

APPLICATION OF FEATURE SELECTION METHOD ON PHYLOGENETIC
TREE CONSTRUCTION USING MOLECULAR AND MORPHOLOGICAL
DATA

by

YASİN BAKIŞ

THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

THE ABANT İZZET BAYSAL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF

DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF BIOLOGY

FEBRUARY 2013

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Yaşar Dürüst
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Ekrem Gürel
Head of Department

This is to certify that I have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. M. Tekin Babaç
Supervisor

Prof. Dr. O. Uğur Sezerman
Cosupervisor

Examining Committee Members

Prof. Dr. M. Tekin Babaç

Prof. Dr. Neş'e Bilgin

Prof. Dr. O. Uğur Sezerman

Doç. Dr. Serap Mutun

Yard. Doç. Dr. Emel Uslu

ABSTRACT

APPLICATION OF FEATURE SELECTION METHOD ON PHYLOGENETIC TREE CONSTRUCTION USING MOLECULAR AND MORPHOLOGICAL DATA

Yasin Bakış

Ph.D., Department of Biology

Supervisor: Prof. Dr. M. Tekin Babaç

Co-supervisor: Prof. Dr. O. Uğur Sezerman

February 2013, 170 pages

This thesis analyzes biological datasets in the form of (morphological data and molecular sequences data) by utilization of Genetic Algorithms as the Feature Selection method. Information contents of the datasets were investigated and discussed. The thesis also traced the congruencies and incongruences within the dataset and seek to answer questions such as what is the minimum number of characters required to reconstruct a desired tree, or which part of the dataset carries what type of information especially for taxonomic applications.

This study is composed of two parts; analysis of morphological data and molecular sequence data. In the analysis of morphological data, I have analyzed a classical dataset based on *Hoplitis* taxa which can be accepted as a benchmark numerical taxonomy dataset and acorn dataset. The proposed method successfully optimized the *Hoplitis* dataset to a set of 67 characters which was very close to predictions of the authors of the dataset. Optimization of the acorn dataset was also quite successful by returning reasonable characters.

Shorter molecular sequences up to 1000 nt were also analyzed by the optimization algorithm. Heme Cupper Oxidase dataset were resulted with positions that are found within the solution subsets frequently. Some of those positions were highly accurate with the current literature. For the longer sequences, a similar algorithm has been designed by the use of Longest Common Substring method. In this method, only the difference is that analyses were performed on small subsets separated by common blocks rather than characters or residues. In this part, we have used a well-known mammalian mitochondrial genome dataset. The algorithm has returned a relatively small subset of data to construct the same phylogeny.

Keywords: Genetic Algorithms, Feature Selection, Numerical Taxonomy, Phylogenetics, Longest Common Substring, Molecular, Morphology

ÖZET

KARAKTER SEÇİMİ METODUNUN MOLEKÜLER VE MORFOLOJİK DATA KULLANILARAK FİLOGENETİK AĞAÇ OLUŞTURULMASINDA UYGULANMASI

Yasin Bakış

Ph.D., Biyoloji Bölümü

Tez Danışmanı: Prof. Dr. M. Tekin Babaç

Yardımcı Tez Danışmanı: Prof. Dr. O. Uğur Sezer

Şubat 2013, 170 sayfa

Bu tezde, Genetik Algoritmalar – Karakter Seçimi metodu kullanılarak biyolojik veri öbegleri (morfolojik ve moleküler dizi) çözümlenmiştir. Verilerin bilgi içeriği araştırılmış ve tartışılmıştır. Bu tez, belirli bir veri öbeği içindeki uyum ve uyumsuzlukları incelemekle beraber özellikle taksonomik uygulamalarda istenen ağacı bulmak için en az kaç karakter kullanmalıdır veya veri kümesinin hangi kısmı nasıl bir bilgi içerir gibi soruların da cevaplarını aramaktadır.

Bu çalışma iki kısımdan oluşmaktadır; morfolojik veri çözümlemesi ve moleküler dizi verisinin çözümlemesi. İlk kısımda, sayısal taksonomide referans noktası olarak da kabul edilen *Hoplitis* veri öbeğini ve daha önceki çalışmamda ortaya koymuş olduğum palamut veri öbeğinin çözümlemesini yaptım. Öne sürülen yöntem, *Hoplitis* veri öbeğini başarılı bir şekilde veri öbeğinin yazarlarının belirttiğine çok yakın olarak 67 karaktere eniyiledi. Palamut veri öbeğinin eniyilemesi ise kabul edilebilir derecede başarılı olarak mantıklı karakterlerle sonuçlandı.

Bin nükleotide kadar olan daha kısa diziler aynı eniyileme algoritması kullanılarak çözümlendi. Heme Cupper Oxidase veri öbeği çözümlemesi sonucunda, önemli pozisyonlar çözüm kümelerindeki sıklıklarına göre belirlendi. Bu pozisyonlardan bazıları güncel çalışmalarla oldukça benzerlik gösterdi. Daha uzun diziler için, En Uzun Ortak Altdizi yöntemi kullanılarak benzer bir algoritma tasarlandı. Bu yöntemde tek fark, analizler karakterler ya da pozisyonlar yerine ortak bloklar tarafından ayrılan küçük altdiziler üzerinde gerçekleştirilmiştir. Bu kısımda iyi bilinen memeli mitokondri genom veri öbeği kullanıldı. Algoritma, aynı topolojiyi oluşturabilmek için, göreceli olarak oldukça küçük bir altdiziyi buldu.

Keywords: Genetik Algoritmalar, Karakter Seçimi, Sayısal Taksonomi, Filogenetik, En Uzun Ortak Altdizi, Moleküler, Morfoloji

ACKNOWLEDGMENTS

I express my sincere appreciation to Prof. Dr. O. Uğur Sezerman for his guidance and encouragement throughout the research and other aspects. His dedication and enthusiasm for scientific research is unsurpassed, and his vast knowledge and incisive insight have been inspiring.

I would like to thank to Dr. Hasan H. Otu for his advice and support in completion of this work. The technical assistance of Cem Meydan is gratefully acknowledged. In addition, I would like to thank to the other members of the Sezerman Lab. for their encouragement and friendship.

Thanks go to the members of the Doctoral Thesis Monitoring Committee for their suggestions and comments.

I must express my gratitude to my wife, Gökçe, for her continued support and encouragement. I also offer sincere thanks for her unshakable faith in me and her willingness to endure with me the vicissitudes of my endeavors.

TABLE OF CONTENTS

ABSTRACT	III
ÖZET	V
ACKNOWLEDGMENTS	VII
TABLE OF CONTENTS.....	VIII
LIST OF TABLES	XI
LIST OF FIGURES.....	XII
LIST OF ABBREVIATIONS	XIII
CHAPTER 1. THESIS OVERVIEW	1
1.1. INTRODUCTION	1
1.1.1. <i>Phenetic, Cladistic and Evolutionary Approaches</i>	1
1.1.2. <i>Operational Taxonomic Units</i>	5
1.1.3. <i>The Role of Morphological Data in Phylogeny Reconstruction</i>	5
1.1.3.1. <i>Character</i>	6
1.1.3.2. <i>The Requisite Number of Characters</i>	6
1.2. MOTIVATION.....	8
1.2.1. <i>New Needs for Morphological Analysis</i>	8
1.2.2. <i>New Needs for Analyzing Molecular Sequence Data</i>	8
1.2.3. <i>A Solution to The Needs</i>	10
1.3. SCOPE AND ORGANIZATION	10
1.3.1. <i>Scope</i>	10
1.3.2. <i>Primary Research Question</i>	11
1.3.3. <i>Thesis Statement</i>	11
1.1. THESIS OVERVIEW	11
CHAPTER 2. GENETIC ALGORITHMS	13
2.1. HISTORY	13
2.2. BASIC DESCRIPTION	14
2.3. HOW DO THEY WORK?	15
2.3.1. <i>Outline of the Basic Genetic Algorithm</i>	16
2.3.2. <i>Representation (Encoding of Chromosomes)</i>	17
2.3.3. <i>Initial population</i>	18
2.3.4. <i>Evaluation (Fitness) function</i>	18
2.3.5. <i>Genetic operators</i>	19
2.3.5.1. <i>Crossover</i>	19
2.3.5.2. <i>Mutation</i>	19
2.3.6. <i>Parameters</i>	20
2.3.6.1. <i>Crossover probability</i>	20
2.3.6.2. <i>Mutation probability</i>	20

2.3.6.3.	Population size	20
2.3.6.4.	Selection	20
2.3.7.	Stopping The Process	22
CHAPTER 3.	IMPLEMENTATION OF THE ALGORITHM	24
3.1.	BACKGROUND	24
3.1.1.	Genetic Algorithms.....	24
3.1.1.1.	Feature Selection.....	25
3.2.	METHODOLOGY	25
3.2.1.	The Island (Migration) Model	25
3.2.2.	The Feature Selection Algorithm	27
3.2.3.	Phylogeny Reconstruction	30
3.2.4.	Phi Coefficient.....	31
3.2.5.	Test for Optimization.....	31
3.2.6.	Perturbance Technique	32
3.2.7.	Relative Complexity Measure Method	34
3.2.8.	Tree Construction and Topology Distance Methods	36
3.2.9.	Longest Common Substring method.....	37
CHAPTER 4.	EVALUATION OF MORPHOLOGICAL DATA.....	38
4.1.	INTRODUCTION	38
4.2.	THE DATASETS	39
4.2.1.	The Hoplitis Dataset.....	39
4.2.2.	The Acorn Dataset.....	43
4.3.	RESULTS.....	45
4.3.1.	Hoplitis Dataset	45
4.3.2.	Acorn Dataset	48
CHAPTER 5.	EVALUATION OF MOLECULAR SEQUENCES.....	51
5.1.	THEORETICAL BACKGROUND.....	51
5.2.	DATA SETS.....	53
5.2.1.	Heme Copper Oxidases	53
5.2.2.	Mammalian Mitochondrial Genome	56
5.2.3.	Simulated Datasets.....	59
5.3.	RESULTS.....	62
5.3.1.	Validation of The RCM Approach	62
5.3.2.	HCO data	63
5.3.3.	MMG data	65
CHAPTER 6.	DISCUSSION & CONCLUSIONS	68
6.1.	EVALUATION OF MORPHOLOGICAL DATA.....	68
6.1.1.	Hoplitis Dataset	68
6.1.2.	Acorn Dataset	69
6.2.	EVALUATION OF MOLECULAR DATA	71
6.2.1.	HCO Dataset.....	71
6.2.2.	MMG Dataset.....	73
6.3.	CONCLUSIONS	74
6.4.	FUTURE WORK.....	75
REFERENCES.....		76

APPENDICES.....	88
A - DATA RELATIVE TO CHAPTER 3	88
B - DATA RELATIVE TO CHAPTER 4	104
C - DATA RELATIVE TO CHAPTER 5	112
CURRICULUM VITAE.....	168

LIST OF TABLES

Table 1.1 Numbers of possible subsets of characters are tabulated with respect to the total number of characters.	7
Table 1.2 Examples of criteria used in choosing characters.	9
Table 6.1 Explanation of LZ-complexity of the sequence “TGATGCGACACA”.	36
Table 7.1 Taxonomic hierarchy of <i>Hoplitis</i> complex (Michener and Sokal 1957).	40
Table 7.2 Taxonomic hierarchy of <i>Quercus</i> genus (Bakış, 2005).	44
Table 7.3 Specifications of the data and obtained optimization and accuracy results.	48
Table 8.1 Structural and functional properties of HCOs and NORs.	54
Table 8.2 Analyzed part of the Mammalian Mitochondrial Genomic data.	57
Table 8.3 Comparison of constructed phylogenies for simulated data (standard deviations in parenthesis).	63
Table 8.4 Important results of the evaluation of HCO dataset.	64
Table 8.5 Results of MMG dataset segments with their size and frequencies.	66
Table A.1 Sample optimization process.	88
Table B.1 Characters that were used in the <i>Hoplitis</i> data.	104
Table B.2 Characters that were used in the Acorn data.	106
Table B.3 Phi coefficient results for <i>Hoplitis</i> dataset.	107
Table B.4 Results of 25 unique analysis were listed with less than or equal to 10 characters per solution sets.	109
Table B.5 Phi coefficient results for Acorn dataset.	111
Table C.1 Detailed information on Heme Copper Oxidase dataset.	112
Table C.2 Detailed information on mitochondrial sequences.	134
Table C.3 Optimization results of HCO sequence data.	139

LIST OF FIGURES

Figure 1.1 The first-known sketch of evolutionary tree drawn by Charles Darwin (Darwin 1837-1838).	4
Figure 1.2 Thesis organization roadmap.....	12
Figure 2.1 Genetic Algorithm logic flow.....	16
Figure 2.2 Types of chromosome encoding.....	17
Figure 2.3 Crossover operation.....	19
Figure 2.4 Mutation operation.....	20
Figure 3.1 The Island (Migration) Model.....	26
Figure 3.2 Optimization Algorithm.....	30
Figure 3.3 Optimization process.....	32
Figure 3.4 Perturbation Technique.....	33
Figure 3.5 Sample for Longest Common Substring method.....	37
Figure 4.1 Frequency of the characters that are found in 100 unique solution subsets.....	48
Figure 4.2 Comparing the tree of the original data with 33 characters (A) and the tree of the optimized data with 6 characters (B).	49
Figure 5.1 Weighted network representation (A) and phylogenetic analysis (B) of subunit I of HCO and NOR enzymes.....	55
Figure 5.2 Phylogenetic time tree of mammalian families deformed from (Meredith et al. 2011).....	58
Figure 5.3 A simple representation of similarities between different groups in Class Mammalia.	59
Figure 5.4 Screenshot of the Rose - Expert Submission module on the BiBiServ Bioinformatics Server (Stoye et al. 1996).	60
Figure 5.5 Balanced phylogeny – “phylogeny a” in (Hillis and Bull 1993) – with 16 OTUs.	61
Figure 5.6 Tree topology randomly generated by ROSE (Stoye et al. 1996) for the use of longer genomic sequence benchmark.	61
Figure 5.7 Alignment of genomic sequences blocks.....	65
Figure 5.8 The tree topology of the best solution produced by the optimization algorithm with TD=14.	66

LIST OF ABBREVIATIONS

GA	Genetic Algorithm
FS	Feature Selection
NT	Numerical Taxonomy
OTU	Operational Taxonomic Units
NC	Number of Characters
TD	Tree (Topology) Distance
HCO	Heme Copper Oxidases
MMG	Mammalian Mitochondrial Genome
nt	Nucleotide
bp	Base Pair
AA	Amino Acid
ROSE	Random Model of Sequence Evolution
MSA	Multiple Sequence Alignment
ML	Maximum Likelihood
RCM	Relative Complexity Measure
LZ	Lempel and Ziv (1976) complexity Measure
POY	Phylogenetic analysis of DNA and other data using dynamic homology

PhyML	Phylogenetic Maximum Likelihood software
DNADist	DNA Distance routine in PHYLIP
DNAML	Maximum Likelihood routine for DNA sequences in PHYLIP
TREEDIST	Tree Distance routine in PHYLIP
NCBI	National Center for Biotechnology Information
RF	Robinson and Faoulds (1981) tree topology distance measure
UPGMA	Unsupervised Pair-Group Mean Average
NJ	Neighbor Joining
F84	Felsenstein (1984) DNA Model
K2P	Kimura 2 Parameter DNA Model

CHAPTER 1. THESIS OVERVIEW

1.1. INTRODUCTION

Systematics is a discipline within biology whose goal is the determination of the evolutionary history and relationships among organisms (termed phylogeny) and the use of phylogeny in classifying organisms. To achieve that goal, a systematist will utilize evidence from a wide variety of sources including paleontology, embryology, morphology, behavior, and molecular biology.

Over the last few centuries systematists have developed a number of different approaches for trying to demonstrate the relationships among organisms. The first-known sketch by Charles Darwin of an evolutionary tree describing the relationships among groups of organisms (Darwin 1837-1838) is in Figure 1.1. Different schemes have emerged for ordering these relationships into a classification scheme. The main approaches in the last 50 years may be classified as; Phenetic (numerical taxonomy), Cladistic (phylogenetic) and Evolutionary (synthetic or phyletic).

1.1.1. PHENETIC, CLADISTIC AND EVOLUTIONARY APPROACHES

More than two decades ago, classification was largely carried out by scientists who based their identifications of organisms upon a few "important" characteristics and a "gut feeling" developed during many years experience. Phenetic classification methods were invented by systematists who thought that this type of classification scheme was too subjective. These scientists argue that taxonomy should not be based on a few characteristics arbitrarily judged to be important, but rather upon the degree

of overall similarity between organisms (Mayr 1965). Therefore, they collect data on as many characteristics of the organisms being classified as possible (Sneath and Sokal 1973). Each organism is then compared with every other for all characters measured, and the number of similarities (or differences) is calculated. The organisms are then clustered in such a way that the most similar are grouped close together and the more different ones are linked more distantly. Because of the enormous amount of data involved in this process, these calculations are usually made using specially programmed computers. The taxonomic clusters (phenograms) that result from such an analysis do not necessarily reflect genetic similarity or evolutionary relatedness (Stuessy 2009). Pheneticists would argue that this is OK, because nobody really knows the true evolutionary histories of these organisms and at least this system is objective (Sackin 1972). However, the lack of evolutionary significance in phenetics consequence interest in and use of phenetics has been declining in recent years.

Phenetics has been criticized because phenograms resulting from such analyses do not necessarily correspond to evolutionary histories between organisms. Therefore, an alternative approach to build relationships between taxa was developed, called cladistics. The basic assumption behind cladistics is that members of a group share a common evolutionary history, and are thus more closely related to one another than they are to other groups of organisms (Hennig 1966). Related groups of organisms are recognized because they share a set of unique features which are shared by most or all of the organisms within the group.

In contrast to phenetics, in cladistics groupings do not depend on whether organisms share physical traits, but on their evolutionary relationships (Hennig 1966). Indeed,

in cladistic analyses two organisms may share numerous characteristics but still be considered members of different groups (McNeill 1982). In cladistics, the emphasis is not upon the presence of all shared traits, but upon the presence of shared derived traits. Cladistic analyses have some pretty strict rules. For example, cladists always assume that the lineage always splits in two. Therefore, new species arise by bifurcations of the original lineage (Hennig 1966). Most cladists assume that the original ancestral species no longer exists after this bifurcation, so each branching event results in two new species. In addition, cladistic groupings possess all species in a grouping must share a common ancestor and all species derived from a common ancestor must be included in the taxon.

Like cladistics, evolutionary systematics is based strongly on evolutionary relationships (Carpenter 2005). Evolutionary Systematics gives an a dynamic evolutionary slant to the static Linnean system by supporting combination of branching and divergence (Hall 1994). However, its supporters suggest that the degree of genetic differences between lineages should be used in addition to their genealogical similarities when developing taxonomic classifications (Buettner-Janusch 1966). When creating groupings, evolutionary taxonomists seek to maximize the way in which the groupings they create communicate information about the group of organisms in that taxon. This is the biggest difference between the two approaches. Evolutionary taxonomists are more tolerant of multiple branches developing synchronously from an ancestral line, and they tend to treat organisms that are unchanged when a new branch arises from their lineage as the same species that existed before the branching event.

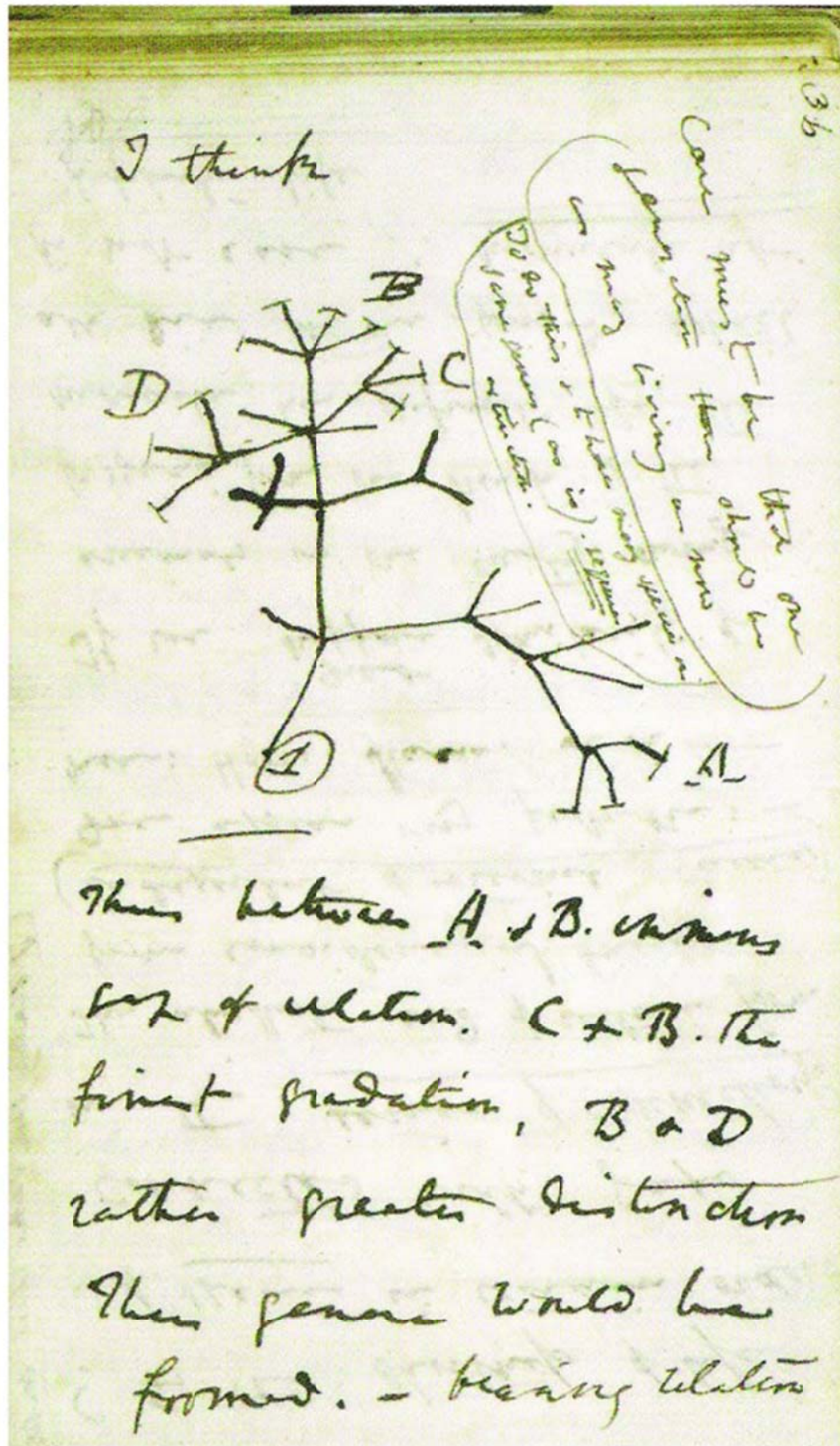


Figure 1.1 The first-known sketch of evolutionary tree drawn by Charles Darwin (Darwin 1837-1838).

1.1.2. OPERATIONAL TAXONOMIC UNITS

In order to restrict one's study to a manageable taxonomic group, one must first make a preliminary selection of specimens. What taxonomic units can be classified by numerical taxonomy? The logical fundamental unit in a large majority of instances is the individual organism. (Hennig 1966) points out that the fundamental element of biological systematics is an individual at a particular point of time.

In some studies individuals are the basic units employed. In many studies the unit taxon will therefore be an individual standing for a race, a species, a genus or even a higher ranked taxon (Sneath and Sokal 1973; Stuessy 2009). The most common unit in classificatory studies will be the species. Therefore, I shall employ taxonomic units of different categorical ranks as the entities to be grouped into more inclusive aggregates during classification. One cannot therefore speak of fundamental taxonomic units, but shall refer to operational taxonomic units (OTUs), which are the lowest ranking taxa employed in a given study (Schloss and Handelsman 2005).

1.1.3. THE ROLE OF MORPHOLOGICAL DATA IN PHYLOGENY RECONSTRUCTION

Most of our knowledge on the Tree of Life, both at lower and higher taxonomic levels, is based on phylogenetic studies of morphological data (Wiens 2004). Despite the growth of molecular systematics, morphology has been and will continue to be one of the most important and widely used sources of characters for phylogeny reconstruction. For example, morphology is the only way to address the phylogenies of fossil taxa and their relationships to living organisms. Similarly, many extant species are known from only a handful of museum specimens and may never be studied using anything but morphology.

1.1.3.1. Character

The term character has been employed in at least two distinct senses by systematists. Its commonest usage is as a distinguishing feature of taxa – a characteristic of one kind of organism that will distinguish it from another kind. This appears to be the sense in which Mayr (1969) defines a taxonomic character as "any attribute of a member of a taxon by which it differs or may differ from a member of a different taxon". Unfortunately, such definitions point out again the dilemma of conventional taxonomic procedure: characters are restricted to differences between members of a taxa, but a taxa cannot be recognized without the characters themselves being first known.

