


Spring 4-13-2017

# Modeling Trait Evolutionary Processes with More Than One Gene

Huan Jiang

Follow this and additional works at: [https://digitalrepository.unm.edu/math\\_etds](https://digitalrepository.unm.edu/math_etds)

 Part of the [Applied Mathematics Commons](#), [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

---

## Recommended Citation

Jiang, Huan. "Modeling Trait Evolutionary Processes with More Than One Gene." (2017). [https://digitalrepository.unm.edu/math\\_etds/106](https://digitalrepository.unm.edu/math_etds/106)

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Mathematics & Statistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

**Huan Jiang**

*Candidate*

**Mathematics and Statistics**

*Department*

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

**James Degnan** \_\_\_\_\_, Chairperson

**Ronald Christensen** \_\_\_\_\_

**Joseph Cook** \_\_\_\_\_

**Li Li** \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

# Modeling Trait Evolutionary Processes with More Than One Gene

BY

HUAN JIANG

M.S., Statistics, University of New Mexico, 2014

DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

**Doctor of Philosophy**

**Statistics**

The University of New Mexico

Albuquerque, New Mexico

**Spring 2017**

©2017, Huan Jiang

## DEDICATION

To my beloved parents and husband.

## ACKNOWLEDGMENTS

Most of all, I would like to thank Professor James Degnan, my dissertation advisor and chair, for his guidance and support over the past three years. Professor Degnan is a great mentor and very helpful.

I would also like to extend my thanks to the members of my dissertation committee: Professor Ronald Christensen, Professor Li Li and Professor Joseph Cook.

I would also like to thank the chair, faculty and staff of the Mathematics and Statistics Department as well as my fellow graduate students.

I would like to thank my parents for their endless support.

Most importantly, I would like to thank my husband with his company through my whole PhD period.

# Modeling Trait Evolutionary Processes with More Than One Gene

by

**Huan Jiang**

## **Abstract**

Phylogenetic comparative methods have been used to test evolutionary signals through trait evolutionary processes. Traditionally, biologists use one phylogenetic tree as a tool to handle dependent data for the traits of interest and hence utilize one gene only. However, it is more informative if the evolutionary processes of a trait are presented by phylogenetic trees reconstructed by the DNA alignments from more than one gene. In this work, we explain and develop two methods involving modeling the trait evolutionary processes: (a) two gene trees via the Brownian motion (BM) model; and (b) two gene trees via the Ornstein-Uhlenbeck (OU) model. For presentation purposes, these two models obtain evolutionary signals from both gene trees and then utilize the variance-covariance matrix of the error terms through the phylogenetic generalized linear model approach. For parameter estimation purposes, we develop the two gene trees model for one trait by using two different Bayesian approaches: (a) the adaptive Metropolis-Hastings (AMH) approach; and (b) the approximate Bayesian computation (ABC) method. The simulations from this study indicate that, under the two gene trees BM model with one trait, the AMH approach performs well on gene trees with longer internal branch lengths regardless of the number of taxa, whereas the ABC method only performs well on small numbers of taxa. In addition,

we find that the AMH approach also performs well on gene trees with larger numbers of taxa regardless of the branch lengths and topologies. To illustrate the two methods, we analyze data of grain width of rice varieties that are associated with two major genes: *GW2* and *qSW5* under artificial selection, which implies that the BM model is a wrong model because it ignores selection. However, when applying model selection, according to both AIC and BIC, between the one gene tree model versus two gene trees model, we find that: (i) the one gene tree BM model has the lowest AIC and BIC; and (ii) the two gene trees OU model is better than the one gene tree OU model. Since the grain width is under artificial selection, the assumptions of the two gene trees OU model seem more applicable to the real case as it could capture more evolutionary signals from the data.

KEY WORDS: Two gene trees trait evolutionary model, phylogenetic trees, Ornstein-Uhlenbeck, Brownian motion, dependency, adaptive Metropolis-Hastings (AMH) approach, approximate Bayesian computation (ABC) method, rice varieties.



# Contents

<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 The Likelihood/Bayesian method for inferring trees from DNA sequences	2
1.2.1 Defining a phylogenetic tree . . . . .	2
1.2.2 Substitution models . . . . .	5
1.2.3 Likelihood/Bayesian method . . . . .	10
1.2.3.1 Maximum Likelihood approach . . . . .	11
1.2.3.2 Bayesian approach . . . . .	15
1.3 Modeling evolutionary changes through traits . . . . .	18
1.3.1 Brownian motion (BM) model . . . . .	18
1.3.2 Ornstein-Uhlenbeck (OU) model . . . . .	21
1.4 The Statistical comparative methods in evolutionary biology . . . . .	23
1.4.1 GLS method . . . . .	25
1.4.2 PIC method . . . . .	26

1.4.3	Application of PIC and GLS methods . . . . .	27
1.5	Summary of background . . . . .	31
<b>2</b>	<b>Two gene trees BM model</b>	<b>32</b>
2.1	Two gene trees BM model with one trait . . . . .	32
2.1.1	Assumptions of two gene trees BM model . . . . .	32
2.1.2	Adaptive Metropolis-Hastings approach . . . . .	35
2.1.3	Approximate Bayesian computation method . . . . .	39
2.2	Two gene trees BM model with more than one trait . . . . .	40
2.3	Simulation study . . . . .	41
2.3.1	Simulation procedures . . . . .	41
2.3.2	Simulation results . . . . .	42
2.3.2.1	Effect of priors . . . . .	42
2.3.2.2	Effect of numbers of taxa . . . . .	44
2.3.2.3	Effect of gene trees incongruence . . . . .	45
<b>3</b>	<b>Two gene trees OU model</b>	<b>51</b>
3.1	Two gene trees OU model with one trait . . . . .	51
3.1.1	Adaptive Metropolis-Hastings approach . . . . .	54
<b>4</b>	<b>Effect of misestimated ultrametric trees</b>	<b>59</b>
4.1	Simulation study based on current one tree trait model . . . . .	59
4.2	Simulation results . . . . .	62
4.2.1	Coverage probabilities with different numbers of taxa . . . . .	62
4.2.2	Different performance in inferring the variance of a BM model using iBPP, RAxML and MrBayes . . . . .	63

4.2.3	Comparison between BM and OU models . . . . .	66
4.2.4	The changes of normalized RF distances with different numbers of taxa . . . . .	79
<b>5</b>	<b>Empirical Study</b>	<b>84</b>
5.1	Data preparation . . . . .	84
5.2	Application of the two gene trees model with one trait . . . . .	86
5.3	Application of the one gene tree model with one trait . . . . .	90
<b>6</b>	<b>Discussion</b>	<b>94</b>
6.1	AMH and ABC methods . . . . .	94
6.2	One gene tree and two gene trees for dependency . . . . .	96
6.3	Simple models versus more complicated models for accuracy . . . . .	98
6.4	Conclusion . . . . .	99
6.5	Potential future work . . . . .	100
<b>A</b>	<b>Data and Command Files in the Empirical Study</b>	<b>101</b>
A.1	The trait data in the empirical study . . . . .	101
A.2	Command files for MrBayes in the empirical study . . . . .	102
	<b>Bibliography</b>	<b>152</b>

# List of Tables

2.1	Performance of the AMH approach of trees with longer internal branch lengths . . . . .	48
2.2	Performance of the ABC method of trees with longer internal branch lengths. . . . .	49
2.3	Performance of the AMH approach of random gene trees. . . . .	50
2.4	Performance of the ABC method of random gene trees. . . . .	50
5.1	Empirical study of the ABC method under the BM model . . . . .	88
5.2	Empirical study of the AMH method under the BM model . . . . .	88
5.3	Empirical study of the AMH method under the OU method . . . . .	89
A.1	The trait data in the empirical study. . . . .	101

# List of Figures

1.1	An example of a phylogenetic tree. The root of the tree is blue square 7. The other blue square numbers 8, 9, 10, and 11 are internal nodes. The yellow square numbers 1, 2, 3, 4, 5, and 6 are tips represented by letters A, B, C, D, E, F and G, respectively. The black square numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 are branch lengths. . . . .	3
1.2	An example to show the difference between rooted (left panel) and unrooted trees (right panel). . . . .	4
1.3	An example to show the difference between balanced (left panel) and unbalanced trees (right panel). . . . .	4
1.4	DNA sequences (left panel) and alignments (right panel); $t_i$ denotes the $i$ species (taxon). . . . .	6
1.5	An example to show how to calculate the likelihood of a given tree in a single site. . . . .	12
1.6	An example of PIC. . . . .	28
1.7	An example of how to compute the branch lengths for contrasts. . . . .	29

- 2.1 An example to calculate the  $V$  matrix based on two gene trees with a trait evolving by the BM model. The trait values  $x_{ijk}$  are at each tip and each node are unknown. Here  $i$  indexes the gene tree,  $j$  indexes the extant species, and  $k$  indexes the trait. The branch lengths  $v_{il}$  are shown next to each branch. For  $v_{il}$ ,  $i$  indexes the gene tree and  $l$  indexes the branch on the tree,  $l = 1, 2, \dots, 2n - 2$  for  $n$  species. Note that in our example, we assume gene tree 1 and 2 are ultrametric. So, in gene tree 1,  $v_{11}$  is the branch length from  $A$  or  $B$  to their most recent common ancestor;  $v_{12}$  is the branch length from  $C$  or  $D$  to their most recent common ancestor. In gene tree 2,  $v_{21}$  is the branch length from  $A$  or  $C$  to their most recent common ancestor;  $v_{22}$  is the branch length from  $B$  or  $D$  to their most recent common ancestor. . . . . 46
- 2.2 An example of trace plots and histograms using the AMH approach for the 60 taxa trees with longer internal branch lengths. . . . . 47

- 3.1 An example to calculate the V matrix based on two gene trees with a trait evolving by the OU model. The trait values  $x_{ijk}$  are at each tip and each node are unknown. Here  $i$  indexes the gene tree,  $j$  indexes the extant species, and  $k$  indexes the trait. The branch lengths  $v_{il}$  are shown next to each branch. For  $v_{il}$ ,  $i$  indexes the gene tree and  $l$  indexes the branch on the tree,  $l = 1, 2, \dots, 2n - 2$  for  $n$  species. Note that in our example, we assume gene tree 1 and 2 are ultrametric. So, in gene tree 1,  $v_{11}$  is the branch length from  $A$  or  $B$  to their most recent common ancestor;  $v_{12}$  is the branch length from  $C$  or  $D$  to their most recent common ancestor. In gene tree 2,  $v_{21}$  is the branch length from  $A$  or  $C$  to their most recent common ancestor;  $v_{22}$  is the branch length from  $B$  or  $D$  to their most recent common ancestor. . . . . 58
- 4.1 A plot of the simulation process of inferring variances from BM models. 60
- 4.2 Coverage probability when variance = 1 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000. . . . . 64
- 4.3 Coverage probability when variance = 2 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000. . . . . 65

- 4.4 Different performance of confidence intervals when variance = 1, birth rate = 1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . . 67
- 4.5 Different performance of confidence intervals when variance = 1, birth rate = 0.1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . . 68
- 4.6 Different performance of confidence intervals when variance = 2, birth rate = 1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . . 69
- 4.7 Different performance of confidence intervals when variance = 2, birth rate = 0.1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . . 70
- 4.8 Different performance of confidence intervals when variance = 1, birth rate = 1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . . 71
- 4.9 Different performance of confidence intervals when variance = 1, birth rate = 0.1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . . 72



4.10	Different performance of confidence intervals when variance = 2, birth rate = 1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . .	73
4.11	Different performance of confidence intervals when variance = 2, birth rate = 0.1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . .	74
4.12	Different performance of confidence intervals when variance = 1, birth rate = 1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . .	75
4.13	Different performance of confidence intervals when variance = 1, birth rate = 0.1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . .	76
4.14	Different performance of confidence intervals when variance = 2, birth rate = 1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . .	77
4.15	Different performance of confidence intervals when variance = 2, birth rate = 0.1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000. . . . .	78

4.16	Percentages of correct decision ( $1 - \alpha$ ) in testing $H_0$ : BM model is adequate vs. $H_1$ : OU model is adequate for simulate data from the BM model when variance = 1 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000. . . . .	80
4.17	Percentages of correct decision ( $1 - \alpha$ ) in testing $H_0$ : BM model is adequate vs. $H_1$ : OU model is adequate for simulate data from the BM model when variance = 2 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000. . . . .	81
4.18	Normalized RF distances when birth rate = 1 (left) and 0.1 (right). . . . .	83
5.1	<i>GW2</i> and <i>qSW5</i> gene trees. . . . .	86
5.2	Trace plots and histograms in our best model. Grey indicated burnin for the trace plots. Histograms don't include the burnin. . . . .	91
5.3	The concatenated gene tree for <i>GW2</i> and <i>qSW5</i> . . . . .	92

# Chapter 1

## Introduction

### 1.1 Background

The history of evolution is a mystery. During evolutionary processes, when species evolve, the adaptation of traits happens with respect to changes from mutations of deoxyribonucleic acid (DNA) alignments and of outside environments. The big question is: how do we trace back the life of living or extinct organisms or species and explore the relationships between them? In *On the Origin of Species*, Charles Darwin used a theoretical tree diagram for species of a large genus in order to present their evolutionary process [1]. Before DNA data become so easy to access under fast and improved computer techniques, traits' values were the main thing used to estimate evolutionary processes. Even in recent days, modeling the evolutionary changes through traits can be thought of as equally important to reconstructing the phylogenetic trees from DNA alignments. Some researchers even combine these two methodologies together to better understand the evolution of some groups of specie [2]. Current trait evolutionary models are based on only one tree, which often

represents the history of a single gene. This might be unrealistic since traits are influenced by multiple genes. Our research is trying to look for a more realistic way to model the trait evolutionary processes by considering more than one gene. To expand on our research interest, we need to introduce the background of three different concepts: (1) how to infer a phylogenetic tree from DNA sequences; (2) how to model evolutionary changes through traits; and (3) how to use statistical comparative methods in evolutionary biology to infer the relationships between different species.

## **1.2 The Likelihood/Bayesian method for inferring trees from DNA sequences**

Inferring a phylogenetic tree from DNA sequences became easy and effective with increased convenience in generating DNA sequences. Three main types of methods have been proposed in the last 50 years to infer phylogenetic trees: (1) Parsimony; (2) Distance and (3) Likelihood/Bayesian methods. Unlike the other two methods, the Likelihood/Bayesian method calculates the probabilities of the DNA data as a function of the tree and is considered the most accurate method in the literature [3, 4]. Before we talk about the Likelihood/Bayesian method, we will introduce related concepts including substitution models.

### **1.2.1 Defining a phylogenetic tree**

If we want to use a phylogenetic tree to describe the evolutionary history, the first thing we should know is its two parts. Any phylogenetic tree is made up of two parts, namely: the topology and branch lengths. Topology is the shape of the tree

through which one connects the root, nodes, and tips together with branches. We denote the *tips* as extant species, *nodes* as common ancestors with either offspring as tips in the tree or interval nodes. These nodes may be either extinct species or extant species. The *clade* is any group of species for which the common ancestor does not have any other descendants in the study. For example, in Figure 1.1, the sets  $\{A, B\}$ ,  $\{C, D\}$ ,  $\{C, D, E\}$  and  $\{A, B, C, D, E\}$  are all clades. The root is the common ancestor of all the species presented in the phylogenetic tree which could go backward millions of years ago. Figure 1.1 shows an example of a phylogenetic tree. Sometimes, a phylogenetic tree is presented without a root, in which case it is called

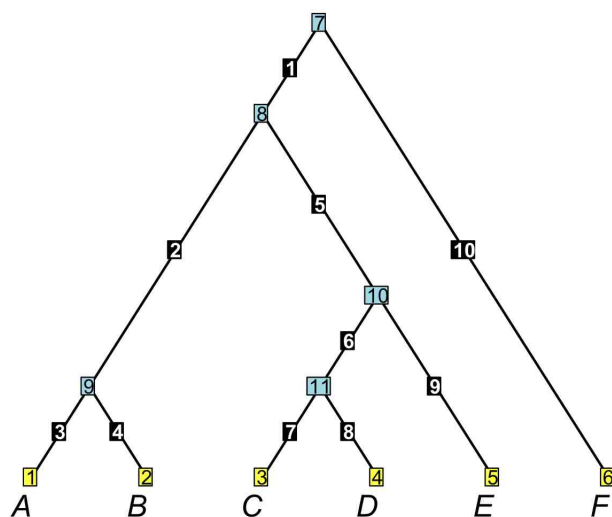


Figure 1.1: An example of a phylogenetic tree. The root of the tree is blue square 7. The other blue square numbers 8, 9, 10, and 11 are internal nodes. The yellow square numbers 1, 2, 3, 4, 5, and 6 are tips represented by letters A, B, C, D, E, F and G, respectively. The black square numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 are branch lengths.

as an unrooted tree. If the phylogenetic tree has a root, we call it a rooted tree. Figure 1.2 shows the difference between a rooted tree on the left and an unrooted tree on the right with four taxa (tips). According to the topology, trees could be

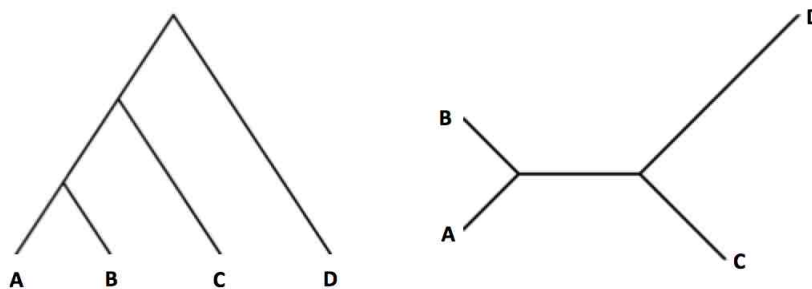


Figure 1.2: An example to show the difference between rooted (left panel) and unrooted trees (right panel).

categorized as either balanced or unbalanced trees. Balanced trees have symmetric shapes; otherwise, they are unbalanced trees. Figure 1.3 shows a balanced 4 taxa tree on the left and an unbalanced 4 taxa tree on the right. Branch lengths generally

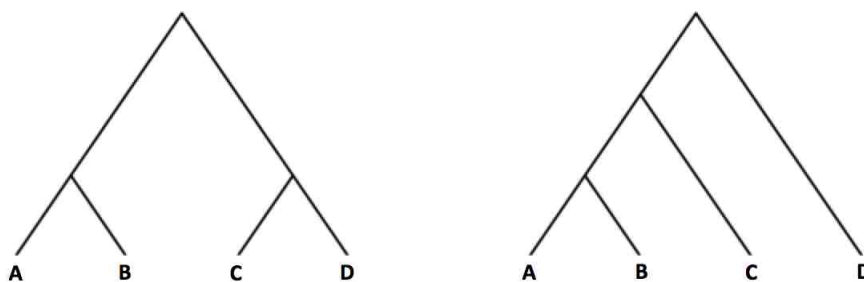


Figure 1.3: An example to show the difference between balanced (left panel) and unbalanced trees (right panel).

represent the time spent in the traits' evolutionary process or the expected amount of mutations, where one mutation is defined as any change in DNA alignments. If we have long branch lengths, it means that species spent more time to speciation. If the sum of branch lengths from the tip to the root are the same for all the species in the tree, we call the tree ultrametric or clock-like. The two trees in Figure 1.3 are ultrametric trees.

### 1.2.2 Substitution models

Since we are using DNA alignments as our data, first we need to explain how DNA is related to evolution. A DNA sequence is a long sequence of four different types of nucleotides made by four nitrogenous bases: adenine, guanine, cytosine, and thymine. In biology, we denote these four bases as letters A, G, C, and T, respectively. To understand the difference between DNA alignments and sequences, we simply consider DNA sequences as the raw data we get directly from data sampling; i.e. such data are not ready for analysis and include missing values and sequencing errors. However, DNA alignments are the data that have been cleaned, imputed for missing values, free of errors and ready to use. As illustrated on the left panel of Figure 1.4, through evolution, there occurs lots of deletion, insertion, and replication in the DNA sequences. On the other hand, as illustrated in the right panel of Figure 1.4, DNA alignments are received after applying sequence alignment techniques to handle these three situations (deletion, insertion, and replication). The very same Figure 1.4 (right panel) shows that the gaps are inserted into alignments and shows that deletion happened in the evolutionary process. Also the figure that shows three mutations have happened in the second species ( $t_2$ ). These DNA alignments are the genetic material that con-

	DNA Sequences		DNA Alignments
<b>t<sub>1</sub></b>	CCGAGTCTG	Apply sequence alignment methods →	<b>t<sub>1</sub></b> - C C G - A G T C - T G - -
<b>t<sub>2</sub></b>	CCCAAATCCT		<b>t<sub>2</sub></b> - C C C A A A T C C T - - -
<b>t<sub>3</sub></b>	GCCGATGTT		<b>t<sub>3</sub></b> G C C G A T G T - - T - - -
<b>t<sub>4</sub></b>	CCGAGTCT		<b>t<sub>4</sub></b> - C C G - A G T C - T - - -
<b>t<sub>5</sub></b>	GAATCTGGA		<b>t<sub>5</sub></b> G - - - A A - T C - T G G A

Figure 1.4: DNA sequences (left panel) and alignments (right panel);  $t_i$  denotes the  $i$  species (taxon).

tains all inherited characteristics in living organisms, which influence how organisms look and behave. When organisms have descendants, their inherited characters are passed on though the replication of their DNAs. In the process of replication over generations, mutations of DNA would happen that may cause changes of all aspects of organisms' lives, including the creation of new species. Evolution occurs with any mutation of the DNA alignments. Without mutation, there would be no evolution [5].

Since mutation is so rare, a mutation may cause big changes of any organisms' lives. For example, a published paper, *Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome*, in Nature in 2005 [6] stated that the mean rate of mutations of a single nucleotide is 1.23% during the replication processes between the human and chimpanzee genome. Here I do not need to explain the differences between human and chimpanzee. This tells us that a small percentage of DNA mutations over generations could cause the huge differences between the human and chimpanzee as an example. If we use DNA alignments from different species to reconstruct the history of evolution according to the mutations of DNA nucleotides, more closely related species tend to have more similar DNA alignments. So the most



important part in reconstructing phylogenetic trees is to figure out probabilities of mutations.

Since we know that a phylogenetic tree uses branch lengths to capture the time of the evolutionary process, then it's natural to think of the evolutionary process as a continuous-time Markov chain with finite states, A, T, C and G; the four nitrogenous bases. In the section, we use the notations that Felsenstein uses in his book *Inferring phylogenies* [47]. The direction of time in the Markov Chain is defined by the topology of the tree from the past to the present, and the time of the Markov Chain process is defined by the branch lengths of the tree. Given the Markov chain process, we need to define the instantaneous rate matrix denoted as  $Q$  and the base frequencies vector denoted as  $\pi$ . The rate matrix  $Q$  shows the instantaneous rate of any mutation from one base changing into another base per unit time. In other words, the entry  $Q_{ij}$  for  $i \neq j$  is the rate at which base  $i$  mutates to base  $j$  with

$$Q_{ii} = - \sum_{\{j \neq i\}} Q_{ij}, \quad (1.1)$$

which means that the sum of each row in the rate matrix equals to zero. Since substitution models are time-reversible and stationary models,

$$\pi_i Q_{ij} = \pi_j Q_{ji}, \quad (1.2)$$

we can estimate  $\pi$  by using the average frequencies of four bases over the entire sequence alignment.

In the tree, branch lengths represent the time spent on a lineage, where a lineage means the path from the tip to the root. If the branch length is short, the probability of any mutation is much smaller than the probability of staying at the same base. Therefore, a transition matrix, called the substitution model, is proposed to measure

probabilities of changes between different bases during any time interval  $t$ . It is denoted as  $P(t)$ , which can be computed from the instantaneous rate matrix:

$$P(t) = e^{Qt}. \quad (1.3)$$

Let's denote the  $ij$  probability in the matrix  $P(t)$  be  $P_{ij}(t)$ , which  $P_{ij}(t)$  gives after time  $t$ , conditioning on the past base  $i$ , the probability of changing to base  $j$ . Since we can consider the base frequencies matrix  $\pi$  as the equilibrium matrix of base frequencies, it could be calculated as  $\lim_{t \rightarrow \infty} P_{ij}(t) = \pi_j$ . So if the time goes to infinity, at some position of the DNA alignment given a base  $i$ , the probability of base  $j$  will get closer and closer to the equilibrium probability of base  $j$  at the position without considering the original base  $i$ . Because of this, some researchers calculate the base frequencies matrix  $\pi$  as the probability of being at any base on the root of the tree if the tree is rooted [7]. According to this scheme, we have

$$\pi_i P(t)_{ij} = \pi_j P(t)_{ji}. \quad (1.4)$$

Note that bases  $i$  and  $j$  can be either the base A, C, G, or T, respectively. Since substitution models use a continuous-time Markov chain to model evolutionary process, it is widely used to construct phylogenetic trees [9, 10, 11].

Due to having different rate matrices, many names for substitution models were introduced. For example, in the substitution model selection program jModelTest2, one can compare as many as 1624 possible models. This program considers in its comparison the changes on invariable sites (+I), rate variation among sites (+G), or both (+I + G) [12, 13, 14]. The simplest substitution model is the JC69 model [15]. The reason we call this model as the simplest model is because it only has one parameter, which means that it assumes equal base frequencies and equal mutation

rates. In other words, the base frequencies are

$$\pi = (\pi_A, \pi_G, \pi_C, \pi_T) = \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right), \quad (1.5)$$

and the rate matrix is [15]

$$Q = \begin{matrix} & A & G & C & T \\ \begin{matrix} A \\ G \\ C \\ T \end{matrix} & \begin{pmatrix} -\alpha & \frac{\alpha}{4} & \frac{\alpha}{4} & \frac{\alpha}{4} \\ \frac{\alpha}{4} & -\alpha & \frac{\alpha}{4} & \frac{\alpha}{4} \\ \frac{\alpha}{4} & \frac{\alpha}{4} & -\alpha & \frac{\alpha}{4} \\ \frac{\alpha}{4} & \frac{\alpha}{4} & \frac{\alpha}{4} & -\alpha \end{pmatrix} \end{matrix}. \quad (1.6)$$

Thus, a transition matrix  $P(t)$  is [15]

$$P = \begin{matrix} & A & G & C & T \\ \begin{matrix} A \\ G \\ C \\ T \end{matrix} & \begin{pmatrix} \frac{1}{4} + \frac{3}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} \\ \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} + \frac{3}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} \\ \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} + \frac{3}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} \\ \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} - \frac{1}{4}e^{-t\alpha} & \frac{1}{4} + \frac{3}{4}e^{-t\alpha} \end{pmatrix} \end{matrix}, \quad (1.7)$$

where  $t$  represents the branch lengths; the expected numbers of changes of nucleotides in each site. The meaning of branch lengths here is the expected number of changes on a branch. If the tree is ultrametric, this is proportional to the amount of time.

The generalized time reversible (GTR) model is the most widely used, which was first described by Simon Tavaré [16]. In the GTR model, the base frequency vector  $\pi$  has four different frequencies and the instantaneous rate matrix  $Q$  has six different parameters since we assume that the rate from base  $i$  to  $j$  equals the rate from base  $j$  to  $i$ , for any base A, C, G and T.

### 1.2.3 Likelihood/Bayesian method

The basic principle of the Likelihood/Bayesian method is to estimate the parameters using the likelihood function over the parameter space. In statistics, if we need to estimate parameters, we can use either the Maximum likelihood (ML) method or the Bayesian approach. The main difference between them is whether to place prior distributions on the parameters of interest or not.

Let's explain ML and Bayesian approaches from the Bayes' Rule:

$$p(\theta | D) = \frac{p(D | \theta) * p(\theta)}{p(D)}, \quad (1.8)$$

where  $D$  is the data and  $\theta$  is the parameter of interest. With the ML approach, we estimate by  $\hat{\theta}$  via maximizing the likelihood of  $p(D | \theta)$ . However, in the Bayesian approach, we treat  $\theta$  as a random variable with some distribution. By using Equation 1.8, a distribution of  $p(\theta | D)$ , posterior, will be calculated and  $\hat{\theta}$  will be estimated according to some measure for the central tendency. In Comparative Biology, it's common to use both the ML and Bayesian methods to infer a phylogenetic tree.

The first thought about inferring a phylogenetic tree was proposed by Cavalli-Sforza and Edwards [28, 17]. They used gene frequency data from human populations to estimate the branch points of a branching process, which was used to infer a phylogenetic tree. They defined a Brownian motion process to model the splitting process of genes in a specific time. The summation of Brownian motion processes would result in speciations that afterward were modeled by a Yule process between populations, where a Yule process is a birth-death process with the death rate equal to zero. At the time (in the 1960-1970), it was mathematically impossible to analyze these data without using computers. But their research raised interest on how to find the likelihood of a tree and how to infer a best fitted phylogenetic tree to given DNA

alignments.

### 1.2.3.1 Maximum Likelihood approach

Influenced by Cavalli-Sforza and Edwards' research [17], Felsenstein ignored the Yule process to simplify their calculation and first proposed maximum likelihood for estimating phylogenetic trees [18] by using discrete characters of extant species. Later in 1981, he applied the same algorithm to nucleotide sequences [9] for inferring phylogenetic trees. The process of inferring a phylogenetic tree includes three steps as follows [9]:

Step 1: Find the likelihood of a possible tree by using a substitution model;

Step 2: Compare current tree's likelihood with previous possible tree's likelihood after small changes in the branch lengths while keeping the larger one for next comparison;

Step 3: Repeat step 1-2 until all possible trees are searched.

To estimate a phylogenetic tree, denoted as  $T$ , which is defined by a specific topology and branch lengths, given the data, DNA alignments, denoted as  $D$ , and a substitution model, denoted as  $M$ , one could calculate the probability  $P(D|T, M)$ . The substitution model,  $M$ , gives values for  $P_{ij}(t)$ , the probability of going from nucleotide  $i$  to  $j$  during time  $t$  (i.e. branch lengths). In order to compute the probability  $P(D|T, M)$ , some assumptions are made [9]. We assume that different sites of DNA sequences evolve independently so that the probabilities of a given set of DNA sequences on a given tree can be computed site by site so that the product of all site probabilities provide the total likelihood of the DNA alignments. We also assume that species evolve independently in the left and the right lineages if we have a binary tree with the same substitution model in all lineages. Thus, if  $D_i$  is the vector of

bases in site  $i$ , the total likelihood is

$$P(D|T, M) = \prod_i P(D_i|T, M). \quad (1.9)$$

Below is an example about how to calculate the likelihood for a site. Figure 1.5 has three parts. Part 1 shows eight sites for four different species, denoted by  $t1$ ,  $t2$ ,  $t3$  and  $t4$ . Each column in the alignment table is a site. Part 2 shows the rooted tree with the first site colored by yellow. Part 3 shows how to calculate the likelihood given the unrooted tree, in which node 0 is suppressed. The branch lengths in part 2 and part 3 are denoted by  $v_1$ ,  $v_2$ ,  $v_3$ ,  $v_4$ ,  $v_5$ , and  $v_6$ . We denote the bases in the

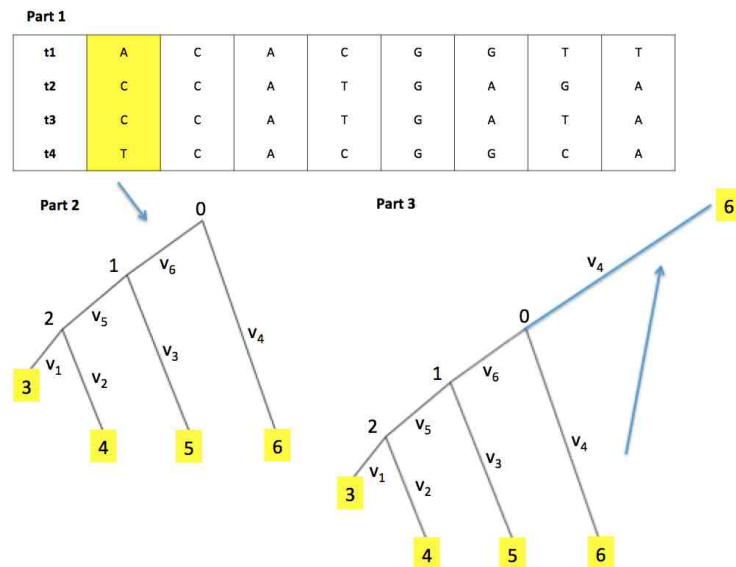


Figure 1.5: An example to show how to calculate the likelihood of a given tree in a single site.

interior nodes 0, 1 and 2, tips 3, 4, 5 and 6 as  $s_0$ ,  $s_1$ ,  $s_2$ ,  $s_3, s_4, s_5$ , and  $s_6$ , respectively.

Since we are using the site 1 as an example, the bases for  $s_3, s_4, s_5$ , and  $s_6$  are A, C,

C and T. If we assume  $s_0, s_1, s_2$  are A, C, and G, the likelihood of site 1 is

$$\begin{aligned}
P(D_1|T, M) &= P(s_3 = A, s_4 = C, s_5 = C, s_6 = T, s_0 = A, s_1 = C, s_2 = G|T, M) \\
&= P(s_0 = A)P(s_6 = T|s_0 = A, v_4) \\
&\quad P(s_1 = C|s_0 = A, v_6)P(s_5 = C|s_1 = C, v_3)P(s_2 = G|s_1 = C, v_5) \\
&\quad P(s_4 = C|s_2 = G, v_2)P(s_3 = A|s_2 = G, v_1). \tag{1.10}
\end{aligned}$$

However, the bases in the interior nodes are unknown, so the sum of the probabilities over all of the possible nucleotides that could exist should be considered. Therefore, the likelihood of any site  $i$  is

$$\begin{aligned}
P(D_i|T, M) &= \sum_{s_0} \sum_{s_1} \sum_{s_2} P(s_0, s_1, s_2, s_3, s_4, s_5, s_6|T, M) \\
&= \sum_{s_0} \sum_{s_1} \sum_{s_2} P(s_0)P(s_6|s_0, v_4) \times \\
&\quad P(s_1|s_0, v_6)P(s_5|s_1, v_3)P(s_2|s_1, v_5) \times \\
&\quad P(s_4|s_2, v_2)P(s_3|s_2, v_1) \times \\
&= \left( \sum_{s_0} P(s_0)P(s_6|s_0, v_4) \right) \times \\
&\quad \left( \sum_{s_1} P(s_1|s_0, v_6)P(s_5|s_1, v_3)P(s_2|s_1, v_5) \right) \times \\
&\quad \left( \sum_{s_2} P(s_4|s_2, v_2)P(s_3|s_2, v_1) \right). \tag{1.11}
\end{aligned}$$

In Equation 1.11, there are 64 different terms. If we increase the number of species in one tree, we will have more terms, calculated by  $2^{(n-2)}$ , which raises a computational issue if we need to calculate all possible trees' likelihoods. Thus, computationally, we need to find a better way to calculate the likelihood of a given tree. Following the approach of Felsenstein [9], we move the summation signs with

its corresponding terms together, we get

$$\begin{aligned}
P(D_i|T, M) &= \left( \sum_{s_0} \pi_{s_0} P_{s_0 s_6}(v_4) \right) \times \\
&\quad \left( \sum_{s_1} P_{s_1 s_0}(v_6) P_{s_5 s_1}(v_3) P_{s_2 s_1}(v_5) \right) \times \\
&\quad \left( \sum_{s_2} P_{s_4 s_2}(v_2) P_{s_3 s_2}(v_1) \right), \tag{1.12}
\end{aligned}$$

where  $P(s_0) = \pi_{s_0}$ , since the equilibrium matrix of base frequencies  $\pi$  is the probability of being any base on the root of the tree. Equation 1.12 shows the calculation of the likelihood for a site as a continuous-time Markov chain process.

The new equation 1.12 has the same form as the tree  $(T, ((C,C),A))$ , which gives us a faster way than our previous algorithm 1.11. If we calculate each summation separately, with small changes in any part of the tree, one can only calculate the part of the tree with different branch lengths or topologies, which will eventually make the algorithm much faster.

When considering the unrooted tree in Part 3 of Figure 1.5, one needs to recall that the evolutionary process is a continuous time Markov Chain process. According to the reversibility of this process, if we just consider the root  $s_0$  as a mutation point between  $s_1$  and  $s_6$  in any location, then the likelihood of the unrooted tree is the same as the rooted one. Therefore,

$$\begin{aligned}
P(D_i|T, M) &= \left( \sum_{s_1} \left( \sum_{s_0} P_{s_0} P_{s_0 s_6}(v_4) \right) P_{s_1 s_0}(v_6) \right) \times \\
&\quad P_{s_5 s_1}(v_3) P_{s_2 s_1}(v_5) \times \\
&\quad \left( \sum_{s_2} P_{s_4 s_2}(v_2) P_{s_3 s_2}(v_1) \right) \\
&= \left( \sum_{s_1} P_{s_1} P_{s_1 s_6}(v_4 + v_6) \right) \times \\
&\quad P_{s_1 s_5}(v_3) P_{s_2 s_1}(v_5) \times
\end{aligned}$$



$$\left(\sum_{s_2} P_{s_4 s_2}(v_2) P_{s_3 s_2}(v_1)\right). \quad (1.13)$$

In order to measure the reliability of the estimated phylogenetic tree, Felsenstein [19] proposed a nonparametric bootstrapping method to show the proportion of bootstrap replicates for each clade inside the estimated tree. First, he considers every site as a unit inside the data, then a sample size of  $n$  DNA sites is drawn from the  $N$  DNA sites with replacement; second, he uses the ML method to infer a bootstrap tree. Repeat the two steps a large number of times (typically 100 to 1000) and construct a 50% majority-rule consensus tree, including only those clades found in a majority ( $\geq 50\%$ ) of the simulated trees. In this overall estimated bootstrap phylogenetic tree, above the line describing each branch, the frequency of such clade is shown in the simulated bootstrap tree.

