at 20°C. No intraspecies variability was observed for the reference species within the limited number of isolates used. For example, the two F. verticillioides isolates were not affected, whilst all three F. oxysporum isolates showed a negative effect (Figure 4.4). Current findings with the F. proliferatum isolates from diverse host and geographical locations are comparable to previous reports that the growth rate varied between 10 – 20 mm per day and that the growth rate was higher between 20 and 25°C compared to 30 and 35°C (Stepien et al., 2011c). In other Fusarium species isolates such as F. oxysporum and F. solani also, the growth rate averaged between 10-15 mm per day, but the optimum growth temperature was 28°C (Gupta et al., 2010). In F. verticillioides isolates, optimum growth temperature was 20°C (Medina et al., 2013).

The growth rate under different water activities ranged between 5 and 20 mm per day and for most isolates the highest growth rate was at 0.995 aw (Figure 4.5). In terms of the effect of water activity on growth, the 14 F. proliferatum isolates again showed three patterns: 1) various isolates showed a strong positive response with maximum growth rate at the highest water activity (aw) level; 2) some isolates showed a moderate response; and 3) at least two isolates e.g. MPVP 328 and ITEM 7595 that did not show any significant difference. In the case of F. verticillioides isolates, optimum water activity for growth was reported as 0.995 aw (Medina et al., 2013). No intraspecies variability was observed among the limited number of reference isolates used with both F. verticillioides isolates showing a moderate response and all three F. oxysporum isolates showing a strong response. Current findings evidence that the influence of water activity on growth rates of the 19 Fusarium isolates (Figure 4.5) was greater than the influence of temperature (Figure 4.4). Present results also concur with previous findings that isolates belonging to Fusarium species including F. verticillioides optimum growth was reported at the highest water activity (Marin et al., 1999b; Hope et al., 2005; Marin et al., 2005; Medina et al., 2013). Similar observations in terms of optimum temperature and water activity on growth have also been reported in the toxigenic indoor mould Stachybotrys chartarum, which is beginning to be recognised as a major health burden in the USA (Frazer et al., 2011).

The level of sporulation varied greatly among the 19 Fusarium isolates and also through time. Within F. proliferatum, three distinct patterns were recorded with high (five isolates), medium (3 isolates) and low (6 isolates) levels of sporulation varying from ~3000 to ~35000 spores per mm². The two F. verticillioides (ITEM 1744 and MPVP 294) sporulated above 40,000 spores per mm², whilst all three F. oxysporum isolates were in the low category. The duration required for maximum
sporulation of *F. proliferatum* isolates also varied with at least three peak points of 10, 15 and 20 days e.g. ITEM 3275 (10 days), 19F.P (15 days) and A8 (20 days), although the general trend appears to be during 15 to 20 days (Figure 4.6). Previous research has reported 27°C and 15th day as the optimum temperature and duration of sporulation in two *F. verticillioides* isolates (Rossi et al., 2009; Frazer et al, 2011). In the present study, there was a 10-fold difference in sporulation rate between the isolates producing the most spores and those producing the least within *F. proliferatum*. It is interesting to note such large intraspecies differences in sporulation, which is clearly greater than the differences recorded in the maximum growth rate.

The study has evidenced considerable differences among four *F. proliferatum* isolates in the production of fumonisins (FB1 and FB2) recognised as a common and economically important group of mycotoxins. The WHO and FAO has jointly specified the maximum tolerable intake of fumonisins at 2 µg per kg body weight per day (JECFA, 2001; Fandohan et al., 2003; Palencia et al., 2003; El-Imam et al., 2012). Two isolates reported here, KF3377 and ITEM 2341, produced high levels of FB1 (8483 and 21382 ppb), respectively. In contrast, the remaining two isolates R16 and MPVP 328 produced low levels of FB1 (149.6 and 1159 ppb), respectively. The level of FB2 production, however, was low in all isolates (e.g. maximum 249 ppb).

Any relationship to the host or geographic origin of the *F. proliferatum* isolates and their mycotoxic potential is not clear at this stage of the research. Highest FB1 producer ITEM 2341 originated from Date palm in Saudi Arabia, followed by KF3377 from garlic in Poland and MPVP 328 from maize in USA. Whilst isolate R16 from onion in the UK was the least producer with a nearly 140-fold difference compared to ITEM 2341 (Figure 4.7). Varying levels of production of fumonisins by different *F. proliferatum* isolates, including the isolates selected for current work have previously been reported through separate studies. For example, isolate KF3377 was identified as the highest (Stepien et al., 2011a), followed by MPVP 328 (Lazzaro et al., 2013), with only trace levels for isolate ITEM 2341 (Abdalla et al., 2000). However, the levels of fumonisin production recorded with ITEM 2341, KF3377 and MPVP 328 in this study under uniform condition of growth and analysis differed considerably to previous reports (Abdalla et al., 2000; Stepien et al., 2011c; Lazzaro et al., 2013). These differences are likely due to the varied nutrition conditions and assays used by different researchers. Nonetheless, it is well recognised that various fungal systems including *Aspergillus* and *Fusarium* species show considerable differences in the quantities of the mycotoxins produced (e.g. Marin et al., 1999; Abdalla et al., 2000; Stepien et al., 2011c; Lazzaro et al., 2013).

It is well documented that, knowledge of the underlying genetic mechanisms involved the regulation of the production of mycotoxins such as fumonisins in
response to various environmental parameters including the host, temperature and water remains incomplete (e.g. Proctor et al., 2013). Additional genomic resources and molecular tool kits are being developed in wider fungal systems including <i>Fusarium</i> species to tackle this complex process in view of the threat posed by mycotoxins such as fumonisins produced by <i>Fusarium</i> pathogens to animal and human health (e.g. Leyva-Madrigal et al., 2015).

In conclusion, the results presented in this chapter provide clear evidence of the extent of biological variability within <i>F. proliferatum</i> reflected by the differences in colony morphology, growth, sporulation and fumonisin production also with reference to key environmental change variables temperature and water activity. Key patterns identified include i) considerable morphological variability existing within <i>F. proliferatum</i> and the overlaps to the morphology of other related species such as <i>F. verticillioides</i> and <i>F. oxysporum</i>; ii) variability in maximum growth rate particularly with reference to increasing temperature among the <i>F. proliferatum</i> isolates; iii) general positive effect of increasing water activity on the maximum growth rate; iv) distinct differences in sporulation among the <i>F. proliferatum</i> isolates with nearly a 10-fold difference among some of isolates; and v) vast differences (e.g. up to 140-fold) among the four <i>F. proliferatum</i> isolates in the production of FB1 and FB2.
5 *Fusarium proliferatum* genome sequencing and FUM gene cluster characterisation

5.1 Introduction

*Fusarium proliferatum* is a prominent producer of fumonisins, a major group of mycotoxins. Fumonisin B group (FBs) is a major cause for concern in view of their high abundance and carcinogenic properties (Desjardins et al., 1996). Exposure of children to multiple mycotoxins including fumonisins through maize-based food products has been reported in parts of Africa (e.g. Kimanya et al., 2014). The level of mycotoxin contamination of food varies based on agricultural production systems (Geary et al., 2016). It is well recognised that different *Fusarium* species as well as isolates within a species vary in their mycotoxigenic potential (e.g. Kokkonen et al., 2010). The FUM gene cluster responsible for the biosynthesis of fumonisins has been identified in *F. verticillioides* and *F. fujikuroi* but not yet in *F. proliferatum*. Differences in the gene cluster size potentially due to deletions and/or mutations in the FUM1 gene has been reported to affect the level of fumonisin production in *F. verticillioides* isolates originating from different hosts (e.g. Glenn et al., 2008; Stepien et al., 2011a; Medina et al., 2013). Furthermore, differences in the genomic location of the FUM gene cluster and consequently the genes flanking the have been reported (Proctor et al., 2013). Production of mycotoxins such as fumonisins by different *Fusarium* species and isolates and their response to diverse biotic and abiotic factors in the laboratory or in nature is a complex process regulated by genetic mechanisms and the genomic architecture.

Rapid development of genomic resources for various *Fusarium* species such as *F. oxysporum*, *F. verticillioides* and *F. fujikuroi* underpinning their biology was enabled by the advent of next generation sequencing (NGS) technologies (e.g. Chiara et al., 2015; King et al., 2015). Comparative genomic analysis has led to the identification of species- and isolate-specific differences and recently, genome sequences of two *F. proliferatum* isolates have become available but the FUM gene cluster has not yet been characterised.
(Niehaus et al., 2016). Furthermore, there is growing interest in developing online resources integrating phylogenetics and comparative genomics to better understand the biology and ecology of Fusarium species. The aim and objectives of the work described in the chapter focused on developing the genome sequence data for four diverse isolates of F. proliferatum, predict the total gene set and their distribution pattern, as well as to characterise the FUM gene cluster responsible for the production of fumonisins.

5.2 Materials and methods

5.2.1 DNA Extraction for Genome Sequencing
To ensure the availability of higher concentration of DNA required for genome sequencing, the process was scaled-up compared to the steps described in 3.2.4. Approximately 2-4 g of air-dried mycelium (the wet mycelial mass was dried on autoclaved cheesecloth at room temperature for 1 hour in a microbiological safety cabinet) was placed in sterile mortar, small amount of chelex® and liquid nitrogen were added. The frozen mycelium was ground for 3 to 5 minutes with sterile pestle to prepare a fine powder. Approximately 300-700 mg of the fine ground mycelial powder was transferred into 2 ml microfuge tubes for the DNA extraction. The column-based kit and method described in 3.2.4 was followed with a slight modification as mentioned below. To extract the genomic DNA for the 4 selected isolates, the volume of lysis solution A and B and the precipitation solution were doubled; an RNase digestion (20 µl) was performed as suggested by the manufacturer to digest the RNA and the elution was performed using the TE buffer. DNA samples were subjected to gel electrophoresis and nanodrop methods to check the quantity, quality and integrity as described in Chapter 3, section 3.2.6.