Another frequent meaning of the term character, which has been espoused by systematists, is that a character is a property or "a feature which varies from one kind of organism to another" (Michener and Sokal 1957) or "anything that can be considered as a variable independent of any other thing considered at the same time" (Cain and Harrison 1958).

Simply, serrated leaves, for example, may distinguish one species of plant from another and hence are called a character. In this example, the nature of the margins of the leaves becomes the character, while entire, serrated, undulating, or any other types of margin become different states of the character (Michener and Sokal 1957). The word "state" may imply qualitative rather than quantitative subdivision.

1.1.3.2. The Requisite Number of Characters

An important question in any numerical taxonomic study is what, number of characters is required in order to obtain stable classifications. In spite of the large amount of work in numerical taxonomy, we are still unable to provide generally

valid answers to this question. An early recommendation that no less than 60 characters should be used, whenever possible, and that if at all feasible, considerably more characters should be employed still seems reasonable, although as shall be seen, no justification for this requirement on either empirical or theoretical grounds can be made. What is the correct number of characters is related to the problem of the congruence of classifications based on sets of characters from different body parts or life history stages. Had the hypothesis of nonspecificity been fully valid, any set of characters would lead to sample estimates of a parametric similarity value and the question of number of characters would simply be a statistical one: what is the requisite number of characters to be sampled to obtain an estimate of similarity with confidence bands of desired width and at a desired probability level? But since it appears obvious that different sets of characters will yield somewhat different phenetic information, I cannot dispose of the problem so simply. Table 1.1 shows how many different sets of characters may overcome for a given number of characters.

Table 1.1 Numbers of possible subsets of characters are tabulated with respect to the total number of characters.

Number of characters	Number of possibilities
1	$2^1 - 1 = 1$
2	$2^2 - 1 = 3$
3	$2^3 - 1 = 7$
4	$2^4 - 1 = 15$
5	$2^5 - 1 = 31$
10	$2^{10} - 1 = 1,023$
20	$2^{20} - 1 = 1,048,575$
40	$2^{40} - 1 = 1,099,511,627,775$
60	$2^{60} - 1 = 1,152,921,504,606,850,000$
80	$2^{80} - 1 = 1,208,925,819,614,630,000,000,000$
100	$2^{100} - 1 = 1,267,650,600,228,230,000,000,000,000,000$
120	$2^{120} - 1 = 1,329,227,995,784,920,000,000,000,000,000,000,000$

1.2. MOTIVATION

1.2.1. *NEW NEEDS FOR MORPHOLOGICAL ANALYSIS*

This part deals with the difficult question of which characters should be chosen as a basis for estimating the similarity between OTU's. In attempting to answer this question, I encounter two relevant problems: (1) What biological factors do the characters represent? (2) Are all characters have equal information content in providing evidence on phenetic similarity? Several researchers have noticed the lack of and stressed the need for explicitness in character selection criteria (Pimentel and Riggins 1987; Stevens 1991; Wiens 1995). However, the extent of this problem had not been formally studied and documented. Table 1.2 shows examples of several criteria used in choosing characters from different studies.

Studies in Numerical Taxonomy are carried out by measuring characters as much as possible. The workload over scientists and labor to perform measurements will increase proportionally with the number of variables (or characters) to be used in the study. However, some part of the data may be irrelevant or sometimes meaningless.

In this thesis, a new algorithm was introduced in order to obtain a subset of data with minimum characters that can represent original data.

1.2.2. *NEW NEEDS FOR ANALYZING MOLECULAR SEQUENCE DATA*

Quantity and size of molecular sequence data increase at an exponential rate. Analysis of complete genomes and huge sized molecular sequences would not be possible with most of the current techniques. Most of the sequences include uninformative or irrelevant bases. Therefore, a need in sequence analysis has been arisen to extract informative and uninformative parts of the sequence, and to compare and understand what kind of information is housed within these parts.

Table 1.2 Examples of criteria used in choosing characters.

Criteria	Reference
“I chose to exclude mensural characters because ... polarities and character states ... are difficult to determine.”	(Livezey 1986)
“Two criteria were used for accepting characters for this analysis. First, the character had to vary discontinuously within the study group. Second, the character states had to be largely invariant within taxa. By 'largely invariant' I mean present in about 80% or more of the adult individuals of the species I examined Despite the second criterion, I occasionally allowed for the use of polymorphic characters in the analysis if the character was polymorphic for only one or two species and was invariant in the remaining species.”	(Kraus 1988)
“Quantitative characters were generally rejected because the character states are circumscribed arbitrarily and because such arbitrary states probably have nothing to do with homology and analysis.”	(Cox and Urbatsch 1990)
“Characters rejected when "known for less than% of OTUs.”	(Bateman et al. 1992)
“Character excluded if "too variable (and either autapomorphic or homoplasious) to be included.”	(Gensel 1992)
“Characters were deleted in the course of the investigation because of excessive missing data or polymorphism within groups.”	(Hufford and Dickson 1992)
“Characters were selected for phylogenetic analysis because of their low intraspecific variation, presence in most taxa, ability to be scored unequivocally, and phylogenetic informativeness.”	(Hood et al. 1994)
“Desired characters distinctive, biologically significant, and sufficiently constant within OTU s to allow for reliable character state assignment.”	(Kesner 1994)

1.2.3. A SOLUTION TO THE NEEDS

The field of bioinformatics shows a tremendous growth at the crossroads of biology, medicine, information science, and computer science. Today, bioinformatics research is as productive as data mining research as a whole. Most bioinformatics research deals with tasks of prediction, classification, and tree or network induction from data. Bioinformatics tasks consist mainly in similarity-based sequence search, microarray data analysis, 2D or 3D macromolecule shape prediction, and phylogenetic classification. Therefore, there are a large number of bioinformatics algorithms that can potentially be applied to domains in biology.

Genetic Algorithms (GAs) optimize search results for problems with large datasets. GAs evolved into complex and strong informatics tools capable of dealing with hard problems of clustering, classification, optimization, and simulation in diverse fields such as Biology, Chemistry and Medicine. GAs have been applied to phylogenetic tree building, gene expression and mass spectrometry data analysis, and many other areas of Bioinformatics that are computationally expensive problems.

In this thesis, Genetic Algorithms based method is developed as the Feature Selection method in optimization of morphological data.

1.3. SCOPE AND ORGANIZATION

1.3.1. SCOPE

This thesis will create two specific GA algorithms to analyze morphological and molecular sequence data. The first algorithm created for analyzing the morphological data has aimed to extract the optimum subset of morphological characters where one can reaches exactly the same phylogenetic tree topology by using this subset. By this

way, it would also be possible to demonstrate the congruence/incongruence among the characters based on their information content.

The second algorithm will be created for analyzing molecular sequence data. By this algorithm it has been aimed to detect informative parts of a set of sequences with a given tree topology. Thereby, these informative parts can be used in phylogenetic analysis and the type of information carried by these parts can be extracted.

1.3.2. PRIMARY RESEARCH QUESTION

The cornerstone for any GA program is the fitness function that quantifies the effectiveness of a potential solution. Therefore, the primary question would be "Given a known correct topology of a dataset, what is the smallest subset of features that can give exactly the same topology?"

1.3.3. THESIS STATEMENT

For a given phylogeny, a Genetic Algorithm – Feature Selection based method can be used to find out the most informative part of a dataset or a subset of data with minimum number of characters which will give the same topology.

1.1. THESIS OVERVIEW

Chapter 2 provides a functional overview of Genetic Algorithms in general. A description of what defines a GA is followed by discussion of the logical structure and mechanisms used in genetic programming.

The methodology section of the Thesis is **Chapter 3**. It begins with a brief overview of Genetic Algorithms and continued by the implementation of the algorithm. Several techniques used during the analyses were explained in rest of the section.

Chapter 4 is devoted to evaluation of morphological data. It begins with the explanations on the data sets and continued by the results that are tabulated briefly.

Evaluation of molecular sequence data were handled in **Chapter 5**. It begins with the explanations on the data sets and continued by the summary of the test results.

Chapter 6 provides a summary of the test results combined with discussions on the findings. Followed by suggestions for further research at the end of the chapter.

Appendix A provides data relative to Chapter 3.

Appendix B provides data relative to Chapter 4.

Appendix C provides data relative to Chapter 5.

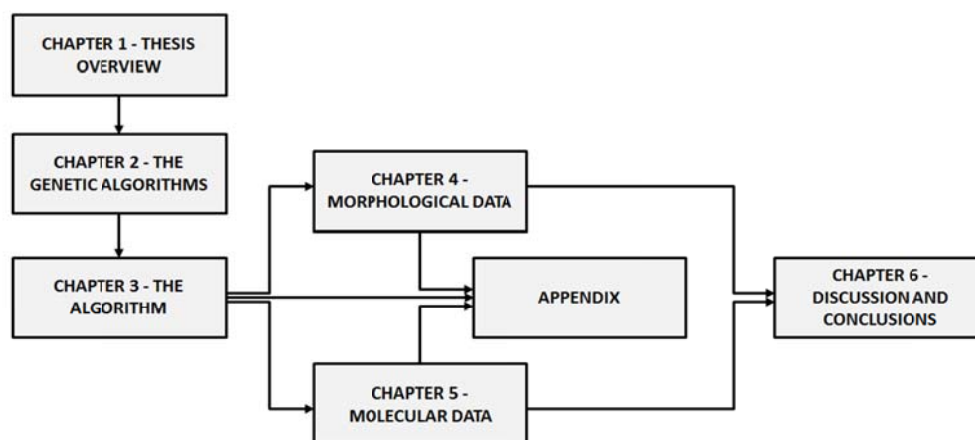


Figure 1.2 Thesis organization roadmap.

CHAPTER 2. GENETIC ALGORITHMS

This chapter introduces some fundamentals of Genetic Algorithms. As the area of genetic algorithms is very wide, it is not possible to cover everything.

2.1. HISTORY

As far back as the 1950's, possibilities of evolving solutions to engineering problems were studied by computer scientists. At that time, computers were not very powerful and computer scientists were motivated to best take advantage of their underpowered machines. To do that, they imported the simple processes of biological evolution that nature uses to solve complex problems (Mitchell 1997).

In the 1960s Rechenberg invented a highly influential set of optimization methods known as evolution strategies as his PhD thesis and he published a book in 1973 on evolution strategies (Rechenberg 1973). Later, his idea was developed further by other researchers. Genetic Algorithms were invented and developed by John Holland. His interest was to study adaptation as it occurs in nature. His book "Adaptation in Natural and Artificial Systems" essentially presented Genetic Algorithms as a tool for environmental modeling. This landmark book introduced the basic structures of GAs that are still used today (Holland 1975).

In 1992 John Koza used GA to evolve programs to perform certain tasks (Koza 1992). He called his method "Genetic Programming". In his book (Koza 1992), Koza shows how this method works and provides substantial empirical evidence that

solutions to a great variety of problems from many different fields can be found by genetically breeding populations of computer programs.

2.2. BASIC DESCRIPTION

Genetic Algorithms are a family of computational models inspired by biological evolution. GAs are a part of evolutionary computing, which is a rapidly growing area of artificial intelligence. GAs are inspired by Darwin's theory on evolution (Darwin 1859). In nature, evolution works by favoring individuals who are most fit for their particular environmental problems. The idea behind GAs is to do what nature does.

GAs encode a potential solution to a specific problem on a simple chromosome like structure and applies evolutionary operators (crossover, mutation and selection) to these structures. An implementation of a GA begins with a population of randomly selected chromosomes. These chromosomes are then evaluated in a way that those which represent better solutions to the target problem are given more chances to reproduce than those which have poorer solutions.

In the rabbit population example (Khouja et al. 1998; Michalewicz 1996), some rabbits are faster and smarter than others. The fast and smart ones will consistently be able to escape from predators. Therefore, those who survive do what rabbits do best - make more rabbits. Some slow and dumb rabbits, however, will survive by luck and will also reproduce. Therefore, their traits will be still alive in the populations gene pool. Nature throws in a "wild hare" once in a while by mutating some of the rabbit genetic material and some eventually become beneficial. The resulting population will (on the average) become faster and smarter than those in the original population because faster and smarter parents survived the foxes. Being fast and smart were the good solutions to the rabbits' problem. This is the model that

Genetic Algorithms are built on. Essentially, when a GA is applied to a problem, the solutions are evolved (Michalewicz 1996). Simply said, solution to a problem solved by GAs is evolved.

2.3. HOW DO THEY WORK?

Genetic Algorithms use a vocabulary borrowed from Genetics. One would talk about *individuals* in a population; quite often these individuals are also called *chromosomes*. Chromosomes are made of *genes* (also *features or characters*) – arranged in linear string. Each single chromosome would represent a potential solution to a specific problem.

During each generation, a genetic algorithm maintains a population of potential solutions (chromosomes). Each solution is evaluated to give some measure of its "fitness". A new population is formed by selecting the more fit individuals. Some members of this new population undergo alterations by means of crossover and mutation, to form new solutions. Crossover combines the features of two parent chromosomes to form two similar offspring by swapping corresponding segments of the parents. An evolution process run on a population of chromosomes – corresponds to a search through a space of potential solutions.

According to (Michalewicz 1996) a GA must have the following five components:

- a genetic representation for potential solutions to a problem,
- a way to create an initial population of potential solutions,
- an evaluation function that plays the role of the environment, rating solutions in terms of their "fitness",
- genetic operators that alter the composition of children,

- values for various parameters that the GA uses (population size, probabilities of applying genetic operators, etc.).

2.3.1. *OUTLINE OF THE BASIC GENETIC ALGORITHM*

Step 1. **Start:** Generate random population of n chromosomes

Step 2. **Fitness:** Evaluate the fitness $f(x)$ of each chromosome in the population

Step 3. **New population:** Create a new population by repeating steps below

3.1. **Selection:** Select two parent chromosomes according to their fitness

3.2. **Crossover:** Cross over the parents to form a new offspring

3.3. **Mutation:** Mutate new offspring

3.4. **Accepting:** Place new offspring in a new population

Step 4. **Replace:** Use new generated population for a further run of algorithm

Step 5. **Test:** If the end condition is satisfied, **stop**, and return the best solution in current population

Step 6. **Loop:** Go to step 2

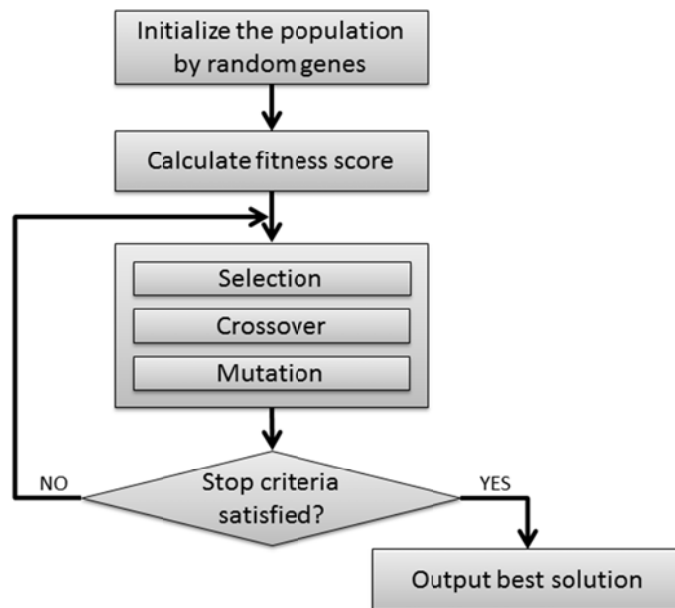


Figure 2.1 Genetic Algorithm logic flow.

2.3.2. REPRESENTATION (ENCODING OF CHROMOSOMES)

In this part some chromosome encodings will be described, which have been already in use. Chromosome encoding is the first problem, when you are starting to solve a problem using GA. Depending on the problem encoding may vary.

Binary encoding is the most common technique. This is mainly because first GA studies used this type of encoding. In binary encoding, string of bits, 0 or 1, composes chromosomes.

Permutation encoding can be used in ordering problems, such as travelling salesman problem or task ordering problem. In permutation encoding, every chromosome is a string of numbers, which represents number in a sequence.

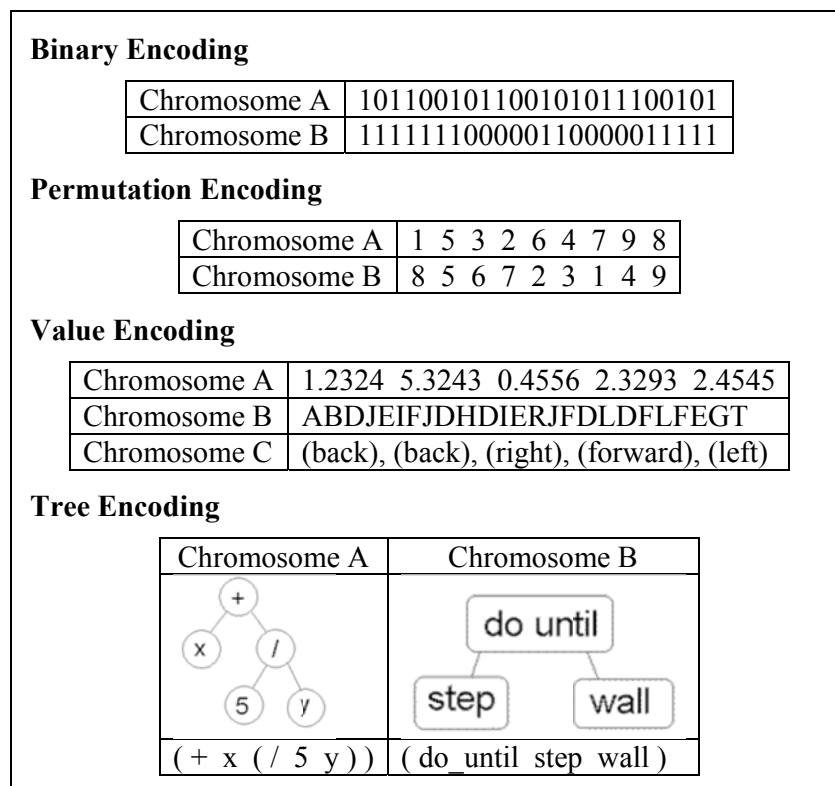


Figure 2.2 Types of chromosome encoding.

Direct value encoding can also be used in problems, where some complicated value, such as real numbers, are involved. Use of binary encoding for this type of problems would be very difficult. In value encoding, every chromosome is a string of some values. Values can be anything connected to problem, form numbers, real numbers or characters to some complicated objects.

Tree encoding is used mainly for evolving programs or expressions, for genetic programming.

2.3.3. *INITIAL POPULATION*

The initial population will be created when chromosome encoding is predicted. The most frequently assessed method to create initial population is creating chromosomes by randomly chosen genes. The population size depends on the nature of the problem, but typically contains a hundred to thousands of possible solutions. For a binary encoded chromosome, for instance, a string composed of random 0's and 1's would be generated. In some cases, to give a good start, one can select a portion of most fit chromosomes from the population. Occasionally, the solutions may be "seeded" in areas where optimal solutions are likely to be found.

2.3.4. *EVALUATION (FITNESS) FUNCTION*

“Fitness” is the basis for selection of individuals from the population of each generation. It is also the criterion to be optimized. In GAs, for each newly generated population, the idea is to delete the 'n' worst solutions (chromosomes), and to breed 'n' new ones from the best solutions. Each solution, therefore, needs to be awarded. This is generated by applying the fitness function to the test results obtained from that solution.

2.3.5. GENETIC OPERATORS

2.3.5.1. Crossover

After deciding what encoding to be used, one can make a step to crossover.

Crossover selects genes from parent chromosomes and creates a new offspring. The simplest way to do this is to choose randomly some crossover point and everything before this point copy from the first parent and then everything after the crossover point copy from the second parent. There are other ways how to make a crossover; for example one can choose several crossover points. Crossovers can be rather complicated and depends on encoding of the chromosome. Specific crossover made for a specific problem can improve performance of the GA. Crossover can then look like in Figure 2.3.

There are also Uniform and Arithmetic crossover types. In Uniform crossover, bits are randomly copied either from the first or from the second parent. In Arithmetic crossover, some arithmetic operation needs to be performed to make a new offspring.

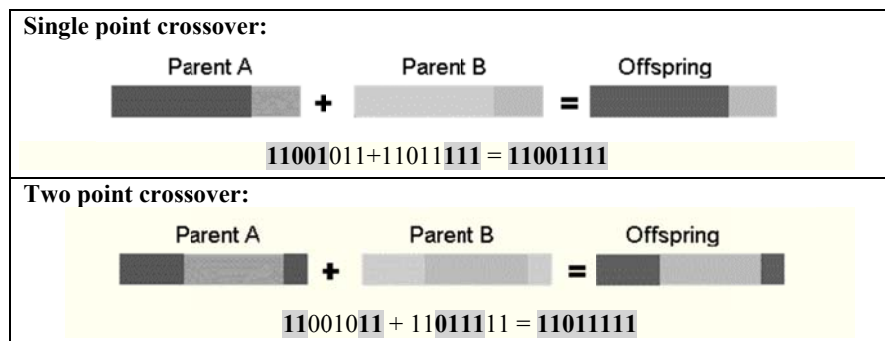


Figure 2.3 Crossover operation.

2.3.5.2. Mutation

Mutation takes place after a crossover is performed. This is to prevent falling all solutions in population into a local optimum of solved problem. Mutation changes randomly the new offspring. For binary encoding one can switch a few randomly

chosen bits from 1 to 0 or from 0 to 1. The mutation depends on the encoding as well as the crossover. For example, when encoding permutations, mutation could be exchanging two genes.

Original offspring 1	110 1 111000011110
Original offspring 2	110110 0 1001101 1 0
Mutated offspring 1	110 0 111000011110
Mutated offspring 2	110110 1 1001101 1 0

Figure 2.4 Mutation operation.

2.3.6. PARAMETERS

2.3.6.1. Crossover probability

Crossover probability describes how often a crossover will be performed. If there is no crossover, then the offspring will be an exact copy of the parents. If there is a crossover, the offspring presents parts of parents' chromosomes.

2.3.6.2. Mutation probability

Mutation probability describes how often parts of a chromosome will be mutated. If there is no mutation, the offspring is taken after crossover (or copy) without any change. If mutation is performed, part of chromosome is changed.

2.3.6.3. Population size

Population size describes how many chromosomes are in a population (in one generation). If there are too few chromosomes, GA have few possibilities to perform crossover and only a small part of search space is explored. On the other hand, if there are too many chromosomes, GA slows down.

2.3.6.4. Selection

As mentioned in the GA outline, chromosomes are selected from a population to be parents to crossover. The problem is how to select these chromosomes. According to

Darwin's evolution theory the best ones should survive and create new offspring. There are many methods for the selection of the best chromosomes. For example, roulette wheel selection, Boltzman selection, tournament selection, rank selection, steady state selection and some others (Gandhi et al. 2012).

Roulette Wheel Selection: Parents are selected according to their fitness. The better the chromosomes are, the more chances to be selected they have. Imagine a roulette wheel where all chromosomes in the population are placed, every chromosome has its place big accordingly to its fitness function. Then a marble is thrown there and selects the chromosome. Chromosome with higher fitness will be selected more times.

Rank Selection: The previous selection will have problems when the fitnesses differ very much. For example, if the best chromosome fitness is 90% of all the roulette wheel then the other chromosomes will have very few chances to be selected. Rank selection first ranks the population and then every chromosome receives fitness from this ranking. The worst will have fitness 1, second worst 2 etc. and the best will have fitness N (number of chromosomes in population).

Steady-State Selection: This is not particular method of selecting parents. Main idea of this selection is that big part of chromosomes should survive to next generation. GA then works in a following way. In every generation are selected a few (good - with high fitness) chromosomes for creating a new offspring. Then some (bad - with low fitness) chromosomes are removed and the new offspring is placed in their place. The rest of population survives to new generation.

Elitism: Idea of elitism has been already introduced. When creating new population by crossover and mutation, I have a big chance, that I will loose the best chromosome. Elitism is the name of the method, which first copies the best chromosome (or a few best chromosomes) to new population. The rest is done in classical way. Elitism can very rapidly increase the performance of GA, because it prevents losing the best found solution.

2.3.7. STOPPING THE PROCESS

The final question that must be answered by the programmer is determining when to stop the process. If the exact solution is known, the code can be designed to stop on its own when it is achieved. However the use of a GA in this case would be unnecessary! Computational constraints, such as programming language and hardware, are not the only factors that limit population size, individual member complexity and fitness calculation intricacy, but they may also limit the allowed run time on scarce computing resources. GAs will rarely arrive at an exact solution anyway, regardless of the amount of time allotted due to embedded encouragement in the process for continued population diversity. The code may include a process by which an operator or the program itself may siphon off and observe the results of the process every few generations in order to determine if the algorithm is tracking in the desired direction or has achieved a solution that is good enough to be within set error limits for the task required. Either the programmer can then interrupt the process, or it may be programmed to jump out of the iterative loop. Other limiting factors may also exist to curtail the run before the optimum solution is achieved, such as cost of constructing a physical device based on a Genetic Algorithm solution. By far the simplest way to end the run is to set a counter and run a pre-specified number of

generations. At the end of these runs, the final output is all or part of the population of the final generation. Records may also be generated of previous generations, so that earlier sub-optimal solutions may be used that keep materials, labor and complexity within the available budget.