### 1.2.3.2 Bayesian approach

After Felsenstein's [9] proposal for the ML method to infer the phylogenetic tree by DNA alignments, lots of computational simulations have shown the ML method is better than the Parsimony method and Distance method for inferring trees in various biological situations [3, 4]. Therefore the ML method is widely used. However, there are two main problems that the ML method is suffering from when it's used in recovering a phylogenetic tree. First, the process of finding MLEs of the branch lengths and topology involves having different likelihood functions for different topologies. The changeable likelihood function brings complications in estimating the parameters of interest and let people wonder whether the estimators satisfy the properties of the ML method in general [3, 20]. Second, it's hard to estimate the accuracy of the estimated tree by the ML method. A nonparametric bootstrapping method to estimate

the accuracy was proposed by Felsenstein in 1985 [19]. However, this method can not give a reliable result under some situations [22, 23, 21, 24]. Consequently, Rannala and Yang proposed an empirical Bayesian analysis for inferring phylogenetic trees in 1996 [7] by using a birth-death process as a prior distribution for one phylogenetic tree with no more than five extant species. Later, this method was improved to a hierarchical Bayesian analysis to include larger numbers of species [25]. The differences between those two Bayesian analysis methods are: (1) In an empirical Bayesian analysis, the parameters are initially estimated by the MLEs and then used as the true parameters to evaluate the posterior probabilities of the trees; (2) In a hierarchical Bayesian analysis, the unknown parameters are assigned priors.

According to Rannala and Yang's paper [25] in 1997, a linear birth-death process could be used to model the dynamic evolutionary process by birth rate  $\lambda$  and death rate  $\mu$  with a prior for the topology  $\tau$  and set of branch lengths  $t$ . They used a Metropolis-Hastings algorithm to infer a phylogenetic tree. The steps of this method are as follows:

Step 1: Define uniform priors on  $\lambda$  and  $\mu$ .

Use the ML method to infer a phylogenetic tree, then estimate speciation times assuming the distance from root to tips is 1. Treat the estimated speciation times as observations to estimate the birth rate  $\lambda$  and death rate  $\mu$  that are used as the means of the uniform priors of birth rate and death rate.

Step 2: Use the DNA alignments to estimate the substitution model with substitution rate matrix  $Q$ .

Step 3: Start from a possible tree topology and calculate the acceptance rate by

$$\alpha_{ij} = \min\left(\frac{\pi_j}{\pi_i}, 1\right)$$

$$= \min \left( \frac{f(X|\tau = j)}{f(X|\tau = i)}, 1 \right), \quad (1.14)$$

where  $\alpha_{ij}$  is the probability of changing from time state  $i$  to time state  $j$  under reversible Markov chains only, and  $X$  is the DNA alignment. If  $\alpha_{ij} = 1$ , the possible tree topology is accepted to the posterior distribution of trees. In the Equation 1.14,  $\pi_i$  states the posterior distribution of phylogenetic trees in speciation time  $i$  given the chosen tree topology and it can be calculated by Monte Carlo integration with the substitution model;

Step 4: Repeat step 3 by searching all possible tree topologies with different combinations of extant and extinct species to find the best tree with the highest posterior probabilities.

In the same paper, the very same phylogenetic tree was found by either the ML method or Bayesian method. They also found that the best tree had a total posterior probability of about 95% to 96%. Kishino and Hasegava [26] used the same data and found that the bootstrap proportion of the best tree was 68%. The Bayesian approach tends to give higher posterior probability of the best tree compared to bootstrapping support used with the ML method. However, it is really hard to judge which method is better. Researchers are still working on new ideas to improve the accuracy of inferring a better phylogenetic trees. For example, some researchers think that if traits and DNA alignments are both considered as inputs, more accurate phylogenetic trees might be inferred which was demonstrated in the iBPP program [2].

### 1.3 Modeling evolutionary changes through traits

In comparative biology, researchers hope that with a phylogeny, the association between traits and their environments could reconstruct the evolutionary history. The question is how could we figure out the distribution of one trait values while these values are changing through the evolutionary processes. In 1985, Felsenstein used a stochastic process, Brownian motion (BM), to model the changes of quantitative traits' values over time [27], a work which was motivated by Edwards and Cavalli-Sforza [28]. Edwards and Cavalli-Sforza assumed that genetic drift of gene frequencies followed approximately a Brownian motion on some scale if the changes on gene frequencies were not much. The three researchers' work opened a new view of using a tractable probabilistic model for the evolution of quantitative traits.

#### 1.3.1 Brownian motion (BM) model

Brownian motion is named after Robert Brown, a botanist, in 1827, which was used to describe particles' random movements inside gas or liquid [29, 30]. The mathematician Norbert Wiener used it as a stochastic process to model any phenomena with independent changes according to disjoint time periods [31], which is also the reason that BM has been used to model traits' evolutionary processes. Here is the mathematical definition of BM [32]:

A scalar Brownian motion, or Wiener process, over  $[0, T]$  is a random variable  $W(t)$  that depends continuously on  $t \in [0, T]$  and satisfies the following three conditions.

1.  $W(t) = 0$  (with probability 1).
2. For  $0 \leq s \leq t \leq T$ , the random variable given by the increment  $W(t) - W(s)$  is

normally distributed with mean zero and variance  $(t - s)\sigma^2$ .

3. For  $0 \leq s \leq t \leq u \leq v \leq T$ , the increments  $W(t) - W(s)$  and  $W(u) - W(v)$  are independent.

With three conditions above, we have

$$W_j = W_{j-1} + dW_j, \quad (1.15)$$

where  $j = 1, 2, \dots, N$  with  $W_0 = 0$  with probability 1. Each  $dW_j$  is an independently normally distributed with mean zero and variance  $(t - s)\sigma^2$ . Therefore, as an evolutionary trait model, we write the BM model as

$$dX(t) = \sigma dB(t), \quad (1.16)$$

where time  $t$  is represented as branch lengths (i.e. edge weights in the graph of the tree) and variance,  $\sigma^2$ , is proportional to time or branch lengths ( $\sigma^2$  is also called as the rate of evolution) . Based on the definition of BM, the trait value on the root of the tree is  $X(0) = 0$  with probability 1; the change of the trait value during a  $t$  time period  $X(t)$  is a continuous random variable; the distribution of the change between trait values between any disjoint time period,  $X(t) - X(s) \sim N(0, (t - s)\sigma^2)$ . Therefore, this model is a purely neutral model with constant rate of evolution in any time period. Note that a purely neutral model [33] means that the most changes within and between species are not influenced by natural selection but by the changes of the frequencies of four nucleotides in genes of the species' populations due to randomization in the evolutionary processes. The mutations that happen in DNA sequences are neither deleterious nor beneficial, thus are assumed to be neutral. Also the BM model offers the advantage that the process of evolution could be modeled probabilistically with normal distributions.

One of the most important concerns about the BM model is whether the evolutionary process follows the BM process [34, 35, 36, 37]. How far is this model away from the real biological world? For example, the rates of evolution are not stable in different time periods, different types of selection exist (i.e. natural selection), and therefore the random walk stochastic process in the BM model can not fully describe the trait adaptation. To solve this issue, multiple transformation methods are used to transform the branch lengths to fit the assumptions of the BM model. Three branch length transformations,  $\lambda$ ,  $\delta$ , and  $\kappa$  has been proposed by Mark Pagel [34, 35, 36].

In the  $\lambda$  transformation,  $\lambda$  presents a measurement of the phylogenetic signal with range from zero to one. This transformation tries to increase or decrease the degree of dependency in evolutionary processes. When  $\lambda = 0$ , it means that there is not any phylogenetic signal and all traits evolve independently; when  $0 \leq \lambda \leq 1$ , the phylogenetic signal increases with more dependent trait evolution; when  $\lambda = 1$ , the model is a BM model with unchanged branch lengths.

In the  $\delta$  transformation,  $\delta$  presents the increased or decreased rates of traits' changes through time. When  $\delta \leq 1$ , the trait change tends to slow through time; when  $\delta \geq 1$ , the rates are tending to increase through time; when  $\delta = 0$ , the model is a BM model.

In the  $\kappa$  transformation,  $\kappa$  is used to measure the location of a trait mutations along branch lengths. When  $\kappa = 0$ , all branch lengths are equal and the trait evolution happens at the speciation point; when  $\kappa = 1$ , the evolution model is a BM model; when  $0 \leq \kappa \leq 1$ , the trait evolution happens more often on longer branch lengths.

### 1.3.2 Ornstein-Uhlenbeck (OU) model

In 1988, Felsenstein [37] stated that even though the BM model is a reasonable model of pure selection and branch length transformations could be done to fit the assumptions, the BM model is still too simple to fit the trait evolutionary processes, and a more realistic model would be some adaptive peak model, which meant that the motion of a population attempts to reach its maximum by wandering back and forth through time with the natural selection under the influence of genetic drift. Therefore, he proposed that the Ornstein-Uhlenbeck (OU) model would be a more realistic model to consider. In mathematics, the Ornstein-Uhlenbeck process [38], named after Leonard Ornstein and George Eugene Uhlenbeck, is a stochastic process that describes the velocity of a massive Brownian particle under the influence of friction with the expected change toward the peak point at a rate proportional to the distance between current point and the peak point. Later, Garland [39] simulated traits' values under Brownian motion, and Ornstein-Uhlenbeck evolutionary models to test the differences of home range areas between two clades (i.e., groups with common ancestors) of mammals.

Tracing back to the use of mathematical models on evolutionary adaptation, Hansen in 1997 [40] stated that the adaptation could happen in any trait through selection and he thought the idea of adaptation could model the evolutionary history in a more realistic way. He used the approach that he and Martins [41] proposed to transform the microevolutionary changes to a statistical variance-covariance matrix between species guided by a phylogenetic tree. At the same time, he fixed the rate of evolution processes and separated the whole history to different regimes with different selective factors and corresponding trait optima in order to make a more realistic

biological situation. Butler and King in 2004 improved the Hansen's model to the Ornstein-Uhlenbeck (OU) Model, which was a more generalized model of stabilizing selection [42] by combining original OU model and evolutionary adaption together. In this method, the adaptive trait evolves by the strength of selection either positively or negatively toward an optimum value through time. In other words, the trait wanders around an optimum following a BM process with selection force which tries to pull the trait back to the optimum position. When the selection force is zero, the OU model is equal to the BM model.

Mathematically, under the same assumptions of the BM model we mentioned above, Butler and King stated the OU model as [42]

$$dX(t) = \alpha[\theta - X(t)]dt + \sigma dB(t), \quad (1.17)$$

where  $\alpha$  is the force or strength of evolutionary selection,  $\theta$  is the optimum trait value, and  $\sigma^2$  is the rate of a BM model. Unlike a BM model with constant rate of evolutionary, an OU model is able to constrain the variation during evolution. This property is a very useful feature of an OU model which allows many interesting traits to be limited in their evolutionary changes. Another advantage is that an OU model lets the accumulative history of traits' values lose their natural tendency as they wander around the adaptive optimum, which also makes the traits for species closely related tend to have similar values. Therefore, the OU model is a more realistic model to describe the evolutionary process under the natural selection. However, since 1985, one concern has been raised about whether one tree can give the correct direction of evolution for a specific trait. In the real biology world, it is very rare to find only one gene that influences a trait. For example, rice seed size determines the rice yield. As an important agronomic trait, it is related to more than one gene, such as the genes



GS3, GW2 and qSW5/GW5 [43, 44]. Therefore, to find a model with more than one tree is necessary.

## 1.4 The Statistical comparative methods in evolutionary biology

The idea of comparison among different species is ancient since Aristotle used the shared characters of species as an example to explain the meaning of classification [45]. The modern comparative methods started on 1983 by the publication of Ridley's book named *The Explanation of Organic Diversity : The Comparative Method and Adaptations for Mating* [46], which follows the same track of comparison to infer more information of the history of biology. For example, one of the most popular questions to ask in biology is how species and environmental factors are influenced by each other. To answer this question by using comparative methods, we have three statistical approaches: the phylogenetic independent contrasts (PIC) [27], the generalized least-squares (GLS) models [49, 50], and Monte Carlo computer simulations [51]. These methods are mainly applied to quantitative traits, such as organisms' weights, the sizes of plants' seeds and so on.

A little after Ridley's book, Felsenstein stated the same idea of modern comparative methods in his article entitled *Phylogenies and the Comparative Method* [27]. Both Ridley and Felsenstein argued that the main difference between modern comparative methods and earlier methods was combining the information from phylogenies and comparative data together. The reason for that is if the quantitative traits are from the same taxon, they share a common phylogenetic history and therefore they

are not independent. If the data are not independent, it is really hard to apply to statistical methods, such as regression, AVOVA and so on, since it violates one of the basic assumptions for majority of basic statistical approaches [52]. In Felsenstein's paper in 1985, as we mentioned above, the first fully phylogenetic comparative method was proposed, which was named as Independent Contrasts by him.

We can use the standard linear regression model to test the significant relationships between different traits in a specific group of species by using the formula:

$$Y = X\beta + \epsilon. \quad (1.18)$$

In Equation 1.18, the matrix  $Y$  contains the phenotypic measurements of extant taxa, the matrix  $X$  may contain some other phenotypic measurements of extant taxa unlike in matrix  $Y$  or measurements from environmental factors and the error term matrix is  $\epsilon$ , where

$$\epsilon \sim N(0, \sigma^2). \quad (1.19)$$

However, there is a big issue in comparative biology. If the traits are from the same taxon, they share a common phylogenetic history and therefore they are not independent. We can not use nonindependent comparative data to do regression analysis since it violates one of the basic assumptions. If we look at Equation 1.18 closely, there are two parts included in the formula: the mean term and the error term. Biologists consider that phylogenetic signals can be present in either one of the two terms. Because of that, two methods have been developed. One method is to transform measured values of traits into independent contrasts under the assumption that the phenotypic evolutionary process can be described by a BM model, which is called the phylogenetical independent contrasts approach (PIC) [27]. Another method is to allow nonindependent comparative data to be fitted by a linear regression model

with a variance-covariance matrix provided by its phylogeny, assuming that the phylogenetic process of trait evolution is a BM process evolving on the tree. So that the phylogenetic signals and environmental factors are translated into a complicated error matrix with multivariate normal distributions, such as in the general least squares approach (GLS) [49, 50].

### 1.4.1 GLS method

One trait values are denoted as a column vector  $Y$  with  $n$  rows, where  $n$  is the numbers of tips in the gene tree. Assume the model of trait evolution is a BM model. According to Felsenstein's 1985 paper [27], the trait values' variance is proportional to the time or branch length when it evolves. A variance-covariance matrix, denoted as  $V$  is needed to construct the variance-covariance matrix for the matrix  $Y$ , which can be calculated by the gene tree. Felsenstein explained variance-covariance matrix  $V$  in chapter 23 in his book *Inferring phylogenies* [47]. The variance-covariance matrix  $V$  shows the relationships between different species under a given gene tree, i.e. whether two species are dependent or independent, therefore the elements in  $V$  matrix are the branch lengths shared by the paths from the root down to the tips. If 0s are shown in the matrix, they represent no shared histories between those species and the corresponding trait values are independent. If the rate of evolution of the trait matrix  $Y$  per unit time  $t$  is  $\sigma^2$  and the root trait value is  $\mu_y$  with the variance-covariance matrix  $V$ , then the likelihood of the matrix  $Y$  given the data under the GLS method [48] is

$$L = (2\pi)^{-\frac{n}{2}} |\sigma^2 V|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (Y - X\mu_y)' (\sigma^2 V)^{-1} (Y - X\mu_y) \right], \quad (1.20)$$

and the log-likelihood is

$$\ell = -\frac{1}{2}[n \log(2\pi) + \log|\sigma^2 V| + (Y - X\mu_y)'(\sigma^2 V)^{-1}(Y - X\mu_y)], \quad (1.21)$$

with the maximum likelihood estimators

$$\hat{\mu}_y = (X'V^{-1}X)^{-1}(X'V^{-1}Y) \quad (1.22)$$

and

$$\hat{\sigma}^2 = \frac{1}{n}(Y - X\hat{\mu}_y)'V^{-1}(Y - X\hat{\mu}_y) \quad (1.23)$$

where  $X$  is the design matrix with all elements equal to 1 since we consider just one trait here.

### 1.4.2 PIC method

Under the PIC method, the process is more complicated. First all of the trait's values are transformed to their corresponding contrasts, denoted as  $Y$ , which are independent to each other; after that, the new branch lengths corresponding to contrasts are calculated. These new branch lengths could be used to construct the variance-covariance matrix, denoted as  $V_y$ . Felsenstein in 1985 [27] explained how to calculate the contrasts and new branch lengths, the steps are as follows:

Step 1: Start from a pair of adjacent tips, denoted as  $i$  and  $j$  with corresponding trait' values  $y_i$  and  $y_j$ , respectively. Their contrast  $u_{ij} = y_i - y_j$  is computed with

$$u_{ij} \sim N(0, \sigma^2(v_i + v_j)), \quad (1.24)$$

where  $v_i$  and  $v_j$  are the branch lengths from tip  $i$  and tip  $j$  to their MRCA, denoted as  $k$ , respectively.

Step 2: Calculate the ancestor  $k$  trait value

$$x_k = \frac{\left(\frac{1}{v_i}x_i + \frac{1}{v_j}x_j\right)}{\frac{1}{v_i} + \frac{1}{v_j}}, \quad (1.25)$$

which means that the ancestor  $k$  trait value is the weighted mean of the trait values from two descendent species, using branch lengths  $v_i$  and  $v_j$  to determine the weights.

Step 3: Remove tip  $i$  and  $j$  and let MRCA  $k$ , the MRCA of  $i$  and  $j$ , become a new tip with an adjusted branch lengths, denoted as  $v'_k$ , where

$$v'_k = v_k + \frac{v_i v_j}{v_i + v_j}. \quad (1.26)$$

Step 4: Repeat step 1 - 3 until all tips are removed in the tree. The root will have a zero contrast with a new branch length.

Based on the steps above, we will have  $n$  contrasts with the root contrast value zero. Then the likelihood of the column vector  $U$ , independent contrasts, given the data as a function of  $\sigma^2$  and  $V_u$  is

$$L = (2\pi)^{-\frac{n}{2}} |\sigma^2 V_u|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (U'(\sigma^2 V_u)^{-1} U) \right] \quad (1.27)$$

and the log-likelihood is

$$\ell = -\frac{1}{2} [n \log(2\pi) + \log |\sigma^2 V_u| + (U'(\sigma^2 V_u)^{-1} U)] \quad (1.28)$$

with the maximum likelihood estimator

$$\hat{\sigma}^2 = \frac{1}{n} U' V_u^{-1} U. \quad (1.29)$$

### 1.4.3 Application of PIC and GLS methods

Below is an example about how to use PIC and GLS methods to calculate the likelihoods given the same trait values with the same tree. In Figure 1.6, there are 4 parts.

Part 1 shows the trait values. Part 2 is the tree we have with four tips,  $t_1$ ,  $t_2$ ,  $t_3$  and  $t_4$ . Branch lengths are shown in the figure. Part 3 shows the ancestors' trait values and part 4 shows the corresponding contrasts values without showing the root value zero. Figure 1.7 shows the steps to calculate the new branch lengths for the contrasts based on Freckleton's paper in 2012 [53].

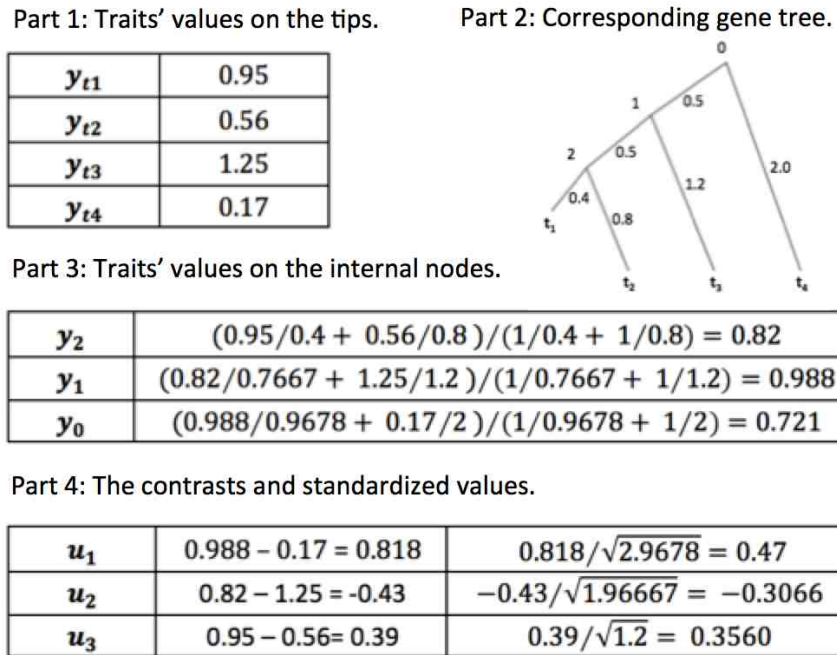


Figure 1.6: An example of PIC.

So based on Figure 1.6 and Figure 1.7 by the PIC method, we have

$$\begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.6522 & 0 & 0 & 0 \\ 0 & 2.9678 & 0 & 0 \\ 0 & 0 & 1.9667 & 0 \\ 0 & 0 & 0 & 1.200 \end{pmatrix} \sigma^2 \right].$$

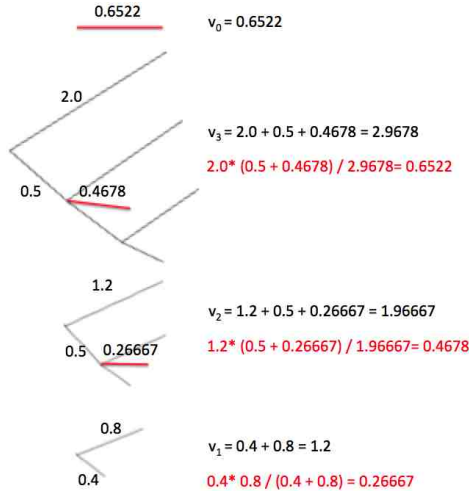


Figure 1.7: An example of how to compute the branch lengths for contrasts.

Therefore, the rate estimator is

$$\hat{\sigma}^2 = \frac{1}{4} \begin{pmatrix} 0 \\ 0.818 \\ -0.43 \\ 0.39 \end{pmatrix}' \begin{pmatrix} 0.6522 & 0 & 0 & 0 \\ 0 & 2.9678 & 0 & 0 \\ 0 & 0 & 1.9667 & 0 \\ 0 & 0 & 0 & 1.200 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0.818 \\ -0.43 \\ 0.39 \end{pmatrix} = 0.1115$$

and the log-likelihood is

$$\begin{aligned} \ell = & -\frac{1}{2} [4 \log(2\pi) + \log|\sigma^2| \begin{pmatrix} 0.6522 & 0 & 0 & 0 \\ 0 & 2.9678 & 0 & 0 \\ 0 & 0 & 1.9667 & 0 \\ 0 & 0 & 0 & 1.200 \end{pmatrix}|] \\ & + \begin{pmatrix} 0 \\ 0.818 \\ -0.43 \\ 0.39 \end{pmatrix}' (\sigma^2 \begin{pmatrix} 0.6522 & 0 & 0 & 0 \\ 0 & 2.9678 & 0 & 0 \\ 0 & 0 & 1.9667 & 0 \\ 0 & 0 & 0 & 1.200 \end{pmatrix})^{-1} \begin{pmatrix} 0 \\ 0.818 \\ -0.43 \\ 0.39 \end{pmatrix} = -2.05. \end{aligned} \quad (1.30)$$

Using the same trait values in Figure 1.6 part 1 and the same gene tree in part 2, under the GLS method, we have

$$\begin{pmatrix} y_{t_1} \\ y_{t_2} \\ y_{t_3} \\ y_{t_4} \end{pmatrix} \sim N \left[ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu_y, \begin{pmatrix} 1.4 & 1.0 & 0.5 & 0 \\ 1.0 & 1.8 & 0.5 & 0 \\ 0.5 & 0.5 & 1.7 & 0 \\ 0 & 0 & 0 & 2.0 \end{pmatrix} \sigma^2 \right].$$

Therefore, the trait value on the root is

$$\hat{\mu}_y = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}' \begin{pmatrix} 1.4 & 1.0 & 0.5 & 0 \\ 1.0 & 1.8 & 0.5 & 0 \\ 0.5 & 0.5 & 1.7 & 0 \\ 0 & 0 & 0 & 2.0 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}' \begin{pmatrix} 1.4 & 1.0 & 0.5 & 0 \\ 1.0 & 1.8 & 0.5 & 0 \\ 0.5 & 0.5 & 1.7 & 0 \\ 0 & 0 & 0 & 2.0 \end{pmatrix}^{-1} \begin{pmatrix} 0.95 \\ 0.56 \\ 1.25 \\ 0.17 \end{pmatrix} = 0.721$$

and the rate of the trait evolution is

$$\hat{\sigma}^2 = \frac{1}{4} \left( \begin{pmatrix} 0.95 \\ 0.56 \\ 1.25 \\ 0.17 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu_y \right)' \begin{pmatrix} 1.4 & 1.0 & 0.5 & 0 \\ 1.0 & 1.8 & 0.5 & 0 \\ 0.5 & 0.5 & 1.7 & 0 \\ 0 & 0 & 0 & 2.0 \end{pmatrix}^{-1} \left( \begin{pmatrix} 0.95 \\ 0.56 \\ 1.25 \\ 0.17 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu_y \right) = 0.1115.$$

So the log-likelihood is

$$\begin{aligned} \ell = & -\frac{1}{2} [4 \log(2\pi) + \log|\sigma^2| \\ & + \left( \begin{pmatrix} 0.95 \\ 0.56 \\ 1.25 \\ 0.17 \end{pmatrix} - \mu_y \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right)' (\sigma^2 \begin{pmatrix} 1.4 & 1.0 & 0.5 & 0 \\ 1.0 & 1.8 & 0.5 & 0 \\ 0.5 & 0.5 & 1.7 & 0 \\ 0 & 0 & 0 & 2.0 \end{pmatrix})^{-1} \left( \begin{pmatrix} 0.95 \\ 0.56 \\ 1.25 \\ 0.17 \end{pmatrix} - \mu_y \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right)] = -2.05. \end{aligned} \quad (1.31)$$

In conclusion, the example above illustrates that there is no difference between PIC and GLS methods even though both methods treat the dependent trait data in different ways [54, 55].

In current comparative biology, simulation is widely used to test new models and hypotheses. Martins and Garland [51] were the first group of researchers who proposed the idea of Monte Carlo computer simulations in this area and used computer simulations to create different circumstances and to test different phylogenetic relationships, such as correlations between two traits, differences between two sets of species. This method can be combined with the PIC method or the GLS method to test lots of different research interests, such as whether a trait evolutionary model with two gene trees is better than one gene tree in my research.



## 1.5 Summary of background

Under current comparative methods [27, 49, 50], such as the BM model, when species evolve, traits evolve with respect to one gene tree and a complicated environment; however, it is rare that a trait depends on one gene in general. Our research uses a simulation study to show that there is a significant error in the inference of the parameters of the BM model when just one gene is responsible for a trait. A new trait model is presented which allows multiple genes with different contribution weights correspond to a trait. We hope through simulations to show that the new trait evolutionary model is a significant improvement over the current one tree model.

# Chapter 2

## Two gene trees BM model

### 2.1 Two gene trees BM model with one trait

#### 2.1.1 Assumptions of two gene trees BM model

Current comparative methods only consider the use of one phylogenetic tree for modeling the relationships between species. To add biological realism, we wish to develop a new trait model considering two gene trees with different contribution weights from the two genes. Our assumptions are as follows:

1. For each gene, branch lengths and topology are known. The two genes are independent.
2. We do not know the trait values corresponding to each gene, but the trait value for each species is known. Here we are assuming that trait values are the summation,  $\mathbf{Y}$ , of two individual traits values from the two genes.  $\mathbf{Y}$  is assumed to have a multivariate normal distribution, such that

$$\mathbf{Y} \sim N(\mathbf{0}, \Sigma_y), \tag{2.1}$$

where  $\Sigma_y$  is the variance-covariance matrix and the trait values are centered.

3. The trait evolutionary process in each gene tree follows a BM model.
4. Each trait evolves independently in different lineages and the trait value at the root of a given phylogenetic tree is zero. The rate of evolution in each gene is the same.

In this section, we only discuss the one trait evolutionary processes that follows the BM model. The trait values on all tips, denoted by  $\mathbf{Y}$ , are made up from  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ , the trait values in each gene tree. Since the sum of trait values on all tips from gene tree 1 (*GT1*) and gene tree 2 (*GT2*) are fixed, we have

$$\mathbf{Y} = p * \mathbf{Y}_1 + (1 - p) * \mathbf{Y}_2, \quad (2.2)$$

where  $p$  is the contribution weight from *GT1* and  $1 - p$  is the contribution weight from *GT2*.  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are independent since they evolve through different genes under the BM model with the variance-covariance matrix from *GT1*

$$\Sigma_{y1} = \sigma^2 V_{y1} \quad (2.3)$$

and from *GT2*

$$\Sigma_{y2} = \sigma^2 V_{y2}, \quad (2.4)$$

where  $\sigma^2$  is the rate of evolution of the trait of interest. Both  $p$  and  $\sigma^2$  are parameters to be estimated. Note that  $V_{y1}$  and  $V_{y2}$  can be defined from *GT1* and *GT2* and they show the relationships between species in each gene tree. Therefore, the joint variance-covariance matrix is

$$\begin{aligned} \Sigma_y &= p^2 \Sigma_{y1} + (1 - p)^2 \Sigma_{y2} \\ &= p^2 \sigma^2 V_{y1} + (1 - p)^2 \sigma^2 V_{y2} \end{aligned}$$

$$\begin{aligned}
&= \sigma^2(p^2V_{y1} + (1 - p)^2V_{y2}) \\
&= \sigma^2V_y,
\end{aligned} \tag{2.5}$$

where  $V_y$  is the joint matrix from  $GT1$  and  $GT2$  and it contains the  $p$  parameter of interest. The variance-covariance matrix  $V$  shows the relationships between different species under a given gene tree, i.e. whether two species are dependent or independent. If 0s are shown in the matrix, they represent no shared history between those species, and the corresponding trait values are deemed independent.

Below is an example for calculating  $V_{y1}$ ,  $V_{y2}$  and the joint matrix  $V_y$ . Figure 2.1 shows two gene trees,  $GT1$  and  $GT2$ , that are used to calculate the variance-covariance matrix.

Let  $x_{ijk}$  denote the trait value for the  $i$ th gene,  $j$ th node species, and  $k$ th trait, where nodes  $1, \dots, n$  represent extant species and nodes with labels greater than  $n$  represent ancestral nodes. Under gene tree 1 in Figure 2.1, we have four trait values from  $GT1$ ,  $x_{11k}$ ,  $x_{12k}$ ,  $x_{13k}$ , and  $x_{14k}$  on the tips  $A$ ,  $B$ ,  $C$  and  $D$ , and two ancestral values  $x_{15k}$  and  $x_{16k}$  on the nodes,  $E$  and  $F$  with different branch lengths. Therefore,

$$V_{y1} = \begin{matrix} & A & B & C & D \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{pmatrix} v_{11} + v_{13} & v_{13} & 0 & 0 \\ v_{13} & v_{11} + v_{13} & 0 & 0 \\ 0 & 0 & v_{12} + v_{14} & v_{14} \\ 0 & 0 & v_{14} & v_{12} + v_{14} \end{pmatrix} \end{matrix}. \tag{2.6}$$

Under gene tree 2 in Figure 2.1, we have four trait values from  $GT2$ ,  $x_{21k}$ ,  $x_{22k}$ ,  $x_{23k}$ , and  $x_{24k}$  on the tips  $A$ ,  $C$ ,  $B$  and  $D$ , and two ancestral values  $x_{25k}$  and  $x_{26k}$  on

the nodes,  $E$  and  $F$  with different branch lengths. Therefore,

$$V_{y2} = \begin{matrix} & A & B & C & D \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{pmatrix} v_{21} + v_{23} & 0 & v_{23} & 0 \\ 0 & v_{22} + v_{24} & 0 & v_{24} \\ v_{23} & 0 & v_{21} + v_{23} & 0 \\ 0 & v_{24} & 0 & v_{22} + v_{24} \end{pmatrix} \end{matrix}. \quad (2.7)$$

Hence, the joint  $V_y$  matrix is

$$V_y = \begin{matrix} & A & B & C & D \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{pmatrix} \beta_1 & p^2 v_{13} & (1-p)^2 v_{23} & 0 \\ p^2 v_{13} & \beta_2 & 0 & (1-p)^2 v_{24} \\ (1-p)^2 v_{23} & 0 & \beta_3 & p^2 v_{14} \\ 0 & (1-p)^2 v_{24} & p^2 v_{14} & \beta_4 \end{pmatrix} \end{matrix}. \quad (2.8)$$

where

$$\beta_1 = p^2(v_{11} + v_{13}) + (1-p)^2(v_{21} + v_{23}),$$

$$\beta_2 = p^2(v_{11} + v_{13}) + (1-p)^2(v_{22} + v_{24}),$$

$$\beta_3 = p^2(v_{12} + v_{14}) + (1-p)^2(v_{21} + v_{23}),$$

$$\beta_4 = p^2(v_{12} + v_{14}) + (1-p)^2(v_{22} + v_{24}).$$

So under model 2.2, there are two parameters  $p$  and  $\sigma^2$  that we need to estimate.

Here we will use two approaches: (1) the adaptive Metropolis-Hastings approach, and

(2) approximate Bayesian computation method to do the estimation.

### 2.1.2 Adaptive Metropolis-Hastings approach

The adaptive Metropolis-Hastings (AMH) algorithm [63, 64] has been used to sample hierarchical models. In this algorithm, the choice of the adaptive proposal distribu-

tions is important for updating the proposal distributions and increasing the efficiency of the sampling process. In our first approach, we use the AMH algorithm to estimate the variance  $\sigma^2$  and the weight contribution  $p$  from the two gene trees. Our data are two different gene trees and a vector of one trait values on all tips with a multivariate normal distribution. For this chapter, we model only one trait, so there is one trait measurement per extant species. From two different gene trees, one can calculate the joint matrix  $V_y$  which includes the two parameters of interest.

Assume two gene trees each with  $n$  species, our model with one trait is

$$\mathbf{Y} = \begin{pmatrix} y_{t_1} \\ y_{t_2} \\ y_{t_3} \\ \vdots \\ y_{t_n} \end{pmatrix} \sim N(\mathbf{0}, \Sigma_y),$$

where

$$\Sigma_y = \sigma^2(p^2 V_{y1} + (1-p)^2 V_{y2}) \quad (2.9)$$

with priors

$$p \sim \text{Beta}(a, b), \quad (2.10)$$

and

$$\sigma \sim \text{Folded - noncentral - } t(df = c, \text{scale} = d), \quad (2.11)$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  are positive real numbers and the parameters  $p$  and  $\sigma^2$  are independent. The posterior distribution of the contribution weight  $p$  and the variance  $\sigma$  under the BM model is

$$f(p, \sigma | \mathbf{Y}, V_{y1}, V_{y2}) = f(\mathbf{Y} | p, \sigma, V_{y1}, V_{y2}) f(p) f(\sigma)$$

$$\begin{aligned}
&= (2\pi)^{\frac{n}{2}} |\sigma^2(p^2V_{y1} + (1-p)^2V_{y2})|^{-\frac{1}{2}} \times \\
&\quad \exp\left(-\frac{1}{2}\mathbf{Y}'(\sigma^2(p^2V_{y1} + (1-p)^2V_{y2}))^{-1}\mathbf{Y}\right) \times \\
&\quad \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \frac{\Gamma(\frac{c+1}{2})}{\sqrt{c\pi}\Gamma(\frac{c}{2})} \frac{1}{d} \left[1 + \frac{1}{c} \frac{\sigma^2}{d^2}\right]^{-\frac{c+1}{2}}.
\end{aligned}$$

We transform  $p$  and  $\sigma^2$  in order to use the normal distribution for adaptive sampling.

Let

$$w = \log\left(\frac{p}{1-p}\right), \quad (2.12)$$

and

$$z = \log \sigma^2. \quad (2.13)$$

We get

$$p = \frac{e^w}{1+e^w}, \quad (2.14)$$

and

$$\sigma^2 = e^z. \quad (2.15)$$

So the Jacobian of the transformation is

$$\begin{aligned}
J &= \begin{vmatrix} \frac{dp}{dw} & \frac{dp}{dz} \\ \frac{d\sigma^2}{dw} & \frac{d\sigma^2}{dz} \end{vmatrix} \\
&= \begin{vmatrix} \frac{d\frac{e^w}{1+e^w}}{dw} & \frac{d\frac{e^w}{1+e^w}}{dz} \\ \frac{de^z}{dw} & \frac{de^z}{dz} \end{vmatrix} \\
&= \begin{vmatrix} \frac{e^w}{(1+e^w)^2} & 0 \\ 0 & e^z \end{vmatrix} = \frac{e^w}{(1+e^w)^2} e^z
\end{aligned}$$

where

$$\begin{aligned}
\frac{d\frac{e^w}{1+e^w}}{dw} &= \frac{(1+e^w)(e^w)' - (1+e^w)'e^w}{(1+e^w)^2} \\
&= \frac{(1+e^w)e^w - e^w e^w}{(1+e^w)^2}
\end{aligned}$$

$$= \frac{e^w}{(1 + e^w)^2}. \quad (2.16)$$

In terms of  $w$  and  $z$ , the posterior distribution is

$$\begin{aligned} f(p, \sigma | \mathbf{Y}, V_{y1}, V_{y2}) &= f(w, z | \mathbf{Y}, V_{y1}, V_{y2}) |J| \\ &= (2\pi)^{\frac{n}{2}} |e^z \left( \left( \frac{e^w}{1 + e^w} \right)^2 V_{y1} + \left( 1 - \frac{e^w}{1 + e^w} \right)^2 V_{y2} \right)|^{-\frac{1}{2}} \times \\ &\quad \exp\left(-\frac{1}{2} \mathbf{Y}' \left( e^z \left( \left( \frac{e^w}{1 + e^w} \right)^2 V_{y1} + \left( 1 - \frac{e^w}{1 + e^w} \right)^2 V_{y2} \right) \right)^{-1} \mathbf{Y}\right) \times \\ &\quad \frac{1}{B(a, b)} \left( \frac{e^w}{1 + e^w} \right)^{a-1} \left( 1 - \frac{e^w}{1 + e^w} \right)^{b-1} \times \\ &\quad \frac{\Gamma(\frac{c+1}{2})}{\sqrt{c\pi}\Gamma(\frac{c}{2})} \frac{1}{d} \left[ 1 + \frac{1}{c} \frac{e^z}{d^2} \right]^{-\frac{c+1}{2}} \frac{e^w}{(1 + e^w)^2} e^z. \end{aligned} \quad (2.17)$$

We use R [65] to program and run the AMH. The procedure is as follows:

Step 1: Give the priors.