5.2.2 Genome Sequencing of four F. proliferatum isolates
DNA sequencing was carried out utilising the service provider facility at the University of Cambridge. Approximately 2 µg of genomic DNA in the TE buffer (Tris-EDTA elution buffer provided in the Sigma Kit) was used for the construction of libraries to generate genome sequence data. Two genomic libraries were prepared using a Truseq Nano and Nextera Mate-Pair library kits (Illumina) with an average insert size of 350 bp to 550 bp. The two libraries were sequenced using the Illumina sequencing platform MiSeq 600 cycles system to generate reads of 301 nucleotides average length from both forward and reverse fragments.
5.2.3 De novo assembly of the genome sequences of F. proliferatum isolates

Quality of the sequence data and the presence of the adaptor sequences were verified using the FASTQC program (FastQC version 0.115). The adaptor sequences where present in the raw reads and any low-quality nucleotide bases with the Phred scores less than Q20 were filtered using BBDuk plugin in the Geneious. De novo assemblies of the genome sequences of the four F. proliferatum isolates (R16, KF3377, ITEM 2341 and MPVP 328) were generated using the Velvet programme (Zerbino and Birney, 2008; Zerbino, 2010). The Velvet assembler uses a range of Kmers from 31 to 121, which was done manually at the interval of 10 initially and later narrowed down to an interval of 2 in order to generate optimal assembly. The default parameters set used a coverage cut off of 10 with the minimum contig length of 1000 and an automatic insert length to allow the programme to use the optimum insert length. The expected coverage was set at the default value of 50%. In addition, SPades was also used for the de novo assembly of the genome sequences of the four isolates (Hittalmani et al., 2016). The quality and completeness of the genomes assembled was assessed using QUAST programme (Gurevich et al., 2013) in terms of the N50 values, number of contigs and the total genome size. In addition, the algorithm Benchmarking Universal Single-Copy Orthologs (BUSCO) was also used to assess the assembly quality and completeness essentially based on the presence of common orthologs in eukaryotes (Waterhouse et al., 2013; Simao et al., 2015).

5.2.4 Identification of the FUM gene cluster and the flanking regions in the genome of F. proliferatum isolates

Annotated FUM gene cluster sequence of the F. fujikuroi (JN807324.1) available in the NCBI database was identified and downloaded. Blast searches of the genome assemblies of the four F. proliferatum isolates generated in this study were performed using the downloaded F. fujikuroi sequence. The FUM gene cluster location was identified in each of the genomes and the data extracted. The FUM gene cluster sequence data extracted for the four F. proliferatum isolates (R16, KF3377, ITEM 2341 and MPVP 328) were aligned to corresponding data from Fusarium species (F. fujikuroi, F. verticillioides and F. oxysporum) available in NCBI using the Mauve genome alignment algorithm plugin in Geneious. Core genes of the FUM cluster, their CDS and mRNA were manually annotated manually in the four F. proliferatum isolates based on the reference species. The annotated data were analysed to determine the cluster architecture and the number, order, orientation of the genes, as well as the SNPs and the amino acid differences in selected genes of the four F. proliferatum isolates.

To identify the flanking regions at the 5’ and 3’ of the FUM gene cluster in the four F. proliferatum isolates, data available in NCBI for Fusarium species F. globosum and F.
were downloaded on the basis of the shared genome context (GC2) of species belonging to the *fujikuroi* complex. The flanking regions of the four *F. proliferatum* isolates characterised in this study were identified by Blast searches of the genome assemblies as described above. Genes present in the 5' and 3' flanking regions, their order, orientation and size, as well as the intergenic regions were determined by performing multiple sequence alignments and manual annotation.

### 5.3 Experimental approaches

*Fusarium proliferatum* isolates R16 (onion, UK), KF3377 (garlic, Poland), ITEM 2341 (date palm, Saudi Arabia) and MPVP 328 (maize, USA) reflecting the biogeographic and genetic diversity within the species were selected for genome sequencing. For each isolate, two genomic fragment libraries were prepared and more than 100x coverage sequencing was carried out using the Illumina MiSeq 600 cycles system. A range of web-accessible software as well as plugins from the licensed package Geneious were utilised to perform *de novo* assembly of the genomes as well as for various analysis including the identification, extraction and annotation of the FUM gene cluster. Further analysis to assess the level of differences among the four *F. proliferatum* isolates at the DNA and protein levels were carried out with the FUM cluster genes.

### 5.4 Results

#### 5.4.1 Genome sequencing, *de novo* assembly and characteristics

For each isolate, a small insert paired-end library and a large insert mate-paired library were prepared using appropriate kits from Illumina as referred to in 5.2.3. Following the generation of fragment clusters using Illumina’s patented technology for the two libraries for each *F. proliferatum* isolate, genome sequencing was carried out. The number of reads with the paired end libraries varied from 6.18 M for isolate ITEM 2341 to 8.75 M for isolate KF3377; and the number of reads with the mate paired libraries varied from 10.50 M for isolate MPVP 328 to 12.25 for isolate KF3377. The total number of reads generated together from the two libraries ranged from 16.73 M for MPVP 328 to 21.00 M for isolate KF3377. The total sequence generated varied from 4.93 billion bases for MPVP 328 to 6.18 billion bases for KF3377 yielding more than 100X coverage at an estimated genome size of 45 MB (Table 5.1).

*De novo* assembly of the curated reads was performed using widely used genome assemblers Velvet and SPades and the key statistics indicative of the assembly quality have been generated. Among the four *F. proliferatum* isolates, the genome size varied...
from 43.96 to 50.00 MB, with isolates R16 and ITEM 2341 showing ~ 45.6 MB. The number of contigs representing the genome assembly varied considerably from 142 with MPVP 328 to 2377 with KF3377. The contig N50 length varied from 0.65 MB with KF3377 to 4.35 MB with MPVP 328 (Table 5.1). BUSCO analysis of the genome assemblies returned a score of 99.7 % coverage for each *F. proliferatum* isolate validating the assembly quality (full data not shown).

To predict the total gene set for each isolate, the assembly was analysed used AUGUSTUS initially (data not shown). This data was further analysed with the OrthoVenn to identify the gene distribution pattern across the four isolates. Based on the OrthoVenn data, the total gene set varied from 13,832 in MPVP 328 to 14,157 in KF3377. The OrthoVenn analysis identified 12,980 genes as common to all four isolates and varying number of genes unique to each isolate (e.g. 0 to 134) and shared by two (e.g. 63 to 238) or three isolates (Figure 5.1).
Table 5.1 Details of the next generation sequencing and assembly process and the genome characteristics of four isolates belonging to *Fusarium proliferatum*

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>F. proliferatum isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R16</td>
</tr>
<tr>
<td>Number of clusters from Paired end library* – Short inserts (M)</td>
<td>3.37</td>
</tr>
<tr>
<td>Number of clusters from Mate paired library** – Long inserts (M)</td>
<td>5.65</td>
</tr>
<tr>
<td>Total clusters</td>
<td>9.03</td>
</tr>
<tr>
<td>Number of reads from Paired end library (M)</td>
<td>6.74</td>
</tr>
<tr>
<td>Number of reads from Mate paired library (M)</td>
<td>11.30</td>
</tr>
<tr>
<td>Total reads (M)</td>
<td>18.04</td>
</tr>
<tr>
<td>Total sequence generated (Billion bases)</td>
<td>5.32</td>
</tr>
<tr>
<td>Coverage (X)</td>
<td>116.47</td>
</tr>
<tr>
<td>Genome assembly size (MB)</td>
<td>45.65</td>
</tr>
<tr>
<td>Contigs</td>
<td>751</td>
</tr>
<tr>
<td>Largest contig length (MB)</td>
<td>6.34</td>
</tr>
<tr>
<td>N50 (MB)</td>
<td>2.37</td>
</tr>
<tr>
<td>N75 (MB)</td>
<td>1.06</td>
</tr>
<tr>
<td>L50</td>
<td>8</td>
</tr>
<tr>
<td>L75</td>
<td>14</td>
</tr>
<tr>
<td>GC content %</td>
<td>48.3</td>
</tr>
<tr>
<td>Predicted genes*</td>
<td>14,117</td>
</tr>
</tbody>
</table>

* The Paired end libraries were prepared using Illumina Truseq Nano kit

** The Mate paired libraries were prepared using the Nextera mate pair kit

* Predicted gene set is based on OrthoVenn output following an initial analysis with Augustus
Figure 5.1 Gene distribution pattern among the *Fusarium proliferatum* isolates R16 (green), KF3377 (blue), ITEM 2341 (red) and MPVP 328 (yellow) based on OrthoVenn analysis of the predicted total gene sets

Venn diagram shows the number of genes that are unique to each isolate, genes shared by two or three isolates and genes common to all four isolates (e.g. 12,980). Output was generated using OrthoVenn at http://probes.pw.usda.gov/OrthoVenn/index.php
5.4.2  FUM gene cluster of *Fusarium proliferatum*

This is first report identifying the FUM gene cluster in the mycotoxigenic pathogen *F. proliferatum*. The FUM gene cluster was identified from four isolates R16, KF3377, ITEM 2341 and MPVP 328 selected in this work based on diverse host and geographic origins. In all four isolates, the FUM cluster consisted of 16 genes and their order as well as orientation was also identical (Figure 5.2). Comparative analysis of the FUM gene cluster data generated for the *F. proliferatum* isolates was carried out with corresponding data of *F. fujikuroi*, *F. verticillioides* and *F. oxysporum* downloaded from NCBI. The number of genes, their order and orientation did not vary across these species (Figure 5.2).

The overall size of the FUM gene cluster among the four isolates of *F. proliferatum* was comparable ranging between 45,300 to 45,320 bp (Table 5.2). FUM cluster in *F. verticillioides* was comparable at 45,586 bp; whereas, the cluster size was more variable in *F. fujikuroi* (46,627 bp) and *F. oxysporum* (47,056 bp). Many of the 16 genes within the FUM cluster showed some variability in size across the four species. For example, Fum8 and FUM17 genes showed 88 and 191 bp differences in size, respectively between *F. proliferatum*, *F. fujikuroi*, *F. verticillioides* and *F. oxysporum* (Table 5.3). FUM3 gene size at 903 bp was the same in *F. proliferatum* as well as the other three *Fusarium* species. In addition, the four *F. proliferatum* isolates possessed the same number of exons and introns in each of the 16 genes within the FUM cluster. Overall, the pattern of exons and introns within the FUM cluster genes also looks similar in the three reference species with the exception of the FUM18 gene, which showed a variation in the number of exons in *F. verticillioides* (Figure 5.3).
Figure 5.2 Order and orientation of the 16 genes located within the FUM cluster in four *F. proliferatum* isolates R16 (onion, UK), MPVP 328 (maize, USA), KF3377 (garlic, Poland) and ITEM 2341 (date palm, Saudi Arabia) characterised in this study.

