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 support 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

List of Tables

List of Figures

List of Abbreviations

ALS	agglutinin-like sequence
aa	amino acid
APS	ammonium persulfate
AWCGS	Alan Wilson Centre Genome services
bsn	boiled supernatant
C. albicans	*Candida albicans*
CFEM	common in several fungal extracellular membrane
CWPs	cell wall proteins
Da	Dalton
DTT	DL-Dithiothreitol
E.coli	*Escherichia coli*
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
S. cerevisiae	*Saccharomyces cerevisiae*
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