CHAPTER 3. IMPLEMENTATION OF THE ALGORITHM

3.1. BACKGROUND

3.1.1. *GENETIC ALGORITHMS*

A Genetic Algorithm is a search technique used in computing to find an exact or an approximate solution for optimization and search problems (Mitchell 1996). Genetic algorithms are categorized as global search heuristics and are a particular class of evolutionary algorithms, also known as evolutionary computation, that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination) (Mitchell 1996).

Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called genes or chromosomes) of candidate solutions (called individuals) to an optimization problem evolves toward better solutions. The evolution usually starts with a population of randomly generated individuals and individuals of next generations were born by evolutionary operations (inheritance, mutation, selection, and crossover) of individuals in previous generations. In each generation, the fitness of every individual in the population is evaluated. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced or a satisfactory fitness level has been reached for the population.

3.1.1.1. Feature Selection

Feature selection, a type of GA, also known as variable selection, feature reduction, attribute selection or variable subset selection, is the technique, commonly used in machine learning, of selecting a subset of relevant features for building robust learning models (Mitchell 1996).

Over the past decade, there has been a massive effort in the expansion of analytical techniques for biological data (Tucker and Duplisea 2012). This has headed to various novel algorithms that are specialized to biological phenomena (Desai et al. 2010; George and Raj 2011; Keleş et al. 2002; Ma and Huang 2008; Tamer and Badr 2012; Tucker and Duplisea 2012). By removing the most irrelevant and redundant features from the data, feature selection techniques have become a necessity in many applications (Guyon and Elisseeff 2003; Saeys et al. 2007). It helps to improve the performance of learning models by; alleviating the effect of the curse of dimensionality, enhancing generalization capability, speeding up learning process, and improving model interpretability (Sivanandam and Deepa 2007). Feature selection also helps people to acquire better understanding of their data, the important features of the dataset and how they are related to each other.

3.2. METHODOLOGY

3.2.1. *THE ISLAND (MIGRATION) MODEL*

A need for solving complex problems drives the Evolutionary Computation community towards advanced models of Evolutionary Algorithms. One such model is the Island Model. The Island Model Genetic Algorithm shows promise as a superior formulation based on considerations from theories of natural evolution and from the efficiencies of coarse-grained parallel computer architectures. The theory of

punctuated-equilibria (Eldredge and Gould 1972) calls for the population to be partitioned into several distinct subpopulations. These subpopulations have extensive periods of isolated evolution, i.e. computation, occasionally interspersed with migration, i.e. communication.

Island Models behave qualitatively different from standard Evolutionary Algorithms. In this model, a single machine may produce different, possibly better solutions for many problems, especially complex ones. Separating individuals spatially from each other results in slowing down the information flow between individuals, which may have both desired and undesired results. A slower information flow may stop temporarily best solution from dominating the population and allow different building blocks or solutions to be discovered and later confronted, which is important in the context of engineering design and creativity. On the other hand one can prevent successful mixing, which could otherwise lead to constructing a novel solution.

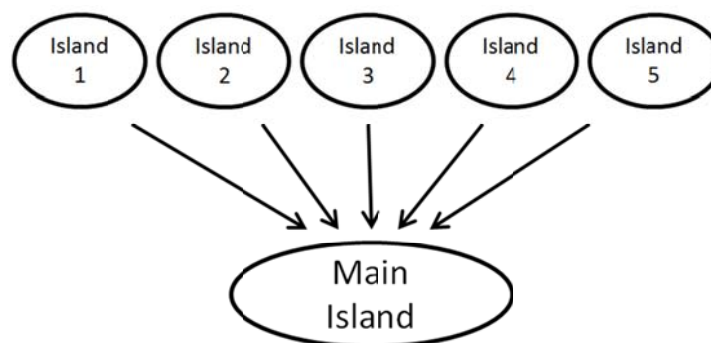


Figure 3.1 The Island (Migration) Model.

In the proposed algorithm, I have applied island model to pool the results in a main island and optimize this island again. For the current study, 5 subpopulations were analyzed separately. Within each resulting subsets, 20% of the individuals with best

fitness scores were chosen to build the main population. Output of the analysis of the main island would give the actual result.

The Island Model was used by creating subpopulations; later subpopulations have been combined to form the actual population in the main island. The procedure is the same for each island except that the main island start with a population composed of the best individuals from the islands.

3.2.2. *THE FEATURE SELECTION ALGORITHM*

The feature selection algorithm involves a basic genetic algorithm feature selection strategy (Figure 2.1). A population let to evolve to give better solutions. Each individual in the population represents a candidate solution to the problem.

Therefore, each candidate solution is a phylogenetic analysis of input data. Input is a data matrix in Phylogenetic Inference Package (PHYLIP) format where columns are characters and rows are OTUs. Phylogenetic analysis of the data is described in the next section. A solution vector is a binary encoding with size n , where n represents number of characters. In this case, size of the binary encoding array is 122.

Therefore, 2^{122} possible individuals exist. The array is composed of 1s and 0s (chromosomes) which predicts whether a character is either selected for the analysis or not, respectively. A combination of parameters were used that yielded optimum results, after attempted a number of combinations.

The main characteristics of the feature subset selection algorithm are as follows:

- population size (p): 200 individuals;
- total number of variables (n): 122;

- number of variables selected in the chromosomes of the starting population (k): 100;
- number of individuals to be selected in Elitist Selection: 0.05%;
- mutation probability: 0.05%;
- cross-over method: randomly chosen couples, two point crossing over;
- cross-over probability: 0.95%;
- stop criterion: change in the average fitness score among the last 20 generations equals to zero or a predefined number of evaluations (generations);

Initialization: The algorithm starts from a population of randomly generated individuals with number of k selected characters. In the analysis of the whole data, 100 characters were chosen as the starting number of characters and population was created by 200 individuals. An individual is generated by creating a random binary array of a given number (k) of 1s and the remaining number (n-k) of 0s. To give a good start, 2000 (10p) individuals are created with random chromosomal configuration, and only 200 individuals with best fitness scores are selected for the evaluation.

Fitness score function: For every individual solution in the population, a fitness score has been calculated by the formula below;

$$fitness\ score = \frac{1}{NC * (TD + 1)^2}$$

NC is the number of characters that are selected to be included in the set. TD is the distance between the tree topologies that are generated from the optimized set and the original data. I have added 1 to the TD to prevent division by 0 when topology distance equals 0. It is crucially important for algorithm to find the closest topology.

If the algorithm starts with decreasing number of characters before finding the best tree, it would never have a chance to get a good tree with lower number of characters. Therefore, the weight is given to the topology distance by taking the second power of it. Since TD and NC are both in the denominator, decreasing either of them will increase the fitness score.

Reproduction: A next generation is generated from those selected through genetic operators: crossover, and/or mutation. A two point cross-over operation starts by selection of a pair of chromosomes of the existing population by tournament selection. Two points on the binary array is predicted randomly. For each cross-over operation, two offspring are generated for the next generation. Mutation is performed by selecting a random unselected character (set 0 to 1 in binary string) and unselecting a random selected character (set 1 to 0 in binary string) for a given number of m randomly selected individuals from current generation, where $es + co + m$ equals to the total number of individuals n . For the analysis of the complete data, the ratio of $es:co:m$ was 5:95:2 respectively for a population with a size of 100.

Selection: The selection function chooses parents for the next generation based on the fitness scores. Individuals with highest fitness scores in current generation are selected as parents and take place of the individuals with lowest fitness scores of next generation. Elitist selection is used to build the 5% of the next generation.

Termination: The algorithm ends in either one of the two conditions; it ends if the change in the average fitness score among the last 100 generations equals to zero, or alternatively, it ends after a given number of iterations (5000) has been reached. When the algorithm ends, it outputs the chromosomes of the best solution.

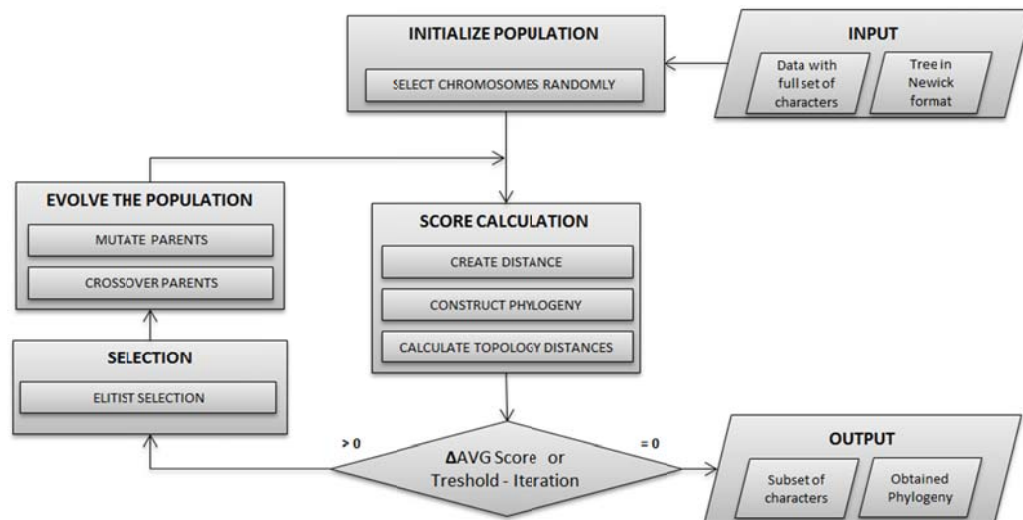


Figure 3.2 Optimization Algorithm

The algorithm starts with a population of randomly generated individuals. Later generations are created by evolving (crossing over, elitist selection and mutation) chromosomes of the previous generations. It ends up at a certain number of iterations or in the case of no change in average fitness score among 10 generations. The procedure is the same for each subpopulation and the main population except that the main population is the collection of subpopulations rather than randomly creating individuals.

3.2.3. *PHYLOGENY RECONSTRUCTION*

PHYLIP version 3.67 was used for phylogeny analysis (Felsenstein 1989; 1998). For each individual in a population, a data matrix is created from chromosomes with smaller subset of data. A distance matrix is calculated by using the similarities between the species. NEIGHBOR routine (Felsenstein 1998) is used to construct UPGMA tree (Sokal and Michener 1958b) from the distance matrix. Characters are considered without giving them weights and no out-group has been set. TREEDIST (Felsenstein 1998) is used to calculate the distance between the tree of the original data and the tree of the optimized data. Only topological distances between two trees

have been calculated since a larger distance is not significantly larger than a smaller one in the calculation of branch length distances (Felsenstein 1998).

3.2.4. PHI COEFFICIENT

I assumed that there might be a congruency among some of the characters. In this case they would occur frequently in the same solution sets. There are also some other characters that would be involved in the sets rotatory, where characters are alternatives of each other. To predict the congruence or incongruence between those character pairs, I need to use Phi coefficient. Phi coefficient is a measure of the degree of association between two binary variables (Dytham 2003). The occurrences of characters in the solution sets are taken as binary values. Thus, the degree of association between two characters can be assessed as a negative or positive value of the association. I stated the rules for phi coefficient for this study as in common: -1.0 to -0.5 strong negative association, -0.5 to -0.3 weak negative association, -0.3 to +0.3 little or no association, +0.3 to +0.5 weak positive association, and +0.5 to +1.0 strong positive association.

3.2.5. TEST FOR OPTIMIZATION

To see that whether the proposed algorithm reaches the optimum solution, results from 20 different analyses for complete data. The best 10 individuals were picked according to fitness score to create a population. By using this population computed optimization algorithm were checked whether any better result subset with fewer characters can be obtained.

On the other hand, the complexity of the problem needs to be shown. After finding an optimized result, it was tested that the same or a better solution could not be obtained randomly. For this reason, 1000 unique subsets with the same number of

characters in the optimized result were created, and phylogenies for each subset were constructed.



Figure 3.3 Optimization process

Graphical representation of a sample analysis of the optimization process for the *Hoplitis* data. Averages of number of characters and fitness scores were plotted with vertical minimum-maximum lines. Average of number of features starts with 40 characters and converges to a solution at 10 characters. Details of the process were tabulated in Table A.1 within the Appendix section.

3.2.6. PERTURBANCE TECHNIQUE

Simulated data were applied for testing the accuracy of both the Relative Complexity Measure (RCM) and the proposed method. However, robustness of the datasets were controversial. Therefore one need to select robust sequence families by which the same topology can be reconstructed even if small changes occur. RCM have been applied in the analysis of molecular sequences (in chapter 5), and it does not rely on Multiple Sequence Alignment (MSA) (Bastola et al. 2004; Otu and Sayood 2003). Since all resampling techniques for molecular sequences (Archie 1989; Felsenstein

1985; Penny and Hendy 1985; Wu 1986; Zharkikh and Li. 1995) were based on MSA, there is a need to prove the robustness of the sequence family by a resampling method that does not require a preceding MSA step. For this aim, a new algorithm have been developed that perform nucleotide mutations in order to create a perturbation from the original dataset. This algorithm generated such samples by randomly mutating for a given $k\%$ of the bases of the original sequence via equal instances of insertion, deletion and substitution mimicking the evolutionary process. Bases that were removed from the original sequence by the deletion process were collected into a pool, and exchanged with the bases from the sequence by substitution. Finally all bases within the pool were inserted into the sequence separately. Using this algorithm, mutations accounting for the 5% of the original sequence were introduced. This procedure has been repeated for 1000 times for each sequence to create any of the 5% degenerations.



Figure 3.4 Perturbation Technique

A thousand datasets were created by perturbation as recommended for most resampling studies (Efron and Tibshirani 1993). Both for nuclear ITS and chloroplast data, perturbed data sets have been created at different mutation rates to find the most appropriate mutation level. Since the biological sequence datasets might contain sequencing errors, this perturbation method enables us to test the robustness of the tree obtained even in the presence of such experimental errors.

3.2.7. *RELATIVE COMPLEXITY MEASURE METHOD*

RCM computes organisms' relatedness based on the relational complexity of the sequences. Overall work flow as explained by Out and Sayood (Otu and Sayood 2003) is as follows: The two step method begins with an exhaustive library of the primary sequence generated by the Lempel and Ziv (LZ) method (Lempel and Ziv 1976). A new library is created from the second sequence by using the exhaustive library of the primary sequence as a starting point. At the heart of the RCM lies the calculation of the LZ-complexity of a sequence, which is obtained by counting the number of steps needed to generate a copy of the sequence starting from a null state. Each step involves a process of copying a nucleotide or a series of nucleotides for a sequence and then adding the next nucleotide from the sequence being analyzed. The number of steps needed to obtain the exhaustive library is identified as the LZ-complexity value of the given sequence. The LZ-complexity of "TGATGCGACACA" is obtained as an example in Table 3.1. Since six steps were needed to generate the exhaustive library, the LZ-complexity for the analyzed sequence is '6'.

In each step of the LZ production process, copying takes place as much as possible from history followed by a single nucleotide addition. For example, in the first three

steps shown in Table 3.1 there is no copy action; only single nucleotide addition is performed. Fourth step starts with search term T. Different from the first three steps, nucleotide T at current position can be copied from the history because T can be found at the first position in the sequence. While search term can be copied from history, current search term - which is T - will be expanded with the nucleotide at the next position. After addition of the next letter, search term becomes TG. Since the updated search term, TG, is also found in the history at position [1, 2], search term will be expanded with the next nucleotide and will be searched again in the history. Updated new search term is TGC, but it is failed at search in the history. Fourth step ends with copying the last successful search term from the history and single nucleotide addition follows.

Given the two sequences S_1 and S_2 , RCM is defined as the number of steps required generating S_1 from S_2 as opposed the generating S_1 from a null state. This measure, which defines a level of closeness between the two sequences, forms the bases of a number of distance metrics that have been shown to be fit to use in phylogenetic analysis. Let $c(S)$ denotes the LZ-complexity of the sequence S . The RCM distance between the two sequences S_1 and S_2 is

$$d(S_1, S_2) = \frac{c(S_1S_2) - c(S_1) + c(S_2S_1) - c(S_2)}{\frac{1}{2}[c(S_1S_2) + c(S_2S_1)]}$$

Here, S_1S_2 represents the concatenation of the two sequences S_1 and S_2 ; therefore, $c(S_2S_1) - c(S_2)$ reflects the number of steps required generating S_1 from S_2 . The calculations were done using the original algorithm and software as previously described (Otu and Sayood 2003) and there were no disagreements between the

topologies rendered by different RCM distance methods as previously reported (Otu and Sayood 2003).

Table 3.1 Explanation of LZ-complexity of the sequence “TGATGCGACACA”.

Step	Copy	Add	Generated Sequence (Z)
1	-	T	T
2	-	G	T . G
3	-	A	T . G . A
4	TG	C	T . G . A . TGC
5	GA	C	T . G . A . TGC . GAC
6	ACA	-	T . G . A . TGC . GAC . ACA

3.2.8. *TREE CONSTRUCTION AND TOPOLOGY DISTANCE METHODS*

A matrix of distances between all OTUs for each dataset were constructed based on the aforementioned RCM distance method. Resulting matrices were used to reconstruct phylogenies using neighbor-joining (NJ) method in NEIGHBOR routine of Phylogeny Inference Package (PHYLIP version 3.67) (Felsenstein 1989; 1998). Consensus of all output trees of NEIGHBOR were computed by CONSENSE routine of PHYLIP using extended Majority Rule (MRe). TREEDIST routine of the PHYLIP package was used to compute topology (symmetric) distances (Robinson and Foulds 1981) between the original tree and the consensus tree of 1000 bootstraps for each mutation level. The branch lengths were not considered as significant in this study. Only the topology distances were calculated in order to present the stability or the change in the phylogeny.

In the comparison of RCM with other methods of phylogeny construction, RCM-NEIGHBOR, DNADIST-NEIGHBOR, DNAML and DNAPARS routines of PHYLIP were used. F84, GTR and K2 DNA models were used to create distance

matrices by DNADIST. Default program parameters were used for the rest of the processes.

3.2.9. LONGEST COMMON SUBSTRING METHOD

The Longest Common Substring (LCS) problem is to find the longest string (or strings) that is a substring (or are substrings) of two or more strings (Gusfield 1999).

In another definition, a common substring of two strings is a substring that appears in both strings. The longest common substring is a common substring of the maximal length. The Longest common substring of a string S , is a set of characters that appear in left to right order, consecutively (Gusfield 1999).

```
TCAAACACCCAAAGCGAGGCACTGA TTAATTAGAGGTAAAATTACACATG AATGCCTAGATGGGCTCACCCAGCCCGTAAACATACA
TTAAACTCTTAAAGCAATACACTGAA TTAATTAGAGGTAAAATTACACATG AATGTTTAGATGATTGTAACCTGAACCCGAGCGCATAA
TTAATTTAAAGCAAAGCACTGAAA TTAATTAGAGGTAAAATTACACATG ATGCTTAGATGGTTTATATGTTAAACCCATAAACACATAG
TTAATGATATCAAAGCAAGGCAC TTAATTAGAGGTAAAATTACACATG GAAAATGCCTAGATGAGTATTCCTACTCCATAAACACATA
TTAAGTCTAAAGCGAGGCACTGA TTAATTAGAGGTAAAATTACACATG ATGCCTAAACGAATCCTAATGATTCCGCAGACACACAGG
TTAAACTAAAGCAAGGCACTGAAA TTAATTAGAGGTAAAATTACACATG CCTAGATGAGTCTGCCTACTCCATAAACATAAAGGTT
TTAAATTAAGCAAAGCACTGAAA TTAATTAGAGGTAAAATTACACATG ATGCTTAGAAGGGCTATACCCCGTAAACATATAATTA
TTAAACAAAGCAAGGCGCTGA TTAATTAGAGGTAAAATTACACATG TGCCTAGATGAGTCATAGACTCCATAAACATACAGGTTT
TTAATTACAAAGCAAGGCACTG TTAATTAGAGGTAAAATTACACATG TGCCTAGATGAGTACGCGCTACTCCATAAACATAAAGG
TTAATAAAAAGTGAGGCACTGAAA TTAATTAGAGGTAAAATTACACATG ATGCCTAGATGAGTATTAACCTCCATAAACATAAAGGTTTG

LCS = TTAATTAGAGGTAAAATTACACATG
```

Figure 3.5 Sample for Longest Common Substring method.

CHAPTER 4. EVALUATION OF MORPHOLOGICAL DATA

4.1. INTRODUCTION

Numerical Taxonomy (NT) is the grouping of taxonomic entities into a taxa by numerical methods on the basis of their character states (Sneath and Sokal 1973). It can also be considered as an attempt to reconstruct evolutionary history of species, usually in molecular, morphology or other traits, by means of Phylogenetics. NT techniques include various forms of clustering and ordination which are sophisticated ways of reducing the variation displayed by organisms to a manageable level. In practice, the process of clustering is made possible by measuring dozens of variables (Cavalli-Sforza and Edwards 1967; Sokal 1963), and presenting the results as cladograms also known as phylogenetic trees.

Much of the technical challenge in NT revolves around balancing the loss of information in such a reduction against the ease of interpreting the resulting trees. Numerical studies in systematics are carried out by the data with a large number of characters (Le Quesne 1969). Some part of this data may be irrelevant or sometimes meaningless. Recent advances in phyloinformatics have made possible to extract uninformative characters and exclude them from the data in parsimony analysis (Swofford 1996). However, most of the techniques were implemented for the studies in molecular bases rather than morphology. Most recently, some new techniques have been described for inferring phylogenetic trees by using answer set programming (Brooks et al. 2007) or for by a particle swarm optimization-aided

fuzzy cloud classifier (Lu et al. 2009). The aim of those methods, different from the current approach used in this thesis work, were the selection of compatible characters in the construction of phylogeny.

Aim of this study is to find optimum set(s) of morphological characters that gives an acceptable approximate or the exact solution to the original topology. By doing so, the workload over the scientists and the labor for gathering data can be decreased by choosing the right characters and eliminating the irrelevant or meaningless ones, while increasing the efficient use of time. A suggestion to give an exact or most approximate solution to find a subset of data with optimum numbers of characters that will represent original data is Genetic Algorithms.

4.2. THE DATASETS

4.2.1. *THE HOPLITIS DATASET*

The dataset used in this study is a classical data used in Michener and Sokal (Michener and Sokal 1957), seniors in Numerical Taxonomy and also founders of the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sokal and Michener 1958b) method. Philosophical bases of their attempts at quantifying systematic relationships have been reported in (Michener and Sokal 1957) and later it was used in the proof of UPGMA method (Sokal and Michener 1958b).

The dataset is based on the American species of the megachilid genera *Ashmeadiella*, *Anthocopa*, *Hoplitis*, and *Proteriades* (Table 4.1). It was consisting of 97 species of solitary bees in the family Megachilidae containing two groups each containing two genera. The data set was composed of 122 characters, each varying among the species so that there were from two to eight alternatives or states per character.

Relationships between the species were compared with relationships as found in the analysis by UPGMA and a good agreement was found in (Sokal and Michener 1958b). The data was found to give a resolution to relationships within the family Megachilidae. However, the data set may contain many characters where some might be irrelevant or uninformative. Therefore, one needs to extract a small portion of the data. The dataset was therefore divided into four pieces representing four genera; *Proteriades*, *Hoplitis*, *Anthocopa* and *Ashmeadiella*. OTUs belong to *Proteriades* genus extend from 1 to 18, *Hoplitis* from 19 to 41, *Anthocopa* from 42 to 62 and *Ashmeadiella* from 63 to 97 (Table 4.1). Details of the data can be available from the report of Michener (Michener and Sokal 1957).

Table 4.1 Taxonomic hierarchy of *Hoplitis* complex (Michener and Sokal 1957).