Step 2: Let  $T = 200,000$ . When  $t = 1$  and  $2$ , we give some relatively small values for  $w$  and  $z$  to start the sampling.

For  $t = 3, \dots, T$ ,

Step 3: Sample  $w^*$  and  $z^*$  from  $N(w^{t-1}, \text{constant} * \text{var}(w^1, w^2, \dots, w^{t-1}) + 0.001)$  and  $N(z^{t-1}, \text{constant} * \text{var}(z^1, z^2, \dots, z^{t-1}) + 0.001)$ , respectively;

Step 4: Accept  $w^*$  and  $z^*$  as  $w^t$  and  $z^t$  with probability

$$\begin{aligned} r &= \min \left( \frac{f(\mathbf{Y} | p^*, \sigma^{2*}, V_{y1}, V_{y2}) f(p^*) f(\sigma^{2*})}{f(\mathbf{Y} | p^{t-1}, \sigma^{2^{t-1}}, V_{y1}, V_{y2}) f(p^{t-1}) f(\sigma^{2^{t-1}})}, 1 \right) \\ &= \min \left( \frac{f(\mathbf{Y} | w^*, z^*, V_{y1}, V_{y2}) f(w^*) f(z^*) \frac{e^{w^*}}{(1+e^{w^*})^2} e^{z^*}}{f(\mathbf{Y} | w^{t-1}, z^{t-1}, V_{y1}, V_{y2}) f(w^{t-1}) f(z^{t-1}) \frac{e^{w^{t-1}}}{(1+e^{w^{t-1}})^2} e^{z^{t-1}}}, 1 \right), \end{aligned} \quad (2.18)$$

otherwise, set  $w^t = w^{t-1}$  and  $z^t = z^{t-1}$ . Then we transform  $p$  and  $\sigma^2$  back;

Step 5: Discard the first 50,000 burnin period. Construct the posterior distributions for  $p$  and  $\sigma^2$ .



### 2.1.3 Approximate Bayesian computation method

It's very common to use the ML approach to estimate parameters. However, in some complicated models, the likelihood function is analytically or computationally intractable. For example, in population genetics, since genetic data are large and complex with many nuisance parameters, the likelihood function has an open form which requires some numerical methods for optimization. Methods such as the approximate Bayesian computation (ABC) approach could provide an alternative to ML method by comparing information from the observed and simulated data [57, 58]. If the distance between the observed data and the simulated data are less than a very small number, called a tolerance value, we accept the proposed candidates into our posterior distributions. Some researchers also use a acceptance percentage instead of a tolerance value as a criterion to accept or reject the proposed candidates for our parameters [58]. The ABC method performs Bayesian inference without explicitly calculating the likelihood function, which is also called likelihood-free method. Because of this, this method has gained attention very quickly [59, 60, 61, 62].

In research, calculating the likelihood is very complicated. So if we could use an alternative method which can give similar estimates, it will save a lot of time. The procedure for the ABC method is as follows:

Step 1: Sample a candidate  $p^*$  and  $\sigma^{2*}$  from prior distributions with

$$p \sim U(0, 1) \tag{2.19}$$

and with different priors for  $\sigma^2$ . Note that the parameters  $p$  and  $\sigma^2$  are independent;

Step 2: Simulate a dataset  $x^*$  from the candidate  $p^*$  and  $\sigma^{2*}$  from prior distributions presented in step 1;

Step 3: Compare the simulated data,  $x^*$ , with the observed data,  $x$ , calculate the

squared Euclidean distance

$$d = (x^* - x)'(x^* - x). \quad (2.20)$$

Step 4: Use the acceptance percentage 0.05% and repeat step 1-3 200,000 times. In this approach, we end up accepting  $p^*$  and  $\sigma^{2*}$  with distance value less than or equal to the 0.05th percentile among the 200,000 distances.

## 2.2 Two gene trees BM model with more than one trait

A more realistic set up involves modeling more than one trait. Assuming that we have  $n$  species with  $k$  different observed traits, two variations are considered: (1) the variation among species within one trait, which can be captured by the joint  $V_y$  matrix and (2) the variation among  $k$  traits in one species, called the evolutionary rate matrix,  $R$  [66, 67, 68]. If all traits are considered to be independent, matrix  $R$  will be diagonal. Therefore, our new variance-covariance matrix for  $n$  species and  $k$  traits involving two gene trees is [67, 47]

$$R \otimes V_y = \begin{pmatrix} R_{11}V_y & R_{12}V_y & \cdots & R_{1k}V_y \\ R_{21}V_y & R_{22}V_y & \cdots & R_{2k}V_y \\ \vdots & \vdots & \ddots & \vdots \\ R_{k1}V_y & R_{k2}V_y & \cdots & R_{kk}V_y \end{pmatrix}. \quad (2.21)$$

The likelihood is

$$L = (2\pi)^{-\frac{nk}{2}} |R \otimes V_y|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} ((\mathbf{Y} - \mu_x X)'(R \otimes V_y)^{-1}(\mathbf{Y} - \mu_x X))\right], \quad (2.22)$$

where  $Y$  is a vector of length  $nk$ . In vector  $Y$ , elements 1 through  $n$  are the  $k$  species values for trait 1 across  $n$  taxa, which is also presented on column 1 in matrix  $X$ ; elements  $n + 1$  through  $2n$  are the  $k$  species values for trait 2 across  $n$  taxa, which is also presented on column 2 in matrix  $X$ . Similarly,  $\mu_x$  is a vector of phylogenetic mean for the  $k$  traits. The log-likelihood is

$$\ell = -\frac{1}{2}[nk \log(2\pi) + \log|R \otimes V_y| + ((Y - \mu_x X)'(R \otimes V_y)^{-1}(Y - \mu_x X))]. \quad (2.23)$$

If we apply the two gene trees BM model with more than one trait in the AMH approach, we just need to update the likelihood function with more than one rate of evolutionary process. In the ABC method, we need to propose more than one proposal distribution for the rate of evolutionary process.

In this work, we only present modeling one trait from two gene trees.

## 2.3 Simulation study

### 2.3.1 Simulation procedures

Simulations are provided to check the performance of the AMH approach and the ABC method. We assume that the trait value at the root of a given phylogenetic tree is zero in the simulation study. In step 4, we chose gene trees with long internal branch lengths and very different topologies to make their contributions easier to distinguish. In step 5, a more realistic approach is used where gene trees are simulated from the same species tree. The simulation procedures are as follows:

Step 1: Simulate the trait values based on two gene trees with  $p = 0.8$  and  $\sigma^2 = 1$ .

Calculate the trait values  $\mathbf{Y}$  by equation 2.2;

Step 2: Use the AMH approach or the ABC method to estimate the parameters  $p$

and  $\sigma^2$  under two the gene trees BM model with one trait;

Step 3: Simulate gene trees with longer internal branch lengths and very different topologies with 4, 8, 16 and 60 taxa. The choice of topologies, for the purpose of this with demonstration, was made so that half of the clades mismatch between the two gene trees;

Step 4: Apply step 1-3 100 times for the two gene trees with 4, 8 and 16 taxa, but 50 times for two gene trees with 60 taxa. We note that two gene trees with 4, 8 and 16 taxa are ultrametric trees with longer internal branch lengths and very different topologies. The simulated gene trees with 60 taxa are gene trees with longer internal branch lengths and very different topologies but are not ultrametric. Save the summary statistics and corresponding credible intervals for both  $p$  and  $\sigma^2$ ;

Step 5: Simulate species trees using the function `sim.bd.taxa()` in the R package *TreeSim* [69] with taxa number 4, 8, 16, 60 and 200 under the birth process with speciation/birth rate 0.1 or 1. Use Hybrid-Lambda [70] to generate two gene trees with similar branch lengths and topologies for different number of taxa;

Step 6: Apply steps 1, 2 and 5 50 times to these two random gene trees with 4, 8, 16, 60 and 200 taxa. Save the summary statistics and corresponding credible intervals for both  $p$  and  $\sigma^2$ .

## 2.3.2 Simulation results

### 2.3.2.1 Effect of priors

Table 2.1 shows the results of the AMH approach with different numbers of taxa and different sets of priors. In the tables shown in the section, Column **Mean** under parameter  $p$  shows the mean of the posterior modes for  $p$  in the simulations. Under

parameter  $\sigma$ , Column **Mean** shows the mean of the posterior modes for  $\sigma^2$  in the simulations. Column **CP** shows the coverage probability of the true values of parameters falling inside the 95% credible intervals, and Column **LC** shows the average length of 95% credible intervals. The best estimates of  $p$  and  $\sigma^2$  should have modes very close in their values to the true values 0.8 and 1. Also it is desirable that they have higher coverage probabilities and narrower average length of credible intervals.

There are three different priors presented in Table 2.1. With the set of priors,  $p \sim \text{Beta}(1,1)$  and  $\sigma \sim \text{FNT}(2,2)$ , the estimates of  $p$  and  $\sigma^2$  performed better than the other two sets of priors,  $p \sim \text{Beta}(0.5,0.5)$  and  $\sigma \sim \text{FNT}(1,1)$  or  $\sigma \sim \text{FNT}(2,2)$ , which means that our model performed better with some priors as opposed to others. Beyond the priors that were shown in Table 2.1 (results are omitted), we also tried  $p \sim U(0,1)$  and  $p \sim \text{LogitN}(0,1)$  for  $p$ ,  $\sigma^2 \sim \text{Gamma}(0.01,0.01)$  and  $\sigma^2 \sim \text{LogN}(0.01,0.01)$ . Presented in Table 2.1 are the ones that performed better. Also those priors in particular gave very fast convergence of  $p$  and  $\sigma^2$ . Figure 2.2 shows that  $p$  and  $\sigma^2$  converge very fast and wiggle around the true values. In Figure 2.2, the two histograms show the distributions of  $p$  and  $\sigma^2$ . Red lines shows the true values and blue lines show the mean of estimated values. In the histogram for the contribution weight  $p$ , we can not find the red line since the estimated value is equal to the true value, therefore, red and blue lines are overlapping.

Table 2.2 shows the results of the ABC method with different numbers of taxa and different sets of priors. The three sets of priors didn't give good estimates for both  $p$  and  $\sigma^2$ . With the set of priors,  $p \sim U(0,1)$  and  $\sigma \sim \text{Exp}(1)$  or  $\sigma \sim \text{FNT}(2,2)$ , the estimates of  $p$  was better than the estimates of  $\sigma^2$ . However, with the set of priors with  $p \sim U(0,1)$  and  $\sigma \sim U(0,50)$ , the estimates of  $\sigma^2$  was better than the

estimates of  $p$ . The poor estimations from the ABC method might be due to having only 200,000 iterations which might not be large enough to accurately estimate the two parameters of interest.

### 2.3.2.2 Effect of numbers of taxa

For the AMH approach with gene trees with longer internal branch lengths, regardless of any priors, with an increased number of taxa, the estimates of  $p$  and  $\sigma^2$  were more accurate relative to trees with small numbers of taxa. However, the coverage probabilities didn't change with different numbers of taxa. With the ABC method, the estimate of  $p$  was good with narrower credible intervals relative to trees with large taxa if the numbers of taxa decreased. However, the coverage probabilities were always high, which means that this estimation method didn't work very well in our case.

Table 2.3 is the simulation results using the AMH approach with random gene trees. We only used priors  $p \sim Beta(1, 1)$  and  $\sigma \sim FNT(2, 2)$  since previous simulation showed that this set of priors worked best. In Table 2.3, the estimates of  $\sigma$  didn't influence by different numbers of taxa. The AMH method worked very well for all different taxa but with increased numbers of taxa and credible intervals became narrower. However, the coverage probabilities gave high probabilities all the time. The estimates of  $p$  was significantly influenced by different numbers of taxa. With an increased numbers of taxa, the estimates of  $p$  improved with more accurate estimate and with narrower credible intervals. Therefore, the AMH approach worked in general case with large numbers of taxa. With the ABC method, using the random gene trees as inputs, the performance of this method was poor for both  $p$  and  $\sigma^2$  shown in

Table 2.4 with priors  $p \sim U(0, 1)$  and  $\sigma^2 \sim FNT(2, 2)$ . So the ABC method didn't work on random gene trees case.

### **2.3.2.3 Effect of gene trees incongruence**

There are two different simulations here with different trees incongruence: random gene trees with small differences and gene trees with longer internal branch lengths and very different topologies. When comparing the AMH approach and the ABC method, the performance of the ABC method was found to be influenced significantly by incongruence. For the gene trees with longer internal branch lengths and very different topologies, the ABC method worked well with smaller numbers of taxa. However, with random gene trees, the ABC method poorly estimated the parameters of interest. The AMH approach for two gene trees with longer internal branch lengths and with very different topologies performed well regardless of numbers of taxa with some set of priors. However, with random gene trees, it just worked for larger numbers of taxa when branch lengths and topologies were similar. Also, comparing the balanced and unbalanced tree topologies with taxa 4 and 8, there was no difference between the two methods in estimating of  $p$  and  $\sigma^2$ .

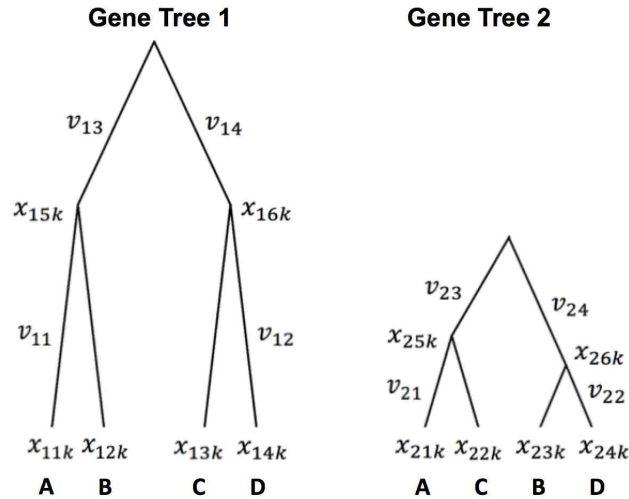


Figure 2.1: An example to calculate the  $V$  matrix based on two gene trees with a trait evolving by the BM model. The trait values  $x_{ijk}$  are at each tip and each node are unknown. Here  $i$  indexes the gene tree,  $j$  indexes the extant species, and  $k$  indexes the trait. The branch lengths  $v_{il}$  are shown next to each branch. For  $v_{il}$ ,  $i$  indexes the gene tree and  $l$  indexes the branch on the tree,  $l = 1, 2, \dots, 2n - 2$  for  $n$  species. Note that in our example, we assume gene tree 1 and 2 are ultrametric. So, in gene tree 1,  $v_{11}$  is the branch length from  $A$  or  $B$  to their most recent common ancestor;  $v_{12}$  is the branch length from  $C$  or  $D$  to their most recent common ancestor. In gene tree 2,  $v_{21}$  is the branch length from  $A$  or  $C$  to their most recent common ancestor;  $v_{22}$  is the branch length from  $B$  or  $D$  to their most recent common ancestor.



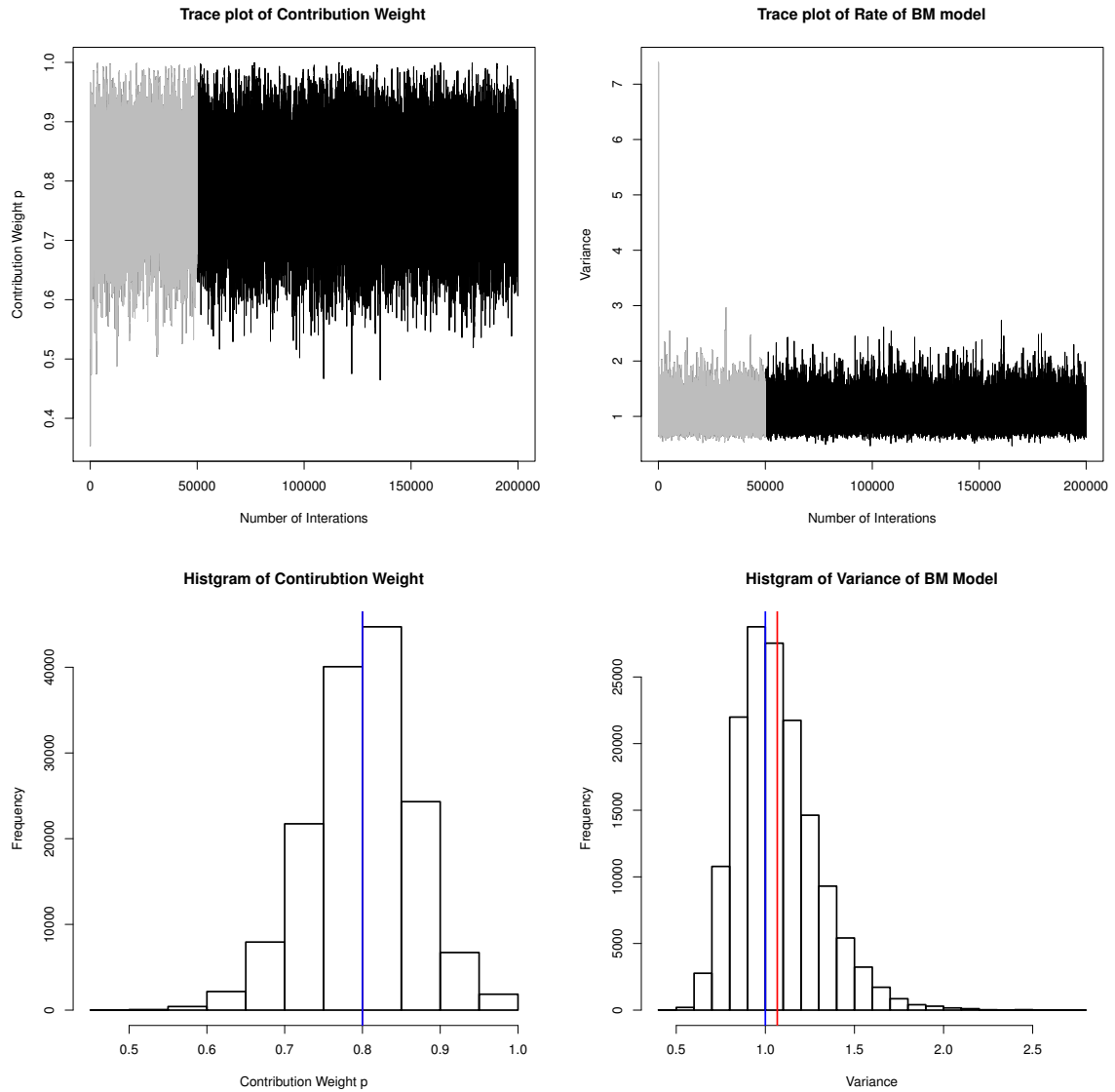


Figure 2.2: An example of trace plots and histograms using the AMH approach for the 60 taxa trees with longer internal branch lengths.

Table 2.1: Performance of the AMH approach of trees with longer internal branch lengths

		$p \sim \text{Beta}(0.5, 0.5)$ $\sigma \sim \text{FNT}(1, 1)$				$p \sim \text{Beta}(0.5, 0.5)$ $\sigma \sim \text{FNT}(2, 2)$				$p \sim \text{Beta}(1, 1)$ $\sigma \sim \text{FNT}(2, 2)$									
		Mean	CP	LC	LC	Mean	CP	LC	LC	Mean	CP	LC	LC						
Taxa 4	Balance	0.58	0.88	0.71	2.30	1.00	10.79	0.49	0.80	0.70	1.73	1.00	10.52	0.76	0.98	0.70	1.25	1.00	10.36
	Unbalance	0.31	0.72	0.81	1.44	1.00	9.59	0.32	0.66	0.81	1.42	1.00	11.45	0.74	1.00	0.79	1.26	1.00	9.14
Taxa 8	Balance	0.60	0.84	0.62	0.97	1.00	3.75	0.61	0.82	0.66	1.06	0.98	4.33	0.82	0.96	0.59	1.01	1.00	4.00
	Unbalance	0.60	0.82	0.62	0.87	0.96	3.37	0.59	0.78	0.66	1.04	0.98	4.31	0.80	0.96	0.59	0.98	0.98	3.81
Taxa 16	Unbalance	0.74	0.98	0.43	0.93	0.98	1.77	0.75	0.96	0.42	0.93	0.98	1.86	0.83	1.00	0.38	0.89	0.98	1.75
Taxa 60	Unbalance	0.77	1.00	0.31	0.96	1.00	0.88	0.79	0.98	0.30	1.00	0.96	0.92	0.82	0.98	0.31	0.99	0.96	0.91

Table 2.2: Performance of the ABC method of trees with longer internal branch lengths.

	$p \sim U(0, 1)$			$\sigma^2 \sim Exp(1)$			$p \sim U(0, 1)$			$\sigma \sim FNT(2, 2)$			$p \sim U(0, 1)$			$\sigma^2 \sim U(0, 50)$			
	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC	
Taxa 4	Balance	0.71	0.88	0.65	0.62	0.95	1.77	0.75	0.92	0.49	0.62	0.86	2.50	0.73	0.97	0.77	2.50	0.95	14.78
	Unbalance	0.72	0.95	0.58	0.66	0.96	1.77	0.73	0.93	0.66	0.62	0.82	2.19	0.67	0.98	0.74	1.73	0.99	23.42
Taxa 8	Balance	0.78	0.95	0.52	0.61	0.87	1.41	0.80	0.93	0.51	0.58	0.75	1.74	0.73	0.97	0.68	1.04	0.97	5.03
	Unbalance	0.77	0.97	0.59	0.57	0.87	1.31	0.76	0.99	0.58	0.49	0.72	1.47	0.73	1.00	0.67	0.98	1.00	4.74
Taxa 16	Unbalance	0.80	0.98	0.51	0.44	0.61	1.02	0.82	0.96	0.49	0.37	0.54	1.00	0.71	0.99	0.67	0.44	0.82	1.67
Taxa 60	Unbalance	0.71	1.00	0.30	0.51	0.90	0.55	0.80	1.00	0.54	0.32	0.20	1.00	0.46	0.98	0.90	0.76	0.96	0.78

Table 2.3: Performance of the AMH approach of random gene trees.

	Birth Rate = 1						Birth Rate = 0.1					
	$p \sim \text{Beta}(1, 1)$			$\sigma \sim \text{FNT}(2, 2)$			$p \sim \text{Beta}(1, 1)$			$\sigma \sim \text{FNT}(2, 2)$		
	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC
Taxa 4	0.53	1.00	0.92	1.32	1.00	9.50	0.48	1.00	0.93	1.06	1.00	8.55
Taxa 8	0.50	0.98	0.87	1.06	1.00	4.18	0.66	1.00	0.87	0.93	1.00	3.63
Taxa 16	0.66	1.00	0.79	1.10	0.96	2.45	0.51	1.00	0.91	1.04	1.00	2.21
Taxa 60	0.76	0.96	0.46	0.96	1.00	0.98	0.60	0.98	0.79	1.19	1.00	1.26
Taxa 200	0.79	1.00	0.30	0.98	0.94	0.57	0.72	0.96	0.59	1.11	1.00	0.85

Table 2.4: Performance of the ABC method of random gene trees.

	Birth Rate = 1						Birth Rate = 0.1					
	$p \sim \text{Beta}(1, 1)$			$\sigma \sim \text{FNT}(2, 2)$			$p \sim \text{Beta}(1, 1)$			$\sigma \sim \text{FNT}(2, 2)$		
	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC	Mean	CP	LC
Taxa 4	0.55	0.94	0.86	0.65	0.82	2.19	0.47	1.00	0.88	0.77	0.96	2.64
Taxa 8	0.64	0.96	0.81	0.54	0.86	1.61	0.67	0.98	0.84	0.46	0.72	1.31
Taxa 16	0.69	0.98	0.76	0.39	0.62	1.04	0.55	1.00	0.87	0.38	0.44	0.96
Taxa 60	0.65	0.96	0.71	0.25	0.06	0.56	0.45	1.00	0.86	0.25	0.12	0.58
Taxa 200	0.76	1.00	0.76	0.18	0.04	0.37	0.48	1.00	0.86	0.17	0.00	0.36

# Chapter 3

## Two gene trees OU model

### 3.1 Two gene trees OU model with one trait

Recall that the BM model is known to be the simplest trait evolutionary model whereas the OU model is known to be a more realistic one. Here we will introduce how to use the OU model for two gene trees with one trait. According to the paper published by Bulter and King in 2004 [42], the OU model related to one phylogenetic tree is

$$dX(t) = \alpha[\theta - X(t)]dt + \sigma dB(t), \quad (3.1)$$

where  $\alpha$  is the strength of evolutionary selection,  $\theta$  is the optimum trait value, and  $\sigma^2$  is the rate of the BM model. If there is no selection,  $\alpha = 0$ , the OU model reduces to the BM model. The OU model enjoys the very same assumptions of the BM model. Although the OU model can allow multiple optimal trait values at different times in the tree, we consider the case that there is one global optimum value for the trait. Therefore, based on Bulter and King's work [42], if we consider we have two gene trees OU model with one trait for  $n$  species, a vector  $\mathbf{Y}$  which contains  $n$  values for

the trait is multivariate normally distributed. We can write the model as

$$\mathbf{Y} \sim N(\theta, \Sigma_y), \quad (3.2)$$

where  $\theta$  is the global optimum,  $\Sigma_y$  is the variance-covariance matrix of the multivariate normal distribution.

The trait values on all tips, denoted as  $\mathbf{Y}$ , are made up from  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ , the trait values from each gene tree. Since the sum of trait values on all tips from gene tree 1 (*GT1*) and gene tree 2 (*GT2*) are fixed, we have, as in the BM model

$$\mathbf{Y} = p * \mathbf{Y}_1 + (1 - p) * \mathbf{Y}_2, \quad (3.3)$$

where  $p$  is the contribution weight from *GT1* and  $1 - p$  is the contribution weight from *GT2*.  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are independent since they evolve through different genes under OU model with the variance-covariance matrix from *GT1*

$$\Sigma_{y1} = \frac{\sigma^2}{2\alpha} V_{y1} \quad (3.4)$$

and from *GT2*

$$\Sigma_{y2} = \frac{\sigma^2}{2\alpha} V_{y2}, \quad (3.5)$$

where  $\sigma^2$  is the rate of evolution, and  $\alpha$  is the strength of selection of *GT1* and *GT2*. Here we assume that the strength of selection from two trees is the same. Note that  $V_{y1}$  and  $V_{y2}$  are related to *GT1* and *GT2*. Let's discuss how to find  $V_{y1}$  and  $V_{y2}$ .

According to Bulter and King [42], we set the branch lengths to zero at the root of the tree. We use  $v_{1T}$  denote the total branch length from the root to the tip. Also we use  $v_{1ij}$  to denote the branch length from the root to a node at which an ancestor diverges to two offsprings,  $i$  and  $j$  given *GT1*. Any element in the matrix  $V_{y1}$  can be calculated using

$$v_{y1ij} = e^{-2\alpha(v_{1T}-v_{1ij})}(1 - e^{-2\alpha v_{1ij}}). \quad (3.6)$$

Given *GT2*, any element in the matrix  $V_{y2}$  can be calculated using

$$v_{y2ij} = e^{-2\alpha(v_{2T}-v_{2ij})}(1 - e^{-2\alpha v_{2ij}}), \quad (3.7)$$

where the total branch length from the root to the tip is  $v_{2T}$ , and  $v_{2ij}$  is the branch length from the root to a node at which an ancestor diverges to two offsprings,  $i$  and  $j$  given *GT2*. Therefore, the joint variance-covariance matrix is

$$\begin{aligned} \Sigma_y &= p^2 \Sigma_{y1} + (1-p)^2 \Sigma_{y2} \\ &= p^2 \frac{\sigma^2}{2\alpha} V_{y1} + (1-p)^2 \frac{\sigma^2}{2\alpha} V_{y2} \\ &= \frac{\sigma^2}{\alpha} (p^2 V_{y1} + (1-p)^2 V_{y2}) \\ &= \frac{\sigma^2}{\alpha} V_y \end{aligned} \quad (3.8)$$

Below is an example of calculating  $V_{y1}$ ,  $V_{y2}$  and their joint variance-covariance matrix  $V_y$ . Figure 3.1, which used as an example in the previous chapter, also is used to calculate the variance-covariance matrix under the OU model.

Therefore,

$$V_{y1} = \frac{1}{2\alpha} \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{pmatrix} 1 - e^{-2\alpha(v_{11}+v_{13})} & e^{-2\alpha v_{11}}(1 - e^{-2\alpha v_{13}}) & 0 & 0 \\ e^{-2\alpha v_{11}}(1 - e^{-2\alpha v_{13}}) & 1 - e^{-2\alpha(v_{11}+v_{13})} & 0 & 0 \\ 0 & 0 & 1 - e^{-2\alpha(v_{12}+v_{14})} & e^{-2\alpha v_{12}}(1 - e^{-2\alpha v_{14}}) \\ 0 & 0 & e^{-2\alpha v_{12}}(1 - e^{-2\alpha v_{14}}) & 1 - e^{-2\alpha(v_{12}+v_{14})} \end{pmatrix} \end{matrix} \quad (3.9)$$

and

$$V_{y2} = \frac{1}{2\alpha} \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{pmatrix} A & B & C & D \\ 1 - e^{-2\alpha(v_{21}+v_{23})} & e^{-2\alpha v_{21}}(1 - e^{-2\alpha v_{23}}) & 0 & 0 \\ e^{-2\alpha v_{21}}(1 - e^{-2\alpha v_{23}}) & 1 - e^{-2\alpha(v_{21}+v_{23})} & 0 & 0 \\ 0 & 0 & 1 - e^{-2\alpha(v_{22}+v_{24})} & e^{-2\alpha v_{22}}(1 - e^{-2\alpha v_{24}}) \\ 0 & 0 & e^{-2\alpha v_{22}}(1 - e^{-2\alpha v_{24}}) & 1 - e^{-2\alpha(v_{22}+v_{24})} \end{pmatrix} \quad (3.10)$$

and then

$$V_y = p^2 V_{y1} + (1-p)^2 V_{y2}. \quad (3.11)$$

### 3.1.1 Adaptive Metropolis-Hastings approach

Assuming two gene trees with  $n$  species and one trait, we have a multivariate normal distribution for the trait values on the tips, our model is as follows:

$$\mathbf{Y} = \begin{pmatrix} y_{t_1} \\ y_{t_2} \\ y_{t_3} \\ \vdots \\ y_{t_n} \end{pmatrix} \sim N \left( \theta, \frac{\sigma^2}{2\alpha} (p^2 V_{y1} + (1-p)^2 V_{y2}) \right),$$

with priors

$$p \sim \text{Beta}(a, b), \quad (3.12)$$

$$\sigma \sim \text{Folded - noncentral - } t(df = c, \text{scale} = d), \quad (3.13)$$

$$\alpha \sim \text{Gamma}(\text{shape} = g, \text{rate} = f), \quad (3.14)$$

and

$$\theta \sim N(\mu = m, \sigma^2 = h^2), \quad (3.15)$$



where  $a, b, c, d, f, g, m$  and  $h$  are positive real numbers and the parameters  $p, \sigma^2, \alpha$  and  $\theta$  are independent. The posterior distribution is

$$\begin{aligned}
f(p, \sigma, \alpha, \theta | \mathbf{Y}, GT_1, GT_2) &= f(\mathbf{Y} | p, \sigma, \alpha, \theta, GT_1, GT_2) f(p, \sigma, \alpha, \theta) \\
&= (2\pi)^{-\frac{n}{2}} \left| \frac{\sigma^2}{\alpha} (p^2 V_{y1} + (1-p)^2 V_{y2}) \right|^{-\frac{1}{2}} \times \\
&\quad \exp\left(-\frac{1}{2} (\mathbf{Y} - \theta)' \left( \frac{\sigma^2}{\alpha} (p^2 V_{y1} + (1-p)^2 V_{y2}) \right)^{-1} (\mathbf{Y} - \theta)\right) \times \\
&\quad \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} \frac{\Gamma(\frac{c+1}{2})}{\sqrt{c\pi} \Gamma(\frac{c}{2})} \frac{1}{d} \left[1 + \frac{1}{c} \frac{\sigma^2}{d^2}\right]^{-\frac{c+1}{2}} \times \\
&\quad \frac{f^g}{\Gamma(g)} \alpha^{g-1} \exp(-f\alpha) (2\pi h^2)^{-\frac{1}{2}} \exp\left(-\frac{(\theta - m)^2}{2h^2}\right).
\end{aligned}$$

We transform  $p, \sigma^2$  and  $\alpha$  in order to use the normal distribution for adaptive sampling. Let

$$w = \log\left(\frac{p}{1-p}\right), \quad (3.17)$$

$$z = \log \sigma^2, \quad (3.18)$$

and

$$s = \log \alpha. \quad (3.19)$$

We get

$$p = \frac{e^w}{1 + e^w}, \quad (3.20)$$

$$\sigma^2 = e^z, \quad (3.21)$$

and

$$\alpha = e^s. \quad (3.22)$$

So the Jacobian of the transformation is

$$J = \begin{vmatrix} \frac{dp}{dw} & \frac{dp}{dz} & \frac{dp}{ds} \\ \frac{d\sigma^2}{dw} & \frac{d\sigma^2}{dz} & \frac{d\sigma^2}{ds} \\ \frac{d\alpha}{dw} & \frac{d\alpha}{dz} & \frac{d\alpha}{ds} \end{vmatrix},$$

$$\begin{aligned}
&= \begin{vmatrix} \frac{d \frac{e^w}{1+e^w}}{dw} & \frac{d \frac{e^w}{1+e^w}}{dz} & \frac{d \frac{e^w}{1+e^w}}{ds} \\ \frac{de^z}{dw} & \frac{de^z}{dz} & \frac{de^z}{ds} \\ \frac{de^s}{dw} & \frac{de^s}{dz} & \frac{de^s}{ds} \end{vmatrix} \\
&= \begin{vmatrix} \frac{e^w}{(1+e^w)^2} & 0 & 0 \\ 0 & e^z & 0 \\ 0 & 0 & e^s \end{vmatrix} = \frac{e^w}{(1+e^w)^2} e^{z+s}. \tag{3.23}
\end{aligned}$$

In terms of  $w$ ,  $z$  and  $s$ , the posterior distribution is

$$\begin{aligned}
f(p, \sigma, \alpha, \theta | \mathbf{Y}, GT_1, GT_2) &= f(w, z, s, \theta | \mathbf{Y}, GT_1, GT_2) |J| \\
&= (2\pi)^{-\frac{n}{2}} |e^{z-s} ((\frac{e^w}{1+e^w})^2 V_{y1} + (1 - \frac{e^w}{1+e^w})^2 V_{y2})|^{-\frac{1}{2}} \times \\
&\quad \exp(-\frac{1}{2} (\mathbf{Y} - \theta)' (e^{z-s} ((\frac{e^w}{1+e^w})^2 V_{y1} + (1 - \frac{e^w}{1+e^w})^2 V_{y2}))^{-1} (\mathbf{Y} - \theta)) \times \\
&\quad \frac{1}{B(a, b)} (\frac{e^w}{1+e^w})^{a-1} (1 - \frac{e^w}{1+e^w})^{b-1} \times \\
&\quad \frac{\Gamma(\frac{c+1}{2})}{\sqrt{c\pi}\Gamma(\frac{c}{2})} \frac{1}{d} [1 + \frac{1}{c} \frac{e^z}{d^2}]^{-\frac{c+1}{2}} \times \\
&\quad \frac{f^g}{\Gamma(g)} e^{s(g-1)} \exp(-f e^s) (2\pi h^2)^{-\frac{1}{2}} \exp(-\frac{(\theta - m)^2}{2h^2}) \times \\
&\quad \frac{e^w}{(1+e^w)^2} e^{z+s}. \tag{3.24}
\end{aligned}$$

We use the R to program and run the AMH. The procedure is as follows:

Step 1: Give the priors;

Step 2: Let  $T = 200,000$ . When  $t = 1$  and  $2$ , we give some relatively small values for  $w$ ,  $z$ ,  $s$  and  $\theta$  to start the sampling;

For  $t = 3, \dots, T$ ,

Step 3: Sample  $w^*$ ,  $z^*$ ,  $s^*$  and  $\theta^*$  from  $N(w^{t-1}, constant * var(w^1, w^2, \dots, w^{t-1}) + 0.001)$ ,  $N(z^{t-1}, constant * var(z^1, z^2, \dots, z^{t-1}) + 0.001)$ ,  $N(s^{t-1}, constant * var(s^1, s^2, \dots, s^{t-1}) + 0.001)$ , and  $N(\theta^{t-1}, constant * var(\theta^1, \theta^2, \dots, \theta^{t-1}) + 0.001)$ , respectively;

Step 4: Accept  $w^*$ ,  $z^*$ ,  $s^*$  and  $\theta^*$  as  $w^t$ ,  $z^t$ ,  $s^t$  and  $\theta^t$  with probability

$$\begin{aligned}
r &= \min \left( \frac{f(\mathbf{Y} | p^*, \sigma^{2*}, \alpha^*, \theta^*, GT_1, GT_2) f(p^*, \sigma^{2*}, \alpha^*, \theta^*)}{f(\mathbf{Y} | p^{t-1}, \sigma^{2^{t-1}}, \alpha^{t-1}, \theta^{t-1}, GT_1, GT_2) f(p^{t-1}, \sigma^{2^{t-1}}, \alpha^{t-1}, \theta^{t-1})}, 1 \right) \\
&= \min \left( \frac{f(\mathbf{Y} | w^*, z^*, s^*, \theta^*, GT_1, GT_2) f(w^*, z^*, s^*, \theta^*) \frac{e^{w^*}}{(1+e^{w^*})^2} e^{(z^*+s^*)}}{f(\mathbf{Y} | w^{t-1}, z^{t-1}, s^{t-1}, \theta^{t-1}, GT_1, GT_2) f(w^{t-1}, z^{t-1}, s^{t-1}, \theta^{t-1}) \frac{e^{w^{t-1}}}{(1+e^{w^{t-1}})^2} e^{(z^{t-1}+s^{t-1})}}, 1 \right),
\end{aligned}$$

(3.25)

otherwise, set  $w^t = w^{t-1}$ ,  $z^t = z^{t-1}$ ,  $s^t = s^{t-1}$ ,  $\theta^t = \theta^{t-1}$ . Then we transform  $p$ ,  $\sigma^2$ , and  $\alpha$  back;

Step 5: Discard the first 50,000 burnin period. Construct the posterior distributions for  $p$ ,  $\sigma^2$ ,  $\alpha$ , and  $\theta$ .