FUM21 and FUM 19 represent the terminal genes. Data from NCBI for *F. fujikuroi* (FF), *F. verticillioides* (FV) and *F. oxysporum* (FO) has been included as reference. Annotation of the genes, their order and orientation was determined based on the alignment of the sequence of the four *F. proliferatum* isolates and three references species using Geneious software. The numbers above the arrow represent the FUM gene numbers (e.g. 1 is FUM1 gene and 13 is FUM13 gene).
### Table 5.2 General information about isolates and their FUM clusters size

<table>
<thead>
<tr>
<th>* Isolate code/Accession number</th>
<th>Species</th>
<th>Cluster size (bp)</th>
<th>Host &amp; Geographic location</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>R16</td>
<td><em>F. proliferatum</em></td>
<td>45,300</td>
<td>Onion, United Kingdom</td>
<td>Present Study</td>
</tr>
<tr>
<td>KF3377</td>
<td><em>F. proliferatum</em></td>
<td>45,318</td>
<td>Garlic, Poland</td>
<td>Present Study</td>
</tr>
<tr>
<td>ITEM 2341</td>
<td><em>F. proliferatum</em></td>
<td>45,321</td>
<td>Date Palm, Saudi Arabia</td>
<td>Present Study</td>
</tr>
<tr>
<td>MPVP 328</td>
<td><em>F. proliferatum</em></td>
<td>45,317</td>
<td>Maize, United States</td>
<td>Present Study</td>
</tr>
<tr>
<td>JN807324.1</td>
<td><em>F. fujikuroi</em></td>
<td>46,627</td>
<td>Not known</td>
<td>NCBI, Unpublished</td>
</tr>
<tr>
<td>AF155773.5</td>
<td><em>F. verticillioides</em></td>
<td>45,586</td>
<td>Maize, California</td>
<td>Desjardins <em>et al.</em>, 2002</td>
</tr>
<tr>
<td>EU4499979</td>
<td><em>F. oxysporum</em></td>
<td>47,056</td>
<td>Not known</td>
<td>NCBI, Unpublished</td>
</tr>
</tbody>
</table>

* NCBI accession number for the FUM gene cluster
Table 5.3 The size range of the FUM cluster genes involved in the biosynthesis of fumonisins in *F. proliferatum*, *F. fujikuroi*, *F. verticillioides* and *F. oxysporum*

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene size (bp)</th>
<th><em>F. proliferatum</em></th>
<th><em>F. fujikuroi</em></th>
<th><em>F. verticillioides</em></th>
<th><em>F. oxysporum</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>FUM21</td>
<td>2586</td>
<td>2553</td>
<td>2595</td>
<td>2589</td>
<td></td>
</tr>
<tr>
<td>FUM1</td>
<td>8140</td>
<td>8140</td>
<td>8163</td>
<td>8152</td>
<td></td>
</tr>
<tr>
<td>FUM6</td>
<td>3612</td>
<td>3612</td>
<td>3593</td>
<td>3617</td>
<td></td>
</tr>
<tr>
<td>FUM7</td>
<td>1263</td>
<td>1263</td>
<td>1275</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>FUM8</td>
<td>2848</td>
<td>2855</td>
<td>2936</td>
<td>2855</td>
<td></td>
</tr>
<tr>
<td>FUM3</td>
<td>903</td>
<td>903</td>
<td>903</td>
<td>903</td>
<td></td>
</tr>
<tr>
<td>FUM10</td>
<td>1833</td>
<td>1835</td>
<td>1786</td>
<td>1863</td>
<td></td>
</tr>
<tr>
<td>FUM11</td>
<td>1161</td>
<td>1161</td>
<td>1164</td>
<td>1171</td>
<td></td>
</tr>
<tr>
<td>FUM2</td>
<td>1696</td>
<td>1696</td>
<td>1735</td>
<td>1694</td>
<td></td>
</tr>
<tr>
<td>FUM13</td>
<td>1104</td>
<td>1104</td>
<td>1110</td>
<td>1104</td>
<td></td>
</tr>
<tr>
<td>FUM14</td>
<td>1950</td>
<td>1951</td>
<td>1950</td>
<td>1954</td>
<td></td>
</tr>
<tr>
<td>FUM15</td>
<td>1864</td>
<td>1864</td>
<td>1856</td>
<td>1861</td>
<td></td>
</tr>
<tr>
<td>FUM16</td>
<td>2219</td>
<td>2220</td>
<td>2223</td>
<td>2224</td>
<td></td>
</tr>
<tr>
<td>FUM17</td>
<td>1220</td>
<td>1045</td>
<td>1236</td>
<td>1222</td>
<td></td>
</tr>
<tr>
<td>FUM18</td>
<td>1485</td>
<td>1487</td>
<td>1524</td>
<td>1490</td>
<td></td>
</tr>
<tr>
<td>FUM19</td>
<td>4760</td>
<td>4751</td>
<td>4806</td>
<td>4792</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.3 The structure of the FUM 18 gene in *Fusarium proliferatum* (FP), *F. fujikuroi* (FF), *F. verticillioides* (FV) and *F. oxysporum* (FO) showing a variation in the number of exons in FV.
5.4.3  The FUM cluster variability among the four *F. proliferatum* isolates

Overall, the number of genes, exons, introns and their sizes were mostly similar across all four *F. proliferatum* isolates characterised (Table 5.4). Whereas most variability existed at the nucleotide base level mainly concentrated within the exons rather than the introns. For example, the number of SNPs within the exons ranged from 10 in the FUM3 gene to 104 in the fumonisin polyketide synthase gene FUM1. The number of SNPs within the introns ranged from none in the FUM10 gene to 23 in FUM14 gene.
Table 5.4 Variability in the FUM cluster genes among the four *Fusarium proliferatum* isolates*

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene size (bp)</th>
<th>Exon number (N)/ Intron number (N-1)</th>
<th>Exon size (bp)</th>
<th>SNPs in Exons</th>
<th>Intron** size</th>
<th>SNPs in Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUM21</td>
<td>2586</td>
<td>9</td>
<td>2067</td>
<td>23</td>
<td>519</td>
<td>16</td>
</tr>
<tr>
<td>FUM1</td>
<td>8140</td>
<td>7</td>
<td>7743</td>
<td>104</td>
<td>397</td>
<td>9</td>
</tr>
<tr>
<td>FUM6</td>
<td>3612</td>
<td>4</td>
<td>3348</td>
<td>37</td>
<td>264</td>
<td>4</td>
</tr>
<tr>
<td>FUM7</td>
<td>1263</td>
<td>1</td>
<td>1263</td>
<td>11</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>FUM8</td>
<td>2848</td>
<td>6</td>
<td>2493</td>
<td>37</td>
<td>355</td>
<td>11</td>
</tr>
<tr>
<td>FUM3</td>
<td>903</td>
<td>1</td>
<td>903</td>
<td>10</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>FUM10</td>
<td>1833</td>
<td>3</td>
<td>1686</td>
<td>16</td>
<td>147</td>
<td>0</td>
</tr>
<tr>
<td>FUM11</td>
<td>1161</td>
<td>5</td>
<td>903</td>
<td>13</td>
<td>258</td>
<td>NA</td>
</tr>
<tr>
<td>FUM2</td>
<td>1696</td>
<td>4</td>
<td>1509</td>
<td>20</td>
<td>187</td>
<td>5</td>
</tr>
<tr>
<td>FUM13</td>
<td>1104</td>
<td>1</td>
<td>1104</td>
<td>37</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>FUM14</td>
<td>1950</td>
<td>5</td>
<td>1656</td>
<td>41</td>
<td>294</td>
<td>23</td>
</tr>
<tr>
<td>FUM15</td>
<td>1864</td>
<td>2</td>
<td>1800</td>
<td>24</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td>FUM16</td>
<td>2219</td>
<td>4</td>
<td>2031</td>
<td>40</td>
<td>188</td>
<td>4</td>
</tr>
<tr>
<td>FUM17</td>
<td>1220</td>
<td>2</td>
<td>1165</td>
<td>35</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>FUM18</td>
<td>1485</td>
<td>4</td>
<td>1242</td>
<td>28</td>
<td>243</td>
<td>7</td>
</tr>
<tr>
<td>FUM19</td>
<td>4760</td>
<td>5</td>
<td>4518</td>
<td>79</td>
<td>242</td>
<td>14</td>
</tr>
</tbody>
</table>

* As reflected by the number of SNPs in the exons and introns
* R16 (Onion, UK); KF3377 (Garlic, Poland); ITEM 2341 (Date palm, Saudi Arabia); and MPVP 328 (Maize, USA)

** FUM1 and FUM2 genes have one nucleotide deletion in intron region; - denotes no introns and NA denotes not applicable
5.4.4  Flanking genes of the FUM cluster in *Fusarium proliferatum*

In the present study, genomic location and flanking genes of the FUM cluster in *F. proliferatum* was identified. All four isolates showed the same genomic location with the same set of genes each flanking the FUM21 at the 5’ and FUM19 at 3’ of the cluster (Figure 5.4). The order and orientation of the flanking genes across the *F. proliferatum* isolates were the same with slight differences in the size of each gene. However, noticeable variability was observed at the nucleotide level reflected by the number of SNPs among the four isolates characterized (Table 5.5). For example, the putative DNA binding protein gene ZCB1 at the 5’ end exhibited the highest number of SNPs (73) with the GAT1 gene at the 3’ end showing the lowest (12) number of SNPs (Table 5.5).
Figure 5.4 Order and orientation of the genes flanking the FUM21 at the 5’ and the FUM19 at the 3’ end of the FUM cluster in the *Fusarium proliferatum* isolates R16, KF3377, ITEM 2314 and MPVP 328

Scale bars below the flanking regions represent 1000 bp. The FUM cluster is represented by the terminal genes FUM21 and FUM19 with the black arrows indicating the 14 other genes not shown.
### Table 5.5 Variability in the genes flanking the FUM cluster among the *Fusarium proliferatum* isolates R16, KF3377, ITEM 2341 and MPVP 328