Kingdom: Animalia
Phylum: Arthropoda
Subphylum: Hexapoda
Class: Insecta
Order: Hymenoptera
Suborder: Apocrita
Superfamily: Apoidea
Family: Megachilidae
Genus group 1
<i>Proteriades</i> (Genus)
<i>Cephalapis</i> (Subgenus)
1. <i>Proteriades jacintana</i>
<i>Proteriades</i>
2. <i>Proteriades semirubra</i>
3. <i>Proteriades deserticola</i>
<i>Xerosmia</i>
4. <i>Proteriades xerophila</i>
5. <i>Proteriades nanula</i>
6. <i>Proteriades pygmaea</i>
7. <i>Proteriades reducta</i>
8. <i>Proteriades cryptanthae</i>
9. <i>Proteriades caudex</i>

10. *Proteriades hamulicornis*
11. *Proteriades truicauda*
12. *Proteriades bidenticauda*
13. *Proteriades nigrella*
14. *Proteriades tristis*
15. *Proteriades palmarum*
16. *Proteriades seminigra*
17. *Proteriades remotula*
18. *Proteriades incanescens*

Hoplitis

Hoplitima

19. *Hoplitis howardi*
20. *Hoplitis bunocephala*
21. *Hoplitis bullifacies*

Acrosmia

22. *Hoplitis laevibullata*
23. *Hoplitis rufina*

Andronicus

24. *Hoplitis cylindrica*

Monumetha

25. *Hoplitis fulgida*
26. *Hoplitis louisae*
27. *Hoplitis viridimicans*
28. *Hoplitis albifrons*

Dasyosmia

29. *Hoplitis biscutellae*
30. *Hoplitis paroselae*

Cyrtosmia

31. *Hoplitis hypocrita*

Alcidamea

32. *Hoplitis producta*
33. *Hoplitis pilosifrons*
34. *Hoplitis grinnelli*
35. *Hoplitis colei*
36. *Hoplitis elongaticeps*
37. *Hoplitis uvulalis*
38. *Hoplitis sambuci*
39. *Hoplitis truncata*

Formicapis

40. *Hoplitis clypeata*

Robertsonella

41. *Hoplitis simplex*

Genus group 2

Anthocopa

Atoposmia

42. *Anthocopa triodonta*
43. *Anthocopa hebitis*
44. *Anthocopa pycnogncatla*
45. *Anthocopa oregona*
46. *Anthocopa anthodyta*
47. *Anthocopa panamintensis*
48. *Anthocopa elongata*
49. *Anthocopa abjecta*

Hexosmia

50. *Anthocopa copelandica*

Eremosmia

51. *Anthocopa hemizoniae*
52. *Anthocopa viguierae*
53. *Anthocopa enceliae*
54. *Anthocopa rubrella*
55. *Anthocopa hurdiana*
56. *Anthocopa daleae*
57. *Anthocopa beameri*
58. *Anthocopa mirifica*
59. *Anthocopa hypostomalis*
60. *Anthocopa nitidivitta*
61. *Anthocopa rupestris*
62. *Anthocopa robustula*

Ashmeadiella

Chilosima

63. *Ashmeadiella holtii*
64. *Ashmeadiella rhodognatha*

Ashmeadiella

65. *Ashmeadiella altadenae*
66. *Ashmeadiella rufitarsis*
67. *Ashmeadiella titusi*
68. *Ashmeadiella femnorata*
69. *Ashmeadiella truncativentris*
70. *Ashmeadiella rufipes*

71. *Ashmeadiella prosopidis*
72. *Ashmeadiella sonora*
73. *Ashmeadiella occipitalis*
74. *Ashmeadiella buconis*
75. *Ashmeadiella cactorunt*
76. *Ashmeadiella bigeloviae*
77. *Ashmeadiella difugita*
78. *Ashmeadiella leucozona*
79. *Ashmeadiella foveata*
80. *Ashmeadiella gillcttei*
81. *Ashmeadiella maxima*
82. *Ashmeadiella meliloti*
83. *Ashmeadiella aridula*
84. *Ashmeadiella californica*
85. *Ashmeadiella opuntiae*

Titusella

86. *Ashmeadiella cubiceps*

Rhamphormia

87. *Ashmeadiella foxiella*
88. *Ashmeadiella timberlakei*
89. *Ashmeadiella eurynorhyncha*
90. *Ashmeadiella stenognatha*
91. *Ashmeadiella salviae*
92. *Ashmeadiella cazieri*
93. *Ashmeadiella clypeodentata*
94. *Ashmeadiella australis*

Corythocila

95. *Ashmeadiella breviceps*
96. *Ashmeadiella inyoensis*

Cubitognatha

97. *Ashmeadiella xenomastax*

4.2.2. THE ACORN DATASET

Acorn data has derived from (Bakış 2005) and is composed of 33 morphological characters extracted from acorns of 14 Turkish *Quercus* species (Table 4.2). Use of acorn data in phylogeny construction of *Quercus* is important since results of Bakış (Bakış 2005) were in contrast with common view that reproductive characters are

unhelpful in classification (Stace 1989). On the other hand, the dataset were tested by several other studies (Bakiş et al. 2009; Çakmak 2010; Çakmak et al. 2011; Erdem 2011) for evaluation of different methods.

I was very familiar with the *Quercus* dataset as I am the author of the dataset.

Therefore, it would be easier to understand the data and the information contents of the characters by the proposed method.

Table 4.2 Taxonomic hierarchy of *Quercus* genus (Bakiş, 2005).

Kingdom Plantae – Plants
Subkingdom Tracheobionta – Vascular plants
Division Magnoliophyta – Angiosperms, flowering plants
Class Magnoliopsida – Dicotyledons
Subclass Hamamelidae
Order Fagales
Family Fagaceae
Genus <i>Quercus</i> L. – Oaks
Section <i>Quercus</i> (Endl.) Örsted: Section Leucobalanus – White Oaks
<i>Q. pontica</i> C. Koch
<i>Q. robur</i> L.
<i>Q. robur</i> subsp. <i>robur</i>
<i>Q. robur</i> subsp. <i>pedunculiflora</i> (C. Koch) Menitsky
<i>Q. hartwissiana</i> Steven
<i>Q. macranthera</i> subsp. <i>syspirensis</i> (C. Koch) Menitsky
<i>Q. frainetto</i> Ten.
<i>Q. petraea</i> (Mattuschka) Liebl.
<i>Q. petraea</i> subsp. <i>petraea</i>
<i>Q. petraea</i> subsp. <i>iberica</i> (Steven ex Bieb.) Krassiln.
<i>Q. petraea</i> subsp. <i>pinnatiloba</i> (C. Koch) Menitsky
<i>Q. vulcanica</i> [Boiss & Heldr. ex] Kotschy
<i>Q. infectoria</i> Olivier
<i>Q. infectoria</i> subsp. <i>infectoria</i>
<i>Q. infectoria</i> subsp. <i>boissieri</i> (Reuter) O. Schwarz
<i>Q. pubescens</i> Willd.
<i>Q. virgiliana</i> Ten.
Section <i>Cerris</i> Loudon – Red or Black Oaks
<i>Q. cerris</i> L.

<i>Q. cerris</i> var. <i>cerris</i>
<i>Q. cerris</i> var. <i>austriaca</i> (Willd.) Loudon
<i>Q. ithaburensis</i> subsp. <i>macrolepis</i> (Kotschy) Hedge and Yalt.
<i>Q. brantii</i> Lindley
<i>Q. libani</i> Olivier
<i>Q. trojana</i> P.B. Webb.
Section <i>Ilex</i> Loudon – Evergreen Oaks
<i>Q. ilex</i> L.
<i>Q. aucheri</i> Jaub. & Spach
<i>Q. coccifera</i> L.

4.3. RESULTS

4.3.1. HOPLITIS DATASET

A feature selection algorithm has been developed for the selection of minimum number of characters that will produce the same topology as the full set of data does. A graphical representation of the optimization algorithm has been shown in Figure 3.2 and the island model used for the algorithm in Figure 3.1. A sample run has also been represented in Figure 3.3. Number of characters has a given initial value of 15 in the beginning and increases up to 23 at iteration 95. During the number of characters increases, topology distances decrease and set to 0 and fitness score increases exponentially. After this point topology distance line becomes stable at 0 and number of characters decreases from 23 to 10 within 150 iterations. Fitness score increase linear during this period. This point forward all series becomes stable for 100 iterations.

Feature selection algorithm was able to find the same topology with smaller number of characters for each dataset. In the evaluation of the complete dataset, minimum number of characters within the solution subsets was 67 in order to construct the same tree topology. The data has been divided into four unequal parts; *Proteriades*,

Hoplitis, *Anthocopa* and *Ashmeadiella*. *Ashmeadiella* dataset was the largest among the subsets with 35 OTUs while others were relatively close together. These subsets were composed of OTU numbers of 18, 23, 21 and 35 and the minimum numbers of characters reached after optimization process were found as 9, 9, 6 and 29 respectively.

Each subset of data were analyzed 100 times, by using the results of these analyses frequency of the characters within the 100 solutions were calculated and shown in Figure 4.1. Especially in *Proteriades*, *Anthocopa* datasets, the dominant characters can easily be recognized. Frequency of the characters within the 100 unique solutions has showed that some characters were chosen all or almost all of the solution subsets while some others were never chosen (Figure 4.1). Besides, frequency of chosen characters differs in subsets, only characters 12 and 13 were found within both *Anthocopa* and *Ashmeadiella* datasets with high frequencies.

Phi coefficient results were tabulated in Table B.3 within the Appendix section. Character couples with negative phi coefficients show incongruence while characters with positive phi coefficients show congruence. There are also some groups of characters were obtained that show strong positive association among each other. And also, characters that are found in some of these groups may show strong negative association with characters of other groups.

In *Proteriades* dataset, two main groups of characters were identified based on high phi coefficients. These characters are in congruence with each other within their own groups. These groups of characters are [97, 8, 2] and [39, 98, 1]. On the other hand, any character from the first group shows incongruence with any other character from the second group by bearing a strong negative value of phi coefficient. There are also

a couple of characters that show strong negative association with characters 56 and 96. In *Hoplitis* dataset, there are two different cases; in the first case, a group of characters [14, 90, 70, 15] are in strong positive association within each other, and all of these characters are in strong negative association with character 111. In the second case there are two groups; [32, 33, 118, 19] and [46, 96, 115]. These groups show strong positive association among each other and strong negative association inter-groups. A group of characters with strong positive association and a couple with strong negative association has been found in *Anthocopa* dataset. The group is [12, 110, 43] and the couple of characters is 53 and 119. There are two main groups of characters in *Ashmeadiella* and a large portion of the characters were not included in phi coefficient due to very high value of frequency (95-100%) would not give a result in phi coefficient. The groups are [37, 74, 98, 64, 23, 35, 71] and [57, 19, 36, 49, 109, 63, 62, 78]. The characters of the two groups show negative association with the characters from the other group.

A test for optimization was performed to see that whether the proposed algorithm reaches to an optimum solution. The result was exactly the same as the best solution of the 20 previous analysis results. The algorithm was not able to return a better result subset with fewer characters. In the second test, one needs to see that the result of the optimization algorithm could not be obtained randomly. In order to show that 1000 random subsets each composed of 67 characters were calculated. The best solution of this test was with 58 topology distance, which has a different topology completely.

Table 4.3 Specifications of the data and obtained optimization and accuracy results.

	Number of OTUs	Number of characters in solution subsets	Accuracy of obtained trees
Complete data	97	67	100%
<i>Proteriades</i> data	18	9	100%
<i>Hoplitis</i> data	23	9	100%
<i>Anthocopa</i> data	21	6	100%
<i>Ashmeadiella</i> data	35	29	100%

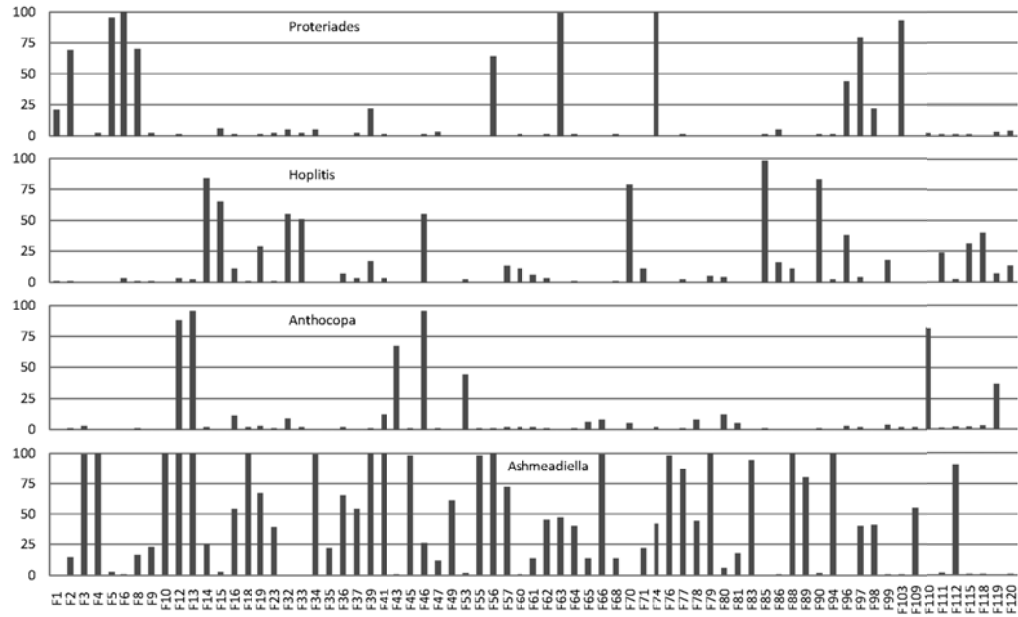


Figure 4.1 Frequency of the characters that are found in 100 unique solution subsets.

4.3.2. ACORN DATASET

A correlation test showed that several characters were in correlation with in Acorn data. However, elimination of any of those characters leads to certain changes in the resulting phylogeny. The algorithm started with sets of 15 randomly chosen characters within the Acorn data and search through smaller sets with similar topology. There is no obligatory rule for starting number of characters. Optimization algorithm has successfully resulted in subsets of the original data starting from 6 acorn features and 15 species distribution features. An expanded list of the resulting

acorn subsets and the occurrences of the characters is given in Table B.4 and Table B.5 in Appendix section.

A phylogeny has been reconstructed for each individual in generations. Figure 4.2 represents the original phylogenetic tree computed with 33 characters (A) and a phylogenetic tree of optimized data with 6 characters (B). The trees have exactly the same topology. The branch lengths are different for many taxa between two trees. The branch lengths of the tree of the optimized data have bigger differences than that of the original tree.

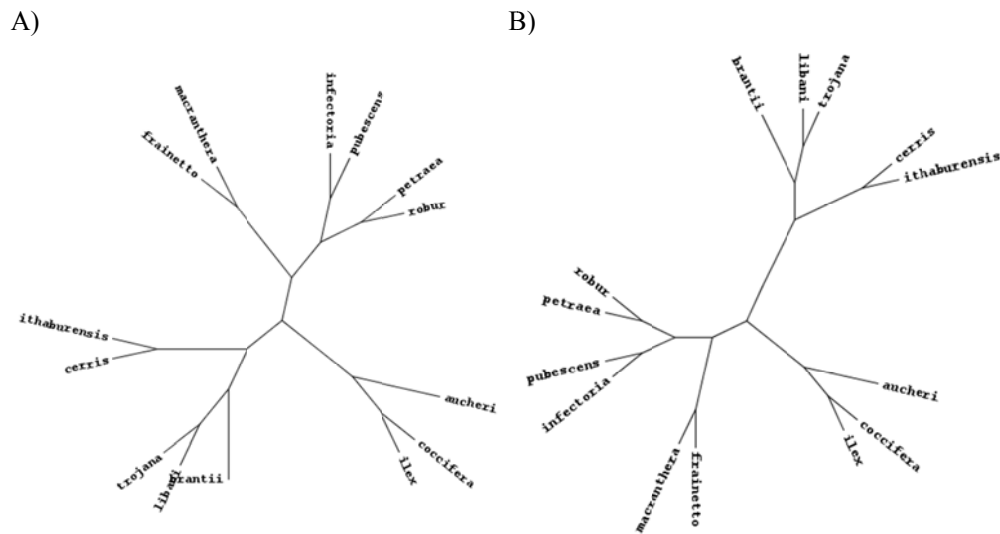


Figure 4.2 Comparing the tree of the original data with 33 characters (A) and the tree of the optimized data with 6 characters (B).

Three different sets of acorn characters gave exactly the same topology with the original phylogeny. All three have 2nd and 32nd characters which are also the most frequently occurred characters among all three sets with 17 occurrences. 16, 30, 1, 13, 23, 5, 31, 4, 6, 14 and 28 are the other characters highly occurred within the solution sets. However, some of the characters, 7, 9 and 24, were never involved in any set. The phi coefficient values of each character couples listed in Table B.5 in the Appendix and Figure 4.1 represents a graphical demonstration of the associations

being involved. 1-2, 1-32, 2-5, 5-32 have strong negative associations, while 2-32 and 5-31 have strong positive associations. All the remaining characters were weakly associated or have no association.

CHAPTER 5. EVALUATION OF MOLECULAR SEQUENCES

5.1. THEORETICAL BACKGROUND

Plenty of phylogenetic analysis techniques exist showing relations among organisms based on molecular sequences. Most commonly utilized methods are distance based methods and evolutionary methods (Felsenstein 1988; Swofford and Olsen 1990). Distance based methods, such as Neighbor-Joining (Saitou and Nei 1987) and UPGMA (Sokal and Michener 1958a), explicitly rely on a measure of genetic distance between OTUs based on their sequence differences. Distance measures are derived from pairwise comparisons of the sequences. Whereas the distance based methods represent a sequence divergence by a single number, the evolutionary methods attempt to infer the phylogeny by fitting individual characters (nucleotides or amino acids) to the tree. The most popular approaches for evolutionary methods are the maximum likelihood (Felsenstein 1973), the maximum parsimony (Mount 2004), and the Bayesian inference (Huelsenbeck and Ronquist 2001). These methods compare various evolutionary trees that could describe the relationships among given taxa. The tree that implies the fewest or most likely evolutionary changes in the characters is taken to be the best estimate of the true phylogenetic tree. A new method in phylogeny reconstruction, namely, simultaneous sequence alignment has recently been introduced by (Yue et al. 2009). This new method is particular in that it can perform true alignment and phylogenetic inference. Mutations can then be included in the overall tree score.

In most distance based and evolutionary methods, one must perform a multiple sequence alignment (MSA) first to infer the relationships between the given set of molecular sequences. MSA is still a notoriously difficult computational problem and different tools often produce different alignments (Chen and Tompa 2010; Löytynoja and Goldman 2008). This leaves users with the problem of choosing the most appropriate alignment method for their use. Distance based or character based phylogeny construction methods that rely on an MSA can lead to misleading trees due to errors in the alignment (Liu et al. 2009). First of all, when more divergent sequences are to be incorporated, generating reliable multiple alignments becomes increasingly difficult (Stropea and Moriyama 2007). The second problem is the difficulty in the choice of suitable scoring matrices and gap penalties for different sequence sets (Mount 2004). Another major problem arises when progressive MSA schemes – which are increasingly popular – are used such as CLUSTALW (Thompson et al. 1994), T-Coffee (Tommaso et al. 2011) and MAFFT (Katoh et al. 2005). In such schemes, the reliance of the final MSA is mostly based on a good alignment of the first two sequences which are generally the most closely related sequences (Thompson et al. 1994). If the error in the alignment of the more closely related sequences cannot be corrected, the error will be increasingly serious within the rest of the alignment procedures. Another disadvantage is that the number of sequences is limited. The computing time complexity is proportional to the exponential of the number of the sequences to be analyzed (Shyu et al. 2003; Thompson et al. 1994; Yue et al. 2009).

To overcome these problems, an alignment-free phylogeny construction seems to be a reasonable way. Application of the non-alignment based methods has gained popularity especially in the analysis of the genome-wide sequences where MSA can

hardly be utilized. Relative Complexity Measure method (Otu and Sayood 2003) has been proposed as a sequence distance measure with applications to phylogeny construction based on Lempel-Ziv complexity (Lempel and Ziv 1976). It is an alignment free method, since it does not require a preceding multiple alignment procedure and can be grouped within the distance based methods. Although the phylogenies generated by RCM were found highly reasonable (Bastola et al. 2004; Otu and Sayood 2003) the statistical significance of the trees have not been tested in detail. Resampling techniques utilizing bootstrap have been used in phylogenetic analyses to assess statistical significance of the tree topology: complete and partial bootstrap (Felsenstein 1985; Shimodaira 2002; Zharkikh and Li. 1995), block bootstrap (Kunsch 1989), jackknife; delete-half (Felsenstein 1985) and delete fraction (Farris et al. 1996), permuting species or characters (Archie 1989). However, most of these methods are based on data with equal-size samples or require a preceding alignment procedure. Since RCM does not rely on alignment, a perturbation technique mimicking the evolutionary process can be proposed in order to test the robustness of the trees obtained by RCM. The accuracy of the Relative Complexity Measure method using different sets of benchmark sequence families generated by ROSE (Stoye et al. 1998), a probabilistic model of the evolution of DNA sequences.

5.2. DATA SETS

5.2.1. *HEME COPPER OXIDASES*

Heme-copper oxygen reductases (HCOs) are membrane bound enzymes that are responsible for the reduction of oxygen to water in respiratory chains (Kim et al. 2004; Pereira et al. 2001). The catalytic subunit of these enzymes is characterized by

the presence of a low-spin heme, a binuclear center composed of a high-spin heme, a copper ion (Cu_B) and a tyrosine residue cross-linked to one of the histidine ligand of the Cu_B (Sousa et al. 2011). Based on their proton conducting channels, these enzymes were proposed to be divided into three different sub-types; A, B and C type enzymes (Pereira et al. 2001). The A type is further divided into A1 and A2. The structural and functional properties of HCOs and Nitric oxide reductases (NORs) has been listed in Table 5.1.

NORs are membrane bound enzymes. They catalyze the reduction of NO to water and N_2O and their catalytic subunit harbors a low spin heme and a binuclear center composed of a high-spin heme and an iron ion (Fe) (Zumft 2005). Their catalytic subunit shares the same predicted structural fold and the same SCOP family classification with HCOs. Therefore, NORs are proposed to have shared a common ancestor with HCOs (Murzin et al. 1995).

Table 5.1 Structural and functional properties of HCOs and NORs.

	Binuclear center				Energy conservation		
	Heme-Cu	Heme-Fe	Tyr-I	Tyr-II	Charge separation	H^+ translocation	
						D-channel ^a	K-channel ^b
A	✓	×	✓	×	✓	✓	✓
B	✓	×	✓	×	✓	×	Alternative
C	✓	×	×	✓	✓	×	Alternative
NOR	×	✓	×	×	×	×	×

HCO database consists of a compilation of HCO and NOR sequences (Pereira and Pereira-Leal 2011; Sousa et al. 2011). Moreover, their available annotations can be retrieved from HCO database. The average length of the sequences is 589

aminoacids. To reduce the complexity of the problem, a total number of 50 sequences were chosen for the analyses, 10 from each type. The sequences were chosen in order to have a representative from each class. A reference phylogenetic sub-tree were reconstructed as in Figure 5.1-B, and the analysis were performed based on this tree.

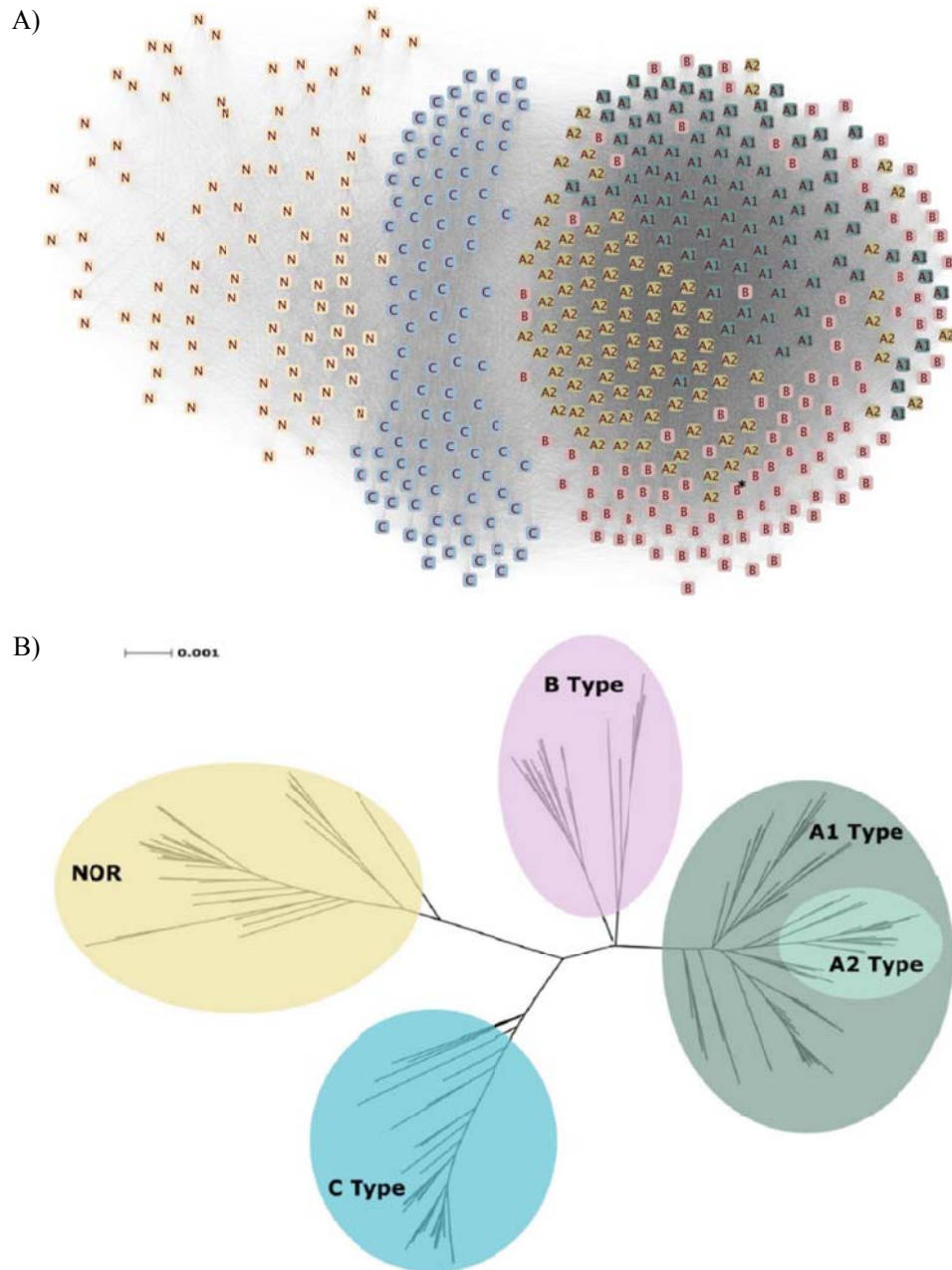


Figure 5.1 Weighted network representation (A) and phylogenetic analysis (B) of subunit I of HCO and NOR enzymes.