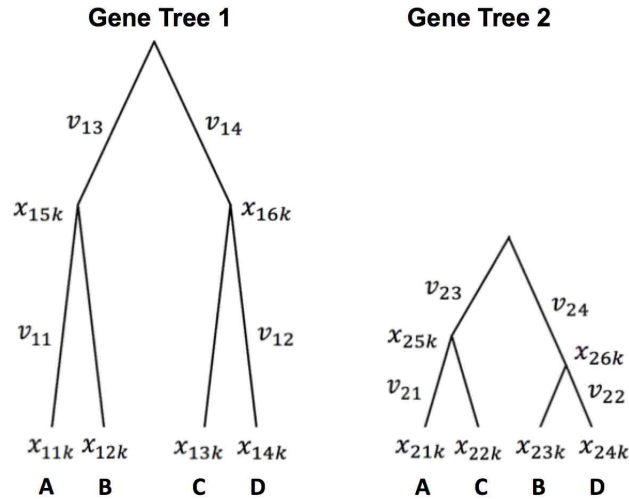


Figure 3.1: An example to calculate the  $V$  matrix based on two gene trees with a trait evolving by the OU model. The trait values  $x_{ijk}$  are at each tip and each node are unknown. Here  $i$  indexes the gene tree,  $j$  indexes the extant species, and  $k$  indexes the trait. The branch lengths  $v_{il}$  are shown next to each branch. For  $v_{il}$ ,  $i$  indexes the gene tree and  $l$  indexes the branch on the tree,  $l = 1, 2, \dots, 2n - 2$  for  $n$  species. Note that in our example, we assume gene tree 1 and 2 are ultrametric. So, in gene tree 1,  $v_{11}$  is the branch length from  $A$  or  $B$  to their most recent common ancestor;  $v_{12}$  is the branch length from  $C$  or  $D$  to their most recent common ancestor. In gene tree 2,  $v_{21}$  is the branch length from  $A$  or  $C$  to their most recent common ancestor;  $v_{22}$  is the branch length from  $B$  or  $D$  to their most recent common ancestor.

# Chapter 4

## Effect of misestimated ultrametric trees

### 4.1 Simulation study based on current one tree trait model

We examine the effect of one phylogenetic tree on the inference from a BM evolutionary trait model. We start by assuming a true phylogenetic tree and then use the BM model to infer the variance of trait evolution. We propose two different methods to obtain this goal. The process of the simulation is presented in Figure 4.1.

Method 1: We estimate ultrametric trees by MrBayes [72, 73] and RAxML [74] using the simulated alignments generated by Seq-Gen [75] from the true phylogenetic tree and simulate values for one trait using the function `fastBM()` in the R package *ape* [76]. Next we use the function `fitContinuous()` in the R package *geiger* [77] to estimate variance of the BM model. Note that for each species, only one DNA

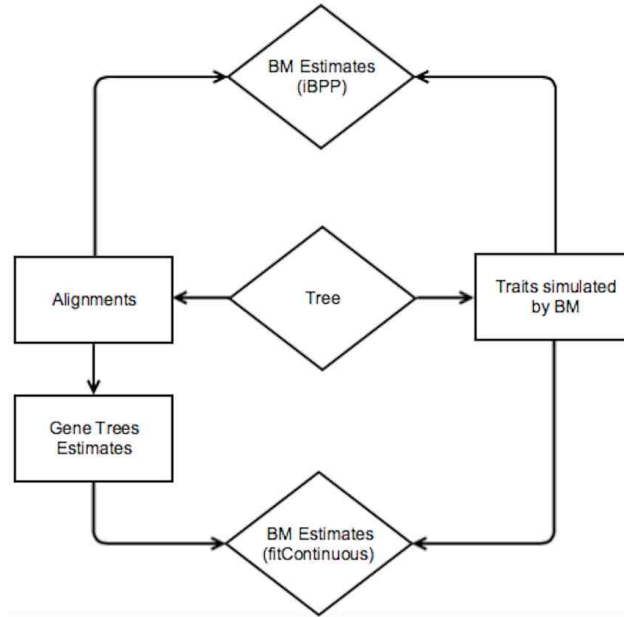


Figure 4.1: A plot of the simulation process of inferring variances from BM models.

sequence will be simulated.

Method 2: We use the simulated alignments and trait values as inputs to estimate the variance of the BM model directly using iBPP [2]. Our final results are coverage probabilities of how many times that true variance is falling inside the 95% confidence interval of the estimated variance of the BM model.

Simulation steps for both methods are as follows:

Step 1: Simulate phylogenetic trees using the function `sim.bd.taxa()` in the R package *TreeSim* [69] with taxa numbers 4, 8, 12, 16, 20, 40, and 60 under the birth process with birth rate 0.1 or 1. For each combination of these two different parameters, we generate one tree and use the same generated tree for future steps.

Step 2: Simulate both an alignment and a trait value for each species in the simulated tree from step 1. We use Seq-Gen to generate sequence alignments of length 500, 1000 and 2000 nucleotides based on the simulated trees from step 1 with the GTR

substitution model. Trait values are generated by `fastBM()` in the R package *ape* [76] with variance 1 or 2.

Step 3 (Specific to method 1): Estimate trees from MrBayes and RAxML using the simulated alignments from step 1. Transform the output trees from MrBayes and RAxML to ultrametric ones and use `fitContinuous()` in R to estimate the variance of the BM models with confidence intervals. Calculate the coverage probability of how many times the true variance falls inside the confidence intervals. In MrBayes, the substitution model is GTR with gamma-distributed rate variation across all the sites. Each simulation runs for one million generations, with a sample frequency of 10. We discard the first 25% of samples as burnin and construct the consensus tree using the last 75% of samples. All trees will be midpoint rooted. In RAxML, a standard hill climbing algorithm is used to search the ML trees with GTRGAMMA model.

Step 4 (Specific to method 2): Use the simulated sequence alignments and trait values as inputs to estimate the variances directly using iBPP with generated guide trees from MrBayes while assuming: Gamma prior for the population size parameter  $\theta_0 \sim G(2, 2000)$ ; Gamma prior for the root divergence time  $\tau_0 \sim G(2, 2000)$ ; and Dirichlet priors for any other divergence time parameters in the tree. Non-informative prior 0 is assigned to the trait parameters and then algorithm 1 is used as described in Yang and Rannala's paper in 2010 [78]. Calculate the coverage probability and 95% highest posterior density (HPD) intervals.

Step 5 (Specific to method 1): Calculate Robinson-Foulds (RF) distances [71] to estimate the accuracy of the estimated trees. Since the output trees from MrBayes are rooted tree, we use the R directly to calculate RF distances and normalized RF distances, where a normalized RF distance is calculated by the ratio of the RF distance

and the maximum RF distance for a rooted tree. The maximum RF distance for a rooted tree is defined as  $2 * (\text{the number of tips} - 2)$ . The output trees from RaxML are unrooted trees, so we use the PHYLIP program [79] to calculate RF distances and normalized RF distances, where a normalized RF distance is calculated by the ratio of the RF distance and the maximum RF distance for an unrooted tree. The maximum RF distance of an unrooted tree is defined as  $2 * (\text{the number of tips} - 3)$ . Step 6: For taxa numbers 4, 8, 12, 16 and 20, repeat step 1 - 5 100 times. For taxa numbers 40 and 60, repeat step 1 - 5 50 times.

## 4.2 Simulation results

The difference between the phylogenetic trees simulated by `sim.bd.taxa()` in R using different birth rates 0.1 or 1 is the difference in time to diverge to sister species. When the birth rate is 0.1, the speciation time is longer than the speciation time with the birth rate 1. So the corresponding simulated trees have longer branch lengths with birth rate 0.1. The goal of the simulation is to see performance of the variance estimation for the BM model under different situations.

### 4.2.1 Coverage probabilities with different numbers of taxa

Figure 4.2 shows the coverage probability of the estimated variances under the BM model when the true variance is 1 with birth rates 1 and 0.1. Under the true variance 1 with the birth rate 1, the coverage probabilities of the estimates of  $\sigma^2$  by the iBPP program with taxa number 4, 8 and 12 are greater than 95% . When the number of taxa increases to 20, the coverage probabilities get worse. However, under the true variance 1 with the birth rate 0.1, the the coverage probabilities of the estimates of



$\sigma^2$  by the iBPP program with taxa number 4 for the lengths of DNA alignments 500 or 1000 are greater than 95%. There is no bias in the estimation when using MrBayes and RAxML. With an increased number of taxa, the performance gets better for all situations except for the lengths of DNA alignments 2000, under the variance 1 with the birth rate 0.1, for which the coverage probabilities decrease to 85% from 93% with taxa number 60.

Figure 4.3 shows the coverage probability of the estimated variances under the BM model when the true variance is 2 with birth rates 1 and 0.1. Under the true variance 2, the coverage probabilities of the estimates of  $\sigma^2$  by the iBPP program is not adequate for being less than 80% with increased numbers of taxa until 20. There is no bias in the estimation when using MrBayes and RAxML and both give high coverage probabilities greater than 85%. This might be due to having an ultrametric tree. However, when the taxa number is less than 10, the coverage probabilities are less than 80%.

#### **4.2.2 Different performance in inferring the variance of a BM model using iBPP, RAxML and MrBayes**

It is not sufficient to only use the coverage probability to make a conclusion about the accuracy of estimation of the variance under a BM model. The length of confidence intervals is another measure to study the performance in inferring the variance. Figures 4.4 to 4.15 show performances of lengths of confidence intervals with taxa 8, 16 and 60. We observe a decreasing trend in the length of the confidence intervals of the variance in the BM model. In particular, with the increased number of taxa until taxa 20, the lengths of confidence intervals decrease. When the taxa number is 60, the

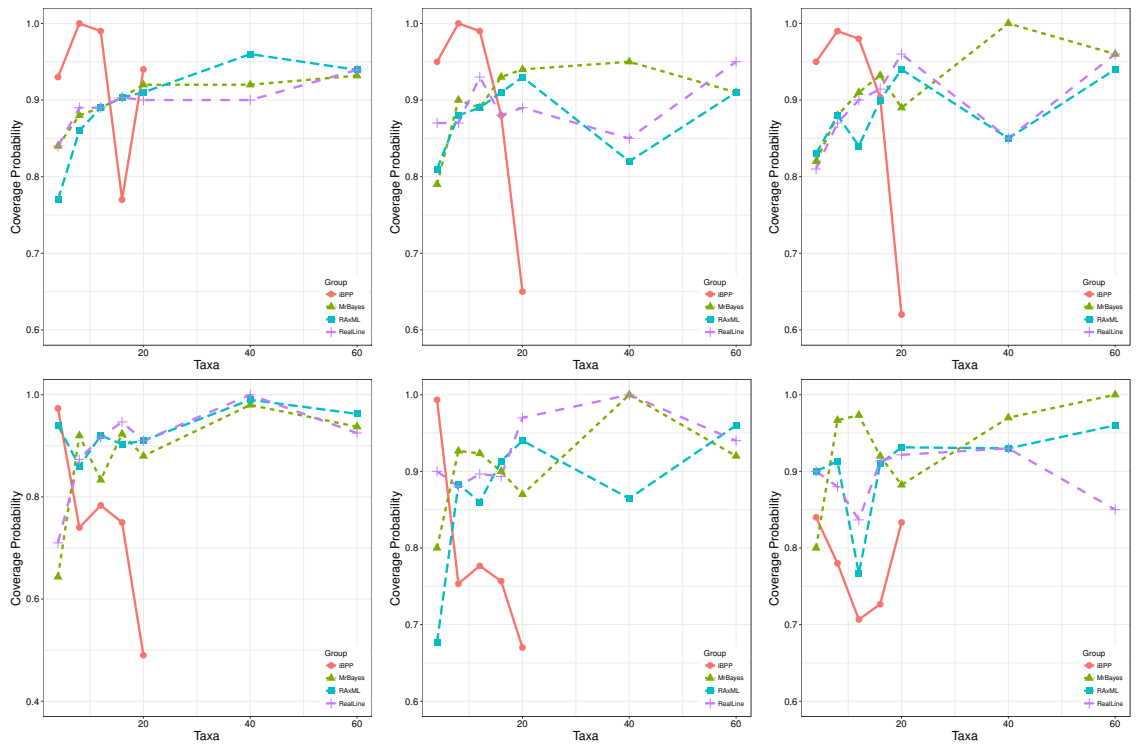


Figure 4.2: Coverage probability when variance = 1 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000.

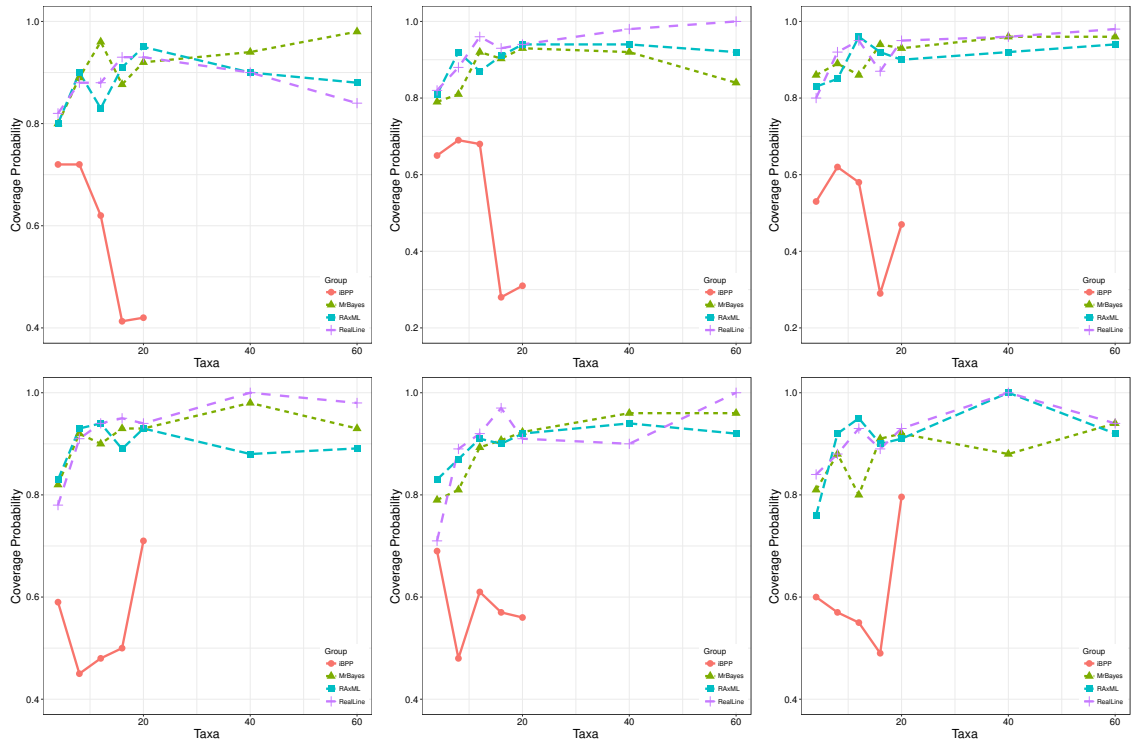


Figure 4.3: Coverage probability when variance = 2 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000.

lengths of confidence intervals become very small. The results mean that the performance of inferring the variance of BM model gets better with the increased number of taxa. The results are consistent with the results from the coverage probabilities.

### 4.2.3 Comparison between BM and OU models

A very interesting question is whether the OU model fits better than the BM model simulated trait values with increased numbers of taxa. For the purpose of this comparison, we calculate the confidence level or percentage of correct decision ( $1 - \alpha$ ) in testing  $H_0$ : BM model is adequate vs.  $H_1$ : OU model is adequate for simulate data from the BM model.

Figure 4.16 shows the percentage of correct decision of the test when the variance is 1 with the birth rate 1 in the first row and 0.1 in the second row. When the variance = 1 with the birth rate 1, given the true trees, the percentages get worse with an increased number of taxa. But given the trees estimated by MrBayes and RAxML, the percentages are greater than 95% with the lengths of DNA alignments 500 and 1000. When the variance = 1 with the birth rate 0.1 for the lengths of DNA alignments 500, the percentages decrease with an increased number of taxa in the left panel. For the lengths of DNA alignments 1000 and 2000, both middle and right panels are complicated with no obvious trends.

Figure 4.17 shows the percentages of correct decision of the test when the variance is 2 with the birth rate 1 in the first row and 0.1 in the second row. When variance = 2 with the birth rate 1, given the true trees, the percentages get worse with an increased number of taxa. But given the trees estimated by MrBayes and RAxML, the percentages are greater than 95% with the lengths of DNA alignments 500, 1000

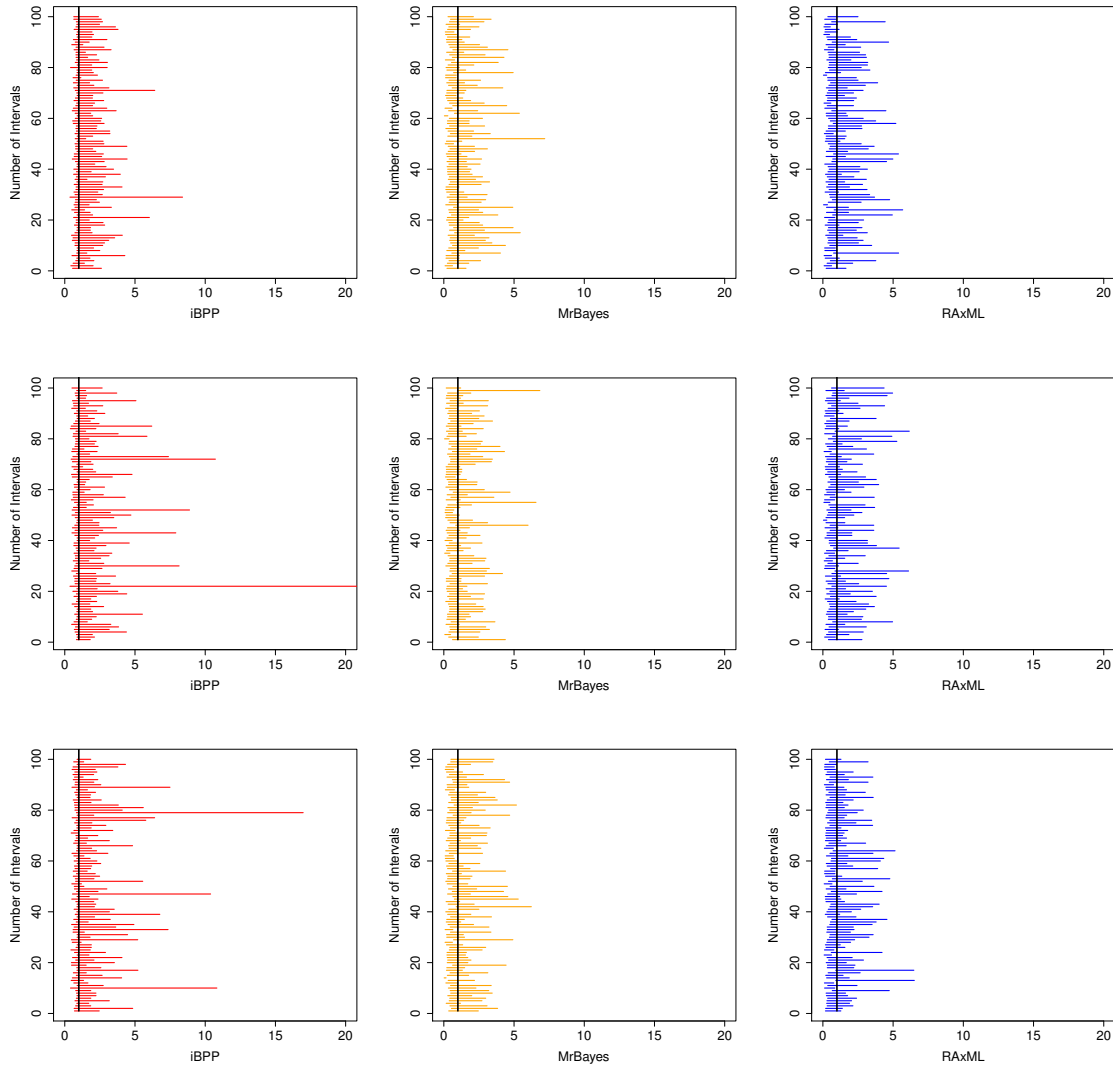


Figure 4.4: Different performance of confidence intervals when variance = 1, birth rate = 1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

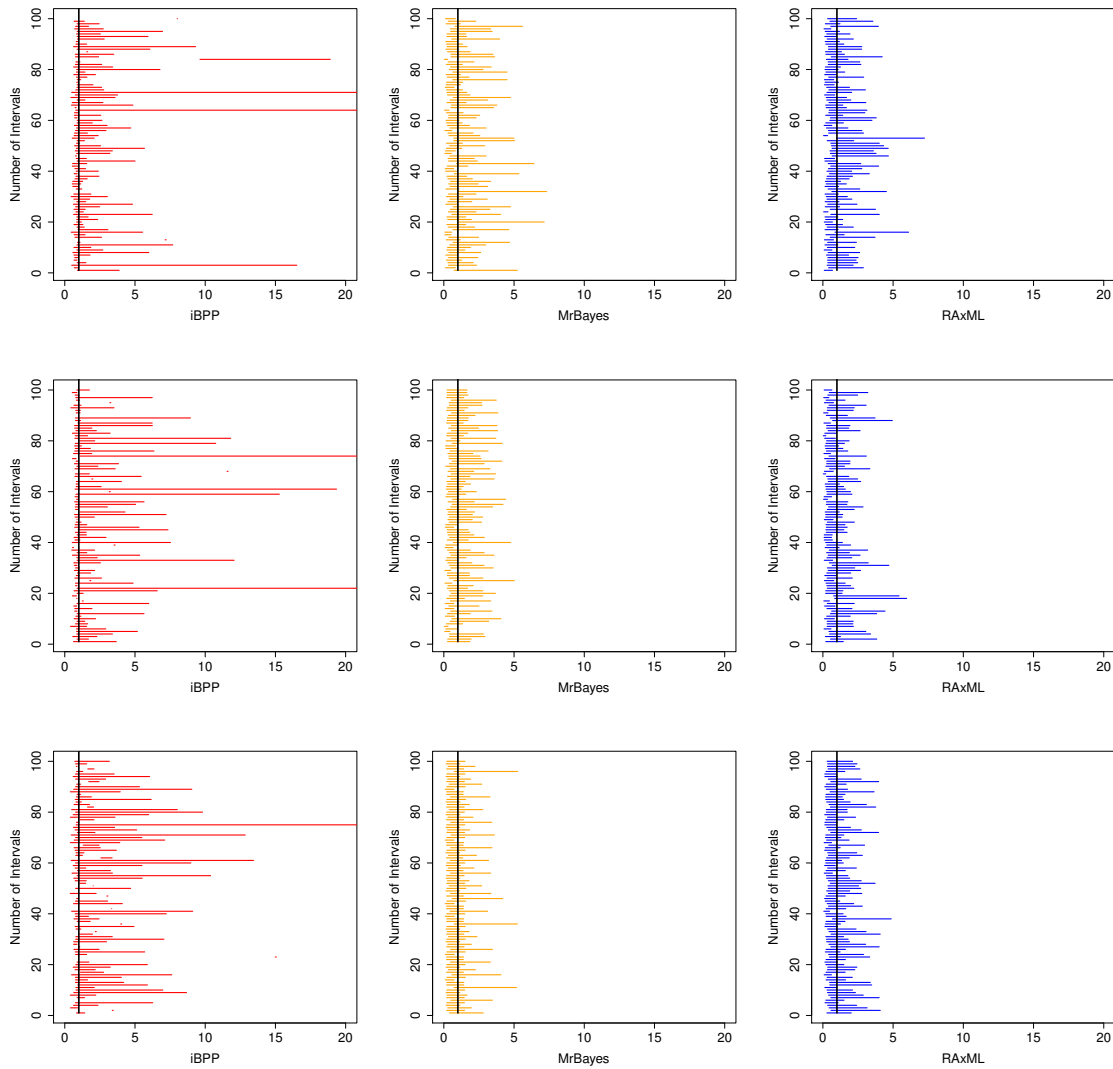


Figure 4.5: Different performance of confidence intervals when variance = 1, birth rate = 0.1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

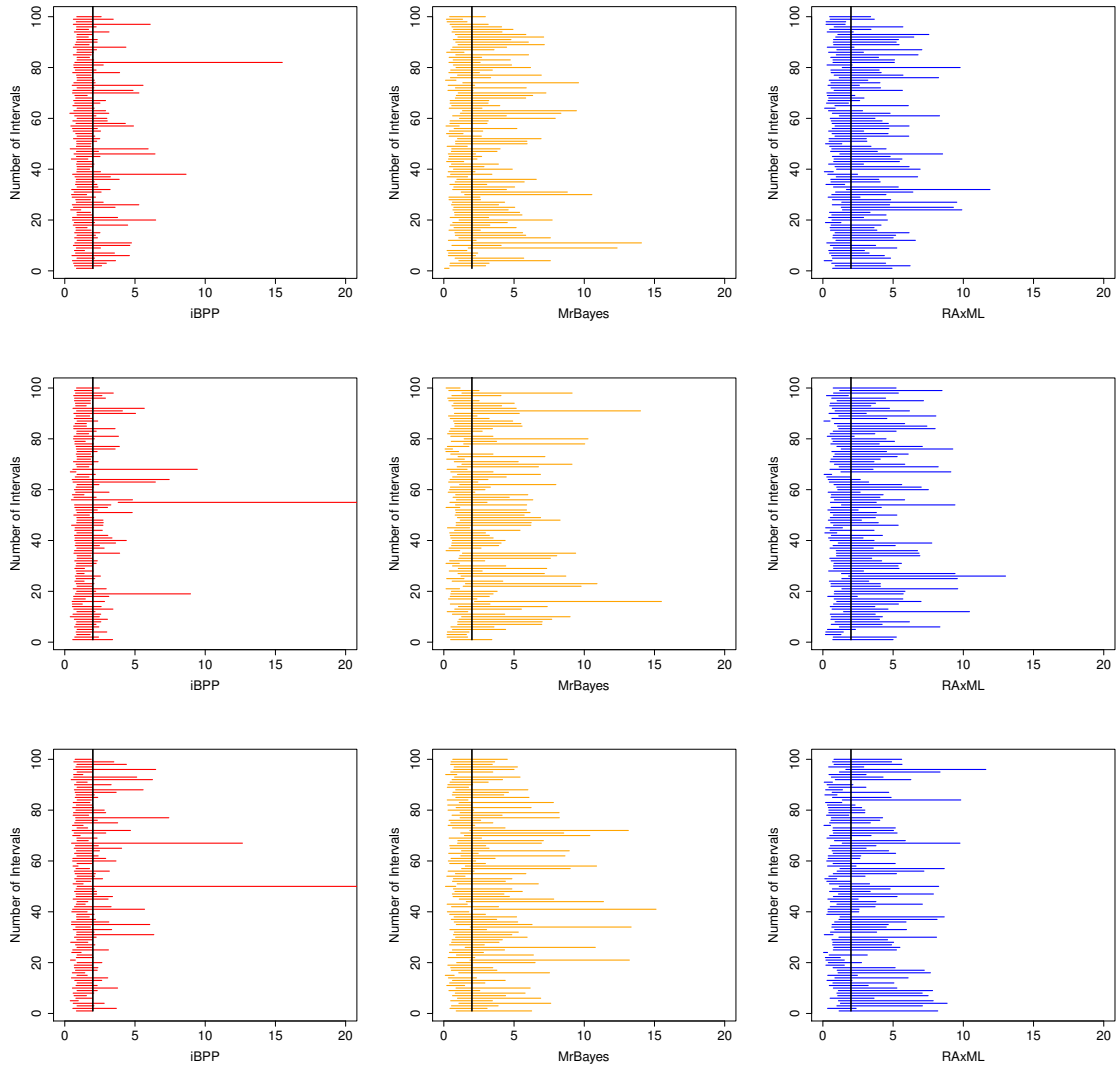


Figure 4.6: Different performance of confidence intervals when variance = 2, birth rate = 1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

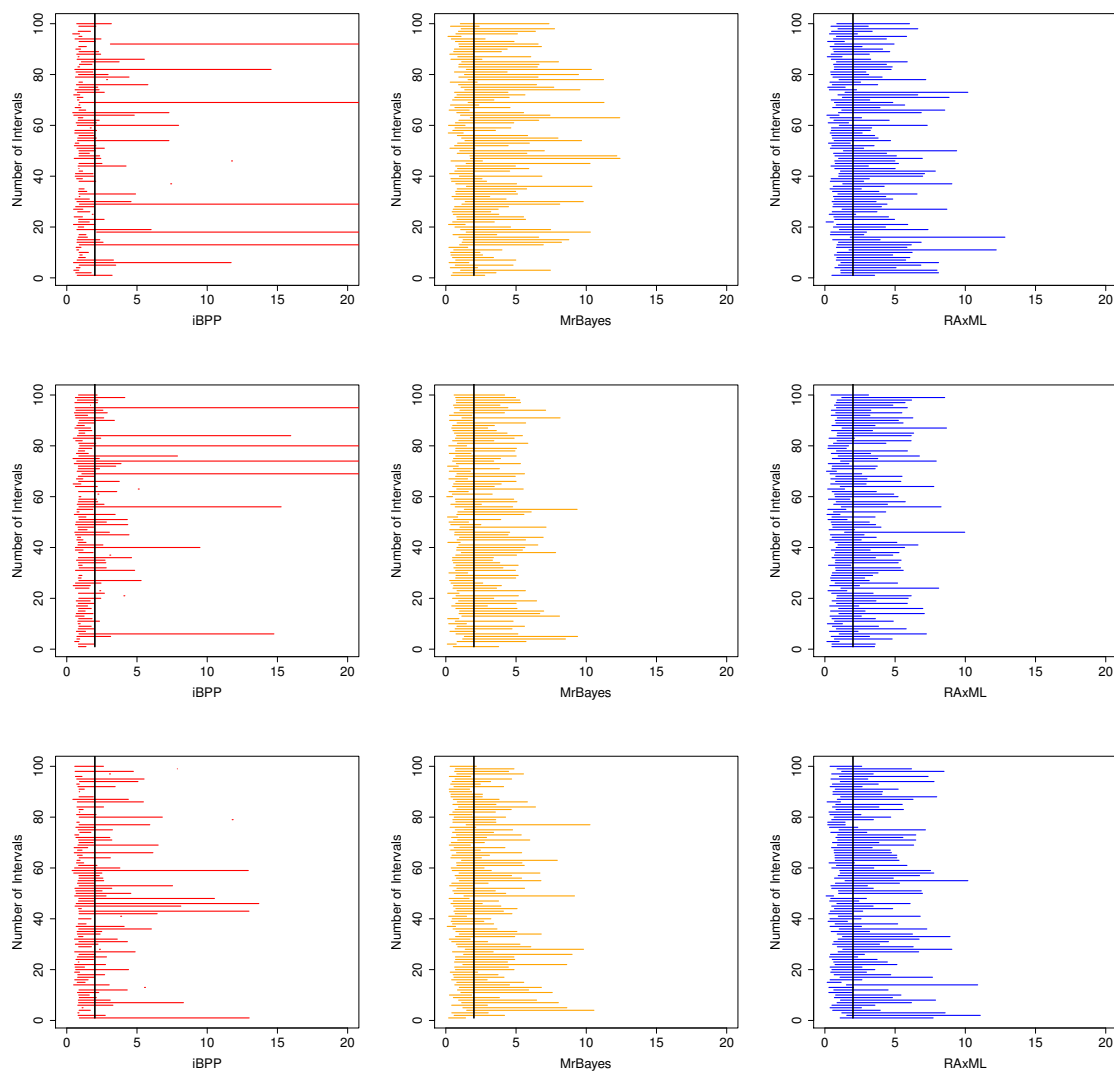


Figure 4.7: Different performance of confidence intervals when variance = 2, birth rate = 0.1, taxa = 8. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.