<table>
<thead>
<tr>
<th>Terminal gene</th>
<th>Gene name</th>
<th><em>Gene size (bp)</em></th>
<th>SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUM21 (5’)</td>
<td>MFS1</td>
<td>1666</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>ZCB1</td>
<td>1509</td>
<td>73</td>
</tr>
<tr>
<td>FUM19 (3’)</td>
<td>ANK1</td>
<td>1288</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>GAT1</td>
<td>792</td>
<td>12</td>
</tr>
</tbody>
</table>

* ANK1 gene size was 1288 bp in MPVP 328 and 1272 bp in KF3377

### 5.4.5 Variability in the protein sequences of selected genes from the FUM cluster

Among the FUM cluster genes, four were selected including two genes (FUM1 and FUM8) reflecting their key role in the biosynthesis of different types of fumonisins as well as the level of SNPs (e.g. 104 in FUM1). The other two represented the terminal genes FUM21 and FUM19, with the FUM19 showing 79 SNPs (Table 5.4). Protein sequences encoded by these genes among the four *F. proliferatum* isolates showed a number of amino acid changes ranging from 12 to 40 (Figure 5.5; 5.6; 5.7). The fumonisin polyketide synthase encoded by the FUM1 gene showed the highest number of amino acid changes (40) and it was noticeable that a number of amino acid changes (e.g. 13) were shared between isolates ITEM 2341 and KF3377 (Figure 5.6). In the proteins encoded by FUM21 and FUM19 genes, 12 and 31 amino acid differences have been identified, respectively in the present study (Figure 5.5). The FUM1 encodes a polyketide synthase and 40 amino acid differences have been identified in the present study. Six putative functional domains have been identified based on previous work in *F. fujikuroi* (Proctor *et al.*, 1999). Among the 4 *F. proliferatum* isolates characterised in this study, the KR domain showed a single amino acid change (Figure 5.6). The FUM8 encodes α-Oxoamine synthase and 13 amino acid differences have been identified in the present study. Three putative functionally important regions of the protein have been identified based on previous work in *Fusarium* and related fungal species (Proctor *et al.*, 2008). Among the 4 *F. proliferatum* isolates characterised in this study, stretch 1 showed a single amino acid change (Figure 5.7).
FUM19

ITEM.2341. (0-Sa-H)  MDFEQLDSASPFGPFWKGCGRGNGFDTFLKFELIIIIIFIAPSCVFTLVFVRIILLVSKQII
R16. (On-Uk-L) MDFEQLDSASPFGPFWKGCGRGNGFDTFLKFELIIIIIFIAPSCVFTLVFVRIILLVSKQII
KF3377. (Gr-P1-H) MDFEQLDSASPFGPFWKGCGRGNGFDTFLKFELIIIIIFIAPSCVFTLVFVRIILLVSKQII
MPVP. 328. (Mz-Ua-L) MDFEQLDSASPFGPFWKGCGRGNGFDTFLKFELIIIIIFIAPSCVFTLVFVRIILLVSKQII

ITEM.2341. (0-Sa-H)  TGNHLPVLGKLVNYAVAFRIRIIILISPGPSNALSFLSISAPALGILVAVSATAVLS
R16. (On-Uk-L) TGNHLPVLGKLVNYAVAFRIRIIILISPGPSNALSFLSISAPALGILVAVSATAVLS
KF3377. (Gr-P1-H) TGNHLPVLGKLVNYAVAFRIRIIILISPGPSNALSFLSISAPALGILVAVSATAVLS
MPVP. 328. (Mz-Ua-L) TGNHLPVLGKLVNYAVAFRIRIIILISPGPSNALSFLSISAPALGILVAVSATAVLS

ITEM.2341. (0-Sa-H)  YAEHWSRSPFILLSTLYCSTLSSLLDLDAHDTLWNLASSLTGETYSSVSVAVAIKAFSTW
R16. (On-Uk-L) YAEHWSRSPFILLSTLYCSTLSSLLDLDAHDTLWNLASSLTGETYSSVSVAVAIKAFSTW
KF3377. (Gr-P1-H) YAEHWSRSPFILLSTLYCSTLSSLLDLDAHDTLWNLASSLTGETYSSVSVAVAIKAFSTW
MPVP. 328. (Mz-Ua-L) YAEHWSRSPFILLSTLYCSTLSSLLDLDAHDTLWNLASSLTGETYSSVSVAVAIKAFSTW

ITEM.2341. (0-Sa-H)  LERSQSEPFDWSSDVEKIQDSTGVYLSNASFWMGLGGLLLYGKVLALSDLPLTDGDML
R16. (On-Uk-L) LERSQSEPFDWSSDVEKIQDSTGVYLSNASFWMGLGGLLLYGKVLALSDLPLTDGDML
KF3377. (Gr-P1-H) LERSQSEPFDWSSDVEKIQDSTGVYLSNASFWMGLGGLLLYGKVLALSDLPLTDGDML
MPVP. 328. (Mz-Ua-L) LERSQSEPFDWSSDVEKIQDSTGVYLSNASFWMGLGGLLLYGKVLALSDLPLTDGDML

ITEM.2341. (0-Sa-H)  GALYERFKYSHTLVTKHTQOQNYRAALLHALSLLANILLPLVLRVAILGSLAQAF
R16. (On-Uk-L) GALYERFKYSHTLVTKHTQOQNYRAALLHALSLLANILLPLVLRVAILGSLAQAF
KF3377. (Gr-P1-H) GALYERFKYSHTLVTKHTQOQNYRAALLHALSLLANILLPLVLRVAILGSLAQAF
MPVP. 328. (Mz-Ua-L) GALYERFKYSHTLVTKHTQOQNYRAALLHALSLLANILLPLVLRVAILGSLAQAF

ITEM.2341. (0-Sa-H)  LTAQRILRYLEDQPHNSYGLTAVLGYICTSCNYFYHERLLCVVRGCLASAIFH
R16. (On-Uk-L) LTAQRILRYLEDQPHNSYGLTAVLGYICTSCNYFYHERLLCVVRGCLASAIFH
KF3377. (Gr-P1-H) LTAQRILRYLEDQPHNSYGLTAVLGYICTSCNYFYHERLLCVVRGCLASAIFH
MPVP. 328. (Mz-Ua-L) LTAQRILRYLEDQPHNSYGLTAVLGYICTSCNYFYHERLLCVVRGCLASAIFH

ITEM.2341. (0-Sa-H)  KTVLLSLNVTSDRTASVMTSTDLSRHKFNLMHEDIPIFIAGLAWFLWRQVGLAFIA
R16. (On-Uk-L) KTVLLSLNVTSDRTASVMTSTDLSRHKFNLMHEDIPIFIAGLAWFLWRQVGLAFIA
KF3377. (Gr-P1-H) KTVLLSLNVTSDRTASVMTSTDLSRHKFNLMHEDIPIFIAGLAWFLWRQVGLAFIA
MPVP. 328. (Mz-Ua-L) KTVLLSLNVTSDRTASVMTSTDLSRHKFNLMHEDIPIFIAGLAWFLWRQVGLAFIA

ITEM.2341. (0-Sa-H)  PIGLVLSLGPVFWALGROYGQVWVMBKIQNRVAITADIKHILVGUIVTPLLEATI
R16. (On-Uk-L) PIGLVLSLGPVFWALGROYGQVWVMBKIQNRVAITADIKHILVGUIVTPLLEATI
KF3377. (Gr-P1-H) PIGLVLSLGPVFWALGROYGQVWVMBKIQNRVAITADIKHILVGUIVTPLLEATI
MPVP. 328. (Mz-Ua-L) PIGLVLSLGPVFWALGROYGQVWVMBKIQNRVAITADIKHILVGUIVTPLLEATI

ITEM.2341. (0-Sa-H)  QKARESRLASGRGIRQRTASLITAIAPAFDITAPGILMAATKQNFTSQGNYTAITALLALL
R16. (On-Uk-L) QKARESRLASGRGIRQRTASLITAIAPAFDITAPGILMAATKQNFTSQGNYTAITALLALL
KF3377. (Gr-P1-H) QKARESRLASGRGIRQRTASLITAIAPAFDITAPGILMAATKQNFTSQGNYTAITALLALL
MPVP. 328. (Mz-Ua-L) QKARESRLASGRGIRQRTASLITAIAPAFDITAPGILMAATKQNFTSQGNYTAITALLALL

ITEM.2341. (0-Sa-H)  TVPFLGSIFRSVPVSPLMLAFACLRQIFAFLDELRKLRHITSITDTSISGKKEIIFEL
R16. (On-Uk-L) TVPFLGSIFRSVPVSPLMLAFACLRQIFAFLDELRKLRHITSITDTSISGKKEIIFEL
KF3377. (Gr-P1-H) TVPFLGSIFRSVPVSPLMLAFACLRQIFAFLDELRKLRHITSITDTSISGKKEIIFEL
MPVP. 328. (Mz-Ua-L) TVPFLGSIFRSVPVSPLMLAFACLRQIFAFLDELRKLRHITSITDTSISGKKEIIFEL

ITEM.2341. (0-Sa-H)  NAPRSGAILIALDFGQWKSQPCQRINLTVNASALTIVIGVPGSKSTLCALGTEL
R16. (On-Uk-L) NAPRSGAILIALDFGQWKSQPCQRINLTVNASALTIVIGVPGSKSTLCALGTEL
KF3377. (Gr-P1-H) NAPRSGAILIALDFGQWKSQPCQRINLTVNASALTIVIGVPGSKSTLCALGTEL
MPVP. 328. (Mz-Ua-L) NAPRSGAILIALDFGQWKSQPCQRINLTVNASALTIVIGVPGSKSTLCALGTEL

ITEM.2341. (0-Sa-H)  FATGKVLDHDSACRGYCQYPFLNCSTQIGFSWNPVYVLWIKASMLPDYLDNE
R16. (On-Uk-L) FATGKVLDHDSACRGYCQYPFLNCSTQIGFSWNPVYVLWIKASMLPDYLDNE
KF3377. (Gr-P1-H) FATGKVLDHDSACRGYCQYPFLNCSTQIGFSWNPVYVLWIKASMLPDYLDNE
MPVP. 328. (Mz-Ua-L) FATGKVLDHDSACRGYCQYPFLNCSTQIGFSWNPVYVLWIKASMLPDYLDNE
Figure 5.5 Multiple sequence alignment of proteins encoded by FUM21 and FUM19 flanking region genes in the four *F. proliferatum* isolates

Isolates originated from diverse hosts and geographic locations and produce variable levels of fumonisins FB1 and FB2: ITEM 2341 (Garlic, Poland); R16 (Onion, UK); KF3377 (Date palm, Saudi Arabia) and MPVP 328 (Maize, USA). FUM21 encodes a transcription factor and FUM19 encodes an ABC transporter. Amino acid residue positions showing changes are highlighted in yellow and the changed amino acids are highlighted in red. Functional domains have not yet been identified and/or analysed adequately in these proteins, in the available literature.
Isolates originated from diverse hosts and geographic locations and produce variable levels of fumonisins FB1 and FB2: ITEM 2341 (Garlic, Poland); R16 (Onion, UK); KF3377 (Date palm, Saudi Arabia) and MPVP 328 (Maize, USA). The FUM1 encodes a putative functional gene in the four F. proliferatum isolates.