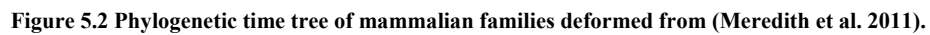
5.2.2. *MAMMALIAN MITOCHONDRIAL GENOME*

Mammals are a group of endothermic (warm-blooded) animals with backbones. They can be distinguished from other animals by the presence of milk-producing skin glands. Class Mammalia is named for these mammary glands, which they are found in every mammal. Taxonomically, Class Mammalia belongs to Kingdom Animalia and belong to the Phylum Chordata, Subphylum Vertebrata. There are two extant subclasses within the Class Mammalia: the Prototheria, – egg-laying monotremes – containing only three modern species; and the Theria – the viviparous mammals. The Theria are also subdivided into the marsupials (infraclass Metatheria), and the placental mammals (Eutheria) (Campbell and Lapointe 2011; Kjer and Honeycutt 2007).

Mammalian Mitochondrial Genomic (MMG) data belonging to 3 Prototherian and 406 Therian (Eutheria 374, Metatheria 32) taxa were obtained from Organelle Genome Resources unit of the NCBI Database (NCBI 1988). The average length of the sequences is 16,658 nt. Within these sequences, 13 protein and 24 RNA regions were identified in the genomic complex (Table 5.2). To reduce the complexity of the problem, a total number of 28 sequences were chosen for the analyses. The sequences were chosen in order to have a representative from each taxonomic class. A reference phylogenetic tree (Figure 5.2) were built for the selected sequences (Table 5.3) based on phylogeny in Figure 5.3(Meredith et al. 2011). Full list of sequences and some other information are available in Table C.2 in the Appendix section.

Table 5.2 Analyzed part of the Mammalian Mitochondrial Genomic data.

OTU	Taxon name	Accession	Length	Protein	RNA
Tachyglossidae	<i>Tachyglossus aculeatus</i>	NC_003321	16,360	13	24
Ornithorhynchidae	<i>Ornithorhynchus anatinus</i>	NC_000891	17,019	13	24
Paucituberculata	<i>Caenolestes fuliginosus</i>	NC_005828	16,629	13	24
Didelphimorphia	<i>Didelphis virginiana</i>	NC_001610	17,084	13	24
Microbiotheria	<i>Dromiciops gliroides</i>	NC_005826	15,350	13	23
Notoryctemorphia	<i>Notoryctes typhlops</i>	NC_006522	15,923	13	23
Dasyuromorphia	<i>Dasyurus hallucatus</i>	NC_007630	17,159	13	24
Peramelemorphia	<i>Perameles gunnii</i>	NC_006521	15,360	13	23
Diprotodontia	<i>Distoechurus pennatus</i>	NC_008145	16,775	13	24
Insectivora	<i>Erinaceus europaeus</i>	NC_002080	17,447	13	24
Pholidota	<i>Manis pentadactyla</i>	NC_016008	16,578	13	24
Carnivora	<i>Canis lupus</i>	NC_008092	16,729	13	24
Chiroptera	<i>Pteropus dasymallus</i>	NC_002612	16,705	13	24
Perissodactyla	<i>Equus asinus</i>	NC_001788	16,670	13	24
Cetartiodactyla	<i>Hippopotamus amphibius</i>	NC_000889	16,407	13	24
Scandentia	<i>Tupaia belangeri</i>	NC_002521	16,754	13	24
Glires	<i>Mus musculus</i>	NC_005089	16,299	13	24
Dermoptera	<i>Galeopterus variegatus</i>	NC_004031	16,748	13	24
Primates	<i>Homo sapiens</i>	NC_012920	16,569	13	24
Cingulata	<i>Dasypus novemcinctus</i>	NC_001821	17,056	13	24
Pilosa	<i>Bradypus tridactylus</i>	NC_006923	16,975	13	24
Tubulidentata	<i>Orycteropus afer</i>	NC_002078	16,816	13	24
Tenrecidae	<i>Echinops telfairi</i>	NC_002631	16,549	13	24
Chrysochloridae	<i>Chrysochloris asiatica</i>	NC_004920	16,652	13	24
Macroscelididae	<i>Macroscelides proboscideus</i>	NC_004026	16,641	13	24
Sirenia	<i>Dugong dugon</i>	NC_003314	16,850	13	24
Proboscidea	<i>Elephas maximus</i>	NC_005129	16,902	13	24
Hyracoidea	<i>Dendrohyrax dorsalis</i>	NC_010301	16,590	13	24



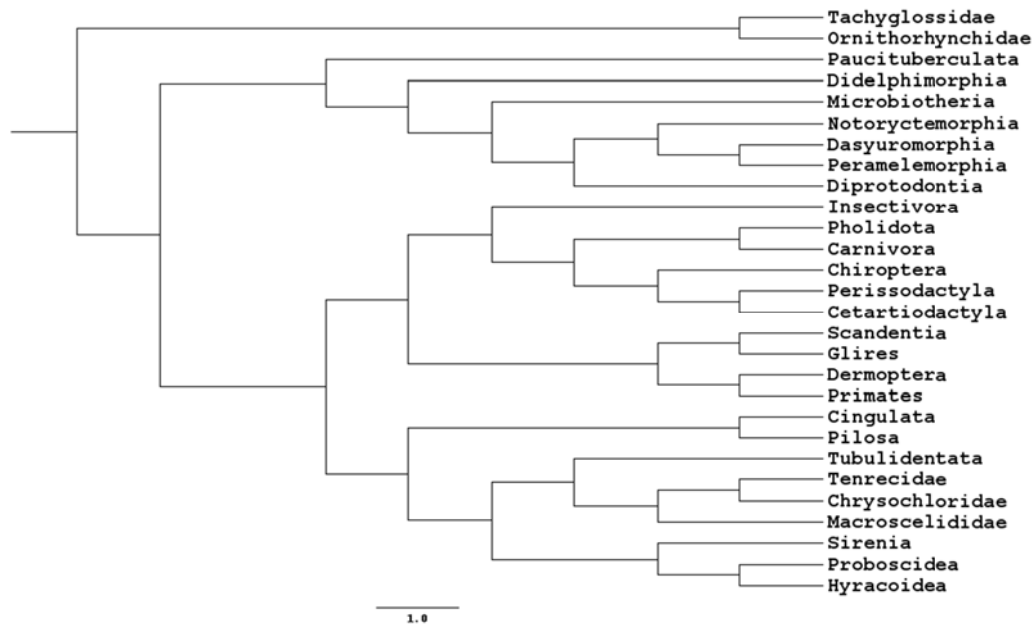


Figure 5.3 A simple representation of similarities between different groups in Class Mammalia.

5.2.3. SIMULATED DATASETS

Random Model of Sequence Evolution (ROSE) (Stoye et al. 1998) implements a probabilistic model of the biological evolution of RNA-, DNA-, or protein-like sequences. A family of related sequences is created for a given evolutionary tree, from a common ancestor sequence by insertion, deletion and substitution of characters. The correct multiple sequence alignment is created by ROSE simultaneously, since the true history of evolutionary process is logged.

First, to test the accuracy of the Relative Complexity Measure method and the proposed algorithm a benchmark sequence families generated by ROSE (version 1.3) was used. The test was performed on the balanced phylogeny (Figure 5.4) – “phylogeny a” in (Hillis and Bull 1993; Ogden and Rosenberg 2006) – with 16 OTUs. 1000 sets of sequence families were generated for each of the two DNA models, F84 and K2P. The DNA sequences have an average length of 500 letters as indicated in (Hillis and Bull 1993) and highest value of relatedness (1000) was given to show the symmetric distances between the topologies. Default values were used

[illegible]

- 60 -

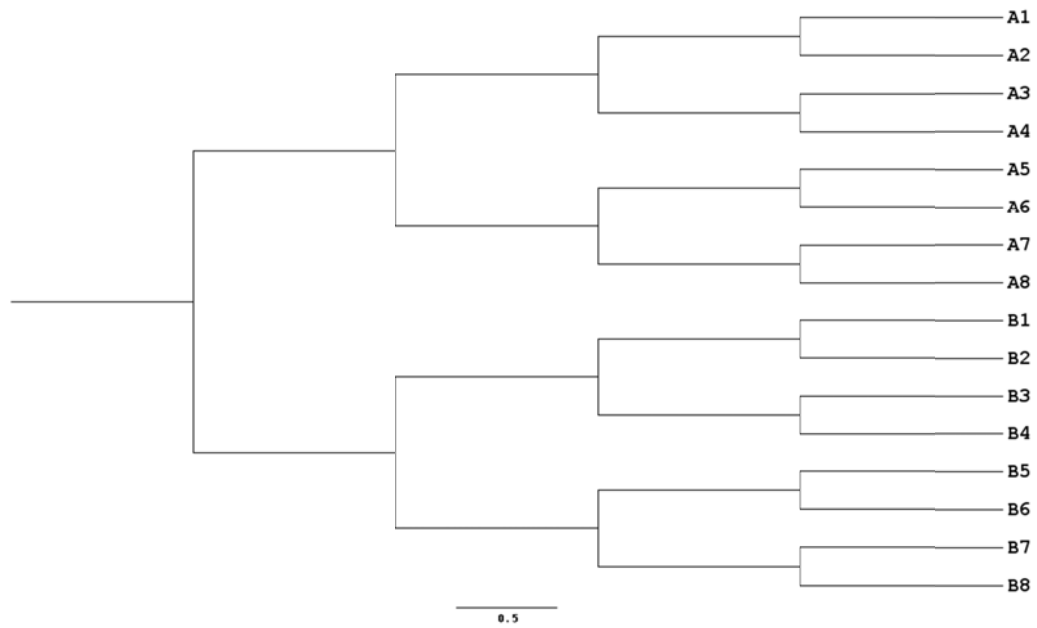


Figure 5.5 Balanced phylogeny – “phylogeny a” in (Hillis and Bull 1993) – with 16 OTUs.

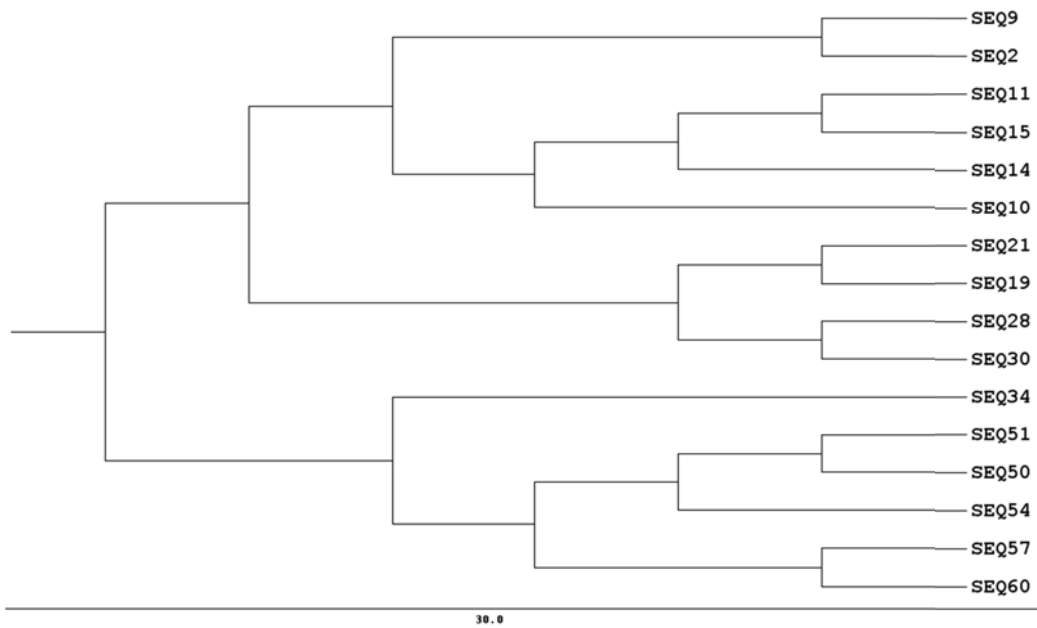


Figure 5.6 Tree topology randomly generated by ROSE (Stoye et al. 1996) for the use of longer genomic sequence benchmark.

5.3. RESULTS

5.3.1. *VALIDATION OF THE RCM APPROACH*

The feature selection algorithm does not rely on aligned sequences when the sequences are relatively long. In this case, it splits data into many distinct segments based on common blocks of data calculated by LCS technique. However, the subsequences that are composing these segments may not be suitable for alignment procedure. An alignment free method – Relative Complexity Measure – was applied to analyze the data and shown that it was working accurately.

The accuracy of the Relative Complexity Measure method was demonstrated by using different sets of benchmark sequence families generated by ROSE (Stoye et al. 1998), a probabilistic model of the evolution of DNA sequences. Simulated data were generated based on two DNA Models with balanced topology. Accuracy of the generated sequence families were also tested by perturbation technique. Only the sequence families that have zero topology distances were included into analysis.

The phylogenies of the collected sets of sequences were constructed by POY (Varón et al. 2009; Wheeler et al. 2006), DNADist, DNAML (Felsenstein 1989), PhyML (Guindon et al. 2010), and RCM (Otu and Sayood 2003) methods and compared. The DNA models that were used to generate sequence families were also given as a parameter within DNADist. The phylogenies for DNADist and RCM methods were calculated by NEIGHBOR routine of Phylogeny Inference Package (PHYLIP version 3.67) (Felsenstein 1998) with default parameters. By using Robinson and Foulds' (RF) method of Symmetric Distance (Robinson and Foulds 1981) in TREEDIST routine, average distances (from the true phylogeny) of 2000 sequence

families for each phylogeny construction method were calculated and tabulated in Table 5.3.

The results of the tree comparisons have shown that the trees constructed by RCM method were the closest to the true topology demonstrating the correctness of the RCM method for both DNA models. POY and DNAML were the second in F84 and K2P models respectively.

Table 5.3 Comparison of constructed phylogenies for simulated data (standard deviations in parenthesis).

	K2P	F84
DNADist	0.213 (0.59)	0.221 (0.68)
DNAML	0.181 (0.61)	0.170 (0.58)
POY	0.151 (0.38)	0.177 (0.42)
RCM	0.116 (0.49)	0.151 (0.60)
PhyML	0.165 (0.61)	0.176 (0.61)

5.3.2. *HCO DATA*

Among the HCO sequences, 10 sequences were chosen from each group, A1, A2, B, C and NOR, where each sequence represents a taxonomic group in phylogeny in Figure 5.1-A. 50 sequences were collected by this way and HCO data were composed. The tree topology were generated from phylogeny in Figure 5.1-A.

HCO sequences were aligned by using Clustal W with default parameters (Thompson et al. 1994). The length of the alignment was 1988 after all sequences were aligned. The aligned sequences with positions, frequencies of most common aminoacids relevant to the alignment positions were also calculated and presented in Table C.3 in Appendix section.

The selected sequences and the tree topology were given as the input to the algorithm. Results of 100 evaluations were collected, and the frequencies of the residues within the solution subsets were calculated as shown in Table 5.4. Complete results were tabulated in Table C.3 in the Appendix section.

All residues listed in Table 5.4 had shown frequently within the solution subsets. In Table 5.4, frequencies, with respect to alignment positions, were listed with common aminoacids for each group. Mentioned residues were also listed through the end of the Table 5.4.

Table 5.4 Important results of the evaluation of HCO dataset.

Alignment	A1	A2	B	C	NOR	Frequency	Residue
518	H	H	H	H	H	0.99	
1003	H	H	H	H	H	0.99	
1005	H	H	H	H	H	0.99	
906	P	P	P	P	P	0.94	
628	Y	Y	Y	Y	Y	0.93	
787	W	W	W	W	W	0.92	
1117	G	G	G	G	A	0.91	
1105	G	G	G	G	G	0.90	
1163	G	G	G	G	G	0.88	
630	P	P	P	E	E	0.87	
1043	P	P	P	P	R	0.87	
1124	R	R	R	W	A	0.86	
404	R	R	R	A	A	0.84	
896	F	F	F	G	G	0.84	
791	H	H	H	N	L	0.83	
800	P	P	P	G	F	0.83	
396	G	G	G	G	Q	0.82	
792	P	P	P	A	W	0.82	
519	G	G	G	T	T	0.81	
887	G	G	G	A	G	0.81	
913	F	F	F	N	E	0.81	
790	G	S	H	H	H	0.80	
982	G	G	G	G	-	0.80	
991	D	D	D	N	N	0.80	
1004	F	F	F	V	G	0.80	
793	E	A	V	V	V	0.79	278
385	Y	Y	Y	F	Y	0.76	35
548	D	D	D	F	K	0.76	124
670	N	N	N	V	R	0.76	199
806	S	S	Y	Y	A	0.76	291
795	Y	Y	-	F	G	0.73	280
537	N	N	A	Y	L	0.70	113
911	K	K	T	M	G	0.70	354
555	N	N	W	K	L	0.68	131
907	T	T	S	S	L	0.61	351
664	T	S	I	V	W	0.50	193
558	S	S	-	A	D	0.46	134

5.3.3. MMG DATA

Tree topology (Figure 5.3) for the analysis were created by using the phylogeny in (Meredith et al. 2011) (Figure 5.2) as reference since the phylogeny revealed strong support for most nodes and it was highly congruent with more recent phylogenies. All the nodes in this phylogeny are strongly supported except for the ones denoted by solid blue circles.

In the alignment of the genomic sequence blocks data (Figure 5.7), 13 proteins and 24 RNAs, some block data found misplaced. Column 9 of *Distoechurus pennatus* should be located in column 7, and similarly column 16 and 17 of *Caenolestes fuliginosus*, *Didelphis virginiana*, *Dromiciops gliroides*, *Notoryctes typhlops*, *Dasyurus hallucatus*, *Perameles gunnii* and *Distoechurus pennatus* need to be located in column 13 and 14 respectively. This would result significant problems in multiple alignment procedure.

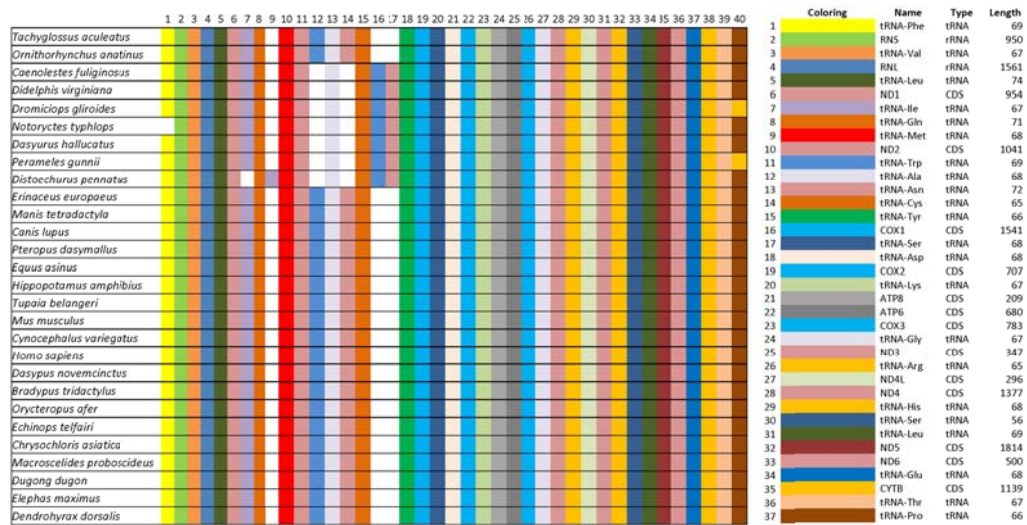


Figure 5.7 Alignment of genomic sequences blocks.

By using the Longest Common Substring technique, 69 segments were identified (Table 5.5). All genomic sequences, except tRNA-Gln, tRNA-Glu, tRNA-Leu, tRNA-Ala and tRNA-Pro were represented by 1 to 6 segments with average size of 33.97. Optimization operation were performed on these segments rather than on the characters or the residues.

The closest tree topology was obtained by the combination of the segments numbered 11, 31,

44, 50, 64 and 67 with a TD of 14 (Figure 5.8). According to the results of 100 evaluations, the frequencies (f) of the segments within the solutions were tabulated in Table 5.5.

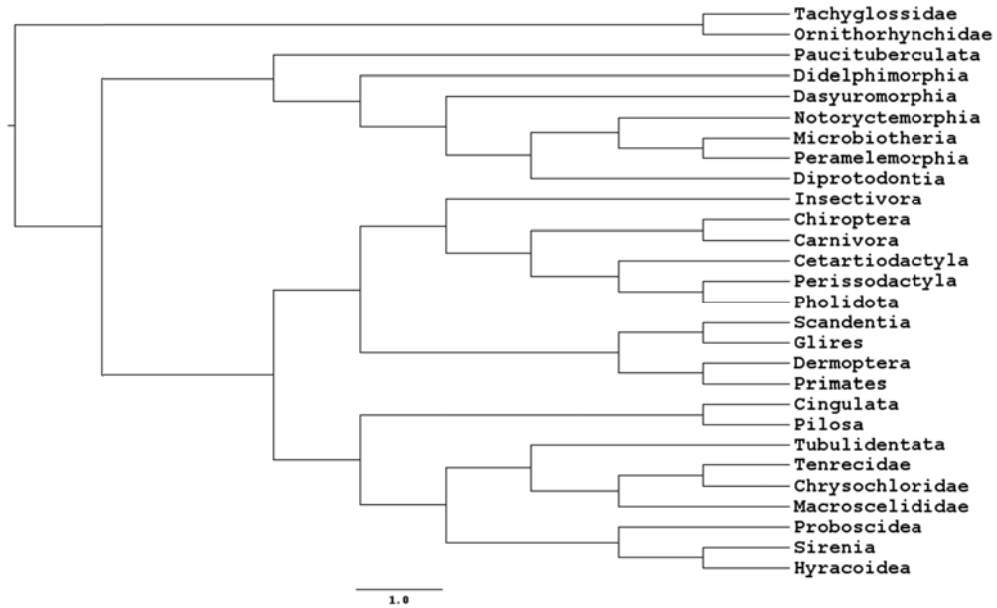


Figure 5.8 The tree topology of the best solution produced by the optimization algorithm with TD=14.

Table 5.5 Results of MMG dataset segments with their size and frequencies.

sequence	segment	Size (avg)	f
tRNA-Phe	1	4	3
RNS	2	11	0
	3	14	4
	4	47	4
	5	54	10
tRNA-Val	6	6	26
RNL	7	36	1
	8	29	3
	9	23	4
	10	96	5
	11	76	28
	12	97	10
ND1	13	31	2
	14	93	14
	15	86	3
	16	28	3
	17	24	37
tRNA-Ile	18	2	32
tRNA-Met	19	8	23
ND2	20	69	12
	21	16	0
	22	61	14
	23	14	41
	24	10	8

tRNA-Trp	25	4	53
tRNA-Trp	26	3	49
tRNA-Cys	27	9	14
tRNA-Tyr	28	8	15
COX1	29	62	9
	30	17	3
	31	61	2
	32	25	29
tRNA-Ser	33	4	17
tRNA-Asp	34	9	24
COX2	35	60	7
	36	20	2
	37	87	8
	38	24	3
tRNA-Lys	39	7	27
ATP8	40	8	17
ATP6	41	39	5
	42	89	19
	43	45	0
COX3	44	71	2
	45	88	21
tRNA-Gly	46	9	32
ND3	47	48	15
tRNA-Arg	48	4	23
ND4L	49	26	27
ND4	50	2	0
	51	19	1
	52	25	4
	53	14	1
tRNA-His	54	9	35
tRNA-Ser	55	3	13
tRNA-Leu	56	6	24
ND5	57	17	11
	58	10	0
	59	21	3
ND6	60	33	18
	61	44	0
	62	55	13
	63	45	17
CYTB	64	99	6
	65	15	14
	66	47	0
	67	22	5
	68	90	9
tRNA-Thr	69	6	2

CHAPTER 6. DISCUSSION & CONCLUSIONS

6.1. EVALUATION OF MORPHOLOGICAL DATA

6.1.1. HOPLITIS DATASET

The fitness function of the optimization algorithm has been designed in order to find, first, the exact tree topology and then decrease the number of characters in the dataset. To do that, second power of topology distance has been formulated within the fitness score function while the power of the number of characters was only one. In Figure 3.3, it can be seen that the algorithm chooses the solution subsets with higher scores which are most probably the ones with smaller topology distances created. Here, the smaller topology distances are thought to be created by larger datasets with larger number of characters since it is most probable to construct the same topology with larger number of characters. The increase in the number of characters was not highly considered in this case due to the second power of the topology distance. This condition continues and fitness score increases until the topology distance becomes 0. The growth of the fitness score is exponential due to the second power of topology distance. After this point, since the topology distance was set to zero, the algorithm starts choosing the solutions with smaller number of characters, therefore the fitness score decreases as the number of characters decreases. The algorithm converges to a solution/solutions with 10 characters at iteration 245 and exits after 100 iterations with no change in the average fitness score.

