



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Using whole genome comparison to detect sequence similarities between plants and microbes

By

BAREND JUAN VORSTER

submitted in partial fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

in the

Faculty of Natural and Agricultural Sciences
Department of Plant Science
University of Pretoria
Pretoria

SUPERVISORS

PROF. KJ. KUNERT

PROF. CA. CULLIS

December 2007

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.



DECLARATION

I declare that this thesis, which I hereby submit for the degree, Philosophiae Doctor (Plant Science-Plant Biotechnology) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at any other University

A handwritten signature in black ink, appearing to read 'A. van der Merwe'.

Signed _____

15 December 2007

Date _____



TABLE OF CONTENTS

	PAGES
Summary	I
Thesis composition	II
Acknowledgments	II
Index of figures	III
Index of tables	X
1 Introduction and literature review	1
1.1 The eukaryotic genome	2
1.2 Origin of the chloroplast and mitochondrion	6
1.3 DNA transfer	8
1.3.1 DNA transfer between the organelles and nucleus	8
1.3.2 DNA transfer between viruses and plants	9
I. The banana streak virus (BSV)	9
II. Tobacco vein clearing virus (TVCV)	10
III. Petunia vein clearing virus (PVCV)	11
IV. Rice tungro bacilliform virus (RTBV)	11
1.3.3 DNA transfer between bacteria and plants	11
1.3.4 DNA transfer between fungi and plants	13
1.3.5 DNA transfer between plants and plant mitochondria	14
1.5 Models for gene transfer and integration	17
2 Materials and Methods	20
2.1 Blast Analysis	21
2.1.1 Chloroplast and Mitochondrial Comparison	23
2.1.2 Viral Comparison	23
2.1.3 Bacterial Comparison	24
2.1.4 Fungal Comparison	24



3	Results	26
3.1	Chloroplast and Mitochondrial comparison	27
3.1.1	Chloroplast homologies in the rice nuclear genome	27
3.1.2	Representation of the chloroplast genome in the rice nuclear genome	28
3.1.3	Chloroplast rRNA genes in the rice nuclear genome	28
3.1.4	Nature of the chloroplast insertions	30
3.1.5	Mitochondrial homologies in the rice nuclear genome	30
3.1.6	Representation of the mitochondrial genome in the rice nuclear genome	31
3.1.7	Mitochondrial ATPase- β genes in the nucleus	34
3.1.8	Co-alignments between the chloroplast and mitochondria	36
3.1.9	Size distribution of insertions	37
3.2	Viral Comparison	38
3.3	Bacterial Comparison	41
3.3.1	<i>Bacillus</i> sequence similarities in the rice nuclear genome	41
3.3.2	<i>Xanthomonas</i> sequence similarities in the rice nuclear genome	43
3.3.3	<i>Pseudomonas</i> sequence similarities in the rice nuclear genome	44
3.4	Fungal comparison	46
3.4.1	<i>Magnaporthe</i> sequence similarities in the rice nuclear genome	46
3.4.2	Dendogram type 1	47
3.4.3	Dendogram type 2	50
3.4.4	Dendogram type 3	54
4	Discussion	57
4.1	Motivation for this study	58
4.2	Discussion of techniques and results	58
4.2.1	Search techniques and parameters	58
4.2.2	DNA homology between the plastid genomes and the rice nuclear genome	59
4.2.3	Sequence similarities between the viral genomes and the rice nuclear genome	62



4.2.4	Sequence similarities between the bacterial genomes and the rice nuclear genome	65
4.2.5	Sequence similarities between the <i>Magnaporthe</i> genome and the rice nuclear genome	67
4.3	Insertion dynamics	70
4.4	Models for DNA transfer	71
4.5	Benefits of insertion	72
4.6	Contribution of this study to science	74
4.7	Future perspectives	76
5	References	78

APPENDIX

Table A-1	Chloroplast DNA homologies in the rice nuclear genome	101
Table A-2	Mitochondrial DNA homologies in the rice nuclear genome	194
Table A-3	Regions in the rice genome that align with the chloroplast and mitochondrial genomes	265
Table B-1	Rice tungro bacilliform virus (RTBV) alignments in the rice genome	270
Table B-2	RNA viral alignments in the rice nuclear genome	285
Table C-1	<i>Bacillus cereus</i> alignments in the rice nuclear genome	287
Table C-2	<i>Xanthomonas oryzae</i> alignments in the rice genome	307
Table C-3	<i>Pseudomonas syringae</i> alignments in the rice genome	310
Table D-1	<i>Magnaporthe grisea</i> alignments in the rice genome	314



Summary

With an increasing amount of whole genome sequence data becoming available on a daily basis we have an opportunity to study the interactions and dynamics of different organisms on a whole genome level. In the past, reports of horizontal gene transfer have focused mainly on the identification of single genes that show distorted phylogenetic profiles to that of the organism it was isolated from. This study firstly did whole genome comparisons between the rice nuclear and plastid genomes to determine the level and dynamics gene transfer and insertion of the chloroplast and mitochondrial genomes into that of the nuclear genome of rice. Secondly, it looked to identify sequence similarities between the rice genome and microbial genomes by performing whole genome comparisons between the rice genome and that of several microbial genomes. These sequences were analyzed further to identify possible instances of horizontal transfer of DNA from microbes to the rice genome. Using this approach, this study reports several fragments in the rice genome with significant sequence similarity to that of microbial DNA fragments. This study also provides evidence supporting horizontal transfer of several of these fragments. This study provides valuable information regarding intra- as well as inter-genome DNA transfer dynamics.