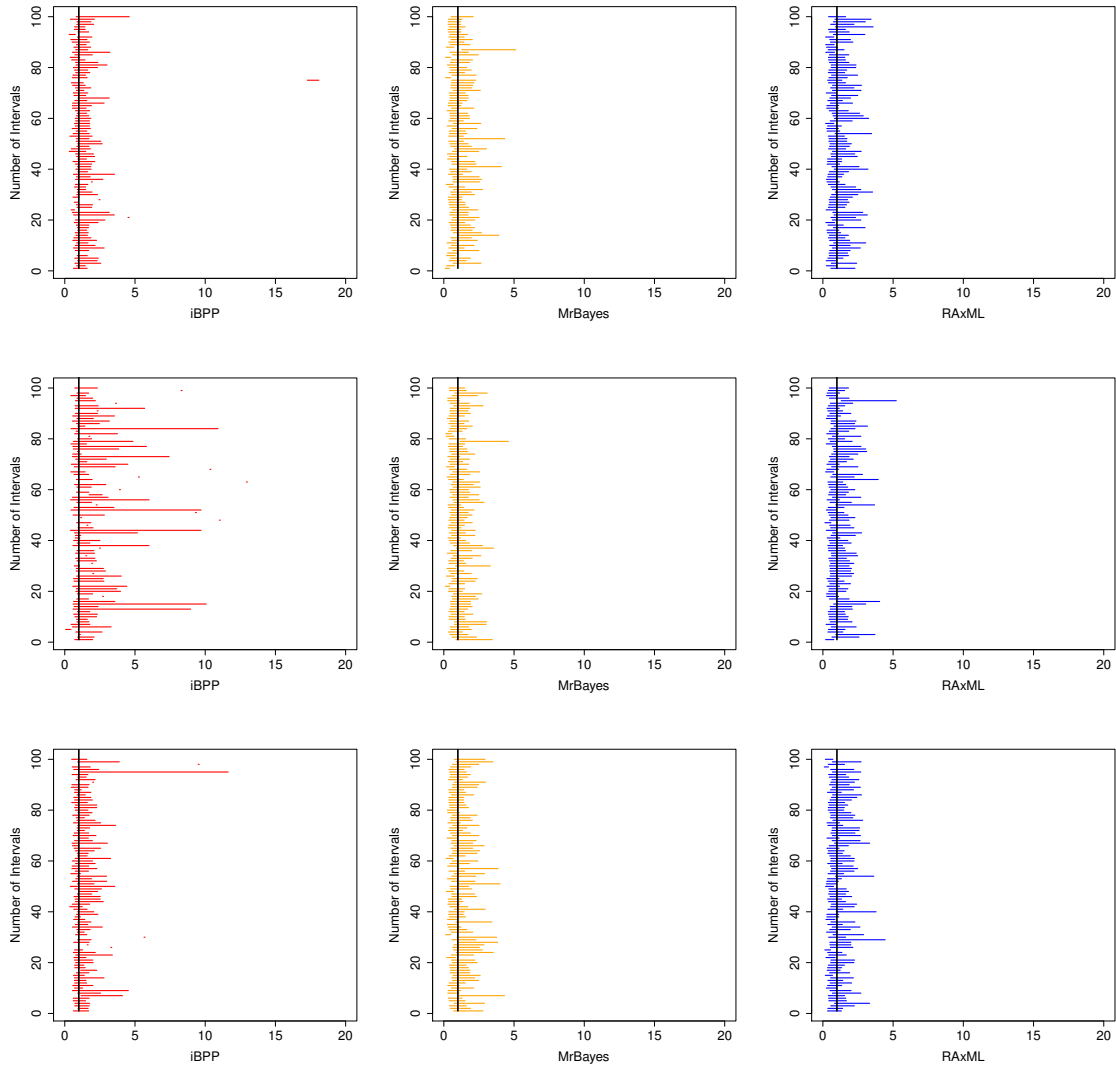


Figure 4.8: Different performance of confidence intervals when variance = 1, birth rate = 1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

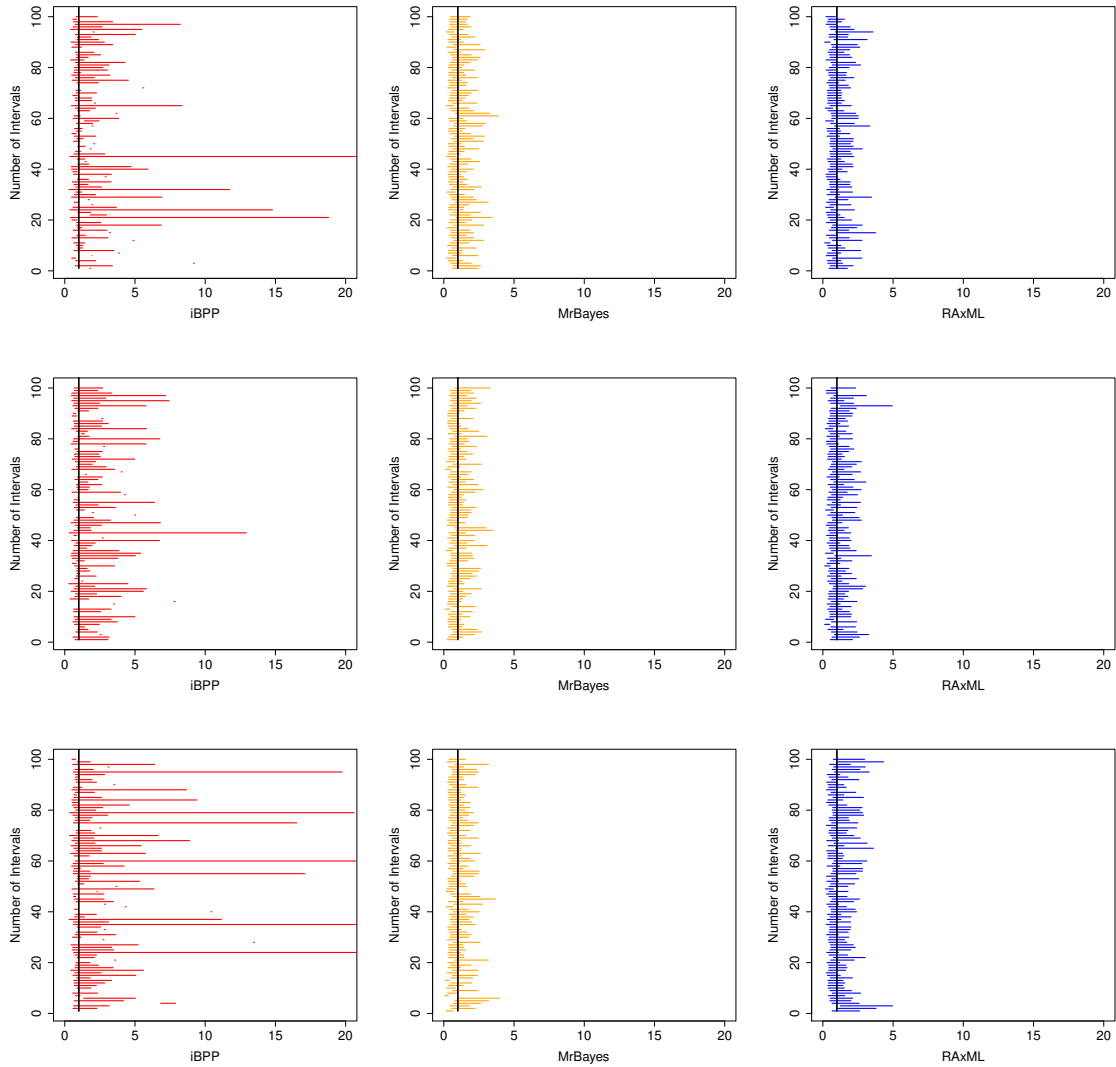


Figure 4.9: Different performance of confidence intervals when variance = 1, birth rate = 0.1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

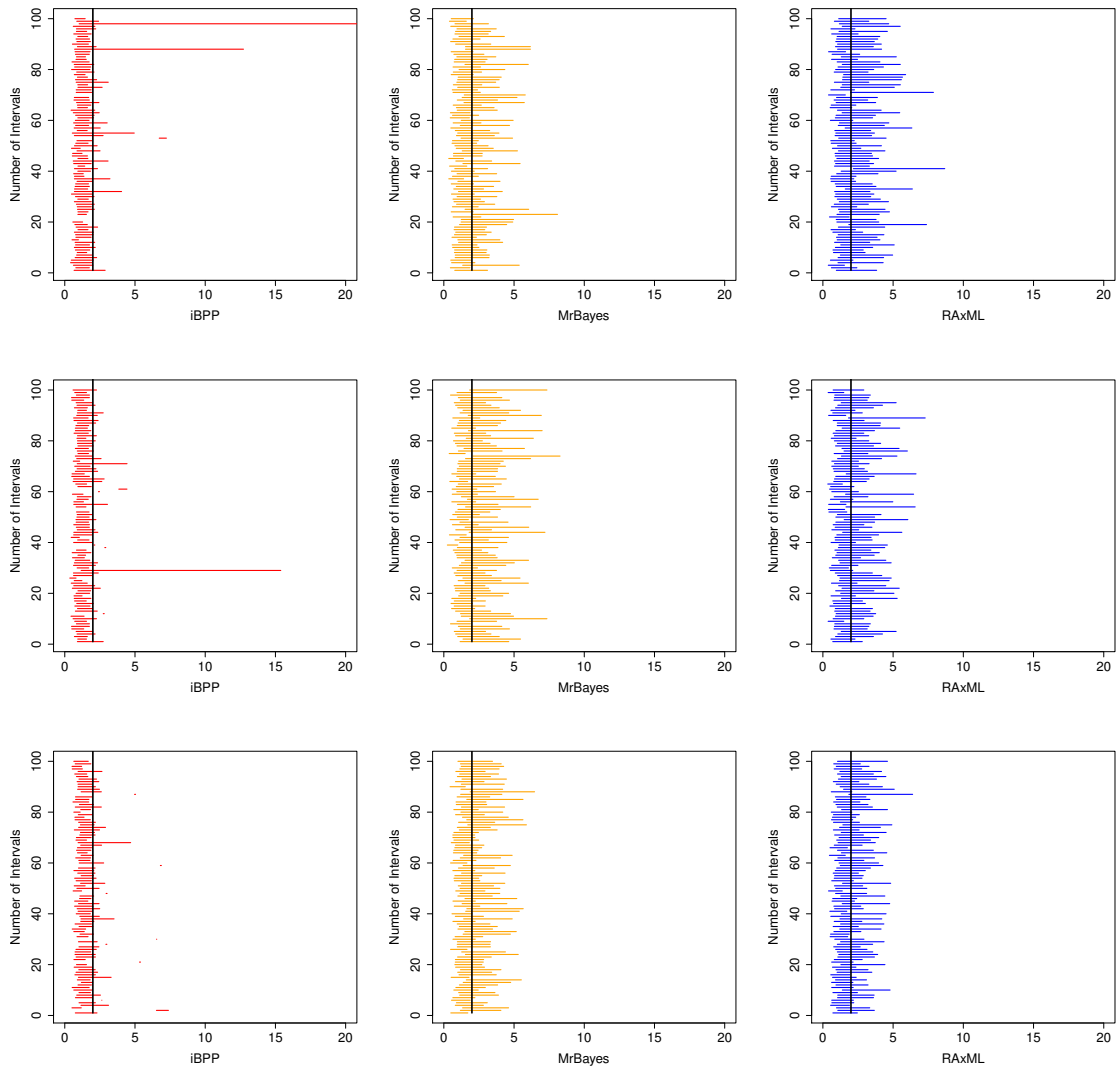


Figure 4.10: Different performance of confidence intervals when variance = 2, birth rate = 1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

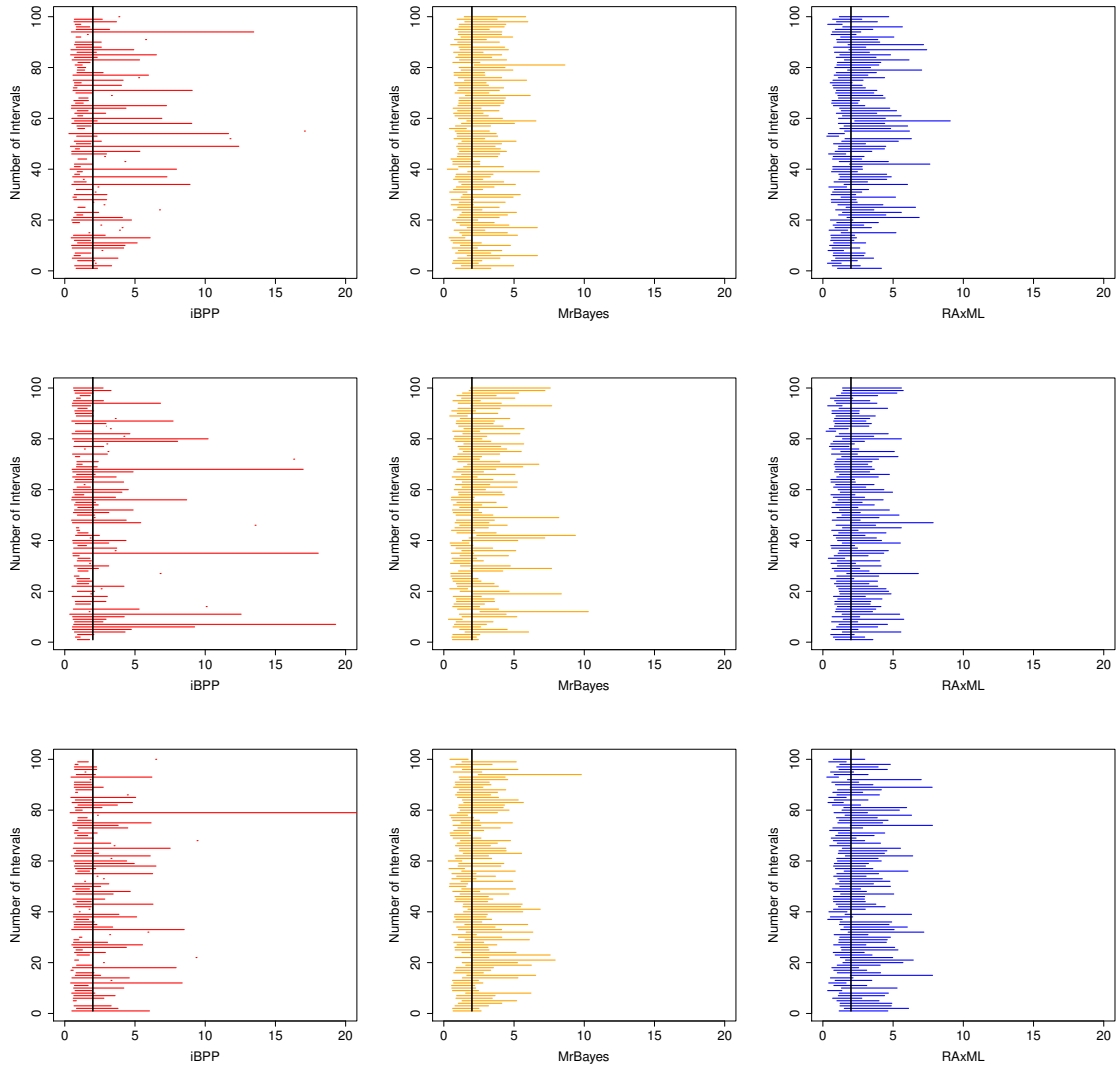


Figure 4.11: Different performance of confidence intervals when variance = 2, birth rate = 0.1, taxa = 16. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

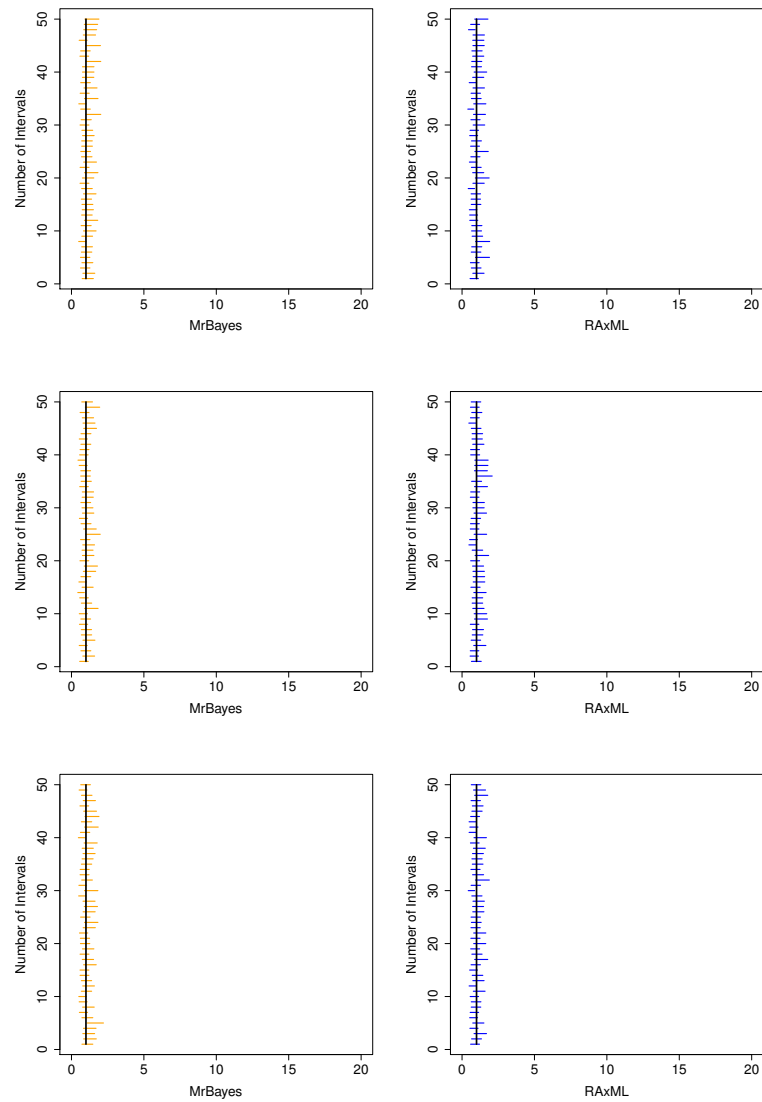


Figure 4.12: Different performance of confidence intervals when variance = 1, birth rate = 1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

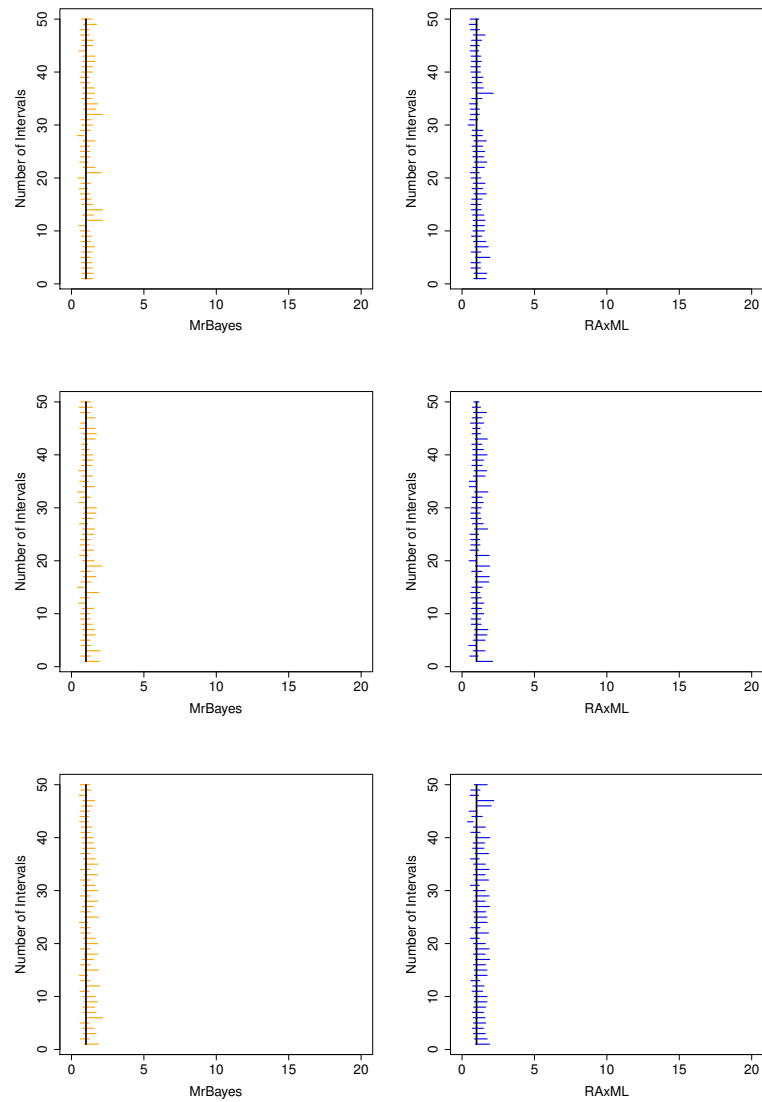


Figure 4.13: Different performance of confidence intervals when variance = 1, birth rate = 0.1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

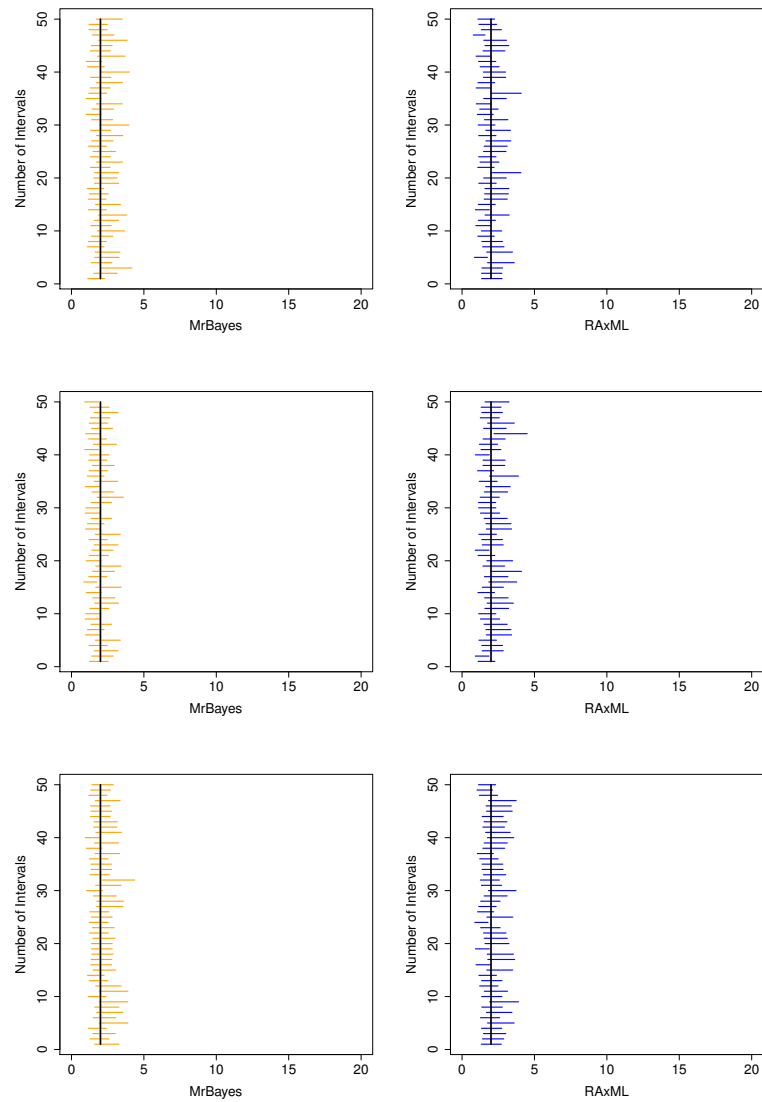


Figure 4.14: Different performance of confidence intervals when variance = 2, birth rate = 1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.

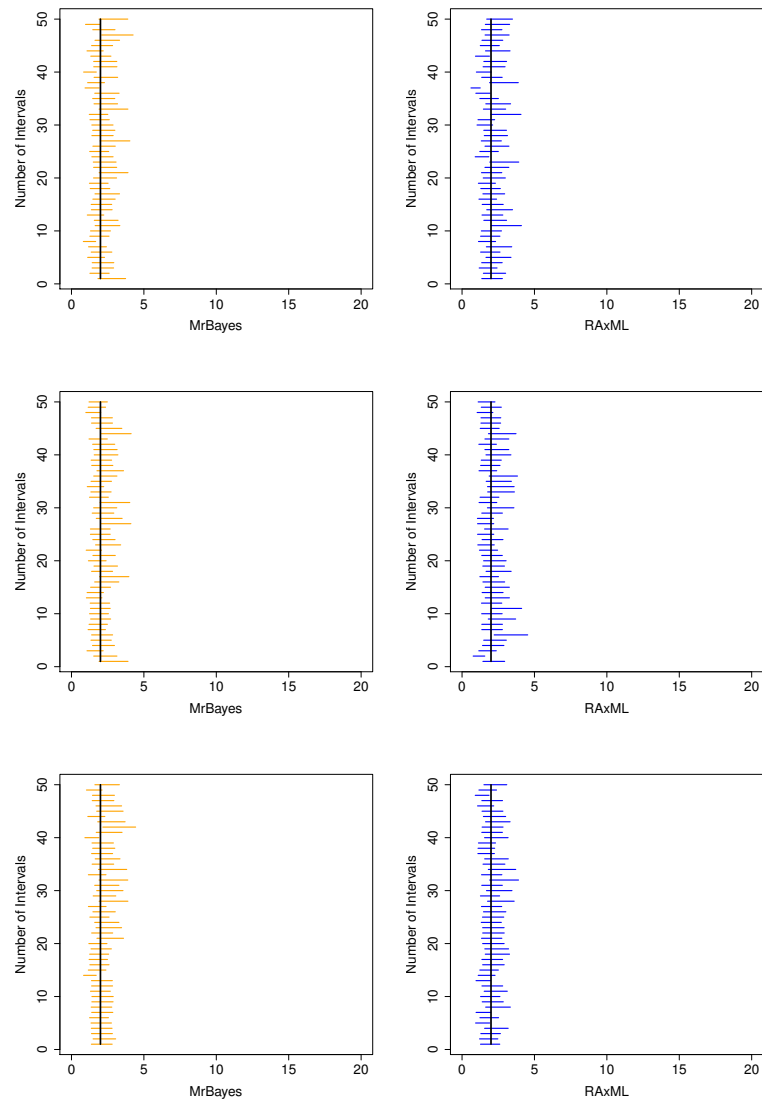


Figure 4.15: Different performance of confidence intervals when variance = 2, birth rate = 0.1, taxa = 60. The first row shows the results for the length of DNA alignments 500, the second shows the results for 1000, and the third row shows the results for 2000.



and 2000. When the variance = 2 with the birth rate 0.1, the percentages get worse with the increased number of taxa until taxa 20 for three different lengths of DNA alignments. When the number of taxa is greater than 20, for the lengths of DNA alignments 1000, the percentages increase with an increased number of taxa. For the length of DNA alignments 500 and 2000, given the true trees and the trees estimated from MrBayes, with an increased number of taxa, the percentages increase or stay the same. But with the trees estimated from RAxML, the percentages decrease when the number of taxa is greater than 20.

Comparing all the plots in Figure 4.16 and Figure 4.17, most percentages of correct decision of the test between the estimated trees from MrBayes or RAxML are greater than 80%. With the birth rate 1, the true trees' percentage of correct decision of the test gets worse under both variances 1 and 2, which means that with an increased number of taxa, the OU model fits the data better if the trees have shorter branch lengths. With the birth rate 0.1, the true trees' percentage of correct decision of the test gets worse when the taxa number is less than 20 under both variances 1 and 2, which means that with the taxa number less than 20, the OU model fits the data better if the trees have longer branch lengths. With the taxa number greater than 20, the situation gets complicated and no obvious trends are observed.

#### **4.2.4 The changes of normalized RF distances with different numbers of taxa**

Normalized RF distances are used to compare the accuracy of the estimated ultrametric trees from MrBayes and RAxML programs such that smaller values indicate better accuracy. The left panel in Figure 4.18 shows the changes of normalized RF distances

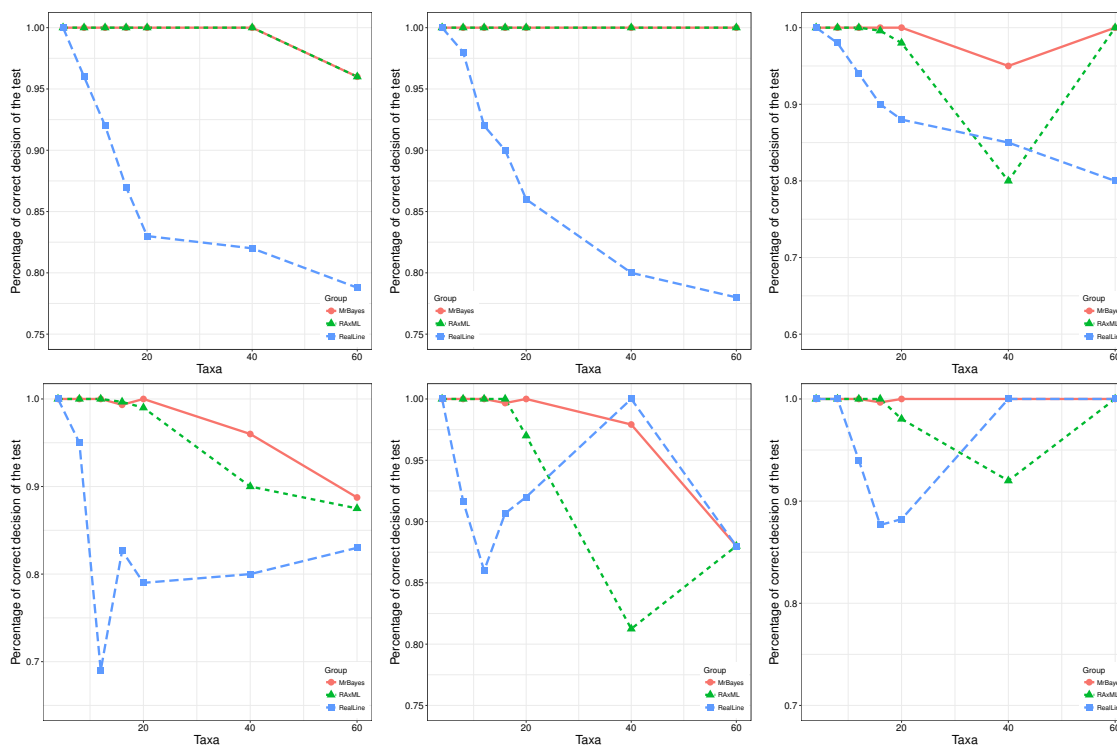


Figure 4.16: Percentages of correct decision ( $1 - \alpha$ ) in testing  $H_0$ : BM model is adequate vs.  $H_1$ : OU model is adequate for simulate data from the BM model when variance = 1 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000.

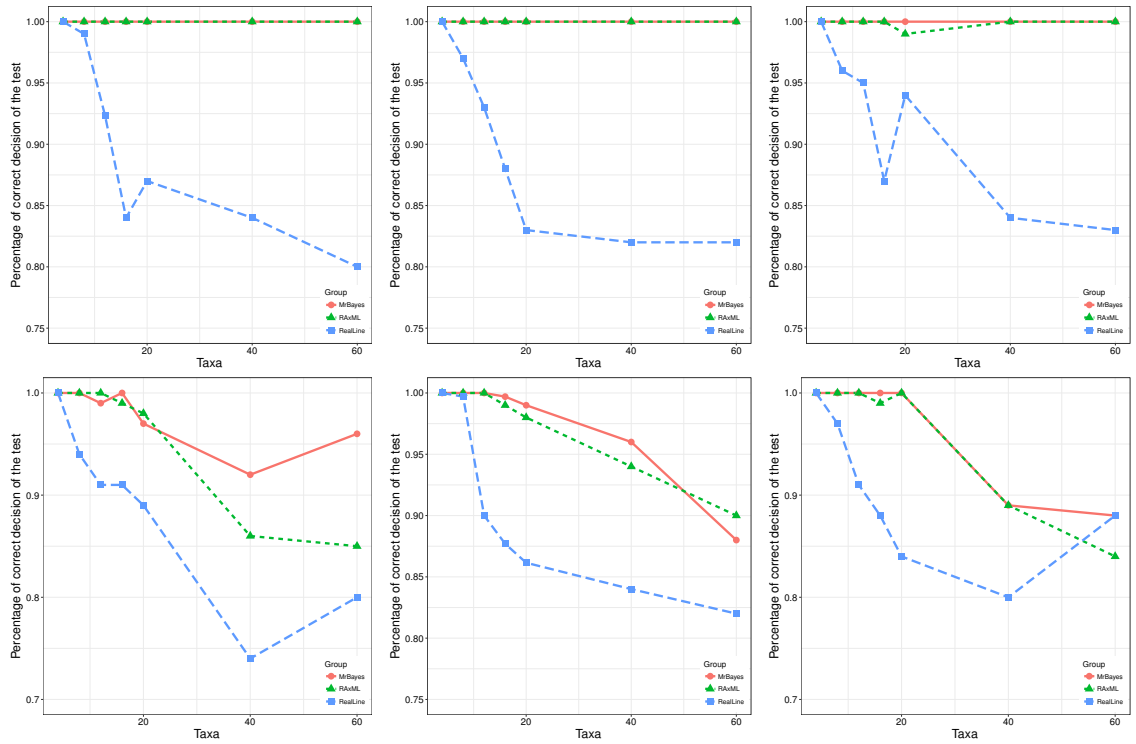


Figure 4.17: Percentages of correct decision ( $1 - \alpha$ ) in testing  $H_0$ : BM model is adequate vs.  $H_1$ : OU model is adequate for simulate data from the BM model when variance = 2 with birth rate = 1 in the first row and 0.1 in the second row. The left panel shows the results for the length of DNA alignments 500, the middle panel shows the results for 1000, and the third panel shows the results for 2000.

under estimated trees from MrBayes and RAxML with the birth rate 1 for different lengths of DNA alignments. There is no difference between the estimated trees from MrBayes and RAxML with the same lengths of DNA alignments which is shown in the figure with two closely tied lines. With an increased number of taxa, there is not a big difference in the overall normalized RF distances. But there are slightly higher values for the length of DNA alignments 500 and 1000 for the taxa number less than 20. For the length of DNA alignments 2000, there is a slight increase in the normalized RF distances with the increased numbers of taxa. All normalized RF distances are around 0.2, which means that 20% of clades are incorrectly estimated.

The right panel in Figure 4.18 shows the changes of normalized RF distances but with the birth rate 0.1. The longer branch lengths make the accuracy of estimated trees more stable under taxa number 40. With the length of DNA alignments 2000, there are no changes between estimated trees from MrBayes and RAxML with increased numbers of taxa. But with the length of DNA alignments 500 and 1000, there is an increasing trend in normalized RF distances for the taxa number greater than 20. All normalized RF distances are less than 0.25.

Comparing the normalized RF distances for the length of DNA alignments 500 and 1000 with the birth rate 1, we observe an increasing trend with taxa numbers less than 20; this trend reverses its direction as we go to taxa numbers larger than 20. With the birth rate 0.1, there is a flat trend with respect to taxa numbers less than 20. After taxa 20, we observe an increasing trend. For the length of DNA alignments 2000 with the birth rate 1, there is an increasing trend with increased numbers of taxa but with the birth rate 0.1, there are almost no changes for the normalized RF distances (less than 0.05) with small numbers of taxa.

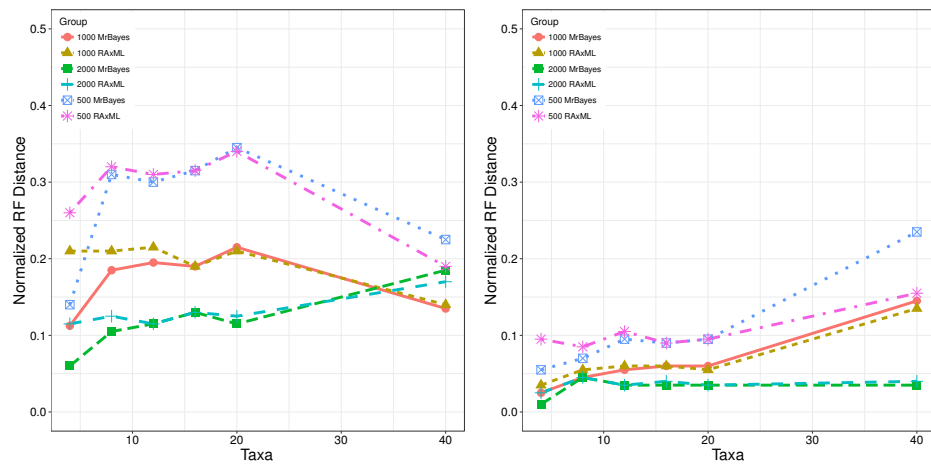


Figure 4.18: Normalized RF distances when birth rate = 1 (left) and 0.1 (right).

# Chapter 5

## Empirical Study

We apply our new methodology to real data in biology. Our data are from the published paper *Natural Variation and Artificial Selection in Four Genes Determine Grain Shape in Rice* in *New Phytologist*, 2013 [44]. The trait of interest in this example is grain width and the corresponding two major genes are *GW2* and *qSW5* [44]. We will use the two gene trees BM model and two gene trees OU model to estimate the parameters of interest.

### 5.1 Data preparation

In the original trait data, there are 127 different rice varieties with the trait (grain width) values. First, we delete these rice varieties with missing DNA sequences. Second, since lots of rice varieties share the same DNA sequences for *GW2* or *qSW5*, we assume that these rice varieties are from the same species, denoted as  $t_i$ , where  $i$  presents the same species of interest in our case. Therefore we combine these rice varieties together as different populations from different species and calculate, for

each species, the mean of trait values. We use this mean as a new measure for the trait value for each species.

In our study, after combing these rice varieties, we get seven species that are the same for *GW2* and *qSW5* and they are presented in Figure 5.1. In this Figure,  $t_1$  represents all the rice varieties that share the same DNA sequences *GW2 - 1* and *qSW5 - K1*,  $t_2$  represents all the rice varieties that share the same DNA sequences *GW2 - 2* and *qSW5 - N1*,  $t_3$  represents all the rice varieties that share the same DNA sequences *GW2 - 3* and *qSW5 - K2*,  $t_4$  represents all the rice varieties that share the same DNA sequences *GW2 - 4* and *qSW5 - N3*,  $t_5$  represents all the rice varieties that share the same DNA sequences *GW2 - 5* and *qSW5 - N7*,  $t_6$  represents all the rice varieties that share the same DNA sequences *GW2 - 7* and *qSW5 - N2*, and  $t_7$  represents all the rice varieties that share the same DNA alignments *GW2 - 8* and *qSW5 - K4*. Note that DNA sequences *GW2 - 1*, *GW2 - 2*, *GW2 - 3*, *GW2 - 4*, *GW2 - 5*, *GW2 - 7*, and *GW2 - 8* are the alleles of the gene *GW2*. DNA sequences *qSW5 - K1*, *qSW5 - N1*, *qSW5 - K2*, *qSW5 - N3*, *qSW5 - N7*, *qSW5 - N2*, and *qSW5 - K4* are the alleles of the gene *qSW5*. The lengths of DNA alignments for *GW2* and *qSW5* are 6792 nucleotides and 3146 nucleotides.

In the original multiple DNA sequences for *GW2* and *qSW5*, there are more than seven DNA sequences for both *GW2* and *qSW5*. First, we use the online multiple sequence alignment program Clustal Omega to align the DNA sequences for *GW2* and *qSW5* [80]. Then, jModeltest 2.1.10 [13, 14] is used to test the best substitution models for MrBayes [72, 73] to infer the gene trees for *GW2* and *qSW5*. After the two gene trees are estimated by MrBayes, we drop some of the tips from the estimated

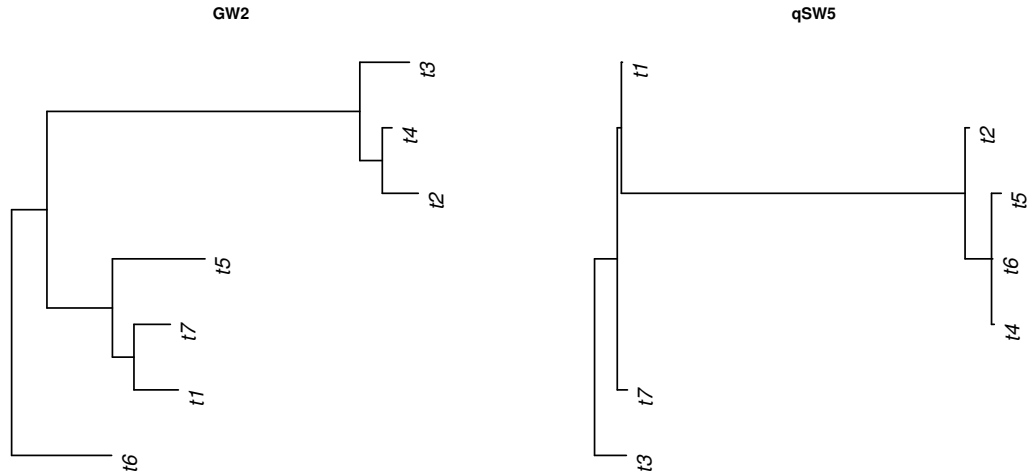


Figure 5.1: *GW2* and *qSW5* gene trees.

trees and let the two gene trees only have the same seven species in the trait data.

## 5.2 Application of the two gene trees model with one trait

Since all of the rice varieties are Asian cultivated rices, their evolutionary processes follow artificial selection, one type of natural selection [5]. Therefore, the BM model is not the correct evolutionary model to use in the study. However, we are still interested in testing both the BM and OU models to see which one fits the data better.

Under the two gene trees BM model, we are interested in three parameters: contribution weight of *GW2* ( $p$ ), the rate of the evolution ( $\sigma^2$ ), and the mean trait value ( $\theta$ ). Be aware that the trait data are not centered. We use two different approaches to estimate the three parameters of interest: the AMH method and ABC method.



Under the AMH method, the priors are

$$p \sim \text{Beta}(a, b), \quad (5.1)$$

$$\sigma \sim \text{Folded} - \text{noncentral} - t(df = c, scale = d), \quad (5.2)$$

and

$$\theta \sim N(\mu = m, \sigma^2 = h^2), \quad (5.3)$$

where  $a = b = 0.5$ , or  $1$ ,  $c = d = 1$  or  $2$ , and  $m$  = the mean of the trait values =  $3.211$  and  $h^2$  = the variance of the trait values =  $0.197$ . The meaning of  $\theta$  is the trait value on the root of both gene trees.

Under the ABC method, the  $p$  prior is

$$p \sim \text{Beta}(1, 1), \quad (5.4)$$

and the prior for the mean value of the trait is

$$\theta \sim N(\mu = 3.211, \sigma^2 = 0.197), \quad (5.5)$$

and the priors for  $\sigma$  are

$$\sigma \sim \text{Folded} - \text{noncentral} - t(df = 2, scale = 2), \quad (5.6)$$

or

$$\sigma^2 \sim \text{Exp}(1), \quad (5.7)$$

or

$$\sigma^2 \sim U(0, 50). \quad (5.8)$$

Table 5.1 and Table 5.2 show the results of the ABC and AMH methods under the two gene trees BM model. Note that in Tables 5.2 and 5.3, **ln** means the maximized

value of the loglikelihood function of our models. The results show that with both methods, under different sets of priors, the estimates for the contribution weight  $p$  were less than 0.5. But with the ABC method, different sets of priors gave very different estimates of the rate of the evolution. For example, the estimate of  $\sigma^2$  was 4.16 under the set of priors,  $p \sim \text{Beta}(1, 1)$  and  $\sigma^2 \sim \text{Exp}(1)$ , while it was 46.60 under the set of priors,  $p \sim \text{Beta}(1, 1)$  and  $\sigma^2 \sim U(0, 50)$ . With the AMH method, different sets of priors gave very similar estimates of the rate of the evolution; however, the estimates of  $\sigma^2$  were more than 400 with very small trait values (the average grain width was 3.211 mm). Therefore, the estimates of the rate of evolution by using the two gene trees BM model were not identical, which may provide the evidence that the BM model doesn't fit our data.