The complete dataset was tested with the proposed feature selection method and a subset with as few as 67 characters was obtained, from which one can obtain exactly the same phylogeny. In their study, Sokal and Michener (1958b) employed 122 characters per species. However, they were assuming significant results might be obtained from as few as 60 characters. Results were fairly close to the expectations of Sokal and Michener.

In the results obtained from the complete and the partial datasets, it was found that the same topology can be obtained with considerably smaller number of characters. With the invention of those characters seems to have, some characters seems to have crucial importance in the construction of the phylogeny of the related group since they have been occurred within the solution subsets with high frequency. This implies that, those characters may have important role in the phylogeny construction where they cannot be replaced by any other character. On the other hand, some other characters have revealed strong positive or strong negative associations based on their occurrences within the same solution subsets. In addition, some characters show strong positive association among each other. Those characters are complementing each other by forming a group carrying a certain information. There might be another group of characters which carry the same information. In this case, those groups are in strong negative association. In this way, it is possible to extract type of information carried by the characters or the groups of characters by using the proposed method.

6.1.2. *ACORN DATASET*

In this part, phylogeny of 14 *Quercus* taxa has been reconstructed with 3 different sets of characters in size of 6 out of 33. Some characters occurred in the sets more frequently than the others implying that these sets may have a critical value for the

data. Thus, algorithm cannot exclude them from the subset since exclusion would lead to a topology distant from the origin. The optimization results showed that, 2nd and 32nd characters are the most occurred characters and they are found in 17 sets out of 25 sets.

When the distribution of these characters' were examined within the sets, they have involved in both strong positive and strong negative associations as well as weak positive and weak negative associations. Two characters are positively associated when they are both found or both not found in the same sets implying that they have congruence. The information they overlaps with each other, thus, they represent much more portion of the data than they separately do. 5-31 and 2-32 are good examples of that. However, some of the couples are in incongruence. They appear within the sets rotatory. These characters include similar information, and when they are involved within the same set there will not be any disability to eliminate one. The 2-31 and the 5-32 sets are examples of incongruence. The 5-31 and the 2-32 sets are character couples which are carrying similar information and alternatives of each other. The 1st character can also be considered in congruence with 5 and 31 and incongruence with 2 and 32. Another most frequently observed character is 16. However, this character shows neither congruence nor incongruence with any of the characters except some weak negative associations and a weak positive association. A possible explanation for this would be the diverse information content of the character. Character 16 should be carrying a kind of information that can match with many other characters in some part and also unique in other parts. According to the Phi coefficients of 16th character, 6, 20, 27 and 28 have weak negative association, but cannot be considered as incongruence. Some more character sets are required to

analyze the situation of 16. However, set of 2-16-30-32 is most frequent data portion while 1-5-30 is alternative for 2-32 and 26 for 30 as observed.

For each solution in the population, algorithm reconstructs a phylogeny for the 14 taxa. A sample phylogenetic tree based on optimization results has been represented in Figure 4.2. The trees in Figure 4.2 have the same structure topologically.

However, branch lengths may differ for each tree. By considering the explanation at PHYLIP home page, branch length distances were not used and the trees were only compared by using the topology distances. Some of the trees have branch lengths quite different from the original tree. This is because a small set of characters may not be carrying each detail among taxa. Thus, the differences between the branch lengths were ignored except they change the topology.

6.2. EVALUATION OF MOLECULAR DATA

6.2.1. *HCO DATASET*

The HCO data were composed by selecting 10 sequences from each group where each sequence representing a taxonomic group in phylogeny in Figure 5.1-A. Total of 50 sequences were collected by this way. Selected sequences and the tree topology that were generated from phylogeny in Figure 5.1-A were given as the input to the algorithm. Frequencies of the residues within solution subsets were calculated based on the results of 100 evaluations (Table 5.4). The aligned sequences with positions, frequencies of most common aminoacids relevant to the alignment positions were also calculated and given in Table C.3 in the Appendix section.

The Table 5.4 shows that the most frequently chosen residues are the ones which have the same aminoacid in every group. However, within group frequencies of these AAs were higher at A1 and A2 groups and lower at NOR and C groups. Therefore,

these residues are chosen due to their grouping property for A1 and A2 and discriminant value from the other groups. Some other residues, like 991, 1004, 793 etc., were chosen due to their grouping value for either A1-A2 couple or for NOR-C separation from the rest.

Some residues – that were previously described (Sousa et al. 2011) for playing an important role for the function of the protein – were included in the Table 5.4.

According to *Paracoccus denitrificans* numbering in (Pereira et al. 2001), the amino acid residues that compose the D-channel were Asp124, Asn199, Asn113, Asn131, Tyr35, Ser134, Ser193 and Glu278 for A1 type enzymes. The corresponding frequencies for these residues were 0.76, 0.76, 0.70, 0.68, 0.76, 0.46, 0.50 and 0.79 respectively. Residues were occurred within the solution subsets from half to 4 out of 5, which were considerably high values. On the other hand, in the D-channel of A2 Type enzymes a tyrosine replaces the glutamate residue. K-channel of A Type enzymes was constituted by Lys354, Thr351, Ser291 and Tyr280 (Tyr-I) residues. In the alternative K-channel of B Type enzymes besides Tyr-I, those residues were replaced by a Thr, a Ser and a Tyr residues. The alternative K-channel from C Type enzymes was constituted by Tyr-II and a Ser and a Tyr residue in the same sequence position of the Thr-351 and SerI-291 residues. The related values for residues numbered 280, 291, 351 and 354 range from 0.61 to 0.76, which is again as high as D-channel.

All residues mentioned above have been found frequently within the solution subsets. In Table 5.4, frequencies that corresponds to the alignment positions were listed with common aminoacids for each group. These residues were also listed at the end of the Table 5.4. frequencies of the residues are as high as 70s except 351, 193 and 134 which are still acceptable. On the other hand, the alignment positions

with highest frequencies, His residues, are important in the C family, specifically, due to their roles in high/low-spin heme ligands, the cation binding site and the Cu_B ligands (Hemp et al. 2007). Gly residues are also important in the structure of the protein complex.

6.2.2. *MMG DATASET*

Multiple sequence alignment of the full genome of the Mammalian Mitochondrial Genomic sequences would be an enormous and erroneous job to be done. In Figure 5.7, it can be seen that the sequence blocks as protein or RNA, are translocated. Several columns belong to some OTUs were located at a different place from its origin. Therefore, a MSA procedure would have failed in these region hence resulting in an inaccurate alignment. Therefore, sequences were splitted into many distinct subsequences by using Longest Common Substring technique (as described in Chapter 3), where they can be compared without a preliminary alignment procedure. By this way 69 segments (subsequences) were generated with average size of 33.97. RCM has been used in order to measure the distances among those subsequences. Closest tree topology was obtained by a TD of 14 as in Figure 5.8 which is still highly accurate. Table 5.5 demonstrates the results of the evaluation based on the MMG data. Here, the segments are found mostly in protein sequences and not in RNA sequences. This may be due to the higher lengths of the protein sequences. On the other hand, an important observation is that the segments in RNA sequences have higher frequencies than the protein segments. This may be due to the highly conserved structure of the protein (Ahituv et al. 2007) allows small sized segments. Therefore, not many information can be gained by these segments. On the other hand, RNA based segments show more variations and sizes of the sequences are larger.

6.3. CONCLUSIONS

In this thesis study, a unique feature selection method have been developed in order to determine the minimum set of characters that can be used to reconstruct the same tree topology as the original tree. The selection of the subset of features is a combinatorial optimization problem. Therefore, it can not be solved in polynomial time. The complexity of the problem increases exponentially with the total number of features.

A GA-FS based optimization algorithm have been used on biological datasets that are based on morphology or molecular sequence. In evaluation of the morphological data using *Hoplitis* data, the algorithm finds the desired tree topology almost half of the characters. The number was very close to the expectations of the experts. On the other hand, number of chosen characters decreases by the decrease in the complexity of the taxonomic hierarchy.

In molecular sequences, there are two cases; aligned sequences and very long sequences which can not be aligned. Analysing the aligned sequences with shorter sizes are similar to the morphological data. All residues are considered as characters. Therefore, optimization is performed on the residues. The results indicated several functionally important residues accurately in accordance with the literature.

However, there are also other residues waiting for the experts' consideration. On the other hand, genomic data with larger size, was analyzed by splitting the data into several subsequences. Optimization process were performed on these subsequences. In this case, to find a topology very close to the desired one was not easy, since the original data cannot give a resolution to the problem in traditional way most probably due to inaccurate MSA in the presence of translocations.

6.4. FUTURE WORK

The proposed algorithm performs biological datasets such as morphological characters and molecular sequences data and exhibits important features and also in/congruencies among the features for a given classification by phylogenetic analysis of the data. As a future work, the proposed optimization algorithm has planned to be capable of performing analysis on several other types of data such as micro array data, complete genome data with very large sizes, landmark and shape based data, and etc. On the other hand, support for different analyzing methods that have been assessed to analyze biological datasets has also been planned as a future work.

REFERENCES

- Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM 2007.
Deletion of ultraconserved elements yields viable mice. *PLoS Biology* 5(9):
e234
- Archie JW 1989. A randomization test for phylogenetic information in systematic
data. *Systematic Zoology* 38: 219-252
- Bakış Y 2005. Morphometric analysis of oak (*Quercus* L.) acorns in Turkey.
Biology. Abant İzzet Baysal University, Bolu, p 72
- Bakış Y, Sezerman OU, Babaç MT, Meydan C 2009. Optimization of morphological
data in numerical taxonomy analysis using genetic algorithms feature selection
method. 11th Genetic and Evolutionary Computation Conference, GECCO
2009, Montreal, Québec, Canada
- Bastola DR, Otu HH, Doukas SE, Sayood K, Hinrichs SH, Iwen PC 2004. Utilization
of the relative complexity measure to construct a phylogenetic tree for fungi.
Mycological Research 108(2): 117-125
- Bateman R, Michele WAd, Willard D 1992. Experimental cladistic analysis of
anatomically preserved arborescent lycopsids from the Carboniferous of
Euramerica: An essay on paleobotanical phylogenetics. *Annals of the Missouri
Botanical Gardens* 79: 500-559

- Brooks DR, Erdem E, Erdoğan ST, Minett JW, Ringe D 2007. Inferring Phylogenetic Trees Using Answer Set Programming. *Journal of Automated Reasoning* 39(4): 471-511
- Buettner-Janusch JOHN 1966. A problem in evolutionary systematics: Nomenclature and classification of Baboons, genus *Papi O*. *Folia Primatologica* 4(4): 288-308
- Cain AJ, Harrison GA 1958. An analysis of the taxonomist's judgment of affinity. *Proc. Zool. Soc. Lond.* 131: 85-98
- Campbell V, Lapointe F-J 2011. Retrieving a mitogenomic mammal tree using composite taxa. *Molecular Phylogenetics and Evolution* 58: 149-156
- Carpenter JM 2005. Random Cladistics. *Cladistics* 8(2): 147-153
- Cavalli-Sforza LL, Edwards AW 1967. Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 19(3 Pt 1): 233-257
- Chen X, Tompa M 2010. Comparative assessment of methods for aligning multiple genome sequences. *Nature Biotechnology* 28(6): 567-572
- Cox PB, Urbatsch L 1990. A phylogenetic analysis of the coneflower genera (Asteraceae: Heliantheae). *Systematic Botany* 15: 394-402
- Çakmak D 2010. *Weighted Phylogenetic Trees and Networks*. Computer Science and Engineering. Sabanci University, İstanbul
- Çakmak D, Erdem E, Erdoğan H 2011. Computing weighted solutions in ASP: representation-based method vs. search-based method. *Annals of Mathematics and Artificial Intelligence* 62(3-4): 219-258
- Darwin C 1837-1838. Notebook B: Transmutation of species.

- Darwin C 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London
- Desai B, Andhale P, Rege M, Yu Q 2010. Biclustering and feature selection techniques in bioinformatics. In: Andres, R.K.a.F. (ed) Second international conference on Data Engineering and Management (ICDEM'10). Springer-Verlag, Berlin, Heidelberg, pp 280-287
- Dytham C 2003. Choosing and Using Statistics : A Biologist's Guide. Blackwell Pub., Malden, MA
- Efron B, Tibshirani R 1993. An introduction to the bootstrap. Chapman & Hall, New York
- Eldredge N, Gould SJ 1972. Punctuated equilibria: An alternative to phyletic gradualism. In: Schopf, T.J.M. (ed) Models in Paleobiology. Freeman, San Francisco, pp 82-115
- Erdem E 2011. Applications of Answer Set Programming in Phylogenetic Systematics. Lecture Notes in Computer Science 6565: 415-431
- Farris J, Albert VA, Källersjö M, Lipscomb D, Kluge AG 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12: 99-124
- Felsenstein J 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet 25(5): 471-492
- Felsenstein J 1985. Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. Evolution 39(4): 783-791
- Felsenstein J 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521-565

- Felsenstein J 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166
- Felsenstein J 1998. PHYLIP Home Page [online]. Available from <http://evolution.genetics.washington.edu/phylip.html> [accessed 21.05.2012]
- Gandhi S, Khan D, Solanki VS 2012. A Comparative Analysis of Selection Scheme. *International Journal of Soft Computing and Engineering* 2(4): 131-134
- Gensel P 1992. Phylogenetic relationships of the zosterophylls and lycopsids: Evidence from morphology, paleoecology, and cladistic methods of inference. *Annals of the Missouri Botanical Gardens* 79: 450-473
- George GVS, Raj VC 2011. Review on Feature Selection Techniques and the Impact of SVM for Cancer Classification using Gene Expression Profile. *International Journal of Computer Science & Engineering Survey (IJCSES)* 2(3):
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59(3): 307-321
- Gusfield D 1999. *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge University Press, USA
- Guyon I, Elisseeff A 2003. An introduction to variable and feature selection. *J. Mach Learn Res.* 3: 1157-1182
- Hall BK 1994. *Homology: The hierarchical basis of comparative biology*. Academic Press, Roma
- Hemp J, Han H, Roh JH, Kaplan S, Martinez TJ, Gennis RB 2007. Comparative genomics and site-directed mutagenesis support the existence of only one input

channel for protons in the C-family (cbb3 oxidase) of heme-copper oxygen reductases. *Biochemistry* 46(35): 9963-9972

Hennig W 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana

Hillis DM, Bull JJ 1993. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology* 42(2): 182-192

Holland JH 1975. *Adaption in Natural and Artificial Systems*. University of Michigan Press, , Ann Arbor, MI

Hood SB, Reznick A, Palmer JD 1994. Phylogenetic relationships in *Anemone* (Ranunculaceae) based on morphology and chloroplast DNA. *Systematic Botany* 19: 169-200

Huelsenbeck JP, Ronquist F 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8): 754-755

Hufford L, Dickson W 1992. A phylogenetic analysis of Cunoniaceae. *Systematic Botany* 17: 181-200

Katoh K, Kuma K-I, Toh H, Miyata T 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33(2): 511-518

Keleş S, van der Laan M, Eisen MB 2002. Identification of regulatory elements using a feature selection method. *Bioinformatics* 18(9): 1167-1175

Kesner MH 1994. The impact of morphological variants on a cladistic hypothesis with an example from a myological data set. *Systematic Biology* 43: 41-57

- Khouja M, Michalewicz Z, Wilmot M 1998. The use of genetic algorithms to solve the economic lot size scheduling problem. *European Journal of Operational Research* 3(1): 509-524
- Kim E, Chufán EE, Kamaraj K, Karlin KD 2004. Synthetic models for heme-copper oxidases. *Chem Rev* 104(2): 1077-1133
- Kjer KM, Honeycutt RL 2007. Site specific rates of mitochondrial genomes and the phylogeny of eutheria. *BMC Evolutionary Biology* 7(8):
- Koza JR 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press, Cambridge, Massachusetts, USA
- Kraus F 1988. An empirical evaluation of the use of the ontogeny polarization criterion in phylogenetic inference. *Systematic Zoology* 37: 106-141
- Kunsch HR 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17: 1217-1241
- Le Quesne WJ 1969. A Method of Selection of Characters in Numerical Taxonomy. *Systematic Zoology* 18(2): 201-205
- Lempel A, Ziv J 1976. On the complexity of finite sequences. *IEEE Transactions on Information Theory* 22: 75-81
- Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T 2009. Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. *Science* 324: 1561-1564
- Livezey BC 1986. Phylogeny and historical biogeography of steamer ducks (Anatidae: Tachyeres). *Systematic Zoology* 35: 458-469

- Löytynoja A, Goldman N 2008. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science* 320: 1632-1635
- Lu H, Pi E, Peng Q, Wang L, Zhang C 2009. A particle swarm optimization-aided fuzzy cloud classifier applied for plant numerical taxonomy based on attribute similarity. *Expert Systems with Applications: An International Journal* 36(5): 9388-9397
- Ma S, Huang J 2008. Penalized feature selection and classification in bioinformatics. *Briefings in Bioinformatics* 9(5): 392-403
- Mayr E 1965. Numerical Phenetics and Taxonomic Theory. *Systematic Zoology* 14(2): 73-97
- Mayr E 1969. Discussion: Footnotes on the philosophy of biology. *Philosophy of Science* 36: 197-202
- McNeill J 1982. Phylogenetic reconstruction and phenetic taxonomy. *Zoological Journal of the Linnean Society* 74(3): 337-344
- Meredith RW, Janec̣ka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science* 334: 521-524
- Michalewicz Z 1996. Genetic Algorithms + Data Structures = Evolution. Springer-Verlag, Berlin
- Michener CD, Sokal RR 1957. A Quantitative Approach to a Problem in Classification. *Evolution* 11(2): 130-162

- Mitchell M 1996. An introduction to genetic algorithms. MIT Press, Cambridge, Mass.
- Mitchell M 1997. An Introduction to Genetic Algorithms. MIT Press, Cambridge
- Mount DW 2004. Using Progressive Methods for Global Multiple Sequence Alignment. In: Mount, D.W. (ed) Bioinformatics: Sequence and Genome Analysis. CSHL Press, New York
- Murzin AG, Brenner SE, Hubbard T, Chothia C 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247(4): 536-540
- NCBI 1988. The National Center for Biotechnology Information [online]. Available from <http://www.ncbi.nlm.nih.gov/> [accessed 2013]
- Ogden TH, Rosenberg MS 2006. Multiple sequence alignment accuracy and phylogenetic inference. Systematic Biology 55(2): 314-328
- Otu HH, Sayood K 2003. A new sequence distance measure for phylogenetic tree construction. Bioinformatics 19: 2122-2130
- Penny D, Hendy MD 1985. Testing methods of evolutionary tree construction. Cladistics 1: 266-278
- Pereira MM, Pereira-Leal JB 2011. HCO: Heme Copper Oxygen reductase classification tool [online]. Available from <http://www.evocell.org/hco/> [accessed 2013]
- Pereira MM, Santana M, Teixeira M 2001. A novel scenario for the evolution of haem-copper oxygen reductases. Biochim Biophys Acta 1505(2-3): 185-208
- Pimentel R, Riggins R 1987. The nature of cladistic data. Cladistics 3: 201-209

- Rechenberg I 1973. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Frommann-Holzboog
- Robinson DF, Foulds LR 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53(1-2): 131-147
- Sackin MJ 1972. "Good" and "Bad" Phenograms. *Systematic Biology* 21(2): 225-226
- Saeys Y, Inza I, Larranaga P 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* 23(19): 2507-2517
- Saitou N, Nei M 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425
- Schloss PD, Handelsman J 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and environmental microbiology* 71(3): 1501-1506
- Shimodaira H 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51(3): 492-508
- Shyu C, Sheneman L, Foster JA 2003. Multiple Sequence Alignment with Evolutionary Computation. *Genetic Programming and Evolvable Machines* 5(2): 121-144
- Sivanandam SN, Deepa SN 2007. Introduction to genetic algorithms. Springer, Berlin ; New York
- Sneath PHA, Sokal RR 1973. Numerical Taxonomy; The principles and practice of numerical classification. W. H. Freeman, San Francisco,

- Sokal R, Michener C 1958a. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: 1409-1438
- Sokal RR 1963. The Principles and Practice of Numerical Taxonomy Taxon 12(5): 190-199
- Sokal RR, Michener CD 1958b. A Statistical Method for Evaluating Systematic Relationships. The University of Kansas Science Bulletin 38(22): 1409-1438
- Sousa FL, Alves RJ, Pereira-Leal JB, Teixeira M, Pereira MM 2011. A Bioinformatics Classifier and Database for Heme-Copper Oxygen Reductases. PLoS ONE 6(4): e19117
- Stace CA 1989. Plant Taxonomy and Biosystematics. Chapman and Hall, London
- Stevens PF 1991. Character states, morphological variation, and phylogenetic analysis: A review. Systematic Botany 16: 553-583
- Stoye J, Evers D, Meyer F 1996. Bielefeld Bioinformatics Service [online]. Available from <http://bibiserv.techfak.uni-bielefeld.de/> [accessed 2012]
- Stoye J, Evers D, Meyer F 1998. Rose: generating sequence families. Bioinformatics 14(2): 157-163
- Stropea PK, Moriyamab EN 2007. Simple alignment-free methods for protein classification: A case study from G-protein-coupled receptors. Genomics 89(5): 602-612
- Stuessy TF 2009. Plant Plant Taxonomy: The Systematic Evaluation of Comparative Data. Columbia University Press, New York
- Swofford DL 1996. Paup 4.0: Phylogenetic Analysis Using Parsimony (and Other Methods). Sinauer Associates Incorporated, Sunderland MA

- Swofford DL, Olsen GJ 1990. Phylogeny reconstruction. In: Moritz, D.M.H.a.C. (ed) Molecular Systematics. Sinauer Associates, Sunderland, pp 411-501
- Tamer A, Badr A 2012. A Comparative Study on Bioinformatics Feature Selection and Classification. International Journal of Computer Applications 43(3): 5-8
- Thompson JD, Higgins DG, Gibson TJ 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research 22(22): 4673-4680
- Tommaso PD, Moretti S, Xenarios I, Orobittg M, Montanyola A, Chang J-M, Taly J-F, Notredame C 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Research 39(Web Server issue): W13-W17
- Tucker A, Duplisea D 2012. Bioinformatics tools in predictive ecology: applications to fisheries. Phil. Trans. R. Soc. B 367(1586): 279-290
- Varón A, Vinh LS, Wheeler WC 2009. POY version 4: phylogenetic analysis using dynamic homologies. Cladistics 26(1): 72 - 85
- Wheeler WC, Aagesen L, Arango CP, Faivovich J, Grant T, D'Haese C, Janies D, Smith WL, Varón A, Giribet G 2006. Dynamic Homology and Phylogenetic Systematics: A Unified Approach Using POY. American Museum of Natural History, New York
- Wiens JJ 1995. Polymorphic characters in phylogenetic systematics. Systematic Biology 44: 482-500

- Wiens JJ 2004. The Role of Morphological Data in Phylogeny Reconstruction. *Systematic Biology* 53(4): 653-661
- Wu CFJ 1986. Jackknife, bootstrap and other resampling plans in regression analysis. *Annals of Statistics* 14: 1261-1295
- Yue F, Shi J, Tang J 2009. Simultaneous phylogeny reconstruction and multiple sequence alignment. *BMC Bioinformatics* 10(1): S11
- Zharkikh A, Li. W-H 1995. Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Molecular Biology and Evolution* 4: 44-63
- Zumft WG 2005. Nitric oxide reductases of prokaryotes with emphasis on the respiratory, heme-copper oxidase type. *J Inorg Biochem* 99(1): 194-215

APPENDICES

A - DATA RELATIVE TO CHAPTER 3

Table 0.1 Sample optimization process.

