

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Evidence that *SSR1* can act as a Hypermutable
Contingency Gene in *Candida albicans*

A thesis presented in partial fulfillment of the requirement for
the degree of

Master of Science

in

Microbiology

at Massey University, Palmerston North,
New Zealand

Zhuo Zhou

2010

Abstract

During adaptation to the host environment, many microorganisms undergo rapid variation in cell surface phenotype through genetic alteration in hypermutable contingency genes. One of the main mechanisms underlying these changes is alteration in the number of DNA repeat units that results in a large and flexible repertoire of similar but non-identical surface proteins. *SSR1*, a gene in the opportunistic pathogen *Candida albicans*, encodes a repeat-containing cell wall protein which may play a role in maintaining cell wall strength. This gene contains 2 regions with multiple 6 bp tandem repeat units, encoding the amino acids serine and alanine, separated by a 200 bp non-repetitive DNA region. This study investigated whether *SSR1* was a hypermutable contingency gene. Among a worldwide collection of 96 infection-causing *C. albicans* strains, 24 alleles and 40 allele combinations were identified by fluorescent-based genotyping of *SSR1* PCR products. Sequencing results confirmed that the differences in allele size were caused by variation in number of tandem repeats. Two very similar allele combinations were overrepresented (30% and 28%) among a cluster of general-purpose genotype (GPG) strains (which is the most widespread cluster) compared with non-GPG strains (Fisher's exact test, $P=0.0001$ and $P<0.0001$). Among a worldwide collection of 36 commensal GPG *C. albicans* strains, 8 allele combinations were identified by genotyping. One of the two allele combinations that were overrepresented in GPG infection-causing strains was found significantly less in GPG commensal strains (Fisher's exact test, $P=0.0004$). After culture of *C. albicans* cells in vitro for 300 generations, mutation of repeats in *SSR1* occurred, giving a high mutation rate of 1.11×10^{-4} per cell division. The results indicate that *SSR1* is a hypermutable gene and that it shows clade-specificity with the GPG cluster. Growth in a rat model did not seem to cause variation in *SSR1* and human host body sites did not seem to be associated with specific *SSR1* alleles, suggesting that *SSR1* is not used for short-term adaptation in these environments. However, the different allele distribution in commensal and infection-causing GPG strains suggest that *SSR1* may have a role in short-term adaptation in GPG strains, contributing to the change between commensalism and infection. In this case, *SSR1* may act as a hypermutable contingency gene.

Acknowledgement

I would like to convey my big thanks to my supervisors Jan Schmid and co-supervisor Zoe Jordens who have supported me the whole way through, without whose guidance and encouragement this would never have happened. You are the best supervisors I could ever have, Thank you!

I would like to thank lab members. Thanks to Ningxin Zhang for her help in the lab. Huge thanks to Shuguang Zhang for his help, without whose assistance and patience to all my questions, I would have struggled to achieve results at all. Thanks to Synodalia Wattimena, Yeying Wang and Panimalar Vijayan for their help, cooperation and all the fun times.

I would also like to thank the Allan Wilson Centre for their great service and assistant.

Thanks to Jing Wang, Melissa Guo and Charlene Zhang for their mental support. Thanks to my home stay family Anne and Satyendra Ram for providing constant encouragement and caring.

Finally, I need to thank my parents who have given the best of themselves and supported me all these years. To them I dedicate this thesis.