Figure 5.6 Multiple sequence alignment of proteins encoded by FUM1 gene in the four F. proliferatum isolates

Isolates originated from diverse hosts and geographic locations and produce variable levels of fumonisins FB1 and FB2: ITEM 2341 (Garlic, Poland); R16 (Onion, UK); KF3377 (Date palm, Saudi Arabia) and MPVP 328 (Maize, USA). The FUM1 encodes a putative functional gene in the four F. proliferatum isolates.

Figure 5.6 Multiple sequence alignment of proteins encoded by FUM1 gene in the four F. proliferatum isolates.
KF3377.(Gr-Pl-H)  AGRLILDDSGRLGKIGPRHLGYLDLMEREHGVSFLKHIGKKLASKTEVVVTGSFFNAF
R16.(On-Uk-L) AGRLILDDSGRLGKIGPRHLGYLDLMEREHGVSFLKHIGKKLASKTEVVVTGSFFNAF
ITEM.2341.(O-Sa-H) AGRLILDDSGRLGKIGPRHLGYLDLMEREHGVSFLKHIGKKLASKTEVVVTGSFFNAF
MPVP.328.(Mz-Us-L) AGRLILDDSGRLGKIGPRHLGYLDLMEREHGVSFLKHIGKKLASKTEVVVTGSFFNAF

************************************************************

KF3377.(Gr-Pl-H)  GQQGGYIISSAPFVEVHTVSSKFVFSTPVTQVAMSNGKVLIELSRTGS
R16.(On-Uk-L)  GQQGGYIISSAPFVEVHTVSSKFVFSTPVTQVAMSNGKVLIELSRTGS
ITEM.2341.(O-Sa-H)  GQQGGYIISSAPFVEVHTVSSKFVFSTPVTQVAMSNGKVLIELSRTGS
MPVP.328.(Mz-Us-L)  GQQGGYIISSAPFVEVHTVSSKFVFSTPVTQVAMSNGKVLIELSRTGS

******************************************************

Figure 5.7 Multiple sequence alignment of proteins encoded by FUM8 genes in the four *F. proliferatum* isolates
Isolates originated from diverse hosts and geographic locations and produce variable levels of fumonisins FB1 and FB2: ITEM 2341 (Garlic, Poland); R16 (Onion, UK); KF3377 (Date palm, Saudi Arabia) and MPVP 328 (Maize, USA). FUM8 encodes α-Oxoamine synthase and the amino acid residue positions showing changes are highlighted in yellow and the changed amino acids are highlighted in red. Three putative functionally important regions of the protein are shown with the amino acid residues underlined and the areas indicated between arrows, stretch 1 showed a single amino acid change.

5.4.6 Evolutionary relationship reflected by the FUM clusters
The FUM cluster sequence data developed in this study for the four *F. proliferatum* isolates and the corresponding for *F. fujikuroi*, *F. verticillioides* and *F. oxysporum* data downloaded from the NCBI were subjected to Bayesian analysis based on the Markov Chain Monte Carlo (MCMC) method. The *F. oxysporum* and *F. verticillioides* were represented distinctively on the Bayesian phylogenetic tree, whilst *F. fujikuroi* and *F. proliferatum* were clustered together with 100 % posterior probability. Among the 4 isolates of *F. proliferatum*, R16 and MPVP 328 were grouped together with 100 % posterior probability (Figure 5.8).
Figure 5.8 Bayesian consensus tree reflecting the evolutionary relationships based on the FUM gene cluster sequences of four species, *F. proliferatum* isolates along with *F. fujikuroi*, *F. verticillioides* and *F. oxysporum*

MPVP 328 (Maize, USA), R16 (Onion, UK), ITEM 2341 (Garlic, Poland) and KF3377 (Date palm, Saudi Arabia) are *F. proliferatum* isolates; FF, FV and FO are *F. fujikuroi*, *F. verticillioides* and *F. oxysporum*, respectively. The numbers above the branch represent posterior probability value.

5.5 Discussion

The work described in this chapter is the first effort to develop the genome sequence for diverse *F. proliferatum* isolates representing different hosts and geographic locations: R16, Onion in UK; KF3377, Garlic in Poland; ITEM 2341, Date palm in Saudi Arabia; and MPVP 328, Maize in USA. The genome sequences of these four isolates were explored to characterise the FUM gene cluster (~45.3 kb) for the first time in *F. proliferatum*.

Use of the Illumina sequencing platform with paired end (short inset) as well as mate paired (long insert) libraries generated total sequence data ranging from 4.93 to 6.18 billion bases for each isolate (Table 5.1). VELVET (Zerbino et al., 2011; Baroncelli et al., 2016) and SPAdes (Hittalmani et al., 2016) were tested to achieve optimum *de novo* genome assembly. SPAdes performed better in terms of the genome assembly size, number of contigs and the N50. Assembly output parameters suggest efficient implementation of the processes as reflected by the contig numbers and coverage. For example, R16 with 751 contigs, 45.65Mb genome
size and 116.47-fold coverage; 311 contigs for ITEM 2341 with 45.67Mb genome size, 111.19-fold coverage (Table 5.1). These parameters compare well with published fungal genome sequences (Chiara et al., 2015; King et al., 2015; Baroncelli et al., 2016; Niehaus et al., 2016). Furthermore, BUSCO analysis of the four genomes of *F. proliferatum* identified 99.7% coverage of the total gene space. This is comparable to other fungal genomes recently sequenced, which ranged from 84.49 - 99.30% (Baroncelli et al., 2016). BUSCO algorithm interrogates the presence of the full set of single copy orthologous genes to assess the quality and to quantify the completeness of a genome sequenced (Waterhouse et al., 2013; Simao et al., 2015).

The genome size deciphered for the *F. proliferatum* isolates R16 (45.65Mb), ITEM 2341 (45.67Mb) and MPVP 328 (43.96Mb) are within the range of the genome size of two *F. proliferatum* isolates recently deposited at the NCBI (43.2Mb to 45.2Mb; Niehaus et al., 2016). However, for the fourth *F. proliferatum* isolate KF3377 from garlic in Poland characterised in this study, the genome size was estimated as 50Mb. Considerable differences in the genome size (47 - 61Mb) have been observed in strains of some related *Fusarium* species including *F. oxysporum* (Ma et al., 2011; Gou et al., 2014; Niehaus et al., 2016; Williams et al., 2016). Similar levels of genome size differences (37 - 45 Mb) are known in other fungal pathogens such as *Magnaporthe oryzae* even though the isolates used were from the same plant host (Dean et al., 2005; Xue et al., 2012; Dong et al., 2015). The observed variation in the genome size might be due the nature of the repetitive regions in the genome or genomic rearrangements (Ma et al, 2011). However, the high number of total raw sequence used and/or the quality of the bases generated cannot be ruled out in the case of KF3377 at this stage, although the genome assembly in comparable to the other three isolates in various quality parameters.

The total gene set for each of the four *F. proliferatum* isolates was predicted using AUGUSTUS to range from 14,235 to 16,046 (data not shown), which is comparable to the number of genes predicted for the ET1 (15,602) and NRRL62905 (16,509) isolates of *F. proliferatum* (NCBI - BioProject). The OrthoVenn algorithm which predicts genes that can be annotated yielded 13,832 - 14,157 for the four *F. proliferatum* isolates in this study (Table 5.1). Among these, 12,980 genes have been predicted to be common to the fours isolates with up to 134 genes potentially unique to an isolate (Figure 5.1). Similar report of genes unique to individual naturally occurring isolates are beginning to emerge with potential relationships to their adaptive divergence (e.g. Xue et al., 2012; Dong et al., 2015). Further analysis with algorithms such as MAKER, InterProScan and SignalP are being used in other fungi to confirm and/or refine the number of genes and their annotation.

The emerging pathogen *F. proliferatum* is well recognised as prolific producer of fumonisins that pose a threat to the health of animals and humans (Nelson et al., 1993; Logrieco et al., 1998; Seefelder et al., 2002). The fumonisin producing genes are known to be clustered together in the genome (FUM gene cluster) and are responsible for the entire fumonisins
biosynthetic pathway and regulation (Brown et al., 2007; Proctor et al., 2008; Rocha et al., 2016). A major focus of this chapter was to identify the FUM gene cluster in *F. proliferatum*, which has not yet been reported. This is the first report of the FUM gene cluster in *F. proliferatum* for four isolates from different geographical locations, as well as characterised biological attributes including levels of fumonisin production, maximum growth rate and sporulation level along with their genome sequence. These integrated resources would enable further assays and comparative analysis to gain new insights into genetic and environmental regulation of fumonisins production.

The FUM gene cluster identified in the genomes of the four *F. proliferatum* is ~45,300 bp and varied by no more 20 nucleotide bases. The FUM gene cluster size in *F. proliferatum* isolates is close to *F. verticillioides* (45,586bp) followed by *F. fujikuroi* (46,627bp) and *F. oxysporum* (47,056bp) with a maximum ~1,735bp difference (Table 5.2). The FUM cluster in *F. proliferatum* contained 16 genes in the same order and orientation (Figure 5.2) as reported in *F. fujikori*, *F. verticillioides* and *Fusarium oxysporum* (Desjardins et al., 2002; Proctor et al., 2008; NCBI, Unpublished).