	Fitness Score			Number of Characters			Tree Distance			Population diversity
	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	
initStep 1	0.0002463	0.0004027	0.0010204	40	40.00	40	6	18.28	28	
initStep 2	0.0003759	0.0004874	0.0010204	40	40.00	40	6	14.44	18	
initStep 3	0.0004202	0.0005517	0.0010204	40	40.00	40	6	12.48	16	
initStep 4	0.0004762	0.0005947	0.0010204	40	40.00	40	6	11.32	14	
initStep 5	0.0004762	0.0006094	0.0010204	40	40.00	40	6	10.96	14	
initStep 6	0.0005495	0.0006434	0.0014286	40	40.00	40	4	10.52	12	
initStep 7	0.0005495	0.0006804	0.0014286	40	40.00	40	4	10.04	12	
initStep 8	0.0005495	0.0006912	0.0014286	40	40.00	40	4	9.84	12	
initStep 9	0.0005495	0.0007150	0.0014286	40	40.00	40	4	9.40	12	
genStep 0	0.0006494	0.0007510	0.0014286	40	40.00	40	4	8.92	10	
genStep 1	0.0006494	0.0007931	0.0014286	38	40.04	43	4	8.40	10	2.0400
genStep 2	0.0006494	0.0008772	0.0023148	37	40.20	44	2	7.64	10	3.2000
genStep 3	0.0007880	0.0010230	0.0024155	36	40.64	46	2	6.44	8	4.6400
genStep 4	0.0007994	0.0011395	0.0024155	35	41.12	47	2	5.68	8	6.1200
genStep 5	0.0009990	0.0013814	0.0071429	37	41.66	54	0	4.84	6	4.6600
genStep 6	0.0010204	0.0015580	0.0071429	36	42.00	54	0	4.16	6	6.0000
genStep 7	0.0010428	0.0017542	0.0071429	36	43.30	55	0	3.44	6	7.3000
genStep 8	0.0013793	0.0020485	0.0071429	36	42.64	52	0	3.08	4	6.6400
genStep 9	0.0014085	0.0024829	0.0071429	36	42.90	52	0	2.60	4	6.9000
genStep 10	0.0014286	0.0030399	0.0071429	35	43.84	52	0	1.84	4	8.8400
genStep 11	0.0022676	0.0033193	0.0071429	35	43.68	52	0	1.56	2	8.6800
genStep 12	0.0022989	0.0037860	0.0072993	35	43.12	52	0	1.36	2	8.1200
genStep 13	0.0023310	0.0045399	0.0073529	35	42.66	52	0	1.04	2	7.6600
genStep 14	0.0023810	0.0056397	0.0073529	35	43.24	52	0	0.56	2	8.2400
genStep 15	0.0065789	0.0069052	0.0073529	36	44.96	52	0	0.00	0	8.9600
genStep 16	0.0067114	0.0069668	0.0073529	36	43.62	49	0	0.00	0	7.6200
genStep 17	0.0068027	0.0070136	0.0073529	36	42.64	47	0	0.00	0	6.6400
genStep 18	0.0068966	0.0070644	0.0073529	36	41.60	45	0	0.00	0	5.6000
genStep 19	0.0069444	0.0071069	0.0073529	36	40.74	44	0	0.00	0	4.7400
genStep 20	0.0069444	0.0071178	0.0073529	36	40.52	44	0	0.00	0	4.5200
genStep 21	0.0070423	0.0071559	0.0073529	36	39.76	42	0	0.00	0	3.7600

genStep 22	0.0070423	0.0071681	0.0073529	36	39.52	42	0	0.00	0	3.5200
genStep 23	0.0070922	0.0071784	0.0073529	36	39.32	41	0	0.00	0	3.3200
genStep 24	0.0070922	0.0072032	0.0073529	36	38.84	41	0	0.00	0	2.8400
genStep 25	0.0071429	0.0072166	0.0073529	36	38.58	40	0	0.00	0	2.5800
genStep 26	0.0071429	0.0072396	0.0074074	35	38.14	40	0	0.00	0	3.1400
genStep 27	0.0071942	0.0072715	0.0075758	32	37.54	39	0	0.00	0	5.5400
genStep 28	0.0071942	0.0072981	0.0075758	32	37.04	39	0	0.00	0	5.0400
genStep 29	0.0072464	0.0073176	0.0077519	29	36.68	38	0	0.00	0	7.6800
genStep 30	0.0072464	0.0073392	0.0077519	29	36.28	38	0	0.00	0	7.2800
genStep 31	0.0072464	0.0073607	0.0077519	29	35.88	38	0	0.00	0	6.8800
genStep 32	0.0072993	0.0074024	0.0077519	29	35.12	37	0	0.00	0	6.1200
genStep 33	0.0072993	0.0074447	0.0078740	27	34.36	37	0	0.00	0	7.3600
genStep 34	0.0073529	0.0075060	0.0078740	27	33.26	36	0	0.00	0	6.2600
genStep 35	0.0074074	0.0075488	0.0078740	27	32.50	35	0	0.00	0	5.5000
genStep 36	0.0074627	0.0076025	0.0078740	27	31.56	34	0	0.00	0	4.5600
genStep 37	0.0075188	0.0076290	0.0078740	27	31.10	33	0	0.00	0	4.1000
genStep 38	0.0075758	0.0076805	0.0080645	24	30.22	32	0	0.00	0	6.2200
genStep 39	0.0076336	0.0077217	0.0080645	24	29.52	31	0	0.00	0	5.5200
genStep 40	0.0076923	0.0077538	0.0080645	24	28.98	30	0	0.00	0	4.9800
genStep 41	0.0076923	0.0077743	0.0080645	24	28.64	30	0	0.00	0	4.6400
genStep 42	0.0076923	0.0077935	0.0080645	24	28.32	30	0	0.00	0	4.3200
genStep 43	0.0077519	0.0078191	0.0080645	24	27.90	29	0	0.00	0	3.9000
genStep 44	0.0077519	0.0078449	0.0080645	24	27.48	29	0	0.00	0	3.4800
genStep 45	0.0078125	0.0078596	0.0080645	24	27.24	28	0	0.00	0	3.2400
genStep 46	0.0078125	0.0078720	0.0080645	24	27.04	28	0	0.00	0	3.0400
genStep 47	0.0078125	0.0079019	0.0080645	24	26.56	28	0	0.00	0	2.5600
genStep 48	0.0078740	0.0079246	0.0081967	22	26.20	27	0	0.00	0	4.2000
genStep 49	0.0078740	0.0079436	0.0081967	22	25.90	27	0	0.00	0	3.9000
genStep 50	0.0078740	0.0079702	0.0081967	22	25.48	27	0	0.00	0	3.4800
genStep 51	0.0079365	0.0080007	0.0081967	22	25.00	26	0	0.00	0	3.0000
genStep 52	0.0079365	0.0080174	0.0081967	22	24.74	26	0	0.00	0	2.7400
genStep 53	0.0079365	0.0080496	0.0081967	22	24.24	26	0	0.00	0	2.2400
genStep 54	0.0080000	0.0080845	0.0081967	22	23.70	25	0	0.00	0	1.7000
genStep 55	0.0080645	0.0081120	0.0083333	20	23.28	24	0	0.00	0	3.2800
genStep 56	0.0080645	0.0081278	0.0083333	20	23.04	24	0	0.00	0	3.0400
genStep 57	0.0081301	0.0081622	0.0083333	20	22.52	23	0	0.00	0	2.5200
genStep 58	0.0081301	0.0081702	0.0083333	20	22.40	23	0	0.00	0	2.4000
genStep 59	0.0081301	0.0081809	0.0083333	20	22.24	23	0	0.00	0	2.2400
genStep 60	0.0081301	0.0081930	0.0083333	20	22.06	23	0	0.00	0	2.0600
genStep 61	0.0081301	0.0082104	0.0083333	20	21.80	23	0	0.00	0	1.8000
genStep 62	0.0081967	0.0082239	0.0083333	20	21.60	22	0	0.00	0	1.6000
genStep 63	0.0081967	0.0082375	0.0084034	19	21.40	22	0	0.00	0	2.4000
genStep 64	0.0081967	0.0082525	0.0084034	19	21.18	22	0	0.00	0	2.1800
genStep 65	0.0081967	0.0082634	0.0084034	19	21.02	22	0	0.00	0	2.0200

genStep 66	0.0081967	0.0082852	0.0084034	19	20.70	22	0	0.00	0	1.7000
genStep 67	0.0082645	0.0082879	0.0084034	19	20.66	21	0	0.00	0	1.6600
genStep 68	0.0082645	0.0082921	0.0084034	19	20.60	21	0	0.00	0	1.6000
genStep 69	0.0082645	0.0083003	0.0084034	19	20.48	21	0	0.00	0	1.4800
genStep 70	0.0082645	0.0083155	0.0084034	19	20.26	21	0	0.00	0	1.2600
genStep 71	0.0082645	0.0083293	0.0084034	19	20.06	21	0	0.00	0	1.0600
genStep 72	0.0083333	0.0083460	0.0084746	18	19.82	20	0	0.00	0	1.8200
genStep 73	0.0083333	0.0083488	0.0084746	18	19.78	20	0	0.00	0	1.7800
genStep 74	0.0083333	0.0083544	0.0084746	18	19.70	20	0	0.00	0	1.7000
genStep 75	0.0083333	0.0083600	0.0084746	18	19.62	20	0	0.00	0	1.6200
genStep 76	0.0083333	0.0083670	0.0084746	18	19.52	20	0	0.00	0	1.5200
genStep 77	0.0083333	0.0083698	0.0084746	18	19.48	20	0	0.00	0	1.4800
genStep 78	0.0083333	0.0083754	0.0084746	18	19.40	20	0	0.00	0	1.4000
genStep 79	0.0083333	0.0083880	0.0084746	18	19.22	20	0	0.00	0	1.2200
genStep 80	0.0083333	0.0083979	0.0084746	18	19.08	20	0	0.00	0	1.0800
genStep 81	0.0084034	0.0084105	0.0084746	18	18.90	19	0	0.00	0	0.9000
genStep 82	0.0084034	0.0084119	0.0084746	18	18.88	19	0	0.00	0	0.8800
genStep 83	0.0084034	0.0084190	0.0084746	18	18.78	19	0	0.00	0	0.7800
genStep 84	0.0084034	0.0084248	0.0085470	17	18.70	19	0	0.00	0	1.7000
genStep 85	0.0084034	0.0084262	0.0085470	17	18.68	19	0	0.00	0	1.6800
genStep 86	0.0084034	0.0084290	0.0085470	17	18.64	19	0	0.00	0	1.6400
genStep 87	0.0084034	0.0084419	0.0086207	16	18.46	19	0	0.00	0	2.4600
genStep 88	0.0084034	0.0084549	0.0086207	16	18.28	19	0	0.00	0	2.2800
genStep 89	0.0084034	0.0084649	0.0086207	16	18.14	19	0	0.00	0	2.1400
genStep 90	0.0084034	0.0084778	0.0086207	16	17.96	19	0	0.00	0	1.9600
genStep 91	0.0084034	0.0084893	0.0086207	16	17.80	19	0	0.00	0	1.8000
genStep 92	0.0084746	0.0085036	0.0086207	16	17.60	18	0	0.00	0	1.6000
genStep 93	0.0084746	0.0085153	0.0086207	16	17.44	18	0	0.00	0	1.4400
genStep 94	0.0084746	0.0085312	0.0086207	16	17.22	18	0	0.00	0	1.2200
genStep 95	0.0084746	0.0085487	0.0086207	16	16.98	18	0	0.00	0	0.9800
genStep 96	0.0085470	0.0085632	0.0086207	16	16.78	17	0	0.00	0	0.7800
genStep 97	0.0085470	0.0085706	0.0086207	16	16.68	17	0	0.00	0	0.6800
genStep 98	0.0085470	0.0085854	0.0086957	15	16.48	17	0	0.00	0	1.4800
genStep 99	0.0085470	0.0085927	0.0086957	15	16.38	17	0	0.00	0	1.3800
genStep 100	0.0085470	0.0086060	0.0086957	15	16.20	17	0	0.00	0	1.2000
genStep 101	0.0085470	0.0086193	0.0086957	15	16.02	17	0	0.00	0	1.0200
genStep 102	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 103	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 104	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 105	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 106	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 107	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 108	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 109	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400

genStep 110	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 111	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 112	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 113	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 114	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 115	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 116	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 117	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 118	0.0086207	0.0086252	0.0086957	15	15.94	16	0	0.00	0	0.9400
genStep 119	0.0086207	0.0086267	0.0086957	15	15.92	16	0	0.00	0	0.9200
genStep 120	0.0086207	0.0086267	0.0086957	15	15.92	16	0	0.00	0	0.9200
genStep 121	0.0086207	0.0086297	0.0086957	15	15.88	16	0	0.00	0	0.8800
genStep 122	0.0086207	0.0086297	0.0086957	15	15.88	16	0	0.00	0	0.8800
genStep 123	0.0086207	0.0086297	0.0086957	15	15.88	16	0	0.00	0	0.8800
genStep 124	0.0086207	0.0086297	0.0086957	15	15.88	16	0	0.00	0	0.8800
genStep 125	0.0086207	0.0086297	0.0086957	15	15.88	16	0	0.00	0	0.8800
genStep 126	0.0086207	0.0086297	0.0086957	15	15.88	16	0	0.00	0	0.8800
genStep 127	0.0086207	0.0086312	0.0086957	15	15.86	16	0	0.00	0	0.8600
genStep 128	0.0086207	0.0086312	0.0086957	15	15.86	16	0	0.00	0	0.8600
genStep 129	0.0086207	0.0086327	0.0086957	15	15.84	16	0	0.00	0	0.8400
genStep 130	0.0086207	0.0086327	0.0086957	15	15.84	16	0	0.00	0	0.8400
genStep 131	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 132	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 133	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 134	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 135	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 136	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 137	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 138	0.0086207	0.0086342	0.0086957	15	15.82	16	0	0.00	0	0.8200
genStep 139	0.0086207	0.0086357	0.0086957	15	15.80	16	0	0.00	0	0.8000
genStep 140	0.0086207	0.0086402	0.0086957	15	15.74	16	0	0.00	0	0.7400
genStep 141	0.0086207	0.0086402	0.0086957	15	15.74	16	0	0.00	0	0.7400
genStep 142	0.0086207	0.0086447	0.0086957	15	15.68	16	0	0.00	0	0.6800
genStep 143	0.0086207	0.0086462	0.0086957	15	15.66	16	0	0.00	0	0.6600
genStep 144	0.0086207	0.0086477	0.0086957	15	15.64	16	0	0.00	0	0.6400
genStep 145	0.0086207	0.0086477	0.0086957	15	15.64	16	0	0.00	0	0.6400
genStep 146	0.0086207	0.0086492	0.0086957	15	15.62	16	0	0.00	0	0.6200
genStep 147	0.0086207	0.0086507	0.0086957	15	15.60	16	0	0.00	0	0.6000
genStep 148	0.0086207	0.0086522	0.0086957	15	15.58	16	0	0.00	0	0.5800
genStep 149	0.0086207	0.0086612	0.0087719	14	15.46	16	0	0.00	0	1.4600
genStep 150	0.0086207	0.0086627	0.0087719	14	15.44	16	0	0.00	0	1.4400
genStep 151	0.0086207	0.0086672	0.0087719	14	15.38	16	0	0.00	0	1.3800
genStep 152	0.0086207	0.0086672	0.0087719	14	15.38	16	0	0.00	0	1.3800
genStep 153	0.0086207	0.0086702	0.0087719	14	15.34	16	0	0.00	0	1.3400

genStep 154	0.0086207	0.0086747	0.0087719	14	15.28	16	0	0.00	0	1.2800
genStep 155	0.0086207	0.0086822	0.0087719	14	15.18	16	0	0.00	0	1.1800
genStep 156	0.0086207	0.0086852	0.0087719	14	15.14	16	0	0.00	0	1.1400
genStep 157	0.0086207	0.0086927	0.0087719	14	15.04	16	0	0.00	0	1.0400
genStep 158	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 159	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 160	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 161	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 162	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 163	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 164	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 165	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 166	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 167	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 168	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 169	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 170	0.0086957	0.0086987	0.0087719	14	14.96	15	0	0.00	0	0.9600
genStep 171	0.0086957	0.0087002	0.0087719	14	14.94	15	0	0.00	0	0.9400
genStep 172	0.0086957	0.0087002	0.0087719	14	14.94	15	0	0.00	0	0.9400
genStep 173	0.0086957	0.0087002	0.0087719	14	14.94	15	0	0.00	0	0.9400
genStep 174	0.0086957	0.0087018	0.0087719	14	14.92	15	0	0.00	0	0.9200
genStep 175	0.0086957	0.0087018	0.0087719	14	14.92	15	0	0.00	0	0.9200
genStep 176	0.0086957	0.0087018	0.0087719	14	14.92	15	0	0.00	0	0.9200
genStep 177	0.0086957	0.0087018	0.0087719	14	14.92	15	0	0.00	0	0.9200
genStep 178	0.0086957	0.0087048	0.0087719	14	14.88	15	0	0.00	0	0.8800
genStep 179	0.0086957	0.0087048	0.0087719	14	14.88	15	0	0.00	0	0.8800
genStep 180	0.0086957	0.0087063	0.0087719	14	14.86	15	0	0.00	0	0.8600
genStep 181	0.0086957	0.0087063	0.0087719	14	14.86	15	0	0.00	0	0.8600
genStep 182	0.0086957	0.0087079	0.0087719	14	14.84	15	0	0.00	0	0.8400
genStep 183	0.0086957	0.0087079	0.0087719	14	14.84	15	0	0.00	0	0.8400
genStep 184	0.0086957	0.0087079	0.0087719	14	14.84	15	0	0.00	0	0.8400
genStep 185	0.0086957	0.0087094	0.0087719	14	14.82	15	0	0.00	0	0.8200
genStep 186	0.0086957	0.0087094	0.0087719	14	14.82	15	0	0.00	0	0.8200
genStep 187	0.0086957	0.0087109	0.0087719	14	14.80	15	0	0.00	0	0.8000
genStep 188	0.0086957	0.0087109	0.0087719	14	14.80	15	0	0.00	0	0.8000
genStep 189	0.0086957	0.0087109	0.0087719	14	14.80	15	0	0.00	0	0.8000
genStep 190	0.0086957	0.0087109	0.0087719	14	14.80	15	0	0.00	0	0.8000
genStep 191	0.0086957	0.0087124	0.0087719	14	14.78	15	0	0.00	0	0.7800
genStep 192	0.0086957	0.0087124	0.0087719	14	14.78	15	0	0.00	0	0.7800
genStep 193	0.0086957	0.0087140	0.0087719	14	14.76	15	0	0.00	0	0.7600
genStep 194	0.0086957	0.0087155	0.0087719	14	14.74	15	0	0.00	0	0.7400
genStep 195	0.0086957	0.0087216	0.0088496	13	14.66	15	0	0.00	0	1.6600
genStep 196	0.0086957	0.0087216	0.0088496	13	14.66	15	0	0.00	0	1.6600
genStep 197	0.0086957	0.0087216	0.0088496	13	14.66	15	0	0.00	0	1.6600

genStep 198	0.0086957	0.0087231	0.0088496	13	14.64	15	0	0.00	0	1.6400
genStep 199	0.0086957	0.0087231	0.0088496	13	14.64	15	0	0.00	0	1.6400
genStep 200	0.0086957	0.0087231	0.0088496	13	14.64	15	0	0.00	0	1.6400
genStep 201	0.0086957	0.0087262	0.0088496	13	14.60	15	0	0.00	0	1.6000
genStep 202	0.0086957	0.0087323	0.0088496	13	14.52	15	0	0.00	0	1.5200
genStep 203	0.0086957	0.0087338	0.0088496	13	14.50	15	0	0.00	0	1.5000
genStep 204	0.0086957	0.0087384	0.0088496	13	14.44	15	0	0.00	0	1.4400
genStep 205	0.0086957	0.0087461	0.0088496	13	14.34	15	0	0.00	0	1.3400
genStep 206	0.0086957	0.0087476	0.0088496	13	14.32	15	0	0.00	0	1.3200
genStep 207	0.0086957	0.0087522	0.0088496	13	14.26	15	0	0.00	0	1.2600
genStep 208	0.0086957	0.0087522	0.0088496	13	14.26	15	0	0.00	0	1.2600
genStep 209	0.0086957	0.0087568	0.0088496	13	14.20	15	0	0.00	0	1.2000
genStep 210	0.0086957	0.0087614	0.0088496	13	14.14	15	0	0.00	0	1.1400
genStep 211	0.0086957	0.0087644	0.0088496	13	14.10	15	0	0.00	0	1.1000
genStep 212	0.0086957	0.0087690	0.0088496	13	14.04	15	0	0.00	0	1.0400
genStep 213	0.0086957	0.0087813	0.0088496	13	13.88	15	0	0.00	0	0.8800
genStep 214	0.0087719	0.0087828	0.0088496	13	13.86	14	0	0.00	0	0.8600
genStep 215	0.0087719	0.0087844	0.0088496	13	13.84	14	0	0.00	0	0.8400
genStep 216	0.0087719	0.0087844	0.0088496	13	13.84	14	0	0.00	0	0.8400
genStep 217	0.0087719	0.0087859	0.0088496	13	13.82	14	0	0.00	0	0.8200
genStep 218	0.0087719	0.0087890	0.0088496	13	13.78	14	0	0.00	0	0.7800
genStep 219	0.0087719	0.0087906	0.0088496	13	13.76	14	0	0.00	0	0.7600
genStep 220	0.0087719	0.0087937	0.0088496	13	13.72	14	0	0.00	0	0.7200
genStep 221	0.0087719	0.0087983	0.0088496	13	13.66	14	0	0.00	0	0.6600
genStep 222	0.0087719	0.0087999	0.0088496	13	13.64	14	0	0.00	0	0.6400
genStep 223	0.0087719	0.0088061	0.0088496	13	13.56	14	0	0.00	0	0.5600
genStep 224	0.0087719	0.0088107	0.0088496	13	13.50	14	0	0.00	0	0.5000
genStep 225	0.0087719	0.0088154	0.0088496	13	13.44	14	0	0.00	0	0.4400
genStep 226	0.0087719	0.0088185	0.0088496	13	13.40	14	0	0.00	0	0.4000
genStep 227	0.0087719	0.0088201	0.0088496	13	13.38	14	0	0.00	0	0.3800
genStep 228	0.0087719	0.0088263	0.0088496	13	13.30	14	0	0.00	0	0.3000
genStep 229	0.0087719	0.0088340	0.0088496	13	13.20	14	0	0.00	0	0.2000
genStep 230	0.0087719	0.0088433	0.0088496	13	13.08	14	0	0.00	0	0.0800
genStep 231	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 232	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 233	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 234	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 235	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 236	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 237	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 238	0.0088496	0.0088496	0.0088496	13	13.00	13	0	0.00	0	0.0000
genStep 239	0.0088496	0.0088511	0.0089286	12	12.98	13	0	0.00	0	0.9800
genStep 240	0.0088496	0.0088527	0.0089286	12	12.96	13	0	0.00	0	0.9600
genStep 241	0.0088496	0.0088543	0.0089286	12	12.94	13	0	0.00	0	0.9400