Thesis Composition

The following thesis comprises of five sections. **1 Introduction and literature review** presents an overview regarding the origin and evolution of the eukaryotic genome as well as the origin of the chloroplast and mitochondrion. It also review previous reports regarding DNA transfer between the organelles and nucleus, as well as those regarding transfer of DNA between various microbes and plants. It also discusses the proposed models whereby transfer occurs. **2 Materials and Methods** outlines the different strategies and methods and sequences used in the whole genome comparisons and the analysis of the results. **3 Results** describes the results obtained from each of the different comparisons. **4 Discussion** provides an overall discussion of the study and provide some future perspectives of research in the field of horizontal DNA transfer. In the **Annex** the tables detailing the sequence similarities found between the rice genome and the various other genomes used are presented.

Acknowledgements

I would like to thank the University of Pretoria and the National Research Foundation for financial assistance that allowed me to conduct and complete my research. Special thanks also to Prof. Karl Kunert for years of patience and support as well as his advice and encouragement. My sincere gratitude and appreciation also goes Prof. Chris Cullis not only for allowed me to work in his laboratory and to use the facilities of Case Western Reserve University to perform the whole genome comparisons, but also for his valuable inputs and for sharing his vast knowledge on genomics and for his guidance. This thesis would not have been completed if not for the help and support from friends, family and lab colleagues. Thanks for all your encouragement and support throughout my project.



Index of figures

Figure 1.1	3
A schematic diagram of the ring of life.	
Figure 1.2	5
The common ancestor of eukaryotes, bacteria, and archaea.	
Figure 1.3	13
A diagrammatic summarization of all major cellular reactions involved in <i>Agrobacterium</i> T-DNA transport.	
Figure 3.1	27
Distribution of chloroplast homologies in the rice nuclear genome.	
Figure 3.2	28
Representation of the chloroplast genome in the nuclear genome of rice.	
Figure 3.3	29
Chloroplast rRNA insertions in the nuclear genome.	
Figure 3.4	30
Chloroplast insertions on rice chromosome 12.	
Figure 3.5	31
Distribution of mitochondrial homologies in the rice nuclear genome.	
Figure 3.6	32
Representation of the mitochondrial genome in the nuclear genome of rice.	
Figure 3.7	33
Localization of plastid NADH dehydrogenase subunits in the nuclear genome of rice.	



Figure 3.8	33
Position of the NADH subunits (red wide box) relative to the chloroplast (green) and mitochondrial (blue) insertions.	
Figure 3.9	35
Location of the five chloroplast ATPase CF ₁ β-subunit copies.	
Figure 3.10	36
Alignment of the translated coding sequences of the two nuclear ATPase-β copies in rice (Rice1 and Rice5) and the rice mitochondrial copy (Mt).	
Figure 3.11	37
Mitochondrion, chloroplast and mitochondrion-chloroplast co-alignments on each of the rice nuclear chromosomes.	
Figure 3.12	38
Size distribution of plastid insertions in the nuclear genome.	
Figure 3.13	40
Positions of the viral homologies in rice.	
Figure 3.14	41
Distribution of <i>Bacillus cereus</i> alignments on the rice karyotype.	
Figure 3.15	43
Distribution of <i>Xanthomonas</i> alignments on the rice karyotype.	
Figure 3.16	45
Distribution of <i>Pseudomonas</i> alignments on the rice karyotype.	
Figure 3.17	46
Distribution of <i>Magnaporthe grisea</i> alignments on the rice karyotype.	
Figure 3.18	48
Sequence dendogram from a multiple sequence alignment with the rice fragment 3:4074435:4074559	



Figure 3.19	49
Sequence dendogram from a multiple sequence alignment with the rice fragment 10:65974:66122.	
Figure 3.20	50
Sequence dendogram from a multiple sequence alignment with the rice fragment 3:11502555 – 11502670.	
Figure 3.21	51
Sequence dendogram showing the relationship between a DNA fragment from a multiple sequence alignment with the rice fragment 6:4668751 – 4669004.	
Figure 3.22	52
Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 8: 5640822 – 5640946.	
Figure 3.23	53
Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 6: 5884317:5884417.	
Figure 3.24	54
Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 2: 3341370: 3342015.	
Figure 3.25	55
Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 4: 23876766:23876875.	
Figure 3.26	56
Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 1: 36368256: 36368446.	



Index of tables

Table 1.1	7
Mitochondrial and chloroplast genome sizes.	
Table 1.2	16
Published accounts of horizontally acquired genes shown or thought to be located in plant mitochondrial genomes.	
Table A-1	101
Chloroplast DNA homologies in the rice nuclear genome.	
Table A-2	194
Mitochondrial DNA homologies in the rice nuclear genome.	
Table A-3	265
Regions in the rice genome that align with the chloroplast and mitochondrial genomes.	
Table B-1	270
Rice tungro bacilliform virus (RTBV) alignments in the rice genome.	
Table B-2	285
RNA viral alignments in the rice nuclear genome.	
Table C-1	287
<i>Bacillus cereus</i> alignments in the rice nuclear genome.	
Table C-2	307
<i>Xanthomonas oryzae</i> alignments in the rice genome.	
Table C-3	310
<i>Pseudomonas syringae</i> alignments in the rice genome.	
Table D-4	314
Magnaporthe grisea alignments in the rice genome.	