Experimental research in *F. verticillioides* has shown that the FUM cluster genes play different roles in relation to fumonisin production and also the expression of some of the genes is related to adapting to different environmental conditions (Medina and Magan, 2010; Medina et al., 2013; Rocha et al., 2016). FUM 21 is the upstream terminal gene in the cluster located adjacent to the fumonisin polyketide synthase gene FUM1 (Figure 5.2). It encodes a transcription factor (DNA-binding domain) involved in regulation of the FUM gene expression (Brown et al., 2007). FUM1 gene speeds up the synthesis of a linear polyketide that forms the backbone structure of fumonisins (Yu et al., 2005), while FUM 8 gene encodes enzymes responsible for the addition of the amino groups (Christen and Methta, 2001; Proctor et al., 2008). The cluster encodes cytochrome P450 oxygenases through at least four genes (Yi et al., 2005; Lazzaro et al., 2012; Medina et al., 2013; Proctor et al., 2013, Rocha et al., 2016). FUM17 and FUM18 genes are similar to the AAL toxin producing genes in *Alternaria alternata* (Brandwagt et al., 2000), encoding longevity assurance factors but their function is not fully understood (Proctor et al., 2003). FUM19 encodes an ABC transporter and disruption of this transporter altered fumonisin production ratio. FUM17, 18, 19 genes are reported to be involved in self-protection function against toxic metabolites in general and whether these genes provide protection from fumonisins remains to be further investigated (Arora and Khachatourians, 2004).

For the first time, detailed data of the FUM gene cluster as well as the genes flanking the cluster at the 5’ and 3’ end in four *F. proliferatum* isolates has been presented. The putative functions of these genes identified in the *F. proliferatum* isolates, that exhibit up to 140-fold differences in the production of fumonisins, have been annotated with reference to the corresponding data from other *Fusarium* species (Tables 5.6 and 5.7; Seo et al., 2001; Proctor et al., 2003; Brown et al., 2005; Proctor et al., 2013).
Table 5.6 FUM cluster genes identified in the *Fusarium proliferatum* in this study and their putative functions with reference to previously known data in other *Fusarium* species

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene</th>
<th>Putative function*</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FUM21</td>
<td>Fumonisin biosynthetic transcription factor</td>
<td>AHC70694.1</td>
</tr>
<tr>
<td>2</td>
<td>FUM1</td>
<td>Polyketide synthase</td>
<td>ADD34559</td>
</tr>
<tr>
<td>3</td>
<td>FUM6</td>
<td>Cytochrome P450 Monoxygenase</td>
<td>BAA82526</td>
</tr>
<tr>
<td>4</td>
<td>FUM7</td>
<td>Dehydrogenase</td>
<td>NP_343836</td>
</tr>
<tr>
<td>5</td>
<td>FUM8</td>
<td>α-Oxoamine synthase</td>
<td>OS4649</td>
</tr>
<tr>
<td>6</td>
<td>FUM9</td>
<td>Dioxygenase</td>
<td>AAK01519</td>
</tr>
<tr>
<td>7</td>
<td>FUM10</td>
<td>Fatty acyl-CoA synthetase</td>
<td>NP_588549</td>
</tr>
<tr>
<td>8</td>
<td>FUM11</td>
<td>Tricarboxylate transporter</td>
<td>NP_594420</td>
</tr>
<tr>
<td>9</td>
<td>FUM12</td>
<td>Cytochrome P450 Monoxygenase</td>
<td>Q00707</td>
</tr>
<tr>
<td>10</td>
<td>FUM13</td>
<td>Short-chain dehydrogenase/reductase</td>
<td>NP_593981/T02760</td>
</tr>
<tr>
<td>11</td>
<td>FUM14</td>
<td>Peptide synthetase condensation domain</td>
<td>CAA61605</td>
</tr>
<tr>
<td>12</td>
<td>FUM15</td>
<td>Cytochrome P450 monoxygenase</td>
<td>NP_354568</td>
</tr>
<tr>
<td>13</td>
<td>FUM16</td>
<td>Fatty acyl-CoA synthetase</td>
<td>T49727</td>
</tr>
<tr>
<td>14</td>
<td>FUM17</td>
<td>Longevity Assurance Factor</td>
<td>NP_596102</td>
</tr>
<tr>
<td>15</td>
<td>FUM18</td>
<td>Longevity assurance factor</td>
<td>NP_593201</td>
</tr>
<tr>
<td>16</td>
<td>FUM19</td>
<td>Fumonisin biosynthetic ABC transporter</td>
<td>AHC70697.1</td>
</tr>
</tbody>
</table>

* Seo et al., 2001; Proctor et al., 2003; Brown et al., 2005; Proctor et al., 2013; and the sequence data accession numbers are from NCBI.

In *Fusarium* species, location of the FUM gene cluster in the genome is known to vary due to the horizontal transfer of the cluster. These have been designated as GC1, GC2, GC3a, GC3b and GC4 based on the genomic context and the genes present in the flanking region spanning the upstream (5’) and downstream (3’) of the cluster (Proctor et al., 2013). In this study, the 5’ and 3’ flanking regions of the FUM gene cluster in the four isolates was located and extracted from the assembled genomes (Table 5.7). MFS1 and ZCB1 on 5’ and ANK1 and GAT1 on the 3’ of the cluster were annotated/identified (Figure 5.4). These genes might influence the regulation and biosynthesis of the fumonisins. In *F. verticillioides*, FCK1, ZFR1, PAC1 and ZFK1 located outside the FUM gene cluster regulate the biosynthesis of FBs and these genes are required for each other to function (Flaherty et al., 2003; Flaherty et al., 2004).
This research is the first report to identify and annotate the 4 complete genes located upstream and downstream of the FUM gene cluster in *F. proliferatum*, building on the partial information available previously (Proctor *et al.*, 2013), confirming its GC2 status. The functions of these genes in the regulation and biosynthesis of the FUM gene cluster in the genome of *F. proliferatum* are still unknown. However, they are all members of important families involved in the transportation, regulation and biosynthesis of secondary metabolites. For example, MFS1 is belongs to a major facilitator superfamily involved in the transport of a wide range of compounds including toxins produced by fungal species. In *F. verticilloides* (GC1 group), the MFS was experimentally shown to be up-regulated in fumonisins-inducing condition but its role remains uncertain (Lopez-Errasquin *et al.*, 2006). ZCB1 is a member of the zinc cluster proteins family and are mostly regulators of both primary and secondary metabolites (Tsuji *et al.*, 2000; Flaherty and Woloshuk, 2004; Todd and Andrianopoulos, 2007). Another member of this family identified as the flanking gene of the FUM cluster in *F. verticilloides* is a positive regulator of FBs biosynthesis (Flaherty and Woloshuk, 2004). Flanking gene ANK1 is a prototype of Ankyrins family involved diverse cellular functions including activation, proliferation, contact and maintenance of specialised membrane domains (Han *et al.*, 2015). The physiological function of this gene in the context of fumonsins production is yet to be characterised. Taken together, the presence of FUM gene cluster containing 16 genes, flanked by 4 genes and the phylogenetic relationship (Figure 5.8) further confirmed *F. proliferatum* as a member of GC2 group alongside *F. fujikuroi* and *F. globosum* (Proctor *et al.*, 2013).

---

**Table 5.7 Genomic context group and the flanking regions genes identified in *Fusarium proliferatum* isolates and their putative functions based on corresponding information from other *Fusarium* species *

<table>
<thead>
<tr>
<th>FUM cluster Group</th>
<th>FUM cluster gene at 5' and 3'</th>
<th>Flanking gene</th>
<th>Putative function</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC2</td>
<td>FUM21</td>
<td>MFS1</td>
<td>Major facilitator superfamily transport protein</td>
<td>AHC70696.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZCB1</td>
<td>Zn2Cys6 binuclear cluster DNA-binding-like protein</td>
<td>AHC70695.1</td>
</tr>
<tr>
<td></td>
<td>FUM19</td>
<td>ANK1</td>
<td>Ankyrin-like protein</td>
<td>AHC70698.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAT1</td>
<td>Glutamine amidotransferase-like domain</td>
<td>AHC70702.1</td>
</tr>
</tbody>
</table>

* Based on Seo *et al.*, 2001; Proctor *et al.*, 2003; Brown *et al.*, 2005; Proctor *et al.*, 2013; and DNA accessions available at NCBI.
Fumonisins are categorised as FBs and FCs and the FUM1 and FUM8 genes have been reported to be very important in their formation (Proctor et al., 1999; Christen and Methta, 2001; Yu et al., 2005; Proctor et al., 2008). In the present study, considerable changes in the nucleotide and amino acid sequences of the FUM21, FUM1, FUM8 and FUM19 genes and their proteins among the four *F. proliferatum* isolates has been established (Table 5.4; Figure 5.5 to 5.7). It is pertinent to highlight that the FUM1 encoded fumonisin polyketide synthase showed a high level of amino acid differences (40) and 13 of these changes are shared by the isolates ITEM 2341 and KF3377 which produce much higher levels of fumonisins particularly FB1 compared to the other isolates. The polyketide synthase encoded by the FUM1 gene is responsible for the synthesis of the backbone structure of fumonisins, and the disruption of this gene reduced fumonisin production by over 99% in *F. fujikuroi* (Proctor et al., 1999, 2013). Based on previous research in *F. fujikuroi* (Proctor et al., 1999), six domains that are functionally important have been identified in the FUM1 protein sequence of *F. proliferatum* isolates in the present study. Among these six domains, the KR domain shows a single amino acid change that distinguishes the high and low fumonisin producing isolates of *F. proliferatum*. FUM8 gene encodes a α-oxoamine synthase and it has been shown in *F. oxysporum* and *F. verticillioides* that different orthologues determine whether FB or FC type fumonisins are produced. Furthermore, 3 regions that are important in the substrate specificity of the enzyme have been reported (Proctor et al., 2008). These regions identified and compared in the *F. proliferatum* isolates in the present study showed an amino acid change within the region 1 (Figure 5.7), but does not distinguish isolates producing different levels of fumonisins. The amino acid changes identified in the proteins encoded by the key genes such as FUM1 and FUM8 require further computational and molecular genetic analyses into the structure function relationships and the potential effects on the quantity and the type of fumonisins produced by *F. proliferatum* isolates.