Table 5.1: Empirical study of the ABC method under the BM model

Prior of $p$	$\hat{p}$	CI	Prior of $\sigma^2$				
$p \sim \text{Beta}(1, 1)$	0.05	(0, 0.28)	$\sigma^2 \sim \text{Exp}(1)$	4.16	(1.71, 6.83)	0.61	(0, 2.27)
$p \sim \text{Beta}(1, 1)$	0.04	(0, 0.28)	$\sigma^2 \sim U(0, 50)$	46.60	(30.97, 49.70)	0.27	(0, 1.74)
$p \sim \text{Beta}(1, 1)$	0.08	(0, 0.35)	$\sigma \sim \text{FNT}(2, 2)$	16.71	(5.30, 78.37)	0.17	(0, 2.11)

Table 5.2: Empirical study of the AMH method under the BM model

Prior of $p$	$\hat{p}$	CI	Prior of $\sigma$	$\hat{\sigma}^2$	CI	$\hat{\theta}$	CI	ln	AIC	BIC
Beta(0.5,0.5)	0.10	(0, 0.59)	FNT(1,1)	506.73	(88.80, 2284.52)	2.30	(2.49, 3.54)	-4.80	15.61	15.44
Beta(0.5,0.5)	0.07	(0, 0.46)	FNT(2,2)	410.22	(99.17, 1505.62)	2.99	(2.49, 3.52)	-3.22	12.45	12.28
Beta(1,1)	0.22	(0, 0.85)	FNT(2,2)	581.21	(103.35, 2884.83)	3.00	(2.50, 3.50)	-2.74	11.48	11.32

Under the two gene trees OU model, we are interested in four parameters: con-

tribution weight of  $GW2$  ( $p$ ), the rate of the evolution ( $\sigma^2$ ), the strength of selection ( $\alpha$ ), and the global optimum of the trait value ( $\theta$ ). The priors are

$$p \sim \text{Beta}(a, b), \quad (5.9)$$

$$\sigma \sim \text{Folded} - \text{noncentral} - t(df = c, \text{scale} = d), \quad (5.10)$$

$$\alpha \sim \text{Gamma}(\text{shape} = g, \text{rate} = f), \quad (5.11)$$

and

$$\theta \sim N(\mu = m, \sigma^2 = h^2), \quad (5.12)$$

where  $a = b = 0.5$ , or  $1$ ,  $c = d = 1$  or  $2$ ,  $g = f = 0.5$  or  $1$ ,  $m =$  the mean of the trait values =  $3.211$  and  $h^2 =$  the variance of the trait values =  $0.197$ .

Table 5.3: Empirical study of the AMH method under the OU method

Prior of $p$	$\hat{p}$	CI	Prior of $\sigma$	$\hat{\sigma}^2$	CI	Prior of $\alpha$	$\hat{\alpha}$	CI	$\hat{\theta}$	CI	ln	AIC	BIC
Beta(0.5,0.5)	0.24	(0, 0.68)	FNT(1,1)	22.89	(3.97, 90.81)	G(0.5,0.5)	0.50	(0, 4.17)	2.30	(2.48, 3.53)	-4.78	13.56	13.45
Beta(0.5,0.5)	0.41	(0, 0.80)	FNT(2,2)	27.56	(5.08, 107.67)	G(0.5,0.5)	0.51	(0, 4.14)	3.01	(2.51, 3.51)	-2.75	9.51	9.40
Beta(1,1)	0.42	(0, 0.80)	FNT(2,2)	27.18	(4.25, 105.78)	G(1,1)	0.72	(0, 3.02)	3.00	(2.50, 3.51)	-2.70	9.40	9.29

Table 5.3 shows the results of the AMH method under the two gene trees OU model. The results show that under different sets of priors, the estimates for the contribution weight  $p$  were less than 0.5. With the sets of priors  $p \sim \text{Beta}(0.5, 0.5)$  or  $p \sim \text{Beta}(1, 1)$  and  $\sigma^2 \sim \text{FNT}(2, 2)$ , the estimates of  $p$  and  $\sigma^2$  were close to each other. We note that the estimates of  $\sigma^2$  were not as big as the estimates from the AMH method with two gene trees BM model.

In order to figure out which model is the best one, we use AIC [81] and BIC [82] as our criteria for model selection. The corresponding AICs and BICs are presented in each table. After comparison, the best model is the two gene trees OU model with

one trait with priors

$$p \sim \text{Beta}(1, 1), \quad (5.13)$$

$$\sigma \sim \text{Folded} - \text{noncentral} - t(df = 2, scale = 2), \quad (5.14)$$

$$\alpha \sim \text{Gamma}(shape = 1, rate = 1), \quad (5.15)$$

and

$$\theta \sim N(\mu = 3.211, \sigma^2 = 0.197). \quad (5.16)$$

To show the convergence of our best model, Figure 5.2 shows the trace plots and histograms of  $p$ ,  $\sigma$ ,  $\alpha$ , and  $\theta$ .

### 5.3 Application of the one gene tree model with one trait

Here we use the same DNA alignments from the online multiple sequence alignment program Clustal Omega for *GW2* and *qSW5* and use the program Sequence Matrix [83] to concatenate both genes sequences. Then jModeltest 2.1.10 [13, 14] is used to test the best substitution model for MrBayes [72, 73] to infer the concatenated gene trees for *GW2* and *qSW5*. Figure 5.3 shows the phylogeny for estimating the parameters of interests.

Since only one gene tree is considered, we don't need to estimate the contribution weight ( $p$ ). Therefore, under the BM model, two parameters must be estimated: the mean and variance. Under the OU model, the strength of selection, the global optimum of trait values, and the variance of pure selection are estimated. We use the R package *OUCH* [42, 85] to estimate the parameters of interest for both the BM and OU models.

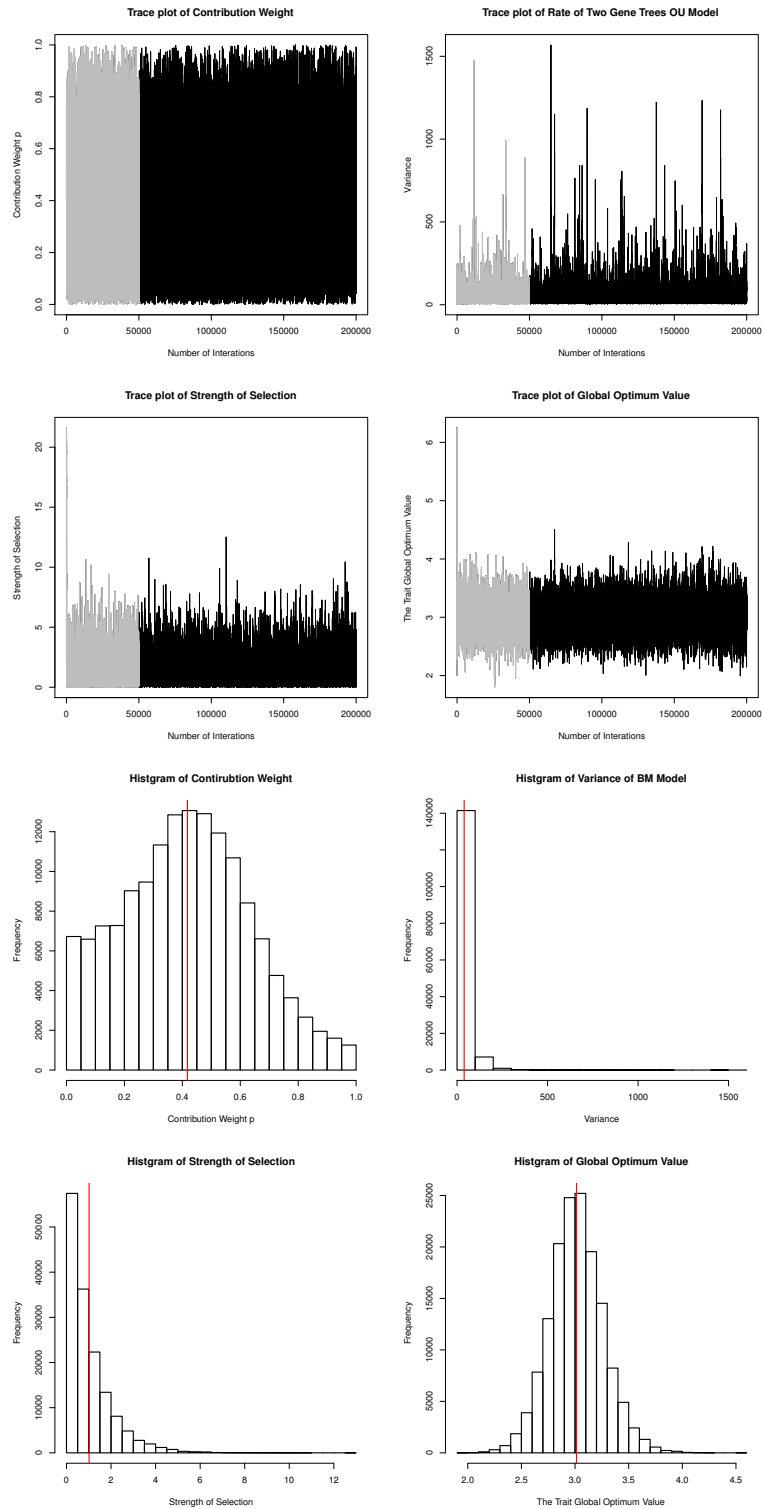


Figure 5.2: Trace plots and histograms in our best model. Grey indicated burnin for the trace plots. Histograms don't include the burnin.

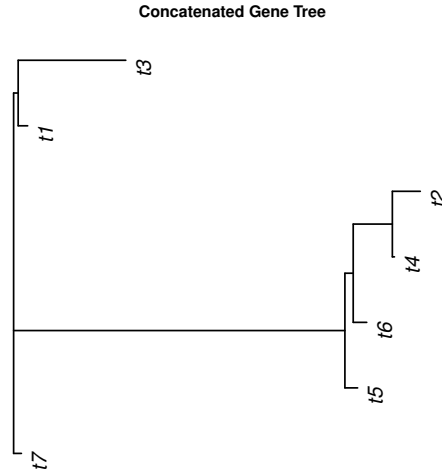


Figure 5.3: The concatenated gene tree for *GW2* and *qSW5*.

Under the BM model,  $\hat{\theta} = 2.68$ ,  $\hat{\sigma}^2 = 0.935$  with  $\log\text{likelihood} = -1.710$ ,  $AIC = 7.42$ , and  $BIC = 7.31$ . Under the OU model,  $\hat{\sigma}^2 = 1.627$ ,  $\hat{\alpha} = 4.989$  and  $\hat{\theta} = 3.208$  with  $\log\text{likelihood} = -2.337$ ,  $AIC = 10.67$ , and  $BIC = 10.51$ . If we use the AIC and BIC to select the best model, the BM model with one gene tree outweighs than the OU model with one gene tree. However, we don't think this comparison is applicable in this situation because the evolution processes of rice varieties are from an artificial selection, which suggests that the OU model should be utilized. Thus, we need to compare the two gene trees OU model with one gene tree OU model to see whether two gene trees model fits better. After comparison, we found that the best model is the two gene trees OU model with one trait. The estimates from our study were found to agree with the statement in the published paper [44] we mentioned in the beginning of this section . Specifically, we found that the sample contribution weight to be 42% and 58% for *GW2* and *qSW5*, respectively. This agrees with the statement in the paper [44] above, which indicated that *qSW5* is the most important quantitative trait

loci (QTL) for grain width.

# Chapter 6

## Discussion

### 6.1 AMH and ABC methods

Bayesian methods have been widely used in inferring both phylogenetic trees [2, 7, 25, 72, 73] and evolutionary processes in comparative methods [86, 87, 88]. In our study, we have proposed the two gene trees evolutionary model for one trait by using two different Bayesian approaches: (1) the adaptive Metropolis-Hastings (AMH) method [63, 64] and (2) approximate Bayesian computation (ABC) method [57, 58, 59, 60, 61, 62]. The main difference between these two methods is whether the likelihood is calculated. Therefore, the ABC method has been called likelihood-free method. There are two reasons that the ABC method has become popular in genetics research. One reason is that without calculating the likelihood, the procedure of this method becomes easy and understandable. The second reason is that the ABC method can handle lots of error in the process of DNA extraction, purification, and alignment without any difficulty. However, this method has some issues. The first one is that the ABC method can only handle small numbers of parameters. Because



of this, the ABC method has not been applied to the two gene trees OU model as it involves a large number of parameters. Therefore, a more generalized ABC method [58] instead of the ABC rejection method, which is used in our study, should be applied in future studies. The second issue is that the ABC method is very sensitive to the size of data of interest [58]. Because of this, the simulation study indicated that the results of the application of the two gene trees BM model with a large number of taxa are worse than the results from the AMH method. So, except considering a more generalized ABC method, increased simulation size and increased tolerance value or acceptance percentile should be considered [62] in the future. The third issue is the lack of model selection for the ABC method. The Bayes factors could be used to compare different models' posterior probabilities [89, 90, 91] under some conditions [92, 93]. Since researchers have not reached an agreement about model selection for the ABC method [92, 93], the rule of thumb for model selection in our study is to check the rate of the trait evolution. In the empirical study, since the estimates of the evolution rates of rice varieties are not identical ranging from a very small number to a large number under the ABC method using different sets of priors, our conclusion is that the estimation from the ABC method is not accurate. In our study, the AMH method performs better than the ABC method. It converges very fast even with small dataset which estimating the same number of parameters of interest when the number of taxa is 4 and 8 with longer internal branch lengths and very different topologies. We will use the AMH method for our future model extension with more parameters.

## 6.2 One gene tree and two gene trees for dependency

After Felsenstein's paper *Phylogenies and the Comparative Method* published in 1985, biologists have used one phylogenetic tree as a tool to handle dependent data for the traits of interest and hence utilized one gene only in comparative methods. Without the phylogeny, the assumption of using the comparative methods to assess the relationships of phenotypic characters is assuming all species evolving independently, which is biologically unrealistic. Constructing a variance-covariance matrix from a phylogeny through a generalized linear model, makes modeling evolutionary signals for traits possible probabilistically. Because of this, biologists have been focusing on developing more complicated models ranging (1) from a simplest BM model [27] to an OU model [42], (2) from a constant evolutionary rate model to a model with multiple evolutionary rates and different strengths of selection in different time regions [94, 95], and (3) from a single model with one trait to a combined model with more than one trait, such as the OUBM or the OUOU model [96, 97]. More established models are reviewed by O'Meara in 2012 [98]. However, there is a blind spot in these studies regarding whether the evolutionary processes of a trait could be presented by only one phylogenetic tree from one gene. If we consider the phylogenetic tree as a gene tree, the answer is negative since it is rare that one trait is influenced by only one gene. If we consider the phylogenetic tree as a species tree, the answer is also negative since there are conflicts between the estimation of species trees and the estimation of different gene trees throughout the genome [99]. This means that species trees are not always the same as the gene trees. Therefore, using one phylogeny is a

simplified way to deal with nonindependent data from biology. The two gene trees model was considered in our study. One benefit of considering the two gene trees model is that more evolutionary signals could be captured through this model. In the empirical study, when applying model selection, according to both AIC and BIC, between one gene tree model versus two gene trees model, it was found that one gene tree BM model has the lowest AIC and BIC. One may think that in this case, the one gene tree model is better than two gene trees model. However, the one gene tree BM model seems implausible because (1) the trait of interest grain width of rice varieties is under artificial selection and (2) BM model is only the case of no selection, which is totally a different type of evolutionary selection relative to artificial selection.

For the two gene trees model, the BM model has two parameters of interest, the contribution weight of one gene,  $p$ , and variance of the BM model,  $\sigma^2$ , under the assumption of equal variance in both genes. The most difficult aspect of the two gene trees model is whether the parameters are identifiable. In particular, if each gene has its own variance, then the effect of higher variance might be similar to a high value of the contribution weight. Therefore, we can easily extend our proposed two gene trees model to a more generalized model without the assumption of equal variance in both genes. We can use the variances from the two genes to weigh the contribution of them. However, it is really hard to figure out whether the higher weight gene is the one with a bigger variance in the BM model. Maybe biologists have different thoughts of this issue. Due to the identifiability issue in the two gene trees model, we have imposed a very strong assumption of equal variances for trait evolutionary model with two genes. One possible solution would be to scale the branch lengths in the two gene trees to transform the unequal variances to equal variances with some

parameters, which would be a topic for future work.

### **6.3 Simple models versus more complicated models for accuracy**

With the increased convenience in collecting and sequencing DNA data, inferring phylogenies became popular, and therefore many related research areas are being developed. For example, a lot of methods for modeling trait evolution have been proposed, such as (1) continuous-time Markov chain models with finite states, (2) the multivariate normal distribution models with continuous traits, and (3) models with birth-death processes reviewed by O’Meara in 2012 [98]. Therefore, given trait data, a phylogeny, and one of the above methods, evolutionary inferences for the past or the ongoing evolutionary processes, such as the rate of a trait evolution or more than one correlated traits evolution, could be estimated. Among so many different models, with increased DNA data, a more complicated model might be preferred in order to correspond to complexity and reality. However, the question is how complex the model should be since more complicated models involve more parameters. For example, moving from one gene tree BM model to one gene tree OU model with consideration of a signal trait, two parameters are added. In another examples, moving from one gene tree BM model to two gene trees OU model with one trait under the assumption of equal variance, equal strength of selection, and equal optimum trait value, four parameters are added. If we want to extend the two gene trees OU model with one trait to a more generalized two gene trees OU model with two correlated traits without the assumption of equal variance, equal strength of selection, and equal optimum trait

value, four different variations need to be considered: (1) the variance among species within one trait, (2) the variation among two correlated traits in one species, (3) the variation of optimum value split between two genes, and (4) the variation of strength selection between two correlated traits. So at least 10 parameters of interest we need to estimate. The increased number of parameters raises an important issue about how complicated models we should have to balance the sufficiency and accuracy of our estimation. It is really hard to simply make a conclusion. My point of view of this issue is that with more data, you may need to have a more complicated model.

## 6.4 Conclusion

There are two new trait evolutionary models proposed: two gene trees BM model and two gene trees OU model, with two different Bayesian approaches for estimating of the parameters of interest: the adaptive Metropolis-Hastings (AMH) method and approximate Bayesian computation (ABC) method. Under the two gene trees BM model with one trait, the AMH approach performed well on gene tree with longer internal branch lengths and very different topologies regardless of the number of taxa, but the ABC method only performed well on small numbers of taxa. With random simulated trees with similar branch lengths and topologies, under the two gene trees BM model with one trait, the AMH approach performed well with larger number of taxa. In the empirical study, the two gene trees OU model with one trait worked better in the real case than the one gene tree OU model with one trait. Therefore, the two gene trees trait evolutionary model is a useful addition to current methods.

## 6.5 Potential future work

One open question related to the current work is how to deal with the identifiability issue in the two gene trees model. Potential work is to propose a method to scale the variation between two gene trees in the same level. Therefore, the higher weight gene will be associated with a bigger variation of BM model. In this case, we can extend the study from the two gene trees BM or OU model with one trait to multiple gene trees BM or OU model with more than one trait. We are also interested in using a more generalized ABC method to estimate the parameters of our new proposed trait models and extending the simulation study to test the effect of misestimated ultrametric trees to misestimated random trees.

# Appendix A

## Data and Command Files in the Empirical Study

### A.1 The trait data in the empirical study

The trait data in the empirical study are as follows:

Table A.1: The trait data in the empirical study.

$t_1$	2.6563
$t_2$	3.7828
$t_3$	3.0804
$t_4$	3.2198
$t_5$	3.4965
$t_6$	3.5860
$t_7$	2.6575

## A.2 Command files for MrBayes in the empirical study

We use program MrBayes to infer the phylogenetic trees for *GW2*, *qSW5* and the concatenated gene tree from *GW2* and *qSW5*. Here we list the commands in program MrBayes to reconstruct the phylogenetic trees for *GW2* in the empirical study.

```
#NEXUS
begin data;
dimensions ntax=10 nchar=6792;
format interleave datatype=DNA missing=N gap=-;

matrix
GW2-9 GATACAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA
GW2-10 GATACAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA
GW2-5 GATAAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA
GW2-8 GATACAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA
GW2-1 GATACAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA
GW2-2 GATACAAAATGCAACCTGGACCCACGTAACCCCTCC——ACCTC
GW2-3 GATACAAAAAGCAACCTAGACCCACGTAACCCCTCC——ACCTC
GW2-4 GATACAAAATGCAACCTGGACCCACGTAACCCCTCC——ACCTC
GW2-6 GATACAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA
GW2-7 GATACAAAAAGCAACCTGGACCCACGTAACCCCTCCACCTCACGGCACA

GW2-9 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-10 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-5 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-8 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-1 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-2 ACTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-3 ACTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-4 ACTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-6 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG
GW2-7 CTTGTACATCCAGCTAGAGATCATCCTACGTAGTACATCTATCTTCACCG

GW2-9 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-10 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-5 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCTGCTTCCGCACGC
GW2-8 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-1 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-2 CCCGCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-3 CCCGCCACCCAAAAACCGTCACGTCTCCGCTGCAATCCGCTTCCGCACGC
GW2-4 CCCGCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-6 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC
GW2-7 CCCACCCACCCAAAAACCGTCACGTCTCCACTGCAATCCGCTTCCGCACGC

GW2-9 CATATATACAAAAAGAGGCTGACCAACACGAAAACATCTAAAACATCTTTA
GW2-10 CATATATACAAAAAGAGGCTGACCAACACGAAAACATCTAAAACATCTTTA
```



GW2-5 CATATATACAAAAGAGGCTGACCAACACGAAAACATCTAAACAATCTTTA  
GW2-8 CATATATACAAAAGAGGCTGACCAACACGAAAACATCTAAACATCTTTA  
GW2-1 CATATATACAAAAGAGGCTGACCAACACGAAAACATCTAAACATCTTTA  
GW2-2 CATATATACAAAAGATGCTGACCAACACGAAAACATCTAAACATCTTTA  
GW2-3 CATATATACAAAAGATGCTGACCAACACGAAAACATCTAAACATCTTTA  
GW2-4 CATATATACAAAAGATGCTGACCAACACGAAAACATCTAAACATCTTTA  
GW2-6 CATATATACAAAAGAGGCTGACCAACACGAAAACATCTAAACATCTTTA  
GW2-7 CATATATACAAAAGAGGCTGACCAACACGAAAACATCTAAACATCTTTA

GW2-9 AGCTAAAATACATATAA—AAAAAGTTAACAAACATGAAAAACACTC  
GW2-10 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-5 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-8 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-1 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-2 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-3 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-4 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-6 AGCTAAAAAACATATAACAAAAAAAAGTTAACAAACATGAAAAACACTC  
GW2-7 AGCTAAAAAACATATAA—AAAAAGTTAACAAACATGAAAAACACTC

GW2-9 TCATCTCACATCTATAAAACCACGGAGAAAACAGGTAGTATAAAGCATAG  
GW2-10 TCATCTCACATCTATAAAACCACGGAGAAAACAGGTAGTATAAAGCATAG  
GW2-5 TCATCTCACATCTATAAAACCACGGAGAAAACAGGTAGTATAAAGCATAG  
GW2-8 TCATCTCACATCTATAAAACCACGGAGAAAACAGGTAGTATAAAGCATAG  
GW2-1 TCATCTCACATCTATAAAACCACGGAGAAAACAGGTAGTATAAAGCATAG  
GW2-2 TCATCTCACATCTATAAAACCACGGAAAAAACAGGTAGTATAAAGCATAG  
GW2-3 TCATCTCACATCTATAAAACCACGGAAAAAACAGGTAGTATAAAGCATAG  
GW2-4 TCATCTCACATCTATAAAACCACGGAAAAAACAGGTAGTATAAAGCATAG  
GW2-6 TCATCTCACATCTATAAAACCACGGAGAAAACAGGTAGTATAAAGCATAG  
GW2-7 TCATCTCACATCTATAAAACCACAGAGAAAACAGGTAGTATAAAGCATAG

GW2-9 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-10 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-5 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-8 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-1 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-2 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-3 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-4 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-6 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC  
GW2-7 ACAAAGGTACCACCAGCATCCTCATCCCCACCACCTCTCTGAAATCTAC

GW2-9 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-10 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-5 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-8 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-1 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-2 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-3 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
GW2-4 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC

GW2-6 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC  
 GW2-7 TACTACCACCGCTATCCCCACAAGCGGCTAGACAGGCAGCCCCCGCCCC

GW2-9 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-10 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-5 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-8 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-1 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-2 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-3 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-4 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-6 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC  
 GW2-7 GTTGCAAACCTGCGGCTATGTGGCTCTGGCACCGAGATTATTATAGCGACC

GW2-9 GCAGCAGCAGCAG—CAGCAGCAGAAGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-10 GCAGCAGCAGCAGCAGCAGCAGA—AGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-5 GCAGCAGCAGCAGCAGCAGCAGA—AGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-8 GCAGCAGCAGCAGCAGCAGCAGA—AAGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-1 GCAGCAGCAGCAGCAGCAGCAGA—AGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-2 GCAGCAGCAGCAGCAGCAGCAGAAGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-3 GCAGCAGCAGCAGCAGCAGCAGAAGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-4 GCAGCAGCAGCAGCAGCAGCAGAAGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-6 GCAGCAGCAGC—AGCAGCAGCAGAAGCTAGGCTACAGAGGTTCGCAGTC  
 GW2-7 GCAGCAGCT—GCAGCAGCAGAAGCTAGGCTACAGAGGTTCGCAGTC

GW2-9 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-10 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-5 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-8 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-1 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-2 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-3 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-4 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-6 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT  
 GW2-7 TCTCTCGCAGTCGCAGCGGCAGAGACAAGGCCGCTCTTCTCTTCTTCTTCT

GW2-9 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-10 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-5 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-8 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-1 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-2 TCCTTTCCTTGTGCATAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-3 TCCTTTCCTTGTGCATAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-4 TCCTTTCCTTGTGCATAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-6 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG  
 GW2-7 TCCTTTCCTTGTGCGTAGCGTAGCGGCTAGCTGTTGCAGCAGCAGAGAGG

GW2-9 TCTCCACTTCCCTTTCCTCCCACCCCGAGAAAGCCAAAAAAGAAAAA  
 GW2-10 TCTCCACTTCCCTTTCCTCCCACCCCGAGAAAGCCAAAAAAGAAAAAT  
 GW2-5 TCTCCACTTCCCTTTCCTCCCACCCCGAGAAAGCCAAAAAAGAAAAAT



GW2-7 GAGGAGGGGGAGTGGTGAGGGTTTCATCTGCGGAGGAGGAGGGAGGAGGA

GW2-9 ——GAGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-10 —————GG

GW2-5 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-8 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-1 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-2 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-3 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-4 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-6 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-7 GGGAGGAGGGTAGATCTGGGAGGGGGATGGGGAACAGGATAGGGGGGAGG

GW2-9 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-10 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-5 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-8 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-1 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-2 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-3 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-4 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-6 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-7 AGGAAGGCGGGGTGGAGGAGAGGTACACGAGGCCGCAGGGGCTGTACGA

GW2-9 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-10 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-5 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-8 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-1 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-2 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-3 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-4 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-6 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-7 GCACAGGGACATCGACCAGAAGAAGCTCCGGAAGCTGATCCTCGAGGCCA

GW2-9 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-10 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-5 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-8 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-1 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-2 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-3 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-4 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-6 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-7 AGCTCGGCCGTGCTACATGGGCGCCGACGACGCCGCCGCCGCCGAC

GW2-9 CTCGAGGAGTGCCCCATCTGCTTCTTGGTATAAAACCGCCTCCGGGGAGA

GW2-10 CTCGAGGAGTGCCCCATCTGCTTCTTGGTATAAAACCGCCTCCGGGGAGA

GW2-5 CTCGAGGAGTGCCCCATCTGCTTCTTGGTATAAAACCGCCTCCGGGGAGA

GW2-8 CTCGAGGAGTGCCCCATCTGCTTCTTGGTATAAAACCGCCTCCGGGGAGA

GW2-1 CTCGAGGAGTGCCCCATCTGCTTCTGGTATAAAAACCGCCTCCGGGGAGA  
GW2-2 CTCGAGGAGTGCCCCATCTGCTTCTGGTATAAAAACCGCCTCCGGGGAGA  
GW2-3 CTCGAGGAGTGCCCCATCTGCTTCTGGTATAAAAACCGCCTCCGGGGAGA  
GW2-4 CTCGAGGAGTGCCCCATCTGCTTCTGGTATAAAAACCGCCTCCGGGGAGA  
GW2-6 CTCGAGGAGTGCCCCATCTGCTTCTGGTATAAAAACCGCCTCCGGGGAGA  
GW2-7 CTCGAGGAGTGCCCCATCTGCTTCTGGTATAAAAACCGCCTCCGGGGAGA

GW2-9 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-10 TGTGGCGGGTGGACGGGAGGAAGAGGTCTCTCCCCTCTTCCCTCACTTCCT  
GW2-5 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-8 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-1 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-2 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-3 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-4 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-6 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT  
GW2-7 TGTGGCGGGTGGACGGGAGGAAGAGATCTCTCCCCTCTTCCCTCACTTCCT

GW2-9 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTTTTTTGGTGGG  
GW2-10 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTT-TTTTGGTGGG  
GW2-5 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTT-TTTTGGTGGG  
GW2-8 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTTTTTTGGTGGG  
GW2-1 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTTTTTTGGTGGG  
GW2-2 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTT-TTTTGGTGGG  
GW2-3 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTT-TTTTGGTGGG  
GW2-4 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTT-TTTTGGTGGG  
GW2-6 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTTTTTTGGTGGG  
GW2-7 CCTCCCGCTTCCCTCCACTTCAGCTTTTTTCTCTCTTTTTTTTTTGGTGGG

GW2-9 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-10 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-5 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-8 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-1 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-2 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-3 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-4 ATTTTGTGCGTAATGCTACTACTAGTACTTGTGTTGCGAGAGTAGGGCGT  
GW2-6 ATTTTGTGCGTAATGCTACTACTAGTACTTATTGTTGCGAGAGTAGGGCGT  
GW2-7 ATTTTGTGCGTAATGCTACTACTAGTACTTATTGTTGCGAGAGTAGGGCGT

GW2-9 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-10 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-5 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-8 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-1 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-2 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-3 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-4 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-6 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG  
GW2-7 GGATTGCGTGTGGCCGGGTGGTGGCGTGGAGCGTTTCGATTAGGGGGTTG

GW2-9 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-10 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-5 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-8 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-1 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-2 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-3 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-4 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-6 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG  
GW2-7 ATTACTTTTGCTCCTGTTTCGATTGTGGAGAGTGCCGGTGGTGTGTGGCCG

GW2-9 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-10 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-5 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-8 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-1 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-2 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-3 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-4 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-6 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC  
GW2-7 TTATGGAGATGGATGACCACGCGCAGTGTTTTGGGGCAAGAATTTTCAGTC

GW2-9 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG  
GW2-10 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG  
GW2-5 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG  
GW2-8 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG  
GW2-1 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG  
GW2-2 GACATTGGCACGGTTTTGGGGTTGTTTGTATTGTTCTTGGATAATTAGTG  
GW2-3 GACATTGGCACGGTTTTGGGGTTGTTTGTATTGTTCTTGGATAATTAGTG  
GW2-4 GACATTGGCACGGTTTTGGGGTTGTTTGTATTGTTCTTGGATAATTAGTG  
GW2-6 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG  
GW2-7 GACATTGGCACGGTTTTGGGGTTGTTTGTGTTGTTCTTGGATAATTAGTG

GW2-9 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTCCT  
GW2-10 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTATT  
GW2-5 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTCCT  
GW2-8 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTCCT  
GW2-1 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTCCT  
GW2-2 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTATT  
GW2-3 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTATT  
GW2-4 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTATT  
GW2-6 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTCCT  
GW2-7 CGGTTTCGTGGAGCTGATTTTGCGAAATGTAGTAGTAGTGTTTGAATTCCT

GW2-9 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-10 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-5 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-8 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-1 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG

GW2-2 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-3 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-4 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-6 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG  
GW2-7 TCTGGTAGTTAGTAGTGGTATCATACTTCGTATGTACATTGTCCGCAATG

GW2-9 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-10 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-5 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-8 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-1 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-2 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-3 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-4 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-6 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA  
GW2-7 TTCTCCTAAATTGTTTTGTTTTCTTAATCATAGTTTACTTACTTTTATTA

GW2-9 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-10 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-5 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-8 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-1 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-2 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-3 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-4 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-6 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT  
GW2-7 TGTAATCTGAGATACGTTTTTTTTGGGATAATTATAGTCTAGGAGGTTTT

GW2-9 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-10 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-5 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-8 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-1 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-2 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-3 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-4 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-6 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC  
GW2-7 GGGTACTCATTATTATTGTCAATATTCTATAGGCCTATCTATTGGCTGC

GW2-9 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-10 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-5 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-8 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-1 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-2 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-3 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-4 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-6 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG  
GW2-7 CTGGCTGTGATGCATTGCATGAAATTTAATGTTTACTGCCACTCCCGGGG

GW2-9 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-10 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-5 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-8 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-1 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-2 TTGGGACATTACAGACCTTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-3 TTGGGACATTACAGACCTTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-4 TTGGGACATTACAGACCTTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-6 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT  
GW2-7 TTGGGACATTACAGACCGTATAGGAAATTTGACTAGTTTCATAAATGAAT

GW2-9 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-10 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-5 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-8 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-1 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-2 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-3 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-4 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-6 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG  
GW2-7 AATTGGCCAACCTGTTCTGCAATGTTGGGTGCCTTCAATTTCAAATAGTGG

GW2-9 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-10 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-5 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-8 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-1 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-2 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-3 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-4 ATGCTCTATAGAGTACGGAGGTACCATGGTGATATTTGTTTGTGTGTGT  
GW2-6 ATGCTCTATAGAGTACGGAGGTACTATGGTGATATTTGTTTGTGTGTGT  
GW2-7 ATGCTCTATAGAGTACGGAGGTACTATGGTGATATTTGTTTGTGTGTGT

GW2-9 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-10 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-5 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-8 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-1 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-2 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-3 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-4 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-6 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT  
GW2-7 AATACTTTATTTTGGTGTATTAATGCTCTTGGGAACCGAGATTGAGAAT

GW2-9 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-10 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-5 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-8 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-1 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-2 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT



GW2-3 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-4 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-6 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT  
GW2-7 CTATTGTAGTGAACCTTGATTCTGTTAGCTGTGCCAGAGTATTTTCGATGT

GW2-9 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTTTTAAAAA  
GW2-10 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTT-AAAAAA  
GW2-5 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTTTTAAAAAA  
GW2-8 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTTTTAAAAAA  
GW2-1 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTTTTAAAAAA  
GW2-2 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTT-AAAAAA  
GW2-3 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTT-AAAAAA  
GW2-4 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTT-AAAAAA  
GW2-6 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTTTTAAAAAA  
GW2-7 CTCTTGTCAAGCCTGCATTTATTTGGCACAAAAGTAGATTTTTTTTAAAAAA

GW2-9 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-10 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-5 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-8 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-1 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-2 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-3 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-4 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-6 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT  
GW2-7 AAATTCTGCAGCTTAAATGCTACTAGCTTACTGCAATTAATGTTATTACT

GW2-9 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-10 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-5 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-8 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-1 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-2 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-3 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-4 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-6 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC  
GW2-7 ATATAGGAATTTAATGCAATTCTATATGGTTGCAATTTATTGCAATTC

GW2-9 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-10 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-5 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-8 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-1 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-2 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-3 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-4 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-6 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA  
GW2-7 TGCTTCATTAGATCTTTAGCTCAAAATCACTCAATAAAATTTGGTTTTAGA

GW2-9 TGAAATCATGAACTTGATTATAGGAGTATATTATATTTGGTGTGAAT

GW2-10 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-5 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-8 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-1 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-2 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-3 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-4 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-6 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT  
GW2-7 TGAAATCATGAAACTTGTATTATAGGAGTATATTATATTTGGTGTGAAT

GW2-9 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-10 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-5 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-8 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-1 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-2 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-3 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-4 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-6 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG  
GW2-7 TATATGTTTGCATTTGTGCTAATAGTGGTATTTAAACAATTATCACTTCTG

GW2-9 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-10 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-5 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-8 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-1 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-2 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-3 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-4 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-6 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT  
GW2-7 GTATTTCTGAGTTCAAACCTGTTTGACAACCACTCCTGTCTGAAATGCAT

GW2-9 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-10 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-5 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-8 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-1 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-2 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-3 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-4 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-6 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC  
GW2-7 GTTTTCTTTCTTTGTAGTACTACCCAAGTCTTAACCGATCAAAGTGTTC

GW2-9 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT  
GW2-10 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT  
GW2-5 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT  
GW2-8 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT  
GW2-1 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT  
GW2-2 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT  
GW2-3 TCAAAGGGATATGCACCGTAATACATCTATCCTATAACTAACAATTGT

GW2-4 TCAAAAGGGATATGCACCGGTAATACATCTATCCTATAACTAACAATTGT  
GW2-6 TCAAAAGGGATATGCACCGGTAATACATCTATCCTATAACTAACAATTGT  
GW2-7 TCAAAAGGGATATGCACCGGTAATACATCTATCCTATAACTAACAATTGT

GW2-9 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-10 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-5 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-8 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-1 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-2 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-3 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-4 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-6 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT  
GW2-7 GTTTTACTTGTTAGGACCATATACCCTTATATTCAATGGTTGGTTTTTTTT

GW2-9 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-10 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-5 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-8 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-1 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-2 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-3 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-4 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-6 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA  
GW2-7 GCAGAGTGCTTTCTCCAAATGAAACCAACTCACACTGCTCAGCCTACACA

