The evolution of selfish genetic elements within bacterial genomes

A thesis submitted in partial fulfilment of the requirements for the degree

of

Ph.D.

in

Molecular Evolution

at Massey University, Auckland, New Zealand.

Frederic Bertels

2012

Abstract

Genes that increase their copy number relative to that of the host genome are termed selfish. Selfish genes are found ubiquitously in bacterial genomes. Within genomes they can often be identified due to their repetitive nature. Short repetitive sequences such as repetitive extragenic palindromic (REP) sequences have been proposed to be selfish genetic elements. However, evidence for the selfishness of REPs is scarce due to the lack of knowledge about their origin, evolution and mechanisms of dispersal. Here, REPs are studied in the model bacterium *Pseudomonas fluorescens* SBW25. The evidence provided suggests that REPs are part of a greater mobile genetic element, which is termed REP doublet forming hairpins (REPINs).

Subsequently, I investigate the cause of REPIN dispersal: a putative transposase. The transposase, named REP-associated tyrosine transposase (RAYT) shares essential motifs with the IS*200* family of insertion sequences. However, unlike insertion sequences, RAYTs are found only as single copy genes. This indicates that RAYTs may not be entirely selfish; instead they may have been co-opted by the host to perform a beneficial function.

Finally, two more repetitive sequence classes are studied in the SBW25 genome. Interestingly, both sequence classes consist of a protein coding sequence and a sequence that forms a stable secondary structure in single stranded DNA or RNA. This arrangement is reminiscent of bacterial toxin-antitoxin (TA) systems. Evidence from sequence analyses suggests that the repetitive nature of these elements in SBW25 may be the result of cooperation between REPINs or other replicative elements and the TA systems.

The presented analyses show that despite the streamlined nature of bacterial genomes selfish genetic elements frequently arise, replicate and probably increase their

persistence and spread through cooperation with addictive and duplicative elements respectively.

Acknowledgements

Foremost, I would like to thank my supervisor Professor Paul Rainey for all the advice, guidance and inspiration he has given me over the past three years. Without him the research presented in this thesis would not have been possible, and I hope we keep in touch for years to come. Further, I would like to thank my co-supervisors Dr Justin O'Sullivan and Professor Allen Rodrigo for not only supporting me during the course of my PhD, but also for their help during my first year in New Zealand. Their fascination for research and science was one of the main reasons why I decided to stay in New Zealand for my PhD.

Another essential factor for the completion of a PhD is funding. Hence, I am very grateful for a doctoral scholarship from the Allan Wilson Centre. It was great being part of the Allan Wilson Centre. I especially enjoyed the scientific exchange and social activities at the annual meetings.

The scientific and social interactions with past and present members of the Rainey Lab also played an important role in my research. I had a great time and I hope to meet all of you again! I would like to especially thank Dr Jenna Gallie for endless advice and discussion. Jenna also proofread countless manuscripts and thesis attempts and taught me how to improve my English on many occasions. Dr Xue-Xian Zhang and Yunhao Liu were of great help during my attempts to conduct lab work.

Furthermore I would like to thank Ben Kerr and the rest of the Kerr lab for stimulating discussions and advice during my stay in Seattle, WA. The time in Seattle and the Kerr lab greatly supported my professional development by exposing me to new ideas and different types of thinking.

Finally, I would like to thank my parents Elke and Ralf Bertels, my sister Helen Bertels, my brothers Felix and Julian Bertels and again my fiancée Jenna Gallie for providing

the support I needed to finish this thesis. I also would like to thank Elaine Riley for helping me move around the world, and numerous other things that made my life much easier.

Table of Contents

Table of Abbreviations

Abbreviation	Meaning
BIMEs	Bacterial interspersed mosaic elements
BLAST	Basic local alignment search tool
BLASTP	BLAST Protein (protein query against protein database)
BLASTN	BLASTN Nucleotide (nucleotide query against nucleotide database)
TBLASTN	Translated BLASTN (protein query against nucleotide database)
bp	base pairs
CAS genes	CRISPR associated genes
CRISPRs	Clustered regularly interspaces short palindromic repeats
ERICs	Enterobacterial repetitive intergenic consensus sequences
IR	Inverted repeat
IS	Insertion sequence
LARDs	Large retrotransposon derivatives
LINEs	Long interspersed elements
LTR	Long terminal repeat
MITE	Miniature inverted repeat transposable elements
NEMISs	*Neisseria* miniature insertion sequences
NGS	Next-Generation Sequencing
PSK	Post segregational killing
PU	Palindromic units
ORF	Open reading frame
RAYTs	REP-associated tyrosine transposase
REPs	Repetitive extragenic palindromic sequences
REPINs	REP doublets forming hairpins
RMS	Restriction modification system
RNAi	RNA interference
RUP	Repeat unit of pneumococcus
SDR	Small dispersed repeats
SINEs	Short interspersed element
ssDNA	Single stranded DNA
TA	Toxin-antitoxin system
TPRT	Target primed reverse transcription
TRIMs	Terminal-repeat retrotransposons in miniature