In conclusion, the research reported in this chapter has made two major contributions by: 1) establishing the assembled genome sequences (average ~46.3 Mb) of four *F. proliferatum* isolates with diverse biological attributes and originating from different geographic locations; and 2) deciphering the entire FUM gene cluster (~45.3 Kb) as well as the flanking regions (~5.8 kb at 5’ and 4.8 kb at 3’) in these isolates including the amino acid differences in the flanking genes FUM19 and FUM21 as well as the core genes FUM1 and FUM8. Emerging tool kits for the identification and analysis of functional domains in key genes through bioinformatics as well as mutagenesis and transformation methodologies of *Fusarium* species would enable further empirical exploration of these data to gain an improved understanding of the production and regulation of fumonisins in the laboratory and in the field.
6 General Discussion, Conclusions and Perspectives

6.1 General Discussion

*Fusarium proliferatum* is distributed in diverse geographic locations affecting a range of important food crops (Yan and Dickman, 1996; Abdalla *et al*., 2000; Stankovic *et al*., 2007; Khudhair *et al*., 2014; Torabi *et al*., 2014). This emerging pathogen poses a serious threat to food security as well as to animal and human health with its strong mycotoxigenic potential (Leslie and Summerell, 2006). There is an increasing reliance on the phylogenetic species concept to accurately identify, characterise and diagnose *Fusarium* pathogens (Summerell *et al*., 2003; Tibayrenc *et al*., 2014; O’Donnell *et al*., 2015). This is due to the inadequacies of the morphological and biological species concepts to cope with the overlapping biological attributes displayed by *Fusarium* species isolates contained within the 20 species complexes (Leslie and Summerell, 2008).

A range of molecular markers including the ITS, TEF1, CAL, ACT, RPB1, RPB2, and FUM1 have been explored for phylogenetic analysis of *Fusarium* species (Geiser *et al*., 2004; Vaquero *et al*., 2004; Nalim *et al*., 2009; Schmitt *et al*., 2009; Stępień *et al*., 2011a; Stockinger *et al*., 2014). Nonetheless, there is a growing recognition of the need to develop new genetic markers with higher resolution to characterize isolates and species in evolving *Fusarium* species complexes such as *F. oxysporum*, *F. solani*, and *F. fujikuroi*, which includes *F. proliferatum* (O’Donnell *et al*., 2015).

In the present study, among the known markers used, CAL proved more effective than TEF1 to decipher the diversity within *F. proliferatum*, and the genetic groups distinguished showed links to the host and geographic origin of the isolates. This is interesting as the CAL marker was not particularly useful at the intraspecies level in a number of *Fusarium* pathogens such as *F. subglutinans*, *F. verticillioides*, and *F. oxysporum* (Mule *et al*., 2004a; Mule *et al*., 2004b; Leyva *et al*., 2015). In other pathogens such as *Magnaporthe oryzae* and...
Colletotrichum species CAL was effective for phylogenetic analysis (Couch and Kohn, 2002; Couch et al., 2005; Liu et al., 2016). Conversely, TEF1 marker was useful to characterise isolates belonging to F. vertcillioides and F. subglutinans (Geiser et al., 2004; Barik and Tayung, 2012; Nugroho et al., 2013), as well as other fungi such as Aspergillus niger and Penicillium (Peterson, 2004; Perrone et al., 2011). With F. proliferatum isolates in this study, genetics groups A1, A2 and A3 identified by the CAL marker were each specifically associated with a single host and geographic location; whereas, A4 included isolates from date palm from more than one country, which could be related to pathogen movement with agricultural trade.

The novel marker developed in this study FG1056 was the most effective in terms of phylogenetic information content with 169 SNPs and 89 ISs among the F. proliferatum isolates as well as in distinguishing the genetic groups. FG1056 distinguished 10 genetic groups, with various groups showing interesting links to the host and geographic origin of the F. proliferatum isolates. Identification of FG1056 as a novel marker for F. proliferatum from FUNYBASE highlights this database as a hugely useful resource for fungal phylogenetics. Similar approach using FUNYBASE proved to be effective for the identification of novel markers for Sardariomycetes (Walker et al., 2012) and Alternaria species (Armitage et al., 2015). RPB1 and RPB2 along with TEF1 are the commonly used phylogenetic markers in many Fusarium species as well as in many fungal pathogens (Hansen et al., 2005; Matheny et al., 2006; Tang et al., 2007; Schoch et al., 2009; Hansen et al., 2013; O'Donnell et al., 2015). The data generated by bioinformatic interrogation of the genome sequences of various Fusarium species revealed the highest level of resolution with FG1056 (Table 6.1). This strongly suggests the potential applicability of FG1056 in wider Fusarium species.

<table>
<thead>
<tr>
<th>Locus/Gen</th>
<th>Size of the locus</th>
<th>SNPs within F. proliferatum</th>
<th>Informative sites (%)</th>
<th>SNPs among representative Fusarium species*</th>
<th>Informative sites (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FG1056</td>
<td>3145</td>
<td>89</td>
<td>2.6</td>
<td>338</td>
<td>9.3</td>
</tr>
<tr>
<td>RPB1</td>
<td>6183</td>
<td>7</td>
<td>0.1</td>
<td>301</td>
<td>4.4</td>
</tr>
<tr>
<td>RPB2</td>
<td>4507</td>
<td>45</td>
<td>0.8</td>
<td>325</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Table 6.1 Comparative data reflecting phylogenetic information content of the novel marker FG1056 relative to known markers RPB1 and RPB2 in four F. proliferatum isolates and six representative Fusarium species.
On the other hand, phylogenies based on multilocus sequence typing (MLST) provided an opportunity to integrate the varying rates of evolution in the genome and achieve a good level of resolution in terms of the genetic groups identified within *F. proliferatum*, and also in their relationship to the host and geographic location. The MLST approach has been successfully applied to elucidate the genetic diversity and phylogenetic relationships in a number of major fungal pathogens including *Alternaria* (Armitage *et al.*, 2015), *Colletotrichum* (Baroncelli *et al.*, 2015) and *Fusarium* species (O'Donnell *et al.*, 2015). Furthermore, international fungal barcoding initiatives such as Assembling the Fungal Tree of Life (AFTOL) database and the quarantine pest identification database QBOL use multiple loci to barcode species (Celio *et al.*, 2006; Bonants *et al.*, 2010) based on highly variable genes (McDonald, 1997; Xu, 2006; Armitage *et al.*, 2015). However, the MLST approach relative to the use of single markers is more resource intensive in view of the time and the costs involved in generating data from four or five markers.

Among the *F. proliferatum* isolates characterised, the morphological characteristics did not show any specific relationships to their origin in terms of the host or geographic location. Furthermore, despite the limited number of reference isolates used, *F. proliferatum* showed overlapping characteristics to *F. verticillioides* and *F. oxysporum* as has been observed by other researchers (Summerell *et al.*, 2003; Mulè *et al.*, 2004a; Proctor *et al.*, 2009).

Compared to temperature, water activity showed a more pronounced effect on the growth of the *F. proliferatum* isolates. In terms of growth and sporulation differences among the *F. proliferatum* isolates, no direct relationship could be deciphered between the groups of isolates showing different growth patterns and the groups of isolates showing different sporulation patterns. However, the intraspecies variability is much higher in sporulation compared to growth. This is extremely interesting in relation to the biological and ecological competence as both growth and sporulation are important factors in functioning as pathogens, endophytes and saprophytes (Barik and Tayung, 2012). Specifically, sporulation is recognised as a critical process to disease initiation, spread and pathogen dispersal. Further pathology assays of these isolates with differing growth and sporulation potential are required to address this area.

In terms of the variability patterns identified in growth and sporulation among the *F. proliferatum* isolates, no direct relationship could be deciphered with the genetic groups identified with single markers or the MLST. A limited exception was the two isolates from maize in USA, which formed a single genetic group with the CAL marker.

Production of fumonisins among the *F. proliferatum* isolates varied significantly, with ITEM 2341 from date palm in Saudi Arabia ranked highest and R16 from onion in the UK the lowest. However, no direct relationship was observed between the fumonisin production levels and the growth and sporulation capabilities of these four isolates (Table 6.2). Interestingly, each of these four isolates belonged to different genetic groups identified by
the CAL marker. For example, high fumonisin producing isolates KF3377 and ITEM 2341 belonged to A3 and A4, respectively and the low fumonisin producing isolates MPVP 328 and R16 belonged to A1 and A5, respectively.

It is well documented that, knowledge of the underlying genetic mechanisms regulating the production of fumonisins under various environmental parameters including the host, temperature and water remains incomplete (e.g. Proctor et al., 2013). Additional genomic resources, and molecular tool kits are being developed in wider fungal systems including *Fusarium* species to tackle this complex process in view of the threat posed by mycotoxins such as fumonisins produced by *Fusarium* pathogens to animal and human health (e.g. Leyva-Madrigal et al., 2015).

This study has developed a range of genomic resources for four *F. proliferatum* isolates originating from four different hosts and geographic locations. This includes establishing fully assembled genomes, which have been annotated and deposited at NCBI, identification of the FUM gene clusters, identification of the genomic context and the genes flanking the clusters. Equally importantly, various biological attributes such as growth and sporulation with reference to temperature and water activity of these four isolates as well as their fumonisins production has been established (Table 6.2). At the genomic level, these four *F. proliferatum* isolates displayed differences in the genome size as well as the total predicted gene set including genes common to all isolates, unique to individual isolates as well as those shared by two or three isolates. However, these differences do not appear to be related to the variability in the growth, sporulation and fumonsin production of the four *F. proliferatum* isolates.
Table 6.2 Details of the genome sequenced *F. proliferatum* isolates along with their biological attributes and the NCBI accession numbers

<table>
<thead>
<tr>
<th>No.</th>
<th>Isolates</th>
<th>Host</th>
<th>Origin</th>
<th>Fumonisin production (FB1 and FB2)</th>
<th>Growth rate</th>
<th>Sporulation</th>
<th>NCBI Genome Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R16</td>
<td>Onion</td>
<td>UK</td>
<td>Low</td>
<td>Fast</td>
<td>Low</td>
<td>PKMG000000000</td>
</tr>
<tr>
<td>2</td>
<td>KF3377</td>
<td>Garlic</td>
<td>Poland</td>
<td>High</td>
<td>Fast</td>
<td>Low</td>
<td>PKMH000000000</td>
</tr>
<tr>
<td>3</td>
<td>ITEM 2341*</td>
<td>Date palm</td>
<td>Saudi Arabia</td>
<td>Very high</td>
<td>Fast</td>
<td>Low</td>
<td>PKMI000000000</td>
</tr>
<tr>
<td>4</td>
<td>MPVP 328</td>
<td>Maize</td>
<td>USA</td>
<td>Low</td>
<td>Slow</td>
<td>Moderate</td>
<td>PKMJ000000000</td>
</tr>
</tbody>
</table>

*Only *F. proliferatum* isolate that showed a positive response in growth to temperature increase with maximum growth rate at 30°C.