Table of Contents

ABSTRACT	I
ACKNOWLEDGEMENT.....	II
TABLE OF CONTENTS	III
LIST OF TABLES.....	VII
LIST OF FIGURES	VIII
LIST OF ABBREVIATIONS.....	X
CHAPTER ONE – INTRODUCTION	1
1.1. Overview of the importance of infections caused by <i>Candida albicans</i>	1
1.2. Phylogeny and population structure of <i>Candida albicans</i>	3
1.3. DNA repeats as mutation units in hypermutable contingency genes	7
1.3.1. Micro-environment evolution and adaptation of microorganisms	7
1.3.2. Genes containing tandem repeats (TRs).....	8
1.3.3. Genes containing tandem repeats (TRs) in <i>Candida albicans</i>	10
1. 4. Cell wall proteins of <i>Candida albicans</i>	12
1. 4. 1. Molecular organization of the cell wall of <i>Candida albicans</i>	12
1. 4.2. Roles of cell wall proteins of <i>Candida albicans</i>	13
1. 5. <i>SSR1</i> gene and Ssr1 protein.....	14
1.5.1. Genomic study of <i>SSR1</i> gene	14
1.5.2. Ssr1p is a repeat-containing cell wall protein of <i>Candida albicans</i> with a role in wall stability.....	15
1. 6. Hypothesis and aim	17
CHAPTER TWO – MATERIALS AND METHODS	18
2.1. Location of tandem repeats	18
2.2. Biological materials	19
2.2.1 <i>Candida albicans</i> strains	19

2.2.2. Other biological materials.....	24
2.3. Media, buffers and solutions.....	24
2.3.1. Media.....	24
2.3.2. Buffers and solutions.....	24
2.4. Growth and maintenance of cultures	25
2.4.1. Growth conditions of <i>Candida albicans</i> cells on agar plates.....	25
2.4.2. Preparation of <i>C. albicans</i> stocks on YPD slants.....	26
2.4.3. Glycerol stock for <i>Candida albicans</i>	26
2.4.4. Growth conditions of <i>Escherichia coli</i> on LB agar plates containing ampicillin (for blue-white selection).....	26
2.5. Preparation of template for PCR	27
2.5.1. Colony PCR.....	27
2.5.2. Extraction of DNA (boiled supernatants).....	27
2.6. PCR.....	27
2.6.1. PCR reagents.....	27
2.6.2. Primers used in experiment.....	28
2.6.3. PCR reaction conditions.....	30
2.6.3.1. Colony PCR of <i>C. albicans</i> : amplification of <i>SSR1</i> with primers <i>SSR1-6</i> and <i>SSR1-7</i>	30
2.6.3.2. Colony PCR of <i>E. coli</i> : amplification of <i>SSR1</i> inserts with primers <i>M13F</i> and <i>M13R</i>	30
2.6.3.3. Colony PCR of <i>C. albicans</i> : amplification of repeat regions 1 and 2 with primers <i>F1</i> and <i>B2</i>	30
2.6.3.4. PCR of extracted DNA (bsn): amplification of repeat region 1 with primers <i>F1</i> and <i>SSR1-1</i> ; amplification of repeat region 2 with primers <i>SSR1-2</i> & <i>SSR1-3</i>	31
2.6.3.5. PCR of extracted DNA (bsn): amplification of repeat region 1 and 2 with primers <i>F1</i> and <i>SSR1-3</i>	31
2.7. Endonuclease digestion of PCR product.....	31
2.8. Generation of competent <i>Escherichia coli</i> cells.....	32
2.9. Gel electrophoresis	33
2.9.1. Agarose gel electrophoresis.....	33
2.9.1.1. Gel preparation.....	33
2.9.1.2. Gel loading and running.....	33
2.9.1.3. Gel staining and illumination.....	33
2.9.2. Polyacrylamide gel electrophoresis (PAGE).....	34
2.9.2.1. Gel preparation.....	34
2.9.2.2. Gel loading and running.....	35
2.9.2.3. Gel staining and illumination.....	35

