Phylogenomics and Plant Evolution

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

Doctor of Philosophy
in Genetics

at Massey University, Manawatu
New Zealand

Bojian Zhong

2013

Abstract

Phylogenomics, the study of evolutionary relationships among groups of organisms using genome-scale data, is central to our understanding of the evolution of life. While large amounts of data are available and methodological developments are increasing at a fast pace, there are basic problems that are overlooked in phylogenomic analyses of molecular sequences, which may impede the accuracy and reliability of tree reconstruction. These problems include: How can we detect the non-phylogenetic signals from genomic data? How can we offer a better statistical fitness between the evolutionary model and data? How can we improve the phylogenetic inference using sophisticated and realistic models? How can we accurately infer the species trees? How can we quantitatively confirm the evolutionary theory? With these goals, this thesis concentrates on phylogenomics of land plants (and their origin) and evolution in general.

- Resolving the phylogenetic position of Gnetales. We show that non-time reversible properties of positions in the chloroplast genomes of Gnetales mislead phylogenetic reconstruction, and highlight that the goodness of fit between substitution model and data should be taken into account when performing phylogenomic analyses.

- Resolving the origin of land plants: 1). The multispecies coalescent model is applied to estimate the species tree of origin of land plants, and it is proved to be able to estimate accurate and congruent species tree in the presence of ancient incomplete lineage sorting from nuclear genes. 2). The chloroplast phylogenomic analyses are conducted using sophisticated and realistic evolutionary models that can account for site-heterogeneity and compositional heterogeneity. These chloroplast phylogenomic results confirm the previous nuclear data analyses.

- We develop a statistical test and demonstrate that evolutionary theory could be tested by convergence of molecular data. It also indicates that the reality of evolution can be tested using standard methods and tools.

Acknowledgements

I would like to thank my supervisors, Professor David Penny and Professor Peter Lockhart for your support, your encouragement and your patience. I would not have had the opportunity to make such an achievement without you.

I also wish to thank my colleagues: Dr. Tim White, who always helps me figure out computing problems; P.A. (Trish) McLenachan and Richard Fong, who have done a lot of experimental work for different projects; Dr. Bennet J. McComish, Dr. Oliver Deusch and Dr. Patrick Biggs, who have provided bioinformatic assistance.

I express my warm thanks to my other co-authors:
Dr. Vadim V. Goremykin (Istituto Agrario San Michele all’Adige Research Center, San Michele all’Adige, Italy)
Dr. Liang Liu (Department of Statistics and Institute of Bioinformatics, University of Georgia, USA)
Dr. Zhenxiang Xi (Department of Organismic and Evolutionary Biology, Harvard University Herbaria, USA)
Dr. Philip M. Novis (Allan Herbarium, Landcare Research, Lincoln 7640, New Zealand)

I thank my friends and my former flatmates: Jingjing, Wesley, Jing Pang, Yanfei, Jian Han, Chuchu, Qinghao, Jinlin, and Hao Zhang who have made New Zealand as my “new home” with a wonderful time.

I thank AWC administrators Wendy and Joy for all your assistance with financial support, and thank to all (former) members in our lab, for their company and discussions.

Lastly, my deepest appreciation to my parents, for their love and support throughout my life.

CONTENTS