GW2-9 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-10 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-5 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-8 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-1 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-2 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-3 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-4 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-6 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA  
GW2-7 GTATCCTTTGTGTACCCCTTGATTATATTTTGTAACCTTGGCAGTTGGCA

GW2-9 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-10 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-5 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-8 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-1 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-2 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-3 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-4 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-6 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT  
GW2-7 ATCTATATTCCTTTTTATGAAAAAATACTTAACCAATAAAGATGTCCAT

GW2-9 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAGACAAAAG  
GW2-10 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAGACAAAAG

GW2-5 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-8 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-1 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-2 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-3 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-4 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-6 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG  
GW2-7 TCTGCAAAACTCCCAGTTATGCTGTGGAGTATCGTGGTGTAAAGACAAAG

GW2-9 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-10 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-5 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-8 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-1 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-2 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-3 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-4 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-6 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT  
GW2-7 GAGGAAAGGAGCATAGAACAATTTGTAAGTAATTTATATTCTGATGTTTT

GW2-9 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-10 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-5 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-8 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-1 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-2 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-3 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-4 TTTTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-6 TT-TTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC  
GW2-7 TT-TTTGTTTTGCCAAATGCATTTTATCTCACATAATATGTGTTAC

GW2-9 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-10 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-5 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-8 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-1 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-2 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-3 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-4 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-6 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA  
GW2-7 AGGAAGAGCAGAAAGTCATAGAAGCACAAATGAGGATGCGCCAGCAAGCA

GW2-9 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-10 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGACCAGGTGCTCTTC  
GW2-5 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-8 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-1 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-2 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-3 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-4 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC

GW2-6 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC  
GW2-7 CTTCAAGATGAAGAAGATAAGATGAAAAGAAAACAGAACAGGTGCTCTTC

GW2-9 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-10 TAGCAAAAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-5 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-8 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-1 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-2 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-3 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-4 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-6 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA  
GW2-7 TAGCAGAACAATCACACCGACCAAAGAAGTGGAGTATAGAGATATTTGCA

GW2-9 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-10 GCACATCCTTTTCAGGTCTGCCCGGGCTGCAACTAACATAGAAATTTAGT  
GW2-5 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-8 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-1 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-2 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-3 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-4 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-6 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT  
GW2-7 GCACATCCTTTTCAGGTCTGCACAGGCTGCAACTAACATAGAAATTTAGT

GW2-9 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-10 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-5 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-8 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-1 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-2 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-3 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-4 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-6 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT  
GW2-7 AATGTACCATTTCTTCTGCTTGATGTGGTAATTTACTATGTGTTATGTCT

GW2-9 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-10 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-5 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-8 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-1 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-2 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-3 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-4 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-6 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC  
GW2-7 GTTTCAGTGCCGTCATACCGATGTGCTGAGCAAGAACTGAATGCTGTTC

GW2-9 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-10 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-5 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA

GW2-8 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-1 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-2 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-3 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-4 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-6 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA  
GW2-7 ATCGGAACCTTCATGCTCTGCCAGACTAGCATGCGCCCTTTCCATTCTA

GW2-9 GGCATAACCGGTATGTTATTATCTTTTCTCTGAGTTTTAGGGTGTCAAT  
GW2-10 GGCATAACCGGTATGTTATTATCTT-TTCTCTGAGTTTTAGGGTGTCAAT  
GW2-5 GGCATAACCGGTATGTTATTATCTT-TTCTCTGAGTTTTAGGGTGTCAAT  
GW2-8 GGCATAACCGGTATGTTATTATCTT-TTCTCTGAGTTTTAGGGTGTCAAT  
GW2-1 GGCATAACCGGTATGTTATTATCTT-TTCTCTGAGTTTTAGGGTGTCAAT  
GW2-2 GGCATAACCGGTATGTTATTATCTTTTCTCTGAGTTTTAGGGTGTCAAT  
GW2-3 GGCATAACCGGTATGTTATTATCTTTTCTCTGAGTTTTAGGGTGTCAAT  
GW2-4 GGCATAACCGGTATGTTATTATCTTTTCTCTGAGTTTTAGGGTGTCAAT  
GW2-6 GGCATAACCGGTATGTTATTATCTTTTCTCTGAGTTTTAGGGTGTCAAT  
GW2-7 GGCATAACCGGTATGTTATTATCTTTTCTCTGAGTTTTAGGGTGTCAAT

GW2-9 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-10 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-5 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-8 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-1 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-2 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-3 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-4 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-6 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT  
GW2-7 GTAGCCTGGTTATGTTGAACTGCAAACCTTACTACACTTCTATGTTAAAT

GW2-9 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-10 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-5 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-8 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-1 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-2 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-3 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-4 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-6 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA  
GW2-7 ACCCTACCTCAACCATGTTGAAGGGAAGCAAATGAGAAGTATTGTTAAAA

GW2-9 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-10 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-5 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-8 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-1 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-2 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-3 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-4 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA  
GW2-6 TAGGATAAAGGAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA

GW2-7 TAGGATAAAGGAAAAGTAACAGTATTATCTATAGAATCAAATCGATAACAA

GW2-9 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-10 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-5 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-8 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-1 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-2 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-3 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-4 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-6 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC  
GW2-7 AGATCAACCGTTGGATTAAATTGGATAGTGGGGACCAGACACCATAAAAC

GW2-9 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-10 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-5 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-8 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-1 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-2 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-3 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-4 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-6 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA  
GW2-7 ACTGGCATTTCGGACAAGATCATTAAATTATTCCACTTTAAGCTTGTGCA

GW2-9 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-10 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-5 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-8 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-1 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-2 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-3 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-4 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-6 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG  
GW2-7 GTCTGTCCTGATCCACTGCACTATTTAAGTGGATGAGAAGCTGTGATTGG

GW2-9 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-10 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-5 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-8 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-1 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-2 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-3 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-4 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-6 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT  
GW2-7 AGTTACTGTTCTAGTGTTAAGGTTGACTTGTATCAGATACAATTGCCTAT

GW2-9 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-10 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-5 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-8 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA

GW2-1 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-2 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-3 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-4 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-6 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA  
GW2-7 CCAGTGACAAAGTTTGAATGTTAATTTGTGAAAGTGTGAATGGAGAGAA

GW2-9 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-10 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-5 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTACTT  
GW2-8 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-1 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-2 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-3 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-4 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-6 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT  
GW2-7 TTTCTGCCCACATCATTGTTGTACCATGATAGTTACAGTAATCTTTAGTT

GW2-9 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-10 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-5 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-8 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-1 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-2 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-3 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-4 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-6 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG  
GW2-7 TATCTATACTCTAGAGGGATAAAAAACACCAAAACAGTTTGATCAAAATTG

GW2-9 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-10 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-5 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-8 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-1 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-2 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-3 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-4 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-6 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA  
GW2-7 TTCGAGGTAAATCCTAAGTATAAAATTGAGAATGGATTTGTTTTCAATA

GW2-9 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-10 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-5 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-8 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-1 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-2 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-3 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-4 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-6 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT  
GW2-7 TGGTTATTTGTTATCGACATTCAGTGTTTTTATTTACACAATGGCCAAT



GW2-9 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-10 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-5 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-8 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-1 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-2 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-3 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-4 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-6 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG  
GW2-7 TCCGAAAAACAATCTGGGTTCTTTTATCCCACTGTCCATACCAAGTTG

GW2-9 ATGTAGCAAGTAACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-10 ATGTAGCAAGTAACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-5 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-8 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-1 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-2 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-3 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-4 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-6 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG  
GW2-7 ATGTAGCAAGTAACTACTACTACACAAAATAGTATATTGTAGCCCTTTACATG

GW2-9 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-10 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-5 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-8 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-1 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-2 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-3 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-4 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-6 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC  
GW2-7 CATCCAATAGATTAAGTGGTCTTCAATCCAAGGATGAAATAAGCAATAAC

GW2-9 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-10 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-5 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-8 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-1 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-2 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-3 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-4 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-6 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG  
GW2-7 GTAGGTTCCGATACCAAATTAGCACAAAACAATGCTAAATGCTTTTAGAG

GW2-9 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-10 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-5 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-8 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-1 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG

GW2-2 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-3 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-4 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-6 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG  
GW2-7 GAAGTAAACCACTCATTGGTCTTTTACCTATAAACTTGTGTTGCGCTTG

GW2-9 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-10 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-5 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-8 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-1 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-2 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-3 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-4 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-6 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT  
GW2-7 ATGTTCTCAATATTGTCAGGCTATTTTATAGACTTGAGGAGTAAACATTT

GW2-9 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-10 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-5 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-8 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-1 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-2 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-3 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-4 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-6 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG  
GW2-7 CTTGATGACGTTATCTTCTCCTATTGTGCTAGACATAAAATACAGTGTAG

GW2-9 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-10 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-5 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-8 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-1 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-2 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-3 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-4 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-6 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA  
GW2-7 CATAACTTGTGTA CTGTGAGTTTTTTTCAAGTTTTTCTTGATGCTGCACA

GW2-9 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-10 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-5 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-8 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-1 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-2 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-3 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-4 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-6 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC  
GW2-7 ATAGTAAACAACAATCGTCTATTTCAGTCACACATGGCACATCCTCACATC

GW2-9 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-10 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-5 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-8 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-1 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-2 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-3 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-4 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-6 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC  
GW2-7 AAACAGAAGTTACTTTTTATGATGTTAATGTGTAGTATAACAGCAGAAGC

GW2-9 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGGGTTC  
GW2-10 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGAGTTC  
GW2-5 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGAGTTC  
GW2-8 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGAGTTC  
GW2-1 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGAGTTC  
GW2-2 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGGGTTC  
GW2-3 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGGGTTC  
GW2-4 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGGGTTC  
GW2-6 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGGGTTC  
GW2-7 TAAACGGTAGAGGGATACAGGATACTCGAGTCAATATGAAGTTGGGGTTC

GW2-9 CATGGAGTATTTGAATCGAATAAAAACATCTTGATGATCAATATTGGTTTT  
GW2-10 CATGGAGTATTTGAATCGAATAAAAATATCTTGATGATCAATATTGGTTTT  
GW2-5 CATGGAGTATTTGAATCGAATAAAAATATCTTGATGATCAATATTGGTTTT  
GW2-8 CATGGAGTATTTGAATCGAATAAAAATATCTTGATGATCAATATTGGTTTT  
GW2-1 CATGGAGTATTTGAATCGAATAAAAATATCTTGATGATCAATATTGGTTTT  
GW2-2 CATGGAGTATTTGAATCGAATAAAAACATCTTGATGATCAATATTGGTTTT  
GW2-3 CATGGAGTATTTGAATCGAATAAAAACATCTTGATGATCAATATTGGTTTT  
GW2-4 CATGGAGTATTTGAATCGAATAAAAACATCTTGATGATCAATATTGGTTTT  
GW2-6 CATGGAGTATTTGAATCGAATAAAAACATCTTGATGATCAATATTGGTTTT  
GW2-7 CATGGAGTATTTGAATCGAATAAAAACATCTTGATGATCAATATTGGTTTT

GW2-9 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-10 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-5 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-8 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-1 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-2 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-3 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-4 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-6 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT  
GW2-7 CATATTGTTTCTATAAGCTAGTTTCATATTTTCTAGAGAACCTATTATT

GW2-9 ATTTTGCTTCCTGTCCCTCAGCAGATTGACATTCTGTACTGGTAGTGA  
GW2-10 ATTTTGCTTCCTGTCCCTCAGCAGATTGACATTCTGTACTGGTAGTGA  
GW2-5 ATTTTGCTTCCTGTCCCTCAGCAGATTGACATTCTGTACTGGTAGTGA  
GW2-8 ATTTTGCTTCCTGTCCCTCAGCAGATTGACATTCTGTACTGGTAGTGA  
GW2-1 ATTTTGCTTCCTGTCCCTCAGCAGATTGACATTCTGTACTGGTAGTGA  
GW2-2 ATTTTGCTTCCTGTCCCTCAGCAGATTGACATTCTGTACTGGTAGTGA

GW2-3 ATTTTGCTTCCTGTCTCCTCAGAAGATTGACATTCTGTACTGGTAGTGA  
GW2-4 ATTTTGCTTCCTGTCTCCTCAGAAGATTGACATTCTGTACTGGTAGTGA  
GW2-6 ATTTTGCTTCCTGTCTCCTCAGAAGATTGACATTCTGTACTGGTAGTGA  
GW2-7 ATTTTGCTTCCTGTCTCCTCAGAAGATTGACATTCTGTACTGGTAGTGA

GW2-9 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-10 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-5 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-8 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-1 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-2 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-3 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-4 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-6 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG  
GW2-7 ATTTGTGTTTCCCTCATACCCTGTTTGGGAGGGTTTCAACGTCAGCTTAG

GW2-9 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-10 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-5 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-8 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-1 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-2 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-3 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-4 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-6 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA  
GW2-7 TGCAGTGCTGCAAAGTTTAAACTCTATGACCAGAGCATAAGTATGGCTTA

GW2-9 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-10 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-5 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-8 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-1 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-2 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-3 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-4 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-6 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC  
GW2-7 AGGACTTCTGATGACAAAGCCATTTTTGAAACCTGAAGCCTATTTTGATC

GW2-9 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-10 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-5 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-8 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-1 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-2 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-3 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-4 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-6 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA  
GW2-7 AACTTATGAACTTATATGGAAAACCAGTTTCTGAAAGAGGCCAAATATCA

GW2-9 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA

GW2-10 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-5 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-8 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-1 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-2 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-3 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-4 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-6 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA  
GW2-7 ATAAATGGTGCCTTAATCTCCCTTCAGTTTTATAGTTTACCATATGGATA

GW2-9 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-10 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-5 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-8 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-1 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-2 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-3 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-4 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-6 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT  
GW2-7 TTTTTCCTGTACTAAGTGTTCCTTGCTTGCTGGTCATGATAATCCTTTAT

GW2-9 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-10 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-5 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-8 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-1 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-2 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-3 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-4 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-6 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC  
GW2-7 GCTGTTGAGATGAAAATGGTTTCTAAAATTATGAAACTGTGTGTTTCCCC

GW2-9 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-10 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-5 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-8 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-1 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-2 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-3 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-4 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-6 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG  
GW2-7 TTACAGTGATGATAACATTGACATGAATATAGAGGATATGATGGTTATGG

GW2-9 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-10 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-5 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-8 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-1 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-2 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-3 AAGCGATTTGGCGTTCATTTCAGGTTAGTAGTTTTCTCACTGGTCCTTAC

GW2-4 AAGCGATTTGGCGTTCATTCCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-6 AAGCGATTTGGCGTTCATTCCAGGTTAGTAGTTTTCTCACTGGTCCTTAC  
GW2-7 AAGCGATTTGGCGTTCATTCCAGGTTAGTAGTTTTCTCACTGGTCCTTAC

GW2-9 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-10 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-5 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-8 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-1 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-2 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-3 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-4 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-6 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA  
GW2-7 ATGAGATTATTGATAATATGCATATGGCATACTGTGGATAATATTATAAA

GW2-9 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-10 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-5 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-8 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-1 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-2 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-3 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-4 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-6 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT  
GW2-7 AGCTGATTTGTTGGGTCAGGTAATGGTTTGTAATGAGCTGCACAAAAAT

GW2-9 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT  
GW2-10 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT  
GW2-5 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT  
GW2-8 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT  
GW2-1 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT  
GW2-2 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCTCCTGGCATGCT  
GW2-3 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCTCCTGGCATGCT  
GW2-4 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCTCCTGGCATGCT  
GW2-6 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT  
GW2-7 AGATTGTTATTTGCATTGAACATGAAAGTTTGCTAATCACCTGGCATGCT

GW2-9 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-10 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-5 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-8 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-1 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-2 TGTTGAATTCAGAGCATTTTGGGAGTTCTGATAAATGCAAGATCTAGTCT  
GW2-3 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-4 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-6 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT  
GW2-7 TGTTGAATTCAGAGCATTTTGGGAGTTCTGACAAATGCAAGATCTAGTCT

GW2-9 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAA-AAAATGTC  
GW2-10 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAA-AAAATGTC

GW2-5 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAA-AAAATGTC  
GW2-8 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAA-AAAATGTC  
GW2-1 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAA-AAAATGTC  
GW2-2 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAAAAAAAATGTC  
GW2-3 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAAAAAAAATGTC  
GW2-4 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAAAAAAAATGTC  
GW2-6 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAAAAAAAATGTC  
GW2-7 GCACAAGCAGAACCTGTATTGTTAAGTGTTCATAATAAAAAAAAAAATGTC

GW2-9 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-10 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-5 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-8 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-1 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-2 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-3 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-4 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-6 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG  
GW2-7 AATTAAATTGTTGAGAAGTGTATTGGCGATAGTGGATAATGTAGGATATG

GW2-9 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-10 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-5 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-8 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-1 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-2 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-3 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-4 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-6 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA  
GW2-7 TTTGGTTTGAGGGACCAACATCAGTTAGTTCATCATTGCCTCATTCCCTCA

GW2-9 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-10 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-5 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-8 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-1 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-2 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-3 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-4 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-6 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT  
GW2-7 GGCACATGCTAGTGAAAAATGAGTAATGCAGTATTCATGCCAAAATCAT

GW2-9 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-10 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-5 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-8 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-1 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-2 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-3 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-4 CCTAGATGGTGTGATTCCCTCAAACAAAACAGGCCAGTAGATGATGGGGTT

GW2-6 CCTAGATGGTGTGATTCTCTCAAACAAAACAGGCCAGTAGATGATGGGGTT  
GW2-7 CCTAGATGGTGTGATTCTCTCAAACAAAACAGGCCAGTAGATGATGGGGTT

GW2-9 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-10 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-5 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-8 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-1 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-2 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-3 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-4 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-6 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG  
GW2-7 GGTAAGATAAGTTATCTTTTTGTCAATTTACTTGAGCAGTACTGCGCTTG

GW2-9 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-10 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-5 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-8 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-1 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-2 GCACAAAAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-3 GCACAAAAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-4 GCACAAAAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-6 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT  
GW2-7 GCATTAATAATATTTCTGGTCATTTTGTGAGAACAAGAGGACTTTTCACAT

GW2-9 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-10 TCCTGCATTCTGAAGGGAACCTATAATCTAACAGCATTGCTAAACTCAG  
GW2-5 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-8 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-1 TCCTGCATTCTGAAGGGAACCTATAATCTAACAGCATTGCTAAACTCAG  
GW2-2 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-3 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-4 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-6 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG  
GW2-7 TCCTGCATTCTGAAGGGAACCTATTATCTAACAGCATTGCTAAACTCAG

GW2-9 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-10 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-5 CTCTTCGCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-8 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-1 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-2 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-3 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-4 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-6 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT  
GW2-7 CTCTTCCCATGCCCTTAGCCATCAAATGATTTGTTATGATGGTTGCATT

GW2-9 GTTAGCTTTGCGGTCCCAATTATTAGCCACTTAACGTTTCAGGTCAGGAAC  
GW2-10 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACGTTTCAGGTCAGGAAC  
GW2-5 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACGTTTCAGGTCAGGAAC



GW2-8 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC  
GW2-1 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC  
GW2-2 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC  
GW2-3 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC  
GW2-4 GTTAGTTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC  
GW2-6 GTTAGCTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC  
GW2-7 GTTAGCTTTGCGGTCCCAATTATTAGCCACTTAACG TTCAGGTCAGGAAC

GW2-9 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA  
GW2-10 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA  
GW2-5 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA  
GW2-8 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA  
GW2-1 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA  
GW2-2 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATACTACCTCCATCCTAA  
GW2-3 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATACTACCTCCATCCTAA  
GW2-4 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATACTACCTCCATCCTAA  
GW2-6 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA  
GW2-7 TAATGGTGT TTTTTTTCTCCGTCCTTTGTGTTATATTACCTCCATCCTAA

GW2-9 AATAAGTGCAGCCATGAGTTTCCAGGCCCAACTTTGATCGTCCGTTTTAT  
GW2-10 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-5 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-8 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-1 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-2 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-3 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-4 AATAAGTGCAGCCATGAGTTTCCGCGCCCAACTTTGATCGTCCGTTTTAT  
GW2-6 AATAAGTGCAGCCATGAGTTTCCAGGCCCAACTTTGATCGTCCGTTTTAT  
GW2-7 AATAAGTGCAGCCATGAGTTTCCAGGCCCAACTTTGATCGTCCGTTTTAT

GW2-9 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-10 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-5 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-8 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-1 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-2 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-3 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-4 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-6 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA  
GW2-7 TTGAAATTTTTTATAATTAGCATTTTTGTTGTTATAAGATGATAAAACA

GW2-9 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-10 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-5 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-8 TGAATAGTACTTTACTCATGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-1 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-2 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-3 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-4 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT  
GW2-6 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT

GW2-7 TGAATAGTACTTTACTCGTGACTTATGTTTTGATTTTTTCAAAAAAATT

GW2-9 TCAAATAAGACAAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-10 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-5 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-8 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-1 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-2 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-3 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-4 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-6 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT  
GW2-7 TCAAATAAGACGAATGGTCAAAGTTGGGCGCTGAAAACCATGGCTGCACT

GW2-9 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA  
GW2-10 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA  
GW2-5 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA  
GW2-8 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA  
GW2-1 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA  
GW2-2 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCTGTTCAAGCTGTA  
GW2-3 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCTGTTCAAGCTGTA  
GW2-4 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCTGTTCAAGCTGTA  
GW2-6 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA  
GW2-7 TATTTTGGGACGGAGGTAGTATCATTATTCATTTGCCGTTCAAGCTGTA

GW2-9 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-10 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-5 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-8 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-1 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-2 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-3 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-4 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-6 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT  
GW2-7 GCTCATCGTGTACTGTTCCCCAATAAGTTGCTTGCCATTTGGTCTATGTT

GW2-9 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-10 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-5 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-8 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-1 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-2 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-3 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-4 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-6 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT  
GW2-7 TCCAGCTTCTTTCTTTTGTGAGCATTTTCCTTCTGTATAGCTTGTTACT

GW2-9 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
GW2-10 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
GW2-5 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
GW2-8 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT

GW2-1 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
 GW2-2 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
 GW2-3 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
 GW2-4 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
 GW2-6 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT  
 GW2-7 AGAAAGATGTAAACTTTATGTAAATGCTGCTTCATCTTTGAATTTAGAT

GW2-9 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-10 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-5 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-8 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-1 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-2 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-3 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-4 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-6 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC  
 GW2-7 TCTTGGGAATAACAATAAGCAAAGTTCATGTATTGCGTATAGAAAAGTC

GW2-9 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-10 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-5 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-8 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-1 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-2 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-3 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-4 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-6 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA  
 GW2-7 TGCATATATTTCTATCCTAATACGGTCAATCTTTTCTCTGAGCAGGAGCA

GW2-9 GGAAGTATAGGGAATCCTGAC—————TGAGC  
 GW2-10 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-5 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-8 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-1 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-2 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-3 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-4 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-6 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC  
 GW2-7 GGAAGTATAGGGAATCCTGTCTGTGGCAACTTTATGCCTGTAAGTGAAGC

GW2-9 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-10 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-5 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-8 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-1 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-2 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-3 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-4 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-6 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA  
 GW2-7 CATCTCCGCGTGAACGCCAGCCATTTCGTTCCAGCTGCTTCTCTAGAAATA

GW2-9 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-10 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-5 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-8 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-1 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-2 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-3 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-4 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-6 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA  
GW2-7 CCTCATGGTGGTGGATTTTCCTGTGCGGTTGCGGCAATGGCTGAGCACCA

GW2-9 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-10 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-5 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-8 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-1 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-2 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-3 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-4 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-6 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG  
GW2-7 GCCACCCAGTATGGACTTCTCTTACATGGCTGGCAGCAGCGCATTCCCAG

GW2-9 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-10 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-5 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-8 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-1 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-2 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-3 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-4 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-6 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT  
GW2-7 TTTTCGACATGTTCCGGCGACCATGCAACATTGCTGGTGGAAAGCATGTGT

GW2-9 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-10 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-5 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-8 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-1 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-2 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-3 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-4 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-6 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG  
GW2-7 AATCTGGAGAGCTCACCGGAGAGCTGGAGCGGGATAGCACCAAGCTGCAG

GW2-9 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCCG  
GW2-10 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCCG  
GW2-5 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCCG  
GW2-8 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCCG  
GW2-1 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCCG

GW2-2 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCGG  
GW2-3 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCGG  
GW2-4 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCGG  
GW2-6 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCGG  
GW2-7 CAGGGAAGTGGTAAGAGAAGAAGGAGAGTGCTCGGCTGACCACTGGTCGG

GW2-9 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCAGAT  
GW2-10 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCGGAT  
GW2-5 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCGGAT  
GW2-8 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCGGAT  
GW2-1 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCGGAT  
GW2-2 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCAGAT  
GW2-3 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCAGAT  
GW2-4 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCAGAT  
GW2-6 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCGGAT  
GW2-7 AGGGTGCAGAGGCCGGAACAAGCTACGCGGGCTCAGACATCGTGGCGGAT

GW2-9 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-10 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-5 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-8 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-1 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-2 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-3 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-4 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-6 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC  
GW2-7 GCCGGGACCATGCCGCAGCTGCCTTTCGCCGAGAACTTCGCCATGGCGCC

GW2-9 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-10 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-5 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-8 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-1 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-2 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-3 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-4 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-6 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG  
GW2-7 AAGCCACTTCCGCCCGGAGAGCATCGAAGAACAGATGATGTTTTCCATGG

GW2-9 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-10 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-5 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-8 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-1 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-2 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-3 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-4 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-6 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA  
GW2-7 CTCTTCTTTTAGCAGATGGTCATGGAAGAACACACTCGCAAGGGTTGGCA

```

GW2-9 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-10 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-5 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-8 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-1 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-2 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-3 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-4 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-6 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
GW2-7 TGGTTGTAGGTAGAGCACTCTAATTTTGACGCCTTGCTGCCC
;

```

```
end;
```

```

begin mrbayes;
set autoclose=yes nowarn=yes;
lset applyto=(all) nst=6 rates=propinv;
mcmc ngen=1000000;
sump relburnin=yes burninfrac=0.25;
sumt relburnin=yes burninfrac=0.25;
end;

```

We also list the commands in program MrBayes to reconstruct the phylogenetic trees for *qSW5* in the empirical study.

```

#NEXUS
begin data;
dimensions ntax=14 nchar=3146;
format interleave datatype=DNA missing=N gap=-;

matrix

S5-K3 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-K2 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
SW5-Z2 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-K4 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
SW5-Z1 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N5 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N3 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N2 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N4 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N8 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N7 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N6 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-K1 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT
S5-N1 ACAATAGAAGTTGGCGATCGAGATGAGGCGGGGCAGCAACATATGAAGAT

S5-K3 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA
S5-K2 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA
SW5-Z2 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA
S5-K4 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA

```

SW5-Z1 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N5 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N3 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N2 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N4 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N8 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N7 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N6 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-K1 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA  
 S5-N1 GGAGATGCTGACATTGAACGGGCACAGTATGCAACCGTCCAAACAGTACA

S5-K3 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGA-  
 S5-K2 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGA-  
 SW5-Z2 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-K4 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 SW5-Z1 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N5 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N3 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N2 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N4 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N8 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N7 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N6 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-K1 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG  
 S5-N1 TCCAAAAAACCGCATATGCATCAGCATCACGCATCTTTCGTGAGAGAGAG

S5-K3 ——GAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-K2 ——GAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 SW5-Z2 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-K4 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 SW5-Z1 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N5 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N3 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N2 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N4 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N8 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N7 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N6 AGAGAGAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-K1 A—GAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA  
 S5-N1 A—GAGAGAGAGACGAAGCTGCAGTAGTTACAATTACCTCAGTTAACA

S5-K3 GTTTATCACCCACGGGATTCGATCGATGAAGAAGGTGAAAGAAGCAGCAGT  
 S5-K2 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTGAAAGAAGCAGCAGT  
 SW5-Z2 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
 S5-K4 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
 SW5-Z1 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
 S5-N5 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
 S5-N3 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
 S5-N2 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
 S5-N4 GTTTATCACCCACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT

S5-N8 GTTTATCACACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
S5-N7 GTTTATCACACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
S5-N6 GTTTATCACACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
S5-K1 GTTTATCACACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT  
S5-N1 GTTTATCACACGGGATTCGATCGATGGAGAAGGTCAAAGAAGCAGCAGT

S5-K3 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-K2 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
SW5-Z2 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-K4 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
SW5-Z1 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N5 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N3 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N2 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N4 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N8 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N7 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N6 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-K1 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG  
S5-N1 GTAGCTCAGCTCACCTCAGCGCAAGCTCAAGCTGGAGATCATCAGTGACG

S5-K3 AGATGTGAGATCATGCATCACACTGATGACTCATCCATACATATATACAG  
S5-K2 AGATGTGAGATCATGCATCACACTGATGACTCATCCAAACATATATACAG  
SW5-Z2 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-K4 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
SW5-Z1 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N5 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N3 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N2 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N4 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N8 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N7 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-N6 AGATGTGAGATCATGCATCATACTGATGACTCATCCAAACATATATACAG  
S5-K1 AGATGTGAGATCATGCATCACACTGATGACTCATCCATACATATATACAG  
S5-N1 AGATGTGAGATCATGCATCACACTGATGACTCATCCATACATATATACAG

S5-K3 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTAGTCGTCACCACAATT  
S5-K2 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
SW5-Z2 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-K4 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
SW5-Z1 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-N5 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-N3 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-N2 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-N4 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAAGT  
S5-N8 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAAGT  
S5-N7 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-N6 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-K1 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT  
S5-N1 CATCTTGACTCGTGGCCACTGTTGCTAACGATCTCGTCGTCACCACAATT





SW5-Z1 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N5 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N3 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N2 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N4 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N8 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N7 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N6 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-K1 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA  
S5-N1 CGCAGGGACGACGTCGTCACCGATCGAGCGATATTGATGAATCGATCGTA

S5-K3 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-K2 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
SW5-Z2 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-K4 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
SW5-Z1 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N5 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N3 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N2 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N4 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N8 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N7 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N6 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-K1 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG  
S5-N1 CTCCTCCTTAGCTAAAGCATAGGGTATGCATCAAATGCAGCTGGTACATG

S5-K3 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-K2 ATCCAGCCTACAGGTTTCTGCAAGCAGATATGGCTAAGCCCCGGCCTTGC  
SW5-Z2 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-K4 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
SW5-Z1 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N5 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N3 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N2 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N4 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N8 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N7 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N6 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-K1 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC  
S5-N1 ATCCAGCCTACAGGTTTCTGCTAGCAGATATGGCTAAGCCCCGGCCTTGC

S5-K3 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-K2 AATTTTTGCAGCTCTTGGGT-ATAGTACCTGCGAAAGGTGACCCCCAAA  
SW5-Z2 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-K4 AATTTTTGCAGCTCTTGGGAATAGTACCTGCGAAAGGTGACCCCCAAA  
SW5-Z1 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N5 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N3 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N2 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N4 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA

S5-N8 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N7 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N6 AATTTTTGCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-K1 AATTTTTTCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA  
S5-N1 AATTTTTTCAGCTCTTGGGGTATAGTACCTGCGAAAGGTGACCCCCAAA

S5-K3 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-K2 AGCCGT-----TTAACAGG  
SW5-Z2 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-K4 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
SW5-Z1 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N5 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N3 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N2 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N4 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N8 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N7 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N6 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-K1 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG  
S5-N1 AGCCGTTTAAATACACACACATTCATGTGTGTGGGGAGGGGAAGAACAGG

S5-K3 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-K2 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
SW5-Z2 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-K4 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
SW5-Z1 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-N5 GCCACGTTATACAGAAAGATGCACTCGCACACATGTATCCAAAGTACAA  
S5-N3 GCCACATTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-N2 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-N4 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-N8 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-N7 GCCACGTTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-N6 GCCACATTATACAGAAAGACGCACTCGCACACATGTATCCAAAGTACAA  
S5-K1 GCCACGTTATACAGAAAGATGCACTCGCACACATGTATCCAAAGTACAA  
S5-N1 GCCACGTTATACAGAAAGATGCACTCGCACACATGTATCCAAAGTACAA

S5-K3 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-K2 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAATAGCTTATTCGTG  
SW5-Z2 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-K4 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
SW5-Z1 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N5 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N3 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N2 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N4 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N8 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N7 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N6 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-K1 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG  
S5-N1 CCTCATCGGTTGACCGAGAGCTGATTTATATACCAAACAGCTTATTCGTG

S5-K3 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAGTTAAGTCTA  
S5-K2 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGTGTAAAGTCTA  
SW5-Z2 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-K4 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
SW5-Z1 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-N5 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAGTTAAGTCTA  
S5-N3 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-N2 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-N4 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-N8 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-N7 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-N6 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAATTAAGTCTA  
S5-K1 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAGTTAAGTCTA  
S5-N1 AAAAATAATTTATGAGTGACCTTTCATATGAGTCTATGGAGTTAAGTCTA

S5-K3 AGATCTGAGCGGTAACCTTCTCGTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-K2 AGATCTGAGCGGTAACCTTCTCGTTTTCTGTTAGTACGTTTTCTAAACTG  
SW5-Z2 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-K4 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
SW5-Z1 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N5 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N3 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N2 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N4 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N8 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N7 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N6 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-K1 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG  
S5-N1 AGATCTGAGCGGTAACCTTCTCATTTTTCTGTTAGTACGTTTTCTAAACTG

S5-K3 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-K2 TTAAACTATGTTTATTTTTT-AAAAATAACTTTTCTACATATAATTAGTT  
SW5-Z2 TTAAACTATGTTTATTTTTT-AAAAATAACTTTTCTACATATAATTAGTT  
S5-K4 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
SW5-Z1 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N5 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N3 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N2 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N4 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N8 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N7 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N6 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-K1 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT  
S5-N1 TTAAACTATGTTTATTTTTTTAAAAAAACTTTTCTACATATAATTAGTT

S5-K3 TCTTTCAAAAAAACCAAAATTTGAGTTTATGATAGTTATAAACCATGTG  
S5-K2 TCTTTCAAAAAAACCAAAATTTGAGTTTATGATAGTTATAAACCATGTG  
SW5-Z2 TCTTTCAAAAAAACCAAAATTTGAGTTTATGATAGTTATAAACCATGTG  
S5-K4 TCTTTCAAAAAAACCAAAATTTGAGTTTATGATAGTTATAAACCATGTG

SW5-Z1 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N5 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N3 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N2 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N4 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N8 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N7 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N6 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-K1 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG  
S5-N1 TCTTTCAAAAAAACCAAATTTTGAGTTTATGATAGTTATAAACCATGTG

S5-K3 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTATTTGTTTGGG  
S5-K2 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGTATGTGTTTTGTTTGGG  
SW5-Z2 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGTATGTGTTTTGTTTGGG  
S5-K4 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
SW5-Z1 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N5 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N3 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N2 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N4 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N8 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N7 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N6 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-K1 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG  
S5-N1 CTCATAAGCTTCTCGTTTTGCGTACAACCTAAGCATGTGTTTTGTTTGGG

S5-K3 GACTATGTGGGATAGGATGAAACCATACATGATTTTAGGGATTGGGTGAT  
S5-K2 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
SW5-Z2 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-K4 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
SW5-Z1 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N5 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N3 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N2 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N4 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N8 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N7 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N6 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-K1 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT  
S5-N1 GACTATGTGGGATAGGATGAAACCATCCCTGATTTTAGGGATTGGGTGAT

S5-K3 CTCACATACTTGTGGTTTATAGGATGAGATGATGCCAGTTTTTTATTT  
S5-K2 CTTACTTACTTGTGGTTTATAGGATGAGATGATGCCAGTTTTTTGTTT  
SW5-Z2 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-K4 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT  
SW5-Z1 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N5 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N3 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N2 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N4 CTCACATACTTGTGGTTTATAAAATGAGATGATGCTAGTTTTTTGTTT

S5-N8 CTCACATACTTGTTTGGTTTATAAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N7 CTCACATACTTGTTTGGTTTATAAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N6 CTCACATACTTGTTTGGTTTATAAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-K1 CTCACATACTTGTTTGGTTTATAAAAATGAGATGATGCTAGTTTTTTGTTT  
S5-N1 CTCACATACTTGTTTGGTTTATAAAAATGAGATGATGCTAGTTTTTTGTTT

S5-K3 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-K2 GGTAGAATAGATAAGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
SW5-Z2 GGTAGAATGGATATGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-K4 GGTAGAATGGATATGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
SW5-Z1 GGTAGAATGGATATGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N5 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N3 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N2 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N4 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N8 GGTAGAATGGATATGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N7 GGTAGAATGGATATGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N6 GGTAGAATGGATATGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-K1 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA  
S5-N1 GGTAGAATGGATAGGGGTGGGGATGGGATGAATGACGTTTGGAGATGTTA

S5-K3 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-K2 GCACCACGTCTACTCCTAGTAAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
SW5-Z2 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-K4 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
SW5-Z1 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N5 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N3 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N2 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N4 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N8 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N7 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N6 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-K1 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG  
S5-N1 GCACCACGTCTACTCCTAGTGAGGAAGGAGCAGCGAACGGCGGGCGGTGG

S5-K3 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-K2 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAACG  
SW5-Z2 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-K4 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
SW5-Z1 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N5 AAGCAGGAGCGGCAAACAACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N3 AAGCAGGAGCGGCAAACAACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N2 AAGCAGGAGCGGCAAACAACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N4 AAGCAGGAGCGGCAAACAACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N8 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N7 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N6 AAGCAGGAGCGGCAAACGACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-K1 AAGCAGGAGCGGCAAACAACCGGCGACCGGCGGCGTGGATGGAAGGAGCG  
S5-N1 AAGCAGGAGCGGCAAACAACCGGCGACCGGCGGCGTGGATGGAAGGAGCG

S5-K3 GTGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-K2 GTGTGCAAGCGAGTAGGCACAGGCGTGAGCAGCGAATGGCAGAATGAGGA  
SW5-Z2 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-K4 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
SW5-Z1 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N5 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N3 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N2 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N4 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N8 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N7 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N6 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-K1 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA  
S5-N1 GCGTGCAAGCGAGCAGGCACAGGCGTGAGCGGCGAACGGCAGAATGAGGA