2.10. Genotyping	35
2.11. Cloning of <i>SSR1</i> for DNA sequencing	38
2.11.1. PCR product purification.....	38
2.11.2. Ligation	38
2.11.2.1. Determination of DNA concentration	38
2.11.2.2. TA-Cloning	39
2.11.3. Transformation.....	40
2.11.4. Blue-white selection	40
2.11.5. DNA Sequencing	41
2.12. Subculturing of <i>Candida albicans</i> for 300 generations	42
2.13. Subculturing of <i>Candida albicans</i> rat samples.....	43
CHAPTER THREE – RESULTS.....	44
3.1. Choice of the <i>SSR1</i> gene for this investigation.....	44
3.2. <i>SSR1</i> has variable repeats that may generate proteins which differ in the number and arrangement of amino acid repeats	44
3.3. Detection of variation in <i>SSR1</i> repeat regions strains by DNA sequencing.....	50
3.4. Detection of <i>SSR1</i> allele variability among <i>C. albicans</i> strains by gel electrophoresis and genotyping.....	55
3.4.1. Identification of PCR product length by agarose gel electrophoresis	55
3.4.2. Identification of PCR product length by polyacrylamide gel electrophoresis	57
3.4.3. Identification of different alleles and allele combinations in <i>C.albicans</i> strains by genotyping	60
3.4.3.1. Identification of length of <i>SSR1</i> repeat regions in <i>C.albicans</i> strains by genotyping	60
3.4.3.2. Identification of allele combinations	64
3.4.3.3. Variation of <i>SSR1</i> repeat regions among infection-causing strains.....	66
3.4.3.4. Different alleles and allele combinations in commensal <i>C. albicans</i> strains	70
3.4.3.5. Site of isolation and alleles and allele combinations.....	73
3.5. Generation of new alleles in laboratory cultures.....	79
3.5.1. Generation of new alleles from the longest <i>SSR1</i> allele using laboratory cultures ...	79
3.5.2. Estimated rates of mutations in other alleles	81
3.6. Rapid mutation is not detected in repeat regions of <i>SSR1</i> in rat models.....	83
3.7. Synonymous point mutations reduce the predicted mutability of repeat regions	87
3.8. Identification of a potential functional domain in Ssr1p using bioinformatics and literature search	89

3.8.1. Ssr1p contains a CFEM domain	89
3.8.2. What is CFEM domain and what does it do?	89
CHAPTER FOUR - CONCLUSIONS, DISCUSSIONS AND FUTURE WORK.....	93
4.1. Key findings of this project.....	93
4.1.1. <i>SSR1</i> is a hypermutable gene	93
4.1.2. <i>SSR1</i> alleles show clade specificity.....	93
4.1.3. <i>SSR1</i> and its role in short-term adaptation	96
4.1.3.1. Massive changes in <i>SSR1</i> allele distribution are not associated with different hosts or body sites.	96
4.1.3.2. A change in alleles may be associated with the switch from commensal to pathogen in GPG strains	96
4.2. A possible history of <i>Candida albicans SSR1</i>	98
4.4. Future Work.....	100
APPENDIX	101
Appendix I: Supplies of Materials and Equipment.....	101
Appendix II: Primers used in this project (M13 primers supplied by Gibco BRL and the others supplied by invitrogen).....	102
Appendix III: Amino acid sequence of <i>SSR1</i> of SC5314 showing CFEM domain	103
Appendix IV: Genotyping results	104
A. GPG infection-causing strains	104
B. Non-GPG infection-causing strains.....	107
C. GPG commensal strains.....	110
REFERENCES.....	112

List of Tables

Table 2.1.	<i>Candida albicans</i> strains used in this study	19
Table 2.2.	Details of reagents used in PCR reactions	28
Table 2.3.	Reagents for restriction of F1/B2 PCR products with <i>Hind</i> III	32
Table 2.4.	Components used for each 8% polyacrylamide gel	34
Table 2.5.	Reagent volumes for ligation reactions	40
Table 2.6.	Reagent volumes for DNA sequencing	41
Table 2.7.	Number of <i>C. albicans</i> cells from rat host	43
Table 3.1.	Short tandem repeats in <i>SSR1</i> determined by Braun et al.	44
Table 3.2.	Short tandem repeats in <i>SSR1</i> determined with SERV software program	45
Table 3.3.	The characteristics of strains used to check size discrimination on PAGE	58
Table 3.4.	Numbers of repeat units and corresponding fragment sizes	61
Table 3.5.	Repeat-containing fragments' sizes in genotyping and in the sequencing results.	62
Table 3.6.	VARscores of 4 repeat regions in RIHO30 <i>SSR1</i> alleles	81
Table 3.7.	Comparison of parent strains and cultured strains in rat hosts	87
Table 3.8.	Comparison of true VARscores with maximum VARscores for repeat regions	88
Table 4.1.	Repeat-containing genes and their clade-specificity	94