Similar differences at the genome level and expansion as well as contraction of different gene families associated with host range have been recently reported in different fungal systems including *Fusarium* (Niehaus *et al*., 2016), *Colletotrichum* (Baroncelli *et al*., 2016) and *Magnaporthe* (Dong *et al*., 2015) species. As referred to above, a range of biological attributes has also been established under uniform laboratory and assay conditions. This along with the genomic data for four *F. proliferatum* isolates developed in this study strongly contributes to the wealth of genomic resources that are being developed for *Fusarium* and offer the scope to gain novel insights into the biology, ecology and adaptive divergence of *F. proliferatum*. For example, in a recent study, the genome sequence of an *F. proliferatum* isolate occurring as an orchid endophyte in Russia has been reported but the FUM gene cluster has not been identified and its fumonisin production potential has not been analysed (Niehaus *et al*., 2016). The FUM gene cluster defined in the present study would enable delineation and further comparative analysis of the FUM gene cluster in the *F. proliferatum* isolate with a distinctive biology and ecology used in the study by Niehaus *et al.* (2016) as well as in other *Fusarium* species such as those belonging to the *F. fujikuroi* complex.

This study has also identified considerable nucleotide and amino acid substitutions among key FUM cluster genes FUM1, FUM8, FUM19, and FUM21 and the proteins encoded by them. It is noteworthy that at least some of these nucleotide/amino acid changes are shared by *F. proliferatum* isolates categorised as high fumonisin producers. For example, in the FUM1 gene, the KR domain showed a single amino acid change that distinguishes the two isolates producing high level of fumonisins from the other two isolates producing low levels (Figure 4.7). This provides an ideal basis for further computational analysis to investigate structure function relationships and identify putative functionally important
domains, regions or specific amico acid residues (e.g. Hvidsten et al., 2009). In parallel, the genomic resources and the characterised isolates developed in this study provide an effective platform for empirical analysis of putative functional domains or specific amino acid residues by molecular genetic analyses based on gene manipulation. For example, a DNA mutation encoding a specific amino acid change such as the one in the KR domain within the FUM1 gene can be carried out in vitro to generate mutant strains and to test their phenotype (e.g. reduced fumonisins production). Subsequent functional complementation analysis of the mutant by the introduction of the wild-type gene to test reversion of the phenotype (e.g. Glenn et al., 2008) would pave the way to unravel the complex mechanisms regulating the production of fumonisins in F. proliferatum and related Fusarium pathogens.

6.2 Conclusions and Future Perspectives

In conclusion, this study led to:

1. An improved understanding of the genetic diversity and phylogenetic relationships in the emerging pathogen Fusarium proliferatum utilising a set of isolates that represents the biogeographic diversity of this species.
2. The identification and development of a novel genetic marker FG1056 that is phylogenetically informative with a high SNP content and applicable for genotyping of diverse F. proliferatum isolates.
3. FG1056 shows potential for application in the phylogenetic analysis of wider Fusarium species addressing a need recognised by the research community in this area.
4. Multilocus genotyping identified genetic groups, clusters and haplotypes of F. proliferatum associated with different hosts and geographic locations.
5. The F. proliferatum genetic groups and clusters identified show different patterns of association to the host and geographic location. These patterns include i) isolates suggestive of adaptive divergence linked to a specific host and/or geographic location; and ii) isolates that are less specific and associated with different hosts and geographic locations, potentially linked to pathogen movement through agricultural and related trades.
6. Clear evidence of the variability among F. proliferatum isolates in biological attributes such as growth rate, sporulation and fumonisins production with
reference to temperature, water activity and duration. Most notably, fumonisins production varied up to 140-fold among the four *F. proliferatum* isolates under identical conditions.

7. Assembled genome sequence data for four *F. proliferatum* isolates representing different host and geographic locations, along with their identified biological characteristics including fumonisin production levels.

8. Annotated genome sequence has been deposited in NCBI with accession numbers referred to in Table 6.2, but the data is currently on-hold from public access.

9. Total predicted gene sets for the four *F. proliferatum* isolates with 12,980 identified as common to all isolates and up to 134 genes unique to an isolate.

10. First report of the identification of the FUM gene cluster in four *F. proliferatum* isolates consisting of 16 genes, their order and orientation as well as putative function.

11. Identification of the genomic location and context of the FUM gene cluster in the four *F. proliferatum* isolates, and designation of the genes flanking the FUM cluster as well as annotation of their putative function.

Overall, this study has developed a clear framework and a significant range of resources for further characterisation of *F. proliferatum* populations associated with diverse crops and prevalent in different geographic locations. Specifically, the knowledge and resources developed in this study provide scope for further research integrating bioinformatics, comparative genomics, and experimental approaches to unravel the genetic basis of the biological variability now documented in this species including the production of hugely variable level of fumonisins. We are in the process of developing the publication(s) from this work following which the genomic sequences will be released by the NCBI open to use by the *Fusarium* research community. Thus, current results and further research advances would contribute to more efficient management of the threat posed to the health of plants, animals and humans by the emerging pathogen *F. proliferatum* with strong mycotoxicogenic potential.


Buscot, F., Munch, J. C. Charcosset, J. Y., Gardes, M., Nehls, U., Hampp, R., 2000. Recent advances in exploring physiology and biodiversity of ectomycorrhizas highlight the functioning of these symbiosis in ecosystems. FEMS microbiology reviews 24, 601-614.


Górna, K., Pawłowicz, I., Waśkiewicz, A., Stepień, Ł., 2016. Fusarium proliferatum strains change fumonisin biosynthesis and accumulation when exposed to host plant extracts. Fungal Biol. 120, 884-93.


167


Peterson, S.W., 2004. Multilocus DNA sequence analysis shows that *Penicillium biourgeianum* is a distinct species closely related to *P. brevicompactum* and *P. olsonii*. Mycol. Res. 108, 434-440.


181


Sobek, E., Munkvold, G., 1999. European corn borer (Lepidoptera: Pyralidae) larvae as vectors of *Fusarium moniliforme*, causing kernel rot and symptomless infection of maize kernels. J. Econ. Entomol. 92, 503-509.


head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. Fungal Genet. Biol. 45, 473–484.


## 8 Appendix

| ITEM, 2336. (D-Sa) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2342. (D-Sa) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| MPVP, 328. (Mz-Us) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 1744. (Mz-It) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| NL70-7. (On-Uk) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| A28. (On-Uk) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 4297. (D-Sp) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2341. (D-Sa) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2337. (D-Sa) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 3268. (Tm-It) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 1952. (On-Uk) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 1453. (As-It) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 4358. (Gr-Yg) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| A8. (On-Uk) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 1950. (On-Uk) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 1477. (As-It) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2979. (Mg-Ml) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| KF3377. (Gr-Po) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 7595. (Mz-Us) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| KF3341. (As-Po) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| R16. (On-Uk) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2343. (D-Sa) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2339. (D-Sa) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 3271. (M-It) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 3274. (M-It) | GGAAGTATGATGCTGACCCCCATGTTGTCTTCGAGCACCCCAAAGTGGAC |
| ITEM, 2336. (D-Sa) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 2342. (D-Sa) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| MPVP, 328. (Mz-Us) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 1744. (Mz-It) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| NL70-7. (On-Uk) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| A28. (On-Uk) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 4297. (D-Sp) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 2341. (D-Sa) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 2337. (D-Sa) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 3268. (Tm-It) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 1952. (On-Uk) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 1453. (As-It) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 4358. (Gr-Yg) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| A8. (On-Uk) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 1950. (On-Uk) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 1477. (As-It) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 2979. (Mg-Ml) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| KF3377. (Gr-Po) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| ITEM, 7595. (Mz-Us) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |
| KF3341. (As-Po) | CACGTCAGCTCCTTCTTTTGTTTACGTGGAGCAACTCTGACTTTATGTCTCGGAGAATTACG |

192
**ITEM.4297.** (D-A28) (On Mz)

**ITEM.1744.** (Mz)

**ITEM.2342.** (D)

**ITEM.2336.** (D)

**ITEM.3274.** (Tm)

**ITEM.3273.** (Tm)

**ITEM.3275.** (Tm)

**ITEM.2339.** (D)

**ITEM.2343.** (D)

**R16.** (On Mz)

**A8.** (On Mz)

**NL70-7. (On-Mz)

**ITEM.4297.** (D-A28) (On Mz)

**ITEM.1453.** (As)

**ITEM.1952.** (On Mz)

**ITEM.3268.** (Tm)

**ITEM.2341.** (D)

**ITEM.4297.** (D-A28) (On Mz)

**ITEM.7595.** (Mz)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3377.** (Gr-Po)

**ITEM.7595.** (Mz-Us)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3341.** (As-Po)

**R16.** (On-Us)

**ITEM.4297.** (D-Sp)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3377.** (Gr-Po)

**ITEM.7595.** (Mz-Us)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3341.** (As-Po)

**R16.** (On-Us)

**ITEM.4297.** (D-Sa)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3341.** (As-Po)

**R16.** (On-Us)

**ITEM.4297.** (D-Sa)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3341.** (As-Po)

**R16.** (On-Us)

**ITEM.4297.** (D-Sa)

**ITEM.1477.** (As-It)

**ITEM.2979.** (Mg-ML)

**RF3341.** (As-Po)

**R16.** (On-Us)

**ITEM.2340.** (D)

**ITEM.1744.** (Mz)

**ITEM.2343.** (D)

**R16.** (On-Mz)

**A8.** (On-Mz)

**ITEM.4297.** (D-Sp)

**ITEM.1744.** (Mz)

**ITEM.2343.** (D)

**R16.** (On-Mz)

**A8.** (On-Mz)

**ITEM.4297.** (D-Sp)

**ITEM.1744.** (Mz)

**ITEM.2343.** (D)

**R16.** (On-Mz)

**A8.** (On-Mz)

**ITEM.4297.** (D-Sp)
Figure 8.1 Multiple sequence alignment for the new genetic marker developed in this study FG1056 data from 26 haplotypes identified among *F. proliferatum* (23), *F. verticillioides* (1) and *F. oxysporum* (2)