Effects of aerobic and anaerobic environments on bacterial mutation rates and mutation spectra assessed by whole genome analyses

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Genetics at Massey University, Palmerston North, New Zealand.

Sonal Shewaramani

2015

Abstract

For organisms that are exposed to different environments, the rates and types of spontaneous mutations that arise in each environment can vary, and potentially impact the direction of evolution as a whole. Oxidative stress is a major cause of mutation, but the effect of oxygen availability on the mutation rates and spectra of organisms grown in aerobic as compared to anaerobic environments is not well understood at the whole genome level. To investigate the mutation rates and spectra of a facultative anaerobic bacterium grown under strictly aerobic or anaerobic conditions, 24 mutation accumulation lineages, derived from *Escherichia coli* REL4536, were established and propagated through 180 and 144 single-colony population bottlenecks, respectively. Spontaneous mutation rates of 2.50×10^{-10} and 4.14×10^{-10} mutations per nucleotide per generation were obtained for aerobically and anaerobically grown cells, respectively. Mutations in the aerobic environment were significantly biased towards G → T mutations and IS*186* transposition, while C → A, T → G, A → C mutations, gross chromosomal rearrangements (GCRs) and IS*150* transposition were significantly more prevalent under anaerobic conditions. Transcriptional profiling, *via* RNAseq, of REL4536 grown under aerobic and anaerobic environments revealed that repair genes, especially those involved in the repair of GCRs, were generally up-regulated in the anaerobic environment, consistent with findings that mutation rates, especially for GCRs, are higher in the anaerobic environment.

GCRs have long been thought to play an important role in the evolutionary process, though their contributions to the process have not been specifically defined. SbcCD, an exonuclease, is involved in the repair of DNA secondary structures, and is thought to help prevent the occurrence of GCRs. Transcriptome analyses showed that in *E. coli*, *sbcC* was up-regulated during growth in an anaerobic environment, as compared to an aerobic environment. To investigate the impact of GCRs on adaptive evolution, an *E. coli* REL4536 strain with disrupted *sbcC* was constructed and evolved under anaerobic conditions for 1,000 generations in glucose-limited media in 14 parallel populations. Mutations that arose during evolution were determined by whole genome re-sequencing of selected clones, and evolved *sbcC* mutant strains displayed more GCRs and enhanced population-level fitness on average. Together, these results suggest that GCRs may play an important role in the rate of adaptation.

Acknowledgements

To begin with, I would like to offer my sincere appreciation and gratitude to my supervisors: Dr Christina Moon, Dr Sinead Leahy, Dr Rees Kassen and Prof Paul Rainey. Thank you for the support, guidance and knowledge you have given during the course of this project. In particular, I would like to thank Christina and Sinead, for their kindness, patience and encouragement over the years. It has truly meant a lot to me and I will always be grateful to both of you for being such great mentors.

I would also like to extend my gratitude towards the Marsden Fund for funding this project, towards Massey University for financial support, and finally, to AgResearch, for providing financial support and excellent research facilities.

This PhD would not have been possible without the people that were around me, with whom I have had the great pleasure of working with. I would like to acknowledge Zaneta Park and Benoit Auvray for their work on the MA simulator programme. Thank you to the Rumen Microbiology team for all the assistance and support over the last few years; it has been a great environment to work in. I would also like to express my thanks to Dr Bill Kelly, Dr Peter Janssen, Dr Ron Ronimus and Dr Dragana Gagic for always being so willing to share their expertise and wisdom. My thanks to Carrie Sang, Dong Li and Rechelle Perry, for all the reagents and technical assistance you have provided over the years. Thank you also to Paul Lovejoy, Dr Eric Altermann, Simon Guest, Paul Maclean and AgResearch Information Services for your invaluable help along the way.

To the members of the student room, both past and present, thank you so much for your friendship. I would especially like to thank Denise Martin, Eva Maier, Filomena Ng, Dr Milica Ciric, Dr Sam Noel and Renee Atua; this experience truly would not have been the same without any of you. Thank you so much for all the laughs, fun, advice and encouragement over the years. To Thomas Finn, thank you so much for being such a great colleague and friend, I truly cannot imagine anyone else who would have made this project as enjoyable as it has been.

Finally, to my family, thank you for always being there for me with all your unwavering love and support, even when you had no clue what my research was about. To my parents especially: I am well aware of all the sacrifices you have made for me over the years, and I thank you for always encouraging my curiosity. I hope you are proud.

Table of Contents

List of tables

List of figures

Non-standard abbreviations

λ-Red	Lambda Red
A	Adenine (nucleotide base)
AE	Aerobic
AHT	Anhydrotetracycline
Amp^R	Ampicillin resistant strain
AN	Anaerobic
Anc	Ancestor
Ara^-	Strain that cannot utilise arabinose
Ara^+	Strain that can utilise arabinose
ATP	Adenosine triphosphate
BER	Base excision repair
BLAST	Basic Local Alignment Sequence Tools
bp	Base pairs
BPS	Base pair substitution
C	Cytosine (nucleotide base)
cDNA	Complementary DNA
CFU	Colony forming units
CL	Confidence limits
Cm^R	Chloramphenicol resistant strain
C_Q	Quantification cycle
DEPC H_2O	Diethylpyrocarbonate treated water
dH_2O	Distilled water
DM	Davis minimal
DNA	Deoxyribonucleic acid
DSB	Double-strand break
dsDNA	Double-stranded DNA
E	Amplification efficiency for qPCR reactions
G	Guanine (nucleotide base)
GCR	Gross chromosomal rearrangement
IR	Inverted repeat
IS	Insertion elements
kb	Kilobase
LB	Luria-Bertani
LTEE	Long-term experimental evolution
MA	Mutation accumulation
Mb	Megabase
MB	Mega bytes

MD	Macrodomain
MGE	Mobile genetic element
MMR	Mismtach repair
mRNA	Messenger RNA
MSS-MLE	Ma-Sandri-Sarkar Maximum Likelihood Estimator method
Nal^R	Nalidixic acid resistant strain
ncRNA	Non-coding RNA
NER	Nucleotide excision repair
NS	Non-structured regions
OD	Optical density
oriC	Origin of replication
PCA	Principal component analysis
PCR	Polymerase chain reaction
pol	Polymerase
ppGpp	Guanosine tetra-phosphates
pppGpp	Guanosine penta-phosphates
qPCR	Quantitative polymerase chain reaction
R^2	Correlation coefficient
Rif^r	Rifampicin resistant reference strain
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPKM	Reads per kilobase of gene per million reads mapped
RPM	Rotations per minute
rRNA	Ribosomal RNA
RT	Room temperature
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SCM	Small colony morphotype
SEM	Standard error of the mean
SSB	Single-strand break
ssDNA	Single-stranded DNA
T	Thymine (nucleotide base)
TCM	Typical colony morphotype
TE	Transposable element
tRNA	Transfer RNA
Ts/Tv	Transition to transversion ratio
v/v	Volume per volume
v/v/v	Volume per volume per volume
w/v	Weight per volume