AN EMPIRICAL ASSESSMENT OF
THE BOOTSTRAP SUPPORT AS AN INDICATOR OF
ROBUSTNESS IN PHYLOGENETIC TREES

SO WEI HUO

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2012

AN EMPIRICAL ASSESSMENT OF
THE BOOTSTRAP SUPPORT AS AN INDICATOR OF
ROBUSTNESS IN PHYLOGENETIC TREES

SO WEI HUO

(SGJ 100005)

SUBMITTED TO

INSTITUTE OF BIOLOGICAL SCIENCES

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

IN PARTIAL FULFILMENT

OF THE REQUIREMENT FOR

THE DEGREE OF MASTER OF BIOINFORMATICS

2012

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate:	SO WEI HUO	(I.C/Passport No: 850701145175)
Registration/Matric No:	SGJ100005	
Name of Degree:	MASTER OF BIOINFORMATICS	

An empirical assessment of the bootstrap support as an indicator of robustness in phylogenetic trees (“this Work”)

Field of Study: Bootstrapping, Phylogenetics, Bioinformatics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;

(2) This Work is original;

(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;

(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

(6) I am fully aware that if in the course of making this Work I have infringed any Copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature Date: 22nd June 2012

Subscribed and solemnly declared before,

Witness’s Signature Date: 22nd June 2012

Name:

Designation:

ABSTRACT

Bootstrapping is a de-facto standard for displaying the validity of a phylogenetic tree. However this may not be the case as bootstrapping assume that the random sequence exist within the genome. In order to examine whether bootstrapping is a valid process, it must be done empirically. Across multiple taxa there are minimal fully curated genomes but there are mitochondrion genomes available to test this. A widely accepted evolutionary tree is chosen with 6 taxa - *Pan paniscus* (bonobo), *Homo sapiens* (human), *Gorilla gorilla* (gorilla), *Pongo pygmaeus* (orangutan), *Hylobates lar* (common gibbon) and *Gallus gallus* (red junglefowl) as the outgroup. A comparison of topology and bootstrap support value is executed between the phylogenetic tree of the whole mitochondrial genome with trees built from all the genes in mitochondrial genome. The result shows that bootstrap support value tend to inflate and larger than the empirical estimate. This suggests that bootstrap support in phylogenetic trees must be interpreted cautiously and not casually accepted at face value.

ABSTRAK

Butstrap ialah piawaian de facto untuk memaparkan kesahihan pokok evolusi. Walau bagaimanapun, ini tidak selalu berlaku kerana butstrap menganggap bahawa jujukan rawak wujud dalam genom. Untuk memeriksa sama ada butstrap adalah satu proses yang sah, ia mesti dilakukan secara empirikal. Terdapat minima genom yang dikemaskini dan dikuratorkan merentasi beberapa taksa terdapat genom mitochondrion disediakan untuk menguji situasi ini. Pokok evolusi yang diterima secara meluas dipilih dengan 6 taksa - *Pan paniscus* (bonobo), *Homo sapiens* (manusia), *Gorilla gorila* (gorila), *Pongo pygmaeus* (orangutan), *Hylobates lar* (siamang biasa) dan *Gallus Gallus* (ayam hutan merah) sebagai outgroup itu. Perbandingan nilai sokongan topologi dan bootstrap yang dimeterai antara pokok filogenetik seluruh genom mitokondria dengan pokok-pokok yang dibina dari semua gen dalam genom mitokondria. Hasil menunjukkan bahawa nilai sokongan butstrap cenderung untuk mengembung dan lebih besar daripada anggaran empirik. Ini menunjukkan bahawa sokongan butstrap dalam pokok evolusi mesti ditafsirkan dengan berhati-hati dan tidak bersahaja diterima pada nilai muka.

ACKNOWLEDGEMENT

First and foremost, I would like to express my utmost gratitude to my respected supervisor, Dr. Khang Tsung Fei for his invaluable guidance, patience and support throughout the duration of the project. I am extremely thankful for his advice, suggestion and time spent during the discussion throughout the duration of this thesis.

I would also like to thank my employer, Dr Suresh Chandran and Dr Wu Sheng Chuan for their understanding and some freedom to complete the thesis. My special thanks to my classmates especially to Chow Kingsley, Kok Eng Piew and Agan Adun Shashi Mohan for their support and lending of a listening ear to complete the thesis.

TABLE OF CONTENTS

Chapter **Pages**

LIST OF FIGURES

Pages

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

APE	Analyses of Phylogenetics and Evolution
ATP	Adenosine Triphosphate
CDS	Coding Sequence
CI	Confidence Interval
GI	GenInfo Identifier
kbp	Kilo-base pair
MEGA5	Molecular Evolutionary Genetics Analysis version 5
ML	Maximum Likelihood
MSA	Multiple Sequence Alignment
mtDNA	Mitochondrion DNA
MUSCLE	Multiple Sequence Comparison by Log- Expectation
NADH	Nicotinamide adenine dinucleotide (reduced form)
NCBI	National Center for Biotechnology Information
ND	NADH dehydrogenase
NNI	Nearest-Neighbour-Interchange
PyCogent	Python the Comparative Genomic Toolkit
RNA	Ribonucleic Acid
rRNA	Ribosome Ribonucleic Acid
SD	Standard Deviation
SE	Standard Error
TN93	Tamura-Nei, 93
tRNA	Transfer Ribonucleic Acid
µm	Micrometer

LIST OF APPENDICES