S5-K3 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-K2 GCGGCGTGGGCGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
SW5-Z2 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-K4 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
SW5-Z1 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N5 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N3 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N2 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N4 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N8 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N7 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N6 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-K1 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG  
S5-N1 GCGGCGCGGGGAGCGAGCGTGTGTAGGGAAGGAGCTGCATGAGCGGGCG

S5-K3 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-K2 AAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGAGGGAGGGAGC  
SW5-Z2 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-K4 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
SW5-Z1 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N5 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N3 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N2 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N4 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N8 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N7 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N6 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-K1 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC  
S5-N1 GAGGAGGACTAACGGGCGCTGGGGGAGGGAGCGGCGCCGGGGAGGGAGC

S5-K3 AAGCGGAGGAGGATGGAGCGGTGCCGGGGAGGGAGTAGGCAGAAGGAGG  
S5-K2 AAGCGGAGGAGGACGGAGCGGTGCCGGGGAGGGAGTAGGCAGAAGGAGG  
SW5-Z2 AAGCGGAGGAGGACGGAGTGGTGCCGGGGAGGGAGTAGGCAGAAGGAGG  
S5-K4 AAGCGGAGGAGGACGGAGTGGTGCCGGGGAGGGAGTAGGCAGAAGGAGG

SW5-Z1 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N5 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGACG  
 S5-N3 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N2 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N4 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N8 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N7 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N6 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-K1 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG  
 S5-N1 AAGCGGAGGAGGACGGAGTGGTGCCGGGGGAGGGAGTAGGCAGAAGGAGG

S5-K3 AGCGGTGCCGGGAGGGAGTAGACAGAAGGAGGAGCGGCGCGATCAGCGA  
 S5-K2 AGCGGTGCCGGGAAGGAGTAGGCAGAAGGAGGAGCGACGCGATCAGCGA  
 SW5-Z2 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGATCAGCGA  
 S5-K4 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGATCAGCGA  
 SW5-Z1 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGATCAGCGA  
 S5-N5 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-N3 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-N2 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-N4 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-N8 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-N7 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-N6 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT  
 S5-K1 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGATCAGCGA  
 S5-N1 AGCGGTGCCGGGAGGGAGTAGGCAGAAGGAGGAGCGGCGCGAAATTAAT

S5-K3 CCGGCGGCGCGGATGGCAGGAGGAGCTGCAC—GAGCGGGCGGTGCAGC  
 S5-K2 CCGACGGCGCGGATGGCAGGAGGAGCTGCAC—GAGCGGGCGGTGCAGC  
 SW5-Z2 CCGGCGGCGCGGATGGCAGGAGGAGCTGCAC—GAGCGGGCGGTGCAGC  
 S5-K4 CCGGCGGCGCGGATGGCAGGAGGAGCTGCAC—GAGCGGGCGGTGCAGC  
 SW5-Z1 CCGGCGGCGCGGATGGCAGGAGGAGCTGCAC—GAGCGGGCGGTGCAGC  
 S5-N5 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-N3 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-N2 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-N4 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-N8 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-N7 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-N6 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC  
 S5-K1 CCGGCGGCGCGGATGGCAGGAGGAGCTGCAC—GAGCGGGCGGTGCAGC  
 S5-N1 TAAACTTGACAGCAGTACTGTTTCATGTAGTAACTGCATGGGAATTTAAAC

S5-K3 CGGCAACTTGCGAGCGGGCAGAGGAGGAGGGACGGGCGCCGGGGGAGGG  
 S5-K2 CGGCAACTTGCGAGCGGGCAGAGGAGGAGGGACGGGCGCCGGGGGAGGG  
 SW5-Z2 CGGCAACTTGCGAGCGGGCAGAGGAGGAGGGACGGGCGCCGGGGGAGGG  
 S5-K4 CGGCAACTTGCGAGCGGGCAGAGGAGGAGGGACGGGCGCCGGGGGAGGG  
 SW5-Z1 CGGCAACTTGCGAGCGGGCAGAGGAGGAGGGACGGGCGCCGGGGGAGGG  
 S5-N5 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT  
 S5-N3 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT  
 S5-N2 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT  
 S5-N4 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT



S5-N8 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT  
 S5-N7 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT  
 S5-N6 TTGCAGGTGTGTGGACGGATGGATAGATGGAT——GGGATTT  
 S5-K1 CGGCAACTTGGCGAGCGGGCAGAGGAGGGACGGGCGCGGGGGAGGG  
 S5-N1 TTGCAGGTGTGTGGACGG——ATGGATAGATGGATG—G-GATTT

S5-K3 AGTAGGCGGAGGAGGACGGAGCGGCATAGGGGGAGGGAGCAGGAAGAGGA  
 S5-K2 AGCAGGCGGAGGAGGACGGAGCGGCATCGAGGGAGGGAGCAGGAAGAGGA  
 SW5-Z2 AGCAGGCGGAGGAGGACGGAGCGGCATCGGGGGAGGGAGCAGGAAGAGGA  
 S5-K4 AGCAGGCGGAGGAGGACGGAGCGGCATCGGGGGAGGGAGCAGGAAGAGGA  
 SW5-Z1 AGCAGGCGGAGGAGGACGGAGCGGCATCGGGGGAGGGAGCAGGAAGAGGA  
 S5-N5 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-N3 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-N2 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-N4 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-N8 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-N7 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-N6 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT  
 S5-K1 AGCAGGCGGAGGAGGACGGAGCGGCATCGGGGGAGGGAGCAGGAAGAGGA  
 S5-N1 GGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAA-TCAT

S5-K3 GGAGGGACGGACGCCGTGGGAGGGAGCAGACGGAGGAGGAGGAACGGGGC  
 S5-K2 GGAGGGACGGACGCCGTGGGAGGGAGCAGACGGAGGAGGAGGAACGGGGC  
 SW5-Z2 GGAGGGACGGACGCCGTGGGAGGGAGCAGACGGAGGAGGAGGAACGGGGC  
 S5-K4 GGAGGGACGGACGCCGTGGGAGGGAGCAGACGGAGGAGGAGGAACGGGGC  
 SW5-Z1 GGAGGGACGGACGCCGTGGGAGGGAGCAGACGGAGGAGGAGGAACGGGGC  
 S5-N5 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-N3 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-N2 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-N4 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-N8 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-N7 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-N6 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT  
 S5-K1 GGAGGGACGGACGCCGTGGGAGGGAGCAGACGGAGGAGGAGGAACGGGGC  
 S5-N1 GAAGATTCGAAGGACATGGAA——GAACAACATCGATGCGTCT

S5-K3 GCCAGTGGCGGAGCCGAGCAGGAGGCAGAAGGAGCTAACGCCATTCGCTC  
 S5-K2 GCCAGTGGCGGAGCCGAGCAGGAGGCAGAAGGAGCTAACGCCATTCGCTC  
 SW5-Z2 GTCAGTGGCGGAGCCGAGCAGGAGGCGGAAGGAGCTAATGCCATTCGCTC  
 S5-K4 GTCAGTGGCGGAGCCGAGCAGGAGGCGGAAGGAGCTAATGCCATTCGCTC  
 SW5-Z1 GTCAGTGGCGGAGCCGAGCAGGAGGCGGAAGGAGCTAATGCCATTCGCTC  
 S5-N5 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-N3 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-N2 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-N4 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-N8 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-N7 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-N6 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC  
 S5-K1 GCCAGTGGCGGAGCCGAGCAGGAGGCGGAAGGAGCTAATGCCATTCGCTC  
 S5-N1 ACCTCTGACGACGC——TGCAGCTTGCTTGCTTGCCCTC

S5-K3 CATCAAAGTGGGATGGATTAGTCCAGCCGATTTTGGCGGACCACCTCGTC  
 S5-K2 CATCAAAGTGGGATGGATTAGTCCAGCCGATTTTGGCGGACCACCTCGTC  
 SW5-Z2 CATCAAAGTGGGATGGATTAGTCCAGCCGATTTTGGCGGACCACCTCGTC  
 S5-K4 CATCAAAGTGGGATGGATTAGTCCAGCCGATTTTGGCGGACCACCTCGTC  
 SW5-Z1 CATCAAAGTGGGATGGATTAGTCCAGCCGATTTTGGCGGACCACCTCGTC  
 S5-N5 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-N3 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-N2 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-N4 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-N8 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-N7 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-N6 CAACGACCGA———CGCATGCAAAAATATGGAAACCAAGTACTG  
 S5-K1 CATCAAAGTGGGATGGATTAGTCCAGCCGATTTTGGCGGACCACCTCGTC  
 S5-N1 CAACG—ACCGACGCATG———CAAAAATATGGAAACCAAGTACTG

S5-K3 CTGGATCTGGGAAGAATATTTCCCATCTAGGATCATCCCGTCTACACATC  
 S5-K2 CTGGATCTGGGAAGAATATTTCCCATCTAGGATCATCCCGTCTACACATC  
 SW5-Z2 CTGGATCTGGGAAGAATATTTCCCATCTAGGATCATCCCGTCTACACATC  
 S5-K4 CTGGATCTGGGAAGAATATTTCCCATCTAGGATCATCCCGTCTACACATC  
 SW5-Z1 CTGGATCTGGGAAGAATATTTCCCATCTAGGATCATCCCGTCTACACATC  
 S5-N5 CATGGAGTAGTAGCAAT———  
 S5-N3 CATGGAGTAGTAGCAAT———  
 S5-N2 CATGGAGTAGTAGCAAT———  
 S5-N4 CATGGAGTAGTAGCAAT———  
 S5-N8 CATGGAGTAGTAGCAAT———  
 S5-N7 CATGGAGTAGTAGCAAT———  
 S5-N6 CATGGAGTAGTAGCAAT———  
 S5-K1 CTGGATCTGGGAAGAATATTTCCCATCTAGGATCATCCCGTCTACACATC  
 S5-N1 CATGGAGTAGTAGCAAT———

S5-K3 CCTCTACCCAAATAGCTCGAAAAATGGGATCGACCCAACCCATTCCATTC  
 S5-K2 CCTCTACCCAAATAGCTCGAAAAATGGGATCGACCCAACCCATTCCATTC  
 SW5-Z2 CCTCTACCCAAACAGCTCGAAAAATAGGATCGACCTAACCCATCTCATTC  
 S5-K4 CCTCTACCCAAACAGCTCGAAAAATAGGATCGACCTAACCCATCTCATTC  
 SW5-Z1 CCTCTACCCAAACAGCTCGAAAAATAGGATCGACCTAACCCATCTCATTC  
 S5-N5 ——  
 S5-N3 ——  
 S5-N2 ——  
 S5-N4 ——  
 S5-N8 ——  
 S5-N7 ——  
 S5-N6 ——  
 S5-K1 CCTCTACCCAAACAGCTCGAAAAATAGGATCGACCTAACCCATCTCATTC  
 S5-N1 ——

S5-K3 CATATCCCACCAACCAACACAAGCTAAAAGTCTCACCCCTAACCTTATAA  
 S5-K2 CATATCCCACCAACCAACACAAGCTAAAAGTCTCACCCCTAACCTTATAA  
 SW5-Z2 CATATCCCACCTAACCAACACAAGCTAAAAGTCTCACCCCTAACCTTATAA  
 S5-K4 CATATCCCACCTAACCAACACAAGCTAAAAGTCTCACCCCTAACCTTATAA

SW5-Z1 CATATCCCACTAACCAAACACAAGCTAAAAGTCTCACCCCTAACCTTATAA

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————

S5-N7 —————

S5-N6 —————

S5-K1 CATATCCCACTAACCAAACACAAGCTAAAAGTCTCACCCCTAACCTTATAA

S5-N1 —————

S5-K3 ACAAACGCACCTTTTAAAGACCAACGCTAGCTTGAAAAAAAAAGAGATACGT

S5-K2 ACAAACGCACCTTTTAAAGACCAACGCTAGCTTGAAAAAAAAAGAGATACGT

SW5-Z2 ACAAACGCACCTTTTAAAGACCAACGCTAGCTTGAAAAAAAAAGAGATACGT

S5-K4 ACAAACGCACCTTTTAAAGACCAACGCTAGCTTGAAAAAAAAAGAGATACGT

SW5-Z1 ACAAACGCACCTTTTAAAGACCAACGCTAGCTTGAAAAAAAAAGAGATACGT

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————

S5-N7 —————

S5-N6 —————

S5-K1 ACAAACGCACCTTTTAAAGACCAACGCTAGCTTGAAAAAAAAAGAGATACGT

S5-N1 —————

S5-K3 TGAGCATATCATAAAATCAACTTCAAAATTTAAATTTACTAAGTTTTAAAT

S5-K2 TGAGCATATCATAAAATCAACTTCGAAATTTAAATTTACTAAGTTTTAAAT

SW5-Z2 TGAGCATATCATAAAATCAACTTCGAAATTTAAATTTACTAAGTTTTAAAT

S5-K4 TGAGCATATCATAAAATCAACTTCGAAATTTAAATTTACTAAGTTTTAAAT

SW5-Z1 TGAGCATATCATAAAATCAACTTCGAAATTTAAATTTACTAAGTTTTAAAT

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————

S5-N7 —————

S5-N6 —————

S5-K1 TGAGCATATCATAAAATCAACTTCGAAATTTAAATTTACTAAGTTTTAAAT

S5-N1 —————

S5-K3 TTAAATTTTACAGTGGCCGATATTATATTACACCCAATCAAAGAGACA

S5-K2 TTAAATTTTACAGTGGCCGATATTATATTACACCCAATCAAAGAGACA

SW5-Z2 TTAAATTTTACAGTGGCCGATATTATATTACACCCAATCAAAGAGACA

S5-K4 TTAAATTTTACAGTGGCCGATATTATATTACACCCAATCAAAGAGACA

SW5-Z1 TTAAATTTTACAGTGGCCGATATTATATTACACCCAATCAAAGAGACA

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 TTTAAATTTTACAGTGGCCGATATTATATTACACCCAATCAAAGAGACA  
 S5-N1 —————  
  
 S5-K3 CAGTAGACCATATTGCGTGATTGATGATGATAGGTGGTACTCTTATCTAC  
 S5-K2 CAGTAGACCATATTGCGTGATTGATGATGATAGGTGGTACTCTTATCTAC  
 SW5-Z2 CAGTAGACCATATTGCGTGATTGATGATGATAGGTGGTACTCTTATCTAC  
 S5-K4 CAGTAGACCATATTGCGTGATTGATGATGATAGGTGGTACTCTTATCTAC  
 SW5-Z1 CAGTAGACCATATTGCGTGATTGATGATGATAGGTGGTACTCTTATCTAC  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————  
 S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 CAGTAGACCATATTGCGTGATTGATGATGATAGGTGGTACTCTTATCTAC  
 S5-N1 —————  
  
 S5-K3 ACGTGCAGCAGGGCCAGGGTGTGTGTACTCCAATGTCGACTTGGTGCCAT  
 S5-K2 ACGTGCAGCAGGGCCAGGGTGTGTGTACTCCAATGTCGACTTGGTGCCAT  
 SW5-Z2 ACGTGCAGCAGGGCCAGGGTGTGTGTACTCCAATGTCGACTTGGTGCCAT  
 S5-K4 ACGTGCAGCAGGGCCAGGGTGTGTGTACTCCAATGTCGACTTGGTGCCAT  
 SW5-Z1 ACGTGCAGCAGGGCCAGGGTGTGTGTACTCCAATGTCGACTTGGTGCCAT  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————  
 S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 ACGTGCAGCAGGGCCAGGGTGTGTGTACTCCAATGTCGACTTGGTGCCAT  
 S5-N1 —————  
  
 S5-K3 GCATCCTACAAGGGGTACAAAGCATGTGTTTCATCACCTGAACTAGCTAGC  
 S5-K2 GCATCCTACAAGGGGTACAAAGCATGTGTTTCATCACCTGAACTAGCTAGC  
 SW5-Z2 GCATCCTACAAGGGGTACAAAGCATGTGTTTCATCACCTGAACTAGCTAGC  
 S5-K4 GCATCCTACAAGGGGTACAAAGCATGTGTTTCATCACCTGAACTAGCTAGC  
 SW5-Z1 GCATCCTACAAGGGGTACAAAGCATGTGTTTCATCACCTGAACTAGCTAGC  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————  
 S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 GCATCCTACAAGGGGTACAAAGCATGTGTTTCATCACCTGAACTAGCTAGC  
 S5-N1 —————

S5-K3 CCTGATCTTAAGTATGTCTATGCATGATTGCACCTCTGGCAATCAAGATA  
S5-K2 CCTGATCTTAAGTATGTCTATGCATGATTGCAACTCTCGCAATCAAGATA  
SW5-Z2 CCTGATCTTAAG-----

S5-K4 CCTGATCTTAAGTATGTCTATGCATGATTGCACCTCTGGCAATCAAGATA  
SW5-Z1 CCTGATCTTAAG-----

S5-N5 -----

S5-N3 -----

S5-N2 -----

S5-N4 -----

S5-N8 -----

S5-N7 -----

S5-N6 -----

S5-K1 CCTGATCTTAAGTATGTCTATGCATGATTGCACCTCTGGCAATCAAGATA

S5-N1 -----

S5-K3 CAGTATGATTCTAGTATGTAGCCTTGCTTGGAAAGCTATATACGGAGTACT

S5-K2 CAGTATGATTCTAGTATGTAGCCTTGCTTGGAAAGCTATATACGGAGTACT

SW5-Z2 -----

S5-K4 CAGTATGATTCTAGTATGTAGCCTTGCTTGGAAAGCTATATACGGAGTACT

SW5-Z1 -----

S5-N5 -----

S5-N3 -----

S5-N2 -----

S5-N4 -----

S5-N8 -----

S5-N7 -----

S5-N6 -----

S5-K1 CAGTATGATTCTAGTATGTAGCCTTGCTTGGAAAGCTATATACGGAGTACT

S5-N1 -----

S5-K3 ACTTTTATTAATGCCCATGCATGGTTAATAATTATGAGTACTGTATATAT

S5-K2 ACTTTTATTAATGCCCATGCATGGTTAATAATTATGAGTACTGTAAATAT

SW5-Z2 -----

S5-K4 ACTTTTATTAATGCCCATGCATGGTTAATAATTATGAGTACTGTATATAT

SW5-Z1 -----

S5-N5 -----

S5-N3 -----

S5-N2 -----

S5-N4 -----

S5-N8 -----

S5-N7 -----

S5-N6 -----

S5-K1 ACTTTTATTAATGCCCATGCATGGTTAATAATTATGAGTACTGTATATAT

S5-N1 -----

S5-K3 GTTCATCTCTCCCGCCGTATTGTACACACATTTTTGAGTTTCGAATTT

S5-K2 GTTCATCTCTCCCGCCGTATTGTACACACATTTTTGAGTTTCGAATTT

SW5-Z2 -----

S5-K4 GTTCATCTCTCCCGCCGTATTGTACACACATTTTTGAGTTTCGAATTT

SW5-Z1 —————  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————  
 S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 GTTCATCTCTCCCGCCGTATTGTACACACATTTTTTGTAGTTTCGAATTT  
 S5-N1 —————  
  
 S5-K3 GAACATGTACCGATCATTATGGTTAGAAGCAGTACCACGATCGAGATCTA  
 S5-K2 GAACATGTACCGATCATTATGGTTAGAAGCAGTACCACGATCGAGATCTA  
 SW5-Z2 —————  
 S5-K4 GAACATGTACCGATCATTATGGTTAGAAGCAGTACCACGATCGAGATCTA  
 SW5-Z1 —————  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————  
 S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 GAACATGTACCGATCATTATGGTTAGAAGCAGTACCACGATCGAGATCTA  
 S5-N1 —————  
  
 S5-K3 GAGTAGGCAAATTTGAACCAAGTGATGGTGACAAGTGATGAGCAATTGCA  
 S5-K2 GAGTAGGCAAATTTGAAC-CAGTGATGGTGACAAGTGATGAGCAATTGCA  
 SW5-Z2 —————  
 S5-K4 GAGTAGGCAAATTTGAAC-CAGTGATGGTGACAAGTGATGAGCAATTGCA  
 SW5-Z1 —————  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————  
 S5-N8 —————  
 S5-N7 —————  
 S5-N6 —————  
 S5-K1 GAGTAGGCAAATTTGAAC-CAGTGATGGTGACAAGTGATGAGCAATTGCA  
 S5-N1 —————  
  
 S5-K3 TGTACTCTCCATCGATCCCTACTCTTGTGCATTATTGTGGTTTACTGTAC  
 S5-K2 TGTACTCTCCATCGATCCCTACTCTTGTGCATTATTGTGGTTTACTGTAC  
 SW5-Z2 —————  
 S5-K4 TGTACTCTCCATCGATCCCTACTCTTGTGCATTATTGTGGTTTACTGTAC  
 SW5-Z1 —————  
 S5-N5 —————  
 S5-N3 —————  
 S5-N2 —————  
 S5-N4 —————

S5-N8 —————  
S5-N7 —————  
S5-N6 —————  
S5-K1 TGTACTCTCCATCGATCCCTACTCTTGTGCATTATTGTGGTTTACTGTAC  
S5-N1 —————

S5-K3 GTGTACGTACGAACGTATTCTGTGTTGCTTTCAGGTGGGAAACATCATGG  
S5-K2 GTGTACGTACGAACGTATTCTGTGTTGCTTTCAGGTGGGAAACATCATGG  
SW5-Z2 —————  
S5-K4 GTGTACGTACGAACGTATTCTGTGTTGCTTTCAGGTGGGAAACATCATGG  
SW5-Z1 —————  
S5-N5 —————  
S5-N3 —————  
S5-N2 —————  
S5-N4 —————  
S5-N8 —————  
S5-N7 —————  
S5-N6 —————  
S5-K1 GTGTACGTACGAACGTATTCTGTGTTGCTTTCAGGTGGGAAACATCATGG  
S5-N1 —————

S5-K3 TCGTACTAGTGGAGTACTTGTACCTAGTACCATTGGAATTCAATTCGTGT  
S5-K2 TCGTACTAGTGGAGTACTTGTACCTAGTACCATTGGAATTCAATTCGTGT  
SW5-Z2 —————  
S5-K4 TCGTACTAGTGGAGTACTTGTACCTAGTACCATTGGAATTCAATTCGTGT  
SW5-Z1 —————  
S5-N5 —————  
S5-N3 —————  
S5-N2 —————  
S5-N4 —————  
S5-N8 —————  
S5-N7 —————  
S5-N6 —————  
S5-K1 TCGTACTAGTGGAGTACTTGTACCTAGTACCATTGGAATTCAATTCGTGT  
S5-N1 —————

S5-K3 GTAGTAGAAATTAATTAACTTGACAGCAGTACTGTTTCATGTAGTAACTG  
S5-K2 GTAGTAGAAATTAATTAACTTGACAGCAGTACTGTTTCATGTAGTAACTG  
SW5-Z2 —————  
S5-K4 GTAGTAGAAATTAATTAACTTGACAGCAGTACTGTTTCATGTAGTAACTG  
SW5-Z1 —————  
S5-N5 —————  
S5-N3 —————  
S5-N2 —————  
S5-N4 —————  
S5-N8 —————  
S5-N7 —————  
S5-N6 —————  
S5-K1 GTAGTAGAAATTAATTAACTTGACAGCAGTACTGTTTCATGTAGTAACTG  
S5-N1 —————

S5-K3 CATGGGAATTTAAACTTGCAGGTGTGTGGACGGATGGATAGATGGATGGG

S5-K2 CATGGGAATTTAAACTTGCAGGTGTGTGGACGGATGGATAGATGGATGGG

SW5-Z2 —————

S5-K4 CATGGGAATTTAAACTTGCAGGTGTGTGGACGGATGGATAGATGGATGGG

SW5-Z1 —————

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————

S5-N7 —————

S5-N6 —————

S5-K1 CATGGGAATTTAAACTTGCAGGTGTGTGGACGGATGGATAGATGGATGGG

S5-N1 —————

S5-K3 ATTTGGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAAT

S5-K2 ATTTGGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAAT

SW5-Z2 —————

S5-K4 ATTTGGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAAT

SW5-Z1 —————

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————

S5-N7 —————

S5-N6 —————

S5-K1 ATTTGGACGGTGTGCGTGAAGAGAACAGTAATTAAGCAGAGAGCGGCAAT

S5-N1 —————

S5-K3 CATGAAGATTCGAAGGACATGGAAGAACAACATCGATGCGTCTACCTCTG

S5-K2 CATGAAGATTCGAAGGACATGGAAGAACAACATCGATGCGTCTACCTCTG

SW5-Z2 —————

S5-K4 CATGAAGATTCGAAGGACATGGAAGAACAACATCGATGCGTCTACCTCTG

SW5-Z1 —————

S5-N5 —————

S5-N3 —————

S5-N2 —————

S5-N4 —————

S5-N8 —————

S5-N7 —————

S5-N6 —————

S5-K1 CATGAAGATTCGAAGGACATGGAAGAACAACATCGATGCGTCTACCTCTG

S5-N1 —————

S5-K3 ACGACGCTGCAGCTTGCTTGCTTGCCTCCAACGACCGACGCAGACAGCAC

S5-K2 ACGACGCTGCAGCTTGCTTGCTTGCCTCCAACGACCGACGCAGACAGCAC

SW5-Z2 —————

S5-K4 ACGACGCTGCAGCTTGCTTGCTTGCCTCCAACGACCGACGCATGC—



```

SW5-Z1 -----
S5-N5 -----
S5-N3 -----
S5-N2 -----
S5-N4 -----
S5-N8 -----
S5-N7 -----
S5-N6 -----
S5-K1 ACGACGCTGCAGCTTGCTTGCTTGCTCCAACGACCGACGCATGC---
S5-N1 -----

```

```

S5-K3 GCATGCAAAAATATGGAAACCAAGTACTGCATGGAGTAGTAGCAAT
S5-K2 GCATGCAAAAATATGGAAACCAAGTACTGCATGGAGTAGTAGCAAT
SW5-Z2 -----
S5-K4 -----AAAAATATGGAAACCAAGTACTGCATGGAGTAGTAGCAAT
SW5-Z1 -----
S5-N5 -----
S5-N3 -----
S5-N2 -----
S5-N4 -----
S5-N8 -----
S5-N7 -----
S5-N6 -----
S5-K1 -----AAAAATATGGAAACCAAGTACTGCATGGAGTAGTAGCAAT
S5-N1 -----

```

```
;
```

```
end;
```

```

begin mrbayes;
set autoclose=yes nowarn=yes;
lset applyto=(all) nst=2 rates=gamma;
mcmc ngen=1000000;
sump relburnin=yes burninfrac=0.25;
sumt relburnin=yes burninfrac=0.25;
end;

```

# Bibliography

- [1] Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (1st ed.)*. London: John Murray.
- [2] Sollís-Lemus, C., Knowles, L.L., Ané, C. (2014). Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* **69(2)**,492-507.
- [3] Yang, Z. (1994). Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. *Syst Biol.* **43**, 329-342.
- [4] Ogden, T. H., Rosenberg, M. S. (2006). Multiple sequence alignment accuracy and phylogenetic inference. *Systematic biology.* **55(2)**, 314-328.
- [5] Freeman, S. (2010). *Biological science (fourth edition)*. Pearson.
- [6] Sequencing, T. C., Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature.* **437(7055)**, 69-87.
- [7] Rannala, B.,Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of molecular evolution. Journal of molecular evolution.* **43(3)**, 304-311.
- [8] Sokal, R.R., Michener, C.D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin.* **38**, 1409-1438.

- [9] Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*. **17(3)**, 368-376.
- [10] Pagel, M. (1994). Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proc. R. Soc. Lond.* **255**, 37-45.
- [11] Yang, Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol.***39**,306-314.
- [12] Posada, D., Crandall, K. A. (2001). Selecting the best-fit model of nucleotide substitution. *Systematic Biology*. **50(4)**, 580-601.
- [13] Darriba, D., Taboada, G. L., Doallo, R., Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*. **9(8)**, 772-772.
- [14] Guindon, S., Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*. **52(5)**, 696-704.
- [15] Jukes T.H., Cantor, C.R. (1969). *Evolution of protein molecules*. In Munro HN, editor, Mammalian Protein Metabolism, pp, 21-132, Academic Press, New York.
- [16] Tavar' S.(1998). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Lectures on Mathematics in the Life Sciences. American Mathematical Society. **17**, 57-86.
- [17] Cavalli-Sforza, L.L., Edwards, A.W. (1967). Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*. **19**,233-257.
- [18] Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*. **25(5)**, 471-492.

- [19] Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. **39**, 783-791.
- [20] Yang, Z. (1996). Phylogenetic analysis by parsimony and likelihood methods. *J Mol Evol*. **42**, 294-307.
- [21] Hillis, D. M., Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic biology*, **42(2)**, 182-192.
- [22] Zharkikh, A., Li, W. H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Molecular Biology and Evolution*, **9(6)**, 1119-1147.
- [23] Zharkikh, A., Li, W. H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: II. Four taxa without a molecular clock. *Journal of molecular evolution*, **35(4)**, 356-366.
- [24] Felsenstein, J., Kishino, H. (1993). Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Systematic Biology*, **42(2)**, 193-200.
- [25] Yang, Z., Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular biology and evolution*, **14(7)**, 717-724.
- [26] Kishino, H., Hasegawa, M. (1989). Evaluation of maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170-179.
- [27] Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.* **125**, 1-15.
- [28] Edwards, A.W., Cavalli-Sforza, L.L. (1964). Reconstruction of evolutionary trees. *Phenetic and Phylogenetic Classification*. London: Systemat. Assoc. **6**, 67-76.

- [29] Brown, R. (1827) A Brief Account of Microscopical Observations, etc., London (not published).
- [30] Ford, B. J. (1992). Brownian movement in clarkia pollen: a reprise of the first observations. *MICROSCOPE-LONDON THEN CHICAGO-*. **40**, 235-235.
- [31] Wiener, N. (1923). Differential Space. *Studies in Applied Mathematics*. **2(1-4)**, 131-174.
- [32] Higham D.J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*. **43**, 525-546.
- [33] Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.
- [34] Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**, 331-348.
- [35] Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*. **401**, 877-884.
- [36] Pagel, M. (1999). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48**, 612-622.
- [37] Felsenstein, J. (1988). Phylogenies from Molecular Sequences: Inference and Reliability. *Annual Review of Genetics*. **22**, 521-565.
- [38] Uhlenbeck, G.E., Ornstein, L.S. (1930). On the theory of Brownian motion. *Phys. Rev.* **36**, 823-41.
- [39] Garland, T., Dickerman, A.W., Janis, C.M., Jones, J.A. (1993). Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*. **42(3)**, 265-292.

- [40] Hansen, T.F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*. **51**, 1341-1351.
- [41] Hansen, T. F., Martins, E. P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* **50**, 1404-1417.
- [42] Butler, M. A., King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*. **164(6)**, 683-695.
- [43] Yan, S., Zou, G., Li, S., Wang, H., Liu, H., Zhai, G., Guo, P., Song, H., Yan, C., Tao, Y. (2011) Seed size is determined by the combinations of the genes controlling different seed characteristics in rice. *Theor Appl Genet*. **123(7)**, 1173-1181.
- [44] Lu, L., Shao, D., Qiu, X., Sun, L., Yan, W., Zhou, X., ... Xing, Y. (2013). Natural variation and artificial selection in four genes determine grain shape in rice. *New Phytologist*. **200(4)**, 1269-1280.
- [45] Johnston, I. (2000). And Still We Evolve: A Handbook on the History of Modern Science. *British Columbia: Malapina University*.
- [46] Ridley, M.(1983). *The Explanation of Organic Diversity : The Comparative Method and Adaptations for Mating*. Oxford, England: Clarendon Press.
- [47] Felsenstein, J. (2004). *Inferring phylogenies* (Vol. 2). Sunderland: Sinauer associates.
- [48] Christensen, R. (2011). *Plane answers to complex questions: the theory of linear models*. Springer Science Business Media.
- [49] Grafen, A. (1989). The Phylogenetic Regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **326(1233)**, 119-157.

- [50] Martins, E. P., Hansen, T. F. (1997). Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. *The American Naturalist*. **149**(4), 646-667.
- [51] Martins, E.P., Garland, T. (1991). Phylogenetic Analyses of the Correlated Evolution of Continuous Characters: A Simulation Study. *Evolution*. **45**(3), 534-557.
- [52] Harvey P.H., Pagel M.D. (1991). *The comparative method in evolutionary biology*. Oxford: Oxford University Press.
- [53] Freckleton, R. P. (2012). Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution*. **3**(5), 940-947.
- [54] Rohlf, F.J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution*. **55**, 2143-2160.
- [55] Rohlf, F.J. 2006. A comment on phylogenetic correction. *Evolution*. **60**,1509-1515.
- [56] Tavaré S., Balding D., Griffith R., Donnelly P.(1997). Inferring coalescence times from DNA sequence data. *Genetics*. **145**(2), 505-518.
- [57] Marin, J. M., Pudlo, P., Robert, C. P., Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*. **22**(6), 1167-1180.
- [58] Beaumont M., Zhang W., Balding D.(2002). Approximate Bayesian computation in population genetics. *Genetics*. **162**(4), 2025-2035.
- [59] Beaumont M., Cornuet J.M., Marin J.M., Robert C.(2009). Adaptive approximate Bayesian computation. *Biometrika*. **96**(4), 983-990.
- [60] Beaumont M., Nielsen R., Robert C., Hey J., Gaggiotti O., Knowles L., Estoup A., Mahesh P., Coranders J., Hickerson M., Sisson S., Fagundes N., Chikhi L., Beerli P.,

- Vitalis R., Cornuet J.M., Huelsenbeck J., Foll M., Yang Z., Rousset F., Balding D., Excoffier L.(2010). In defense of model-based inference in phylogeography. *Molecular Ecology*. **19(3)**, 436-446.
- [61] Sisson SA, Fan Y, Tanaka M (2007) Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*. **104(6)**, 1760-1765.
- [62] Fan, H. H., Kubatko, L. S. (2011). Estimating species trees using approximate Bayesian computation. *Molecular phylogenetics and evolution*. **59(2)**, 354-363.
- [63] Haario, H., Saksman, E., Tamminen, J.(2001). An adaptive Metropolis algorithm. *Bernoulli*. 223-242.
- [64] Roberts, G. O., Rosenthal, J. S.(2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*. **18(2)**, 349-367.
- [65] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [66] Felsenstein J.(2008). Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.* **171**, 713-725.
- [67] Revell L.J., Harmon L.J. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evol. Ecol. Res.* **10**, 311-331.
- [68] Adams, D.C.(2012). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic biology*.sys083.
- [69] Matthews, S. J. (2015, November). Accurate simulation of large collections of phylogenetic trees. In Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on (pp. 113-118). IEEE.



- [70] Zhu, S., Degnan, J. H., Goldstien, S. J., Eldon, B. (2015). Hybrid-Lambda: simulation of multiple merger and Kingman gene genealogies in species networks and species trees. *BMC bioinformatics*. **16(1)**, 292.
- [71] Robinson, D. F., Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical biosciences*. **53(1-2)**, 131-147.
- [72] Huelsenbeck J.P., Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. **17**,754-755.
- [73] Ronquist, F., Huelsenbeck J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. **(19)**,1572-1574.
- [74] Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30(9)**, 1312-1313.
- [75] Rambaut, A., Grass, N. C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer applications in the biosciences: CABIOS*. **13(3)**, 235-238.
- [76] Paradis, E., Claude, J., Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. **20(2)**, 289-290.
- [77] Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics*. **24(1)**, 129-131.
- [78] Yang, Z., and B. Rannala. (2010). Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci.* **107**, 9264-9269.
- [79] Felsenstein J. (1993). *Phylogeny Inference Package (PHYLIP)*. Version 3.5. University of Washington, Seattle.

- [80] <http://www.ebi.ac.uk/Tools/msa/clustalo/>.
- [81] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*. **19(6)**, 716-723.
- [82] Wit, E., Heuvel, E. V. D., Romeijn, J. W. (2012). "All models are wrong...": an introduction to model uncertainty. *Statistica Neerlandica*. **66(3)**, 217-236.
- [83] Vaidya, G., Lohman, D. J., Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multigene datasets with character set and codon information. *Cladistics*. **27(2)**, 171-180.
- [84] Butler, M. A., O'Meara, B. C., Pienaar, J. (2008). Comparative methods and data analysis in R. University of Hawaii.
- [85] King, A. A. and Butler, M. A. (2009). ouch: Ornstein-Uhlenbeck models for phylogenetic comparative hypotheses (R package), <http://ouch.r-forge.r-project.org>.
- [86] Huelsenbeck, J. P., Ronquist, F., Nielsen, R., Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*. **294(5550)**, 2310-2314.
- [87] Huelsenbeck, J. P., Rannala, B. (2003). Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution*. **57(6)**, 1237-1247.
- [88] Huelsenbeck, J. P., Larget, B., Alfaro, M. E. (2004). Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular biology and evolution*. **21(6)**, 1123-1133.
- [89] Wilkinson, R. G. (2007). Bayesian estimation of primate divergence times. *PhD thesis, University of Cambridge*.

- [90] Grelaud, A., Robert, C. P., Marin, J. M., Rodolphe, F., Taly, J. F. (2009). ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*. **4(2)**, 317-335.
- [91] Toni, T., Stumpf, M. P. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*. **26(1)**, 104-110.
- [92] Didelot, X., Everitt, R. G., Johansen, A. M., Lawson, D. J. (2011). Likelihood-free estimation of model evidence. *Bayesian analysis*. **6(1)**, 49-76.
- [93] Robert, C. P., Cornuet, J. M., Marin, J. M., Pillai, N. S. (2011). Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*. **108(37)**, 15112-15117.
- [94] O'Meara, B. C., Ané, C., Sanderson, M. J., Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*. **60(5)**, 922-933.
- [95] Beaulieu, J. M., Jhwueng, D. C., Boettiger, C., O'Meara, B. C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*. **66(8)**, 2369-2383.
- [96] Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S., Hansen, T. F. (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology*. **314**, 204-215.
- [97] Jhwueng, D. C., Maroulas, V. (2016). Adaptive trait evolution in random environment. *Journal of Applied Statistics*. **43(12)**, 2310-2324.
- [98] O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*. **43**, 267-285.

- [99] Degnan, J. H., Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in ecology evolution*. **24(6)**, 332-340.