List of Figures

Figure 1.1.	Budding <i>Candida albicans</i> cells	1
Figure 1.2.	An un-rooted phylogenetic tree of 266 <i>C. albicans</i> isolates	6
Figure 1.3.	Four main domains in <i>SSR1</i>	15
Figure 2.1.	The relative annealing positions of the primers in the <i>SSR1</i> gene of strain SC5314	29
Figure 2.2.	The relative positions of primers used for genotyping and their PCR products	37
Figure 2.3.	pLUG®-Multi TA-Cloning Vector and vector-borne primers M13F and M13R	39
Figure 3.1.	Gene structure and repeat regions identified in <i>SSR1</i>	47
Figure 3.2.	Sequence of <i>SSR1</i> gene and translated sequence showing repeat units in strain SC5314	49
Figure 3.3.	Example of <i>SSR1</i> heterozygosity in strains	51
Figure 3.4.	Examples of heterozygous alleles separated by TA cloning	52
Figure 3.5.	Fourteen sequenced <i>SSR1</i> alleles from this investigation compared with strain SC5314	53
Figure 3.6.	PCR products of <i>SSR1</i> with primers F1/B2 on 1.5% agarose gel	56
Figure 3.7.	PCR products of <i>SSR1</i> digested with <i>HindIII</i> and separated in a 2% agarose gel	57
Figure 3.8.	PCR products of <i>SSR1</i> with primers F1/ <i>SSR1</i> -1 and <i>SSR1</i> -2/ <i>SSR1</i> -3 separated in a 8% polyacrylamide gel	59
Figure 3.9.	The repeat-containing PCR products from SC5314 and HUN64 separated in 8% PAGE	60
Figure 3.10.	Matching of 36 fragments' sequencing results and their genotyping readings	63
Figure 3.11.	The allele/fragment combinations and the relationship with the colour peaks from genotyping a heterozygous strain	62

Figure 3.12.	Proportions of repeat numbers in repeat region 1 and repeat region 2	65
Figure 3.13.	Frequencies of <i>SSR1</i> alleles in infection-causing strains	67
Figure 3.14.	Frequencies of <i>SSR1</i> allele combinations in infection-causing strains	66
Figure 3.15.	Allele combinations in commensal and infection-causing GPG strains	72
Figure 3.16.	Site of isolation of infection-causing <i>SSR1</i> alleles	74
Figure 3.17.	Site of isolation of <i>SSR1</i> alleles	77
Figure 3.18.	Genotyping of repeat region 2 in mutant alleles generated after 300 generations	80
Figure 3.19.	Correlation between VARscore and mutation rates	82
Figure 3.20.	PCR analysis of colony mixtures from a culture obtained from a rat model	85
Figure 3.21.	PCR analysis of 7 of the 60 colonies from the culture from a rat sample	86
Figure 3.22.	A protein family that has significant homology with <i>Ssr1p</i>	90
Figure 3.23.	The CFEM domain and the two repeat regions in the <i>SSR1</i> gene	91

List of Abbreviations

ALS	agglutinin-like sequence
aa	amino acid
APS	ammonium persulfate
AWCGS	Alan Wilson Centre Genome services
bsn	boiled supernatant
<i>C. albicans</i>	<i>Candida albicans</i>
CFEM	common in several fungal extracellular membrane
CWPs	cell wall proteins
Da	Dalton
DTT	DL-Dithiothreitol
E.coli	<i>Escherichia coli</i>
EDTA	Ethylene Diamine Tetraacetic Acid
GPG	general purpose group
GPI-CWPs	glycosylphosphatidylinositol-linked cell wall proteins
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB Medium	Luria-Bertani medium
Non-GPG(ngpg)	non-general purpose group
ORF	open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
RRs	Repeat regions
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	sodium dodecyl sulfate
SERV	Sequence-based Estimation of Repeat Variability
SSR	short sequence repeats
Ssr1p	Ssr1 protein
STRs	short tandem repeats
TBE	Tris/Borate/EDTA
TEMED	Tetramethylethylenediamine
TRs	tandem repeats
X-gal (BCIG)	bromo-chloro-indolyl-galactopyranoside
YPD Medium	Yeast Extract Peptone Dextrose Medium