genStep 242	0.0088496	0.0088543	0.0089286	12	12.94	13	0	0.00	0	0.9400
genStep 243	0.0088496	0.0088543	0.0089286	12	12.94	13	0	0.00	0	0.9400
genStep 244	0.0088496	0.0088543	0.0089286	12	12.94	13	0	0.00	0	0.9400
genStep 245	0.0088496	0.0088559	0.0089286	12	12.92	13	0	0.00	0	0.9200
genStep 246	0.0088496	0.0088575	0.0089286	12	12.90	13	0	0.00	0	0.9000
genStep 247	0.0088496	0.0088590	0.0089286	12	12.88	13	0	0.00	0	0.8800
genStep 248	0.0088496	0.0088590	0.0089286	12	12.88	13	0	0.00	0	0.8800
genStep 249	0.0088496	0.0088590	0.0089286	12	12.88	13	0	0.00	0	0.8800
genStep 250	0.0088496	0.0088606	0.0089286	12	12.86	13	0	0.00	0	0.8600
genStep 251	0.0088496	0.0088622	0.0089286	12	12.84	13	0	0.00	0	0.8400
genStep 252	0.0088496	0.0088638	0.0089286	12	12.82	13	0	0.00	0	0.8200
genStep 253	0.0088496	0.0088685	0.0089286	12	12.76	13	0	0.00	0	0.7600
genStep 254	0.0088496	0.0088701	0.0089286	12	12.74	13	0	0.00	0	0.7400
genStep 255	0.0088496	0.0088717	0.0089286	12	12.72	13	0	0.00	0	0.7200
genStep 256	0.0088496	0.0088717	0.0089286	12	12.72	13	0	0.00	0	0.7200
genStep 257	0.0088496	0.0088717	0.0089286	12	12.72	13	0	0.00	0	0.7200
genStep 258	0.0088496	0.0088717	0.0089286	12	12.72	13	0	0.00	0	0.7200
genStep 259	0.0088496	0.0088733	0.0089286	12	12.70	13	0	0.00	0	0.7000
genStep 260	0.0088496	0.0088748	0.0089286	12	12.68	13	0	0.00	0	0.6800
genStep 261	0.0088496	0.0088748	0.0089286	12	12.68	13	0	0.00	0	0.6800
genStep 262	0.0088496	0.0088748	0.0089286	12	12.68	13	0	0.00	0	0.6800
genStep 263	0.0088496	0.0088796	0.0089286	12	12.62	13	0	0.00	0	0.6200
genStep 264	0.0088496	0.0088812	0.0089286	12	12.60	13	0	0.00	0	0.6000
genStep 265	0.0088496	0.0088843	0.0089286	12	12.56	13	0	0.00	0	0.5600
genStep 266	0.0088496	0.0088859	0.0089286	12	12.54	13	0	0.00	0	0.5400
genStep 267	0.0088496	0.0088906	0.0089286	12	12.48	13	0	0.00	0	0.4800
genStep 268	0.0088496	0.0088954	0.0089286	12	12.42	13	0	0.00	0	0.4200
genStep 269	0.0088496	0.0088954	0.0089286	12	12.42	13	0	0.00	0	0.4200
genStep 270	0.0088496	0.0088954	0.0089286	12	12.42	13	0	0.00	0	0.4200
genStep 271	0.0088496	0.0088954	0.0089286	12	12.42	13	0	0.00	0	0.4200
genStep 272	0.0088496	0.0088986	0.0090090	11	12.38	13	0	0.00	0	1.3800
genStep 273	0.0088496	0.0089033	0.0090090	11	12.32	13	0	0.00	0	1.3200
genStep 274	0.0088496	0.0089049	0.0090090	11	12.30	13	0	0.00	0	1.3000
genStep 275	0.0088496	0.0089049	0.0090090	11	12.30	13	0	0.00	0	1.3000
genStep 276	0.0088496	0.0089065	0.0090090	11	12.28	13	0	0.00	0	1.2800
genStep 277	0.0088496	0.0089096	0.0090090	11	12.24	13	0	0.00	0	1.2400
genStep 278	0.0088496	0.0089128	0.0090090	11	12.20	13	0	0.00	0	1.2000
genStep 279	0.0088496	0.0089144	0.0090090	11	12.18	13	0	0.00	0	1.1800
genStep 280	0.0088496	0.0089160	0.0090090	11	12.16	13	0	0.00	0	1.1600
genStep 281	0.0088496	0.0089160	0.0090090	11	12.16	13	0	0.00	0	1.1600
genStep 282	0.0088496	0.0089223	0.0090090	11	12.08	13	0	0.00	0	1.0800
genStep 283	0.0088496	0.0089223	0.0090090	11	12.08	13	0	0.00	0	1.0800
genStep 284	0.0088496	0.0089254	0.0090090	11	12.04	13	0	0.00	0	1.0400
genStep 285	0.0088496	0.0089286	0.0090090	11	12.00	13	0	0.00	0	1.0000

genStep 506	0.0090090	0.0090467	0.0090909	10	10.54	11	0	0.00	0	0.5400
genStep 507	0.0090090	0.0090483	0.0090909	10	10.52	11	0	0.00	0	0.5200
genStep 508	0.0090090	0.0090483	0.0090909	10	10.52	11	0	0.00	0	0.5200
genStep 509	0.0090090	0.0090500	0.0090909	10	10.50	11	0	0.00	0	0.5000
genStep 510	0.0090090	0.0090500	0.0090909	10	10.50	11	0	0.00	0	0.5000
genStep 511	0.0090090	0.0090516	0.0090909	10	10.48	11	0	0.00	0	0.4800
genStep 512	0.0090090	0.0090516	0.0090909	10	10.48	11	0	0.00	0	0.4800
genStep 513	0.0090090	0.0090549	0.0090909	10	10.44	11	0	0.00	0	0.4400
genStep 514	0.0090090	0.0090549	0.0090909	10	10.44	11	0	0.00	0	0.4400
genStep 515	0.0090090	0.0090549	0.0090909	10	10.44	11	0	0.00	0	0.4400
genStep 516	0.0090090	0.0090565	0.0090909	10	10.42	11	0	0.00	0	0.4200
genStep 517	0.0090090	0.0090581	0.0090909	10	10.40	11	0	0.00	0	0.4000
genStep 518	0.0090090	0.0090598	0.0090909	10	10.38	11	0	0.00	0	0.3800
genStep 519	0.0090090	0.0090614	0.0090909	10	10.36	11	0	0.00	0	0.3600
genStep 520	0.0090090	0.0090631	0.0090909	10	10.34	11	0	0.00	0	0.3400
genStep 521	0.0090090	0.0090647	0.0090909	10	10.32	11	0	0.00	0	0.3200
genStep 522	0.0090090	0.0090647	0.0090909	10	10.32	11	0	0.00	0	0.3200
genStep 523	0.0090090	0.0090647	0.0090909	10	10.32	11	0	0.00	0	0.3200
genStep 524	0.0090090	0.0090663	0.0090909	10	10.30	11	0	0.00	0	0.3000
genStep 525	0.0090090	0.0090663	0.0090909	10	10.30	11	0	0.00	0	0.3000
genStep 526	0.0090090	0.0090663	0.0090909	10	10.30	11	0	0.00	0	0.3000
genStep 527	0.0090090	0.0090663	0.0090909	10	10.30	11	0	0.00	0	0.3000
genStep 528	0.0090090	0.0090696	0.0090909	10	10.26	11	0	0.00	0	0.2600
genStep 529	0.0090090	0.0090729	0.0090909	10	10.22	11	0	0.00	0	0.2200
genStep 530	0.0090090	0.0090729	0.0090909	10	10.22	11	0	0.00	0	0.2200
genStep 531	0.0090090	0.0090729	0.0090909	10	10.22	11	0	0.00	0	0.2200
genStep 532	0.0090090	0.0090729	0.0090909	10	10.22	11	0	0.00	0	0.2200
genStep 533	0.0090090	0.0090745	0.0090909	10	10.20	11	0	0.00	0	0.2000
genStep 534	0.0090090	0.0090762	0.0090909	10	10.18	11	0	0.00	0	0.1800
genStep 535	0.0090090	0.0090762	0.0090909	10	10.18	11	0	0.00	0	0.1800
genStep 536	0.0090090	0.0090778	0.0090909	10	10.16	11	0	0.00	0	0.1600
genStep 537	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 538	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 539	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 540	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 541	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 542	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 543	0.0090090	0.0090794	0.0090909	10	10.14	11	0	0.00	0	0.1400
genStep 544	0.0090090	0.0090811	0.0090909	10	10.12	11	0	0.00	0	0.1200
genStep 545	0.0090090	0.0090827	0.0090909	10	10.10	11	0	0.00	0	0.1000
genStep 546	0.0090090	0.0090827	0.0090909	10	10.10	11	0	0.00	0	0.1000
genStep 547	0.0090090	0.0090827	0.0090909	10	10.10	11	0	0.00	0	0.1000
genStep 548	0.0090090	0.0090827	0.0090909	10	10.10	11	0	0.00	0	0.1000
genStep 549	0.0090090	0.0090827	0.0090909	10	10.10	11	0	0.00	0	0.1000

genStep 638	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 639	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 640	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 641	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 642	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 643	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 644	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 645	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 646	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 647	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 648	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 649	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 650	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 651	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 652	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 653	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 654	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 655	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 656	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 657	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000
genStep 658	0.0090909	0.0090909	0.0090909	10	10.00	10	0	0.00	0	0.0000

B - DATA RELATIVE TO CHAPTER 4

Table 0.2 Characters that were used in the *Hoplitis* data.

Id	Character
0	Size
1	Extent of projection of apex of clypeus over labrum
2	Degree to which clypeal apex projects forward rather than downward
3	Breadth of clypeal truncation (broad to narrow and produced)
4	Ends of clypeal truncation (almost unrecognizable to much produced)
5	Shape of clypeal truncation (convex to concave)
6	Sublateral angles of clypeus
7	Emargination between truncation and sublateral angle of clypeus
8	Width of impunctate margin of clypeus
9	Convexity of clypeus
10	Spaces between clypeal punctures (close to clypeus impunctate)
11	Longitudinal median line of clypeus (absent to distinct and raised)
12	Coarseness of clypeal punctures (finer than to coarser than those of frons)
13	Coarseness of punctures of vertex (finer than to coarser than those of mesoscutum)
14	Ratio of distance from antennal bases to anterior ocellus to distance from anterior ocellus to posterior margin of vertex
15	Ratio of distance between posterior ocelli to distance from one of them to eye margin
16	Ratio of distance between posterior ocelli to distance from them to posterior margin of vertex
17	Ratio of genal width to eye width
18	Degree of convergence (or divergence) of inner margins of eyes
19	Clypeal brushes (absent to large)
20	Mandibular hairs (sparse to dense brush)
21	Width of mandibular apex
22	Width of mandible at narrowest median region
23	Length of mandible
24	Size of upper mandibular tooth (absent to large)
25	Inner subapical swelling of mandible (absent to tooth-like)
26	Ratio of distance from upper to middle mandibular tooth to distance from middle to lower tooth
27	Shape of upper mandibular tooth (acute to absent)
28	Height of hypostomal carinae
29	Posterior angles of hypostomal carinae
30	Shape of hypostomal areas (long to short)
31	Width of proboscoidal fossa
32	Segments of maxillary palpus (5 to 2)
33	Ratio of length of first to second segment of labial palpus
34	Setae on labial palpi and galeae (few straight, to many wavy)
35	Shape of labrum (broad rounded, to narrow bidentate)
36	Length of proboscis
37	Ratio of length to breadth of antennal pedicel
38	Ratio of lengths of first to second flagellar segments
39	Ratio of length of antennal scape to distance between antennal bases
40	Shape of thorax (elongate to nearly spherical)
41	Pre-episternal carina (absent to long)
42	Carina of pronotal lobe (absent to distinct)
43	Hind tibial spur (slender to robust)
44	Hind coxal carina (absent to high lamella)
45	Coarseness of punctures of mesoscutum (by comparison to those of episterna)
46	Coarseness of punctures of episterna (by comparison to those of genal areas)
47	Spaces between mesoscutal punctures
48	Pattern of pubescence at anterior end of mesoscutum (none, pair of spots, merged spots)

49	Size of papillae on wings
50	Wing color (clear to dusky)
51	Depth and definition of concavity of first metasomal tergum
52	Number of fasciae on metasomal terga
53	Coarseness of largest metasomal punctures (by comparison to those of mesoscutum)
54	Coarseness of metasomal punctures in general
55	Width of smooth margins of metasomal terga
56	Distinctness of basal grooves of metasomal terga
57	Length of scopa
58	Amount of scopa on sixth metasomal sternum
59	Size of depression of sixth metasomal sternum (absent to broad)
60	Amount of projection of sixth metasomal sternum beyond corresponding tergum
61	Profile of sixth metasomal tergum (straight to strongly concave)
62	Width of apical flange of sixth metasomal tergum
63	Integumental color (black to metallic)
64	Metasomal color (black to red)
65	Femoral color (black to red)
66	Color of tegulae (black to red)
67	Color of pubescence (white to black)
68	Breadth of clypeal truncation
69	Ends of clypeal truncation
70	Shape of clypeal truncation (smooth to crenulate)
71	Sublateral angles of clypeus
72	Width of impunctate margin of clypeus
73	Convexity of clypeus
74	Spaces between clypeal punctures
75	Coarseness of clypeal punctures (by comparison to those of frons)
76	Direction of subantennal sutures (parallel to diverging upward)
77	Degree of convergence (or divergence) of inner margins of eyes
78	Density of clypeal vestiture
79	Clypeal brushes (absent to distinct)
80	Length of mandible
81	Size of upper mandibular tooth (absent to large)
82	Length of lower mandibular tooth in relation to others
83	Apex of labrum (rounded to emarginate)
84	Antennal scape (slender to swollen)
85	Antennal apex (rounded, pointed, hooked, flattened)
86	Length of hair of antennal flagellum
87	Ratio of length to breadth of antennal pedicel
88	Thickness of antennal flagellum
89	Ratio of length to breadth of median flagellar segments
90	Outer carina of hind coxa
91	Ventral tooth of middle coxa
92	Lateral teeth of sixth metasomal tergum (absent to acute)
93	Median teeth of sixth tergum (absent to twice as long as broad)
94	Apical flange of sixth tergum (absent to broad)
95	Lateral margins of sixth tergum (straight to convex)
96	Extent that seventh tergum is exerted
97	Shape of apex of seventh tergum (medially pointed, rounded, to bilobed)
98	Profile of first metasomal sternum (flat to protuberant)
99	Size of second sternum in relation to third
100	Profile of disc of second sternum (flat to protuberant)
101	Apex of second sternum (flat to transversely elevated)
102	Margin of second sternum (concave to medially pointed)
103	Margin of third sternum (concave to convex)
104	Fringe of third sternum (absent to long across entire margin)
105	Margin of fourth sternum (concave to convex)
106	Fringe of fourth sternum (absent to long)

107	Amount that fifth sternum is exposed
108	Margin of fifth sternum (concave to convex)
109	Fringe of fifth sternum (absent to long)
110	Amount that sixth sternum is exposed
111	Longitudinal ridge of sixth sternum (absent to produced and expanded)
112	Sixth sternum (planate to duplicated)
113	Margin of sixth sternum (convex, to concave, to concave with median point)
114	Subapical lobes of sixth sternum (present or absent)
115	Margin of seventh sternum (convex to concave)
116	Projection of eighth sternum behind lateral apodemes (absent to over twice as long as broad)
117	Curvature of gonocoxites (none to angular)
118	Apices of gonocoxites (pointed to truncate)
119	Shape of volsellae (short rounded, to long pointed)
120	Fusion of volsellae to gonocoxites (none to complete)
121	Distribution (desert only to high montane only)

Table 0.3 Characters that were used in the Acorn data.

Id	Character
1	Maximum number of twins observed
2	Cup scales differ in each other
3	Cup scales' hardness
4	Spiny cup scales
5	Cup scales fibrous
6	Soft cup scales
7	Hairs at cup scales
8	Cupule with fringe
9	Hairs at nut outside
10	Cup scale shape
11	Scale length at cupule tip
12	Nut length
13	Nut diameter
14	Nut bottom diameter
15	Nut shell thickness
16	Cupule depth
17	Cupula length
18	Cupula inner diameter
19	Cupula outer diameter
20	Cupula thickness
21	Interval between cup scales
22	Cup length / nut length
23	Cup depth / cup length
24	Cupula inner diameter / cupula outer diameter
25	Cupula inner diameter / nut diameter
26	Nut diameter / nut length
27	Exterior nut / nut length
28	Interior nut / nut length
29	Cup thickness / cup outer diameter
30	Nut bottom diameter / nut diameter
31	Stalk length
32	Stalk thickness
33	Stalk diameter / stalk length

Table 0.4 Phi coefficient results for *Hoplitis* dataset.

[illegible]

[illegible]

Table 0.5 Results of 25 unique analysis were listed with less than or equal to 10 characters per solution sets

	2 13 14 23 28 32	2 5 16 19 30 32	2 16 19 21 30 32	1 2 13 16 23 30 32	1 2 16 19 23 30 32	2 4 13 20 23 28 32	2 6 18 23 28 30 32	2 10 13 16 23 30 32	2 10 14 16 26 29 32	2 13 16 22 23 30 32	1 4 5 12 16 19 30 31	1 4 5 6 14 19 20 31	1 2 14 17 23 28 32 33	1 2 4 13 14 23 29 32	2 4 16 18 21 28 30 32	2 6 16 18 23 28 30 32	2 6 15 18 26 27 30 32	2 6 12 27 28 30 31 32	2 10 13 16 23 29 30 32	1 5 6 10 13 16 26 31 33	1 3 4 5 12 13 16 25 30 31	1 4 5 6 10 13 22 26 31 33	1 4 5 12 13 16 22 25 30 31	1 5 6 10 13 14 16 26 31 33	1 5 8 14 16 18 23 26 31 33	Occurrence within solutions	
CH1				X	X						X	X	X	X						X	X	X	X	X	X	12	
CH2	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X								17
CH3																						X					1
CH4						X					X	X		X	X						X	X	X	X			8
CH5		X									X	X									X	X	X	X	X	X	9
CH6							X					X				X	X	X		X		X		X	X		8
CH7																											0
CH8																									X		1
CH9																											0
CH10								X	X											X	X		X		X		6
CH11																											0
CH12											X							X						X			4
CH13	X			X		X		X		X				X						X	X	X	X	X	X		12
CH14	X								X			X	X	X											X	X	7
CH15																	X										1
CH16		X	X	X	X			X	X	X	X				X	X			X	X	X		X	X	X	X	16
CH17													X														1
CH18						X									X	X	X								X		5
CH19		X	X		X						X	X															5
CH20						X						X															2
CH21			X												X												2
CH22										X												X	X				3
CH23	X			X	X	X	X	X		X			X	X		X			X							X	12
CH24																											0
CH25																					X		X				2
CH26									X								X					X		X	X	X	6
CH27																	X	X		X							2
CH28	X					X	X						X		X	X		X									7
CH29									X					X						X							3
CH30		X	X	X	X		X	X		X	X				X	X	X	X	X	X		X		X			15
CH31											X	X						X		X	X	X	X	X	X	X	9
CH32	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X	X							17
CH33													X							X		X		X	X	X	5

Table 0.6 Phi coefficient results for Acorn dataset.

	33	32	31	30	29	28	27	26	25	23	22	21	20	19	18	17	16	15	14	13	12	10	8	6	5	4	3	2	1						
1																													0.00						
2																												0.00	-0.71						
3																											0.00	-0.30	0.21						
4																										0.00	0.30	-0.45	0.37						
5																									0.00	0.38	0.27	-0.92	0.61						
6																							0.00	0.20	-0.10	-0.14	-0.27	0.03							
8																						0.00	-0.14	0.27	-0.14	-0.04	-0.30	0.21							
10																						0.00	-0.12	0.22	0.16	-0.19	-0.12	-0.22	0.02						
12																					0.00	-0.25	-0.09	-0.07	0.36	0.40	0.47	-0.40	0.24						
13																				0.00	0.02	0.40	-0.20	-0.14	0.11	0.20	0.21	-0.20	0.20						
14																			0.00	-0.06	-0.27	0.07	0.33	-0.05	0.09	-0.05	-0.13	-0.15	0.29						
15																		0.00	-0.13	-0.20	-0.09	-0.12	-0.04	0.30	-0.15	-0.14	-0.04	0.14	-0.20						
16																	0.00	-0.27	-0.04	0.33	-0.20	0.10	0.23	0.15	-0.38	0.22	-0.20	0.15	-0.16	0.05					
17																0.00	-0.10	-0.04	0.41	-0.09	-0.48	-0.22	-0.28	0.41	0.30	-0.17	-0.13	-0.10	0.13	-0.28					
18															0.00	-0.25	-0.10	0.17	-0.10	-0.09	-0.48	0.06	-0.28	-0.10	0.13	0.25	0.09	-0.10	-0.09	0.12					
19															0.00	0.22	-0.15	-0.06	-0.39	-0.06	0.14	0.01	-0.13	-0.17	-0.06	0.11	0.09	0.43	-0.06	-0.11	0.01				
20														0.00	-0.09	0.22	-0.06	0.22	-0.06	-0.18	-0.28	-0.13	-0.17	-0.06	-0.20	-0.22	0.11	-0.06	0.20	-0.28					
21												0.00	-0.11	-0.19	-0.19	-0.08	0.02	-0.08	-0.23	0.38	0.18	0.08	-0.08	0.01	0.24	0.27	-0.08	-0.27	0.14						
22										0.00	-0.11	-0.28	0.01	-0.28	0.12	0.21	-0.11	-0.20	0.11	0.20	-0.42	-0.17	0.21	-0.32	-0.55	-0.32	-0.20	0.49	-0.12						
23										0.00	-0.11	-0.28	0.01	-0.28	0.12	0.21	-0.11	-0.20	0.11	0.20	-0.42	-0.17	0.21	-0.32	-0.55	-0.32	-0.20	0.49	-0.12						
25									0.00	-0.28	0.35	-0.09	-0.15	-0.15	-0.15	-0.06	0.22	-0.06	-0.18	0.31	0.68	-0.17	-0.06	-0.20	0.39	0.43	0.69	-0.43	0.31						
26									0.00	-0.17	-0.35	0.08	-0.17	-0.28	0.19	-0.12	0.03	0.36	0.28	0.02	-0.25	0.56	0.36	0.42	0.36	-0.19	-0.12	-0.42	0.21						
27							0.00	0.18	-0.09	-0.28	-0.11	-0.09	-0.15	0.22	-0.06	-0.39	0.69	-0.18	-0.28	0.02	-0.25	0.56	0.36	0.42	0.36	-0.19	-0.12	-0.42	0.21						
28						0.00	0.14	-0.35	-0.18	0.29	-0.23	0.14	0.14	-0.31	0.36	0.33	-0.46	-0.13	0.01	-0.24	-0.03	-0.35	-0.13	0.15	-0.47	-0.05	-0.13	0.43	-0.42						
29						0.00	-0.23	-0.11	0.08	-0.11	0.14	-0.14	-0.11	-0.19	-0.19	-0.08	0.02	-0.08	0.32	0.14	-0.16	0.37	-0.08	-0.25	-0.28	0.01	-0.08	0.25	-0.11						
30						0.00	-0.20	-0.04	0.24	-0.50	0.24	-0.03	0.05	0.24	-0.22	0.09	0.04	-0.17	-0.15	0.09	0.11	0.58	0.16	0.27	0.38	0.83	0.38	0.27	-0.92	0.61					
31						0.00	-0.24	-0.28	0.09	0.36	0.39	-0.55	0.24	-0.42	-0.43	0.49	-0.27	0.20	-0.11	-0.09	0.13	0.14	-0.16	0.14	-0.15	-0.20	-0.40	-0.22	-0.30	-0.27	-0.92	-0.45	-0.30	1.00	-0.71
32						0.00	0.32	0.25	0.43	0.20	-0.42	-0.43	0.49	-0.27	0.20	-0.11	-0.09	0.13	0.14	-0.16	0.14	-0.15	-0.20	-0.40	-0.22	-0.30	-0.27	-0.92	-0.45	-0.30	1.00	-0.71			
33	0.00	-0.51	0.46	-0.61	-0.19	-0.09	-0.15	0.66	-0.15	-0.08	0.12	-0.15	-0.15	-0.25	0.00	0.41	-0.04	-0.10	0.36	0.12	-0.22	0.42	0.41	0.30	0.46	-0.13	-0.10	-0.51	0.52						

C - DATA RELATIVE TO CHAPTER 5

Table 0.7 Detailed information on Heme Copper Oxidase dataset.

Taxon	Accession	Type	Length (aa)
Azospirillum brasilense	AAC72071	C	498
Helicobacter pylori	BAA33526	C	488
Achromobacter cycloclastes	CAC27381	NOR	448
Paracoccus denitrificans	CAA29274	A1	554
Thermus thermophilus	AAB00370	B	568
Pseudomonas putida	BAA76357	A1	672
Pseudomonas putida	AAL37189	A1	676
Azotobacter vinelandii	ZP_00417290	A1	530
Azotobacter vinelandii	ZP_00418276	A1	835
Azotobacter vinelandii	ZP_00418645	C	499
Agrobacterium tumefaciens	CAA86308	C	539
Mesorhizobium loti	CAD31422	C	536
Acetobacter pasteurianus	BAB97174	A1	663
Alcaligenes faecalis	BAA90782	NOR	448
Desulfovibrio vulgaris str. Hildenborough	YP_011033	A2	541
Desulfovibrio vulgaris str. 'Miyazaki F'	YP_002434961	A2	541
Desulfovibrio vulgaris str. 'Miyazaki F'	BAA06976	A2	541
Hydrogenobacter thermophilus	BAE79481	NOR	469
Rhodobacter capsulatus	AAC46108	C	532
Rhodobacter sphaeroides	AAB02556	C	535
Synechococcus elongatus PCC 7942	YP_401620	A2	541
Synechococcus elongatus PCC 7942	YP_399221	C	483
Synechocystis sp. PCC 6803	NP_441291	A2	544
Synechocystis sp. PCC 6803	NP_440609	A2	551
Synechocystis sp. PCC 6803	NP_442983	NOR	770
Bacillus firmus	AAA22365	A1	625
Geobacillus stearothermophilus	BAA11112	A1	615
Geobacillus stearothermophilus	BAA35056	B	549
Geobacillus stearothermophilus	BAG66139	NOR	792
Bacillus subtilis	CAA38077	A1	621
Corynebacterium glutamicum	BAB64406	A1	584
Haloferax volcanii	AAK14377	A1	586
Haloferax volcanii	AAK14378	A1	578
Acidianus ambivalens	CAA69980	B	587
Roseobacter denitrificans	BAB84303	NOR	434
Rubrivivax gelatinosus	AAV66315	A1	531
Rubrivivax gelatinosus	AAW66132	C	479
Halomonas halodenitrificans	BAA24701	NOR	451
Rhizobium etli	AAC15888	C	540
Rhodothermus marinus	CAC08532	A2	565
Rhodothermus marinus	AAY23178	B	561
Clavibacter michiganensis subsp. sepedonicus	YP_001710118	A1	581
Synechococcus sp. PCC 7002	YP_001733987	A2	551
Synechococcus sp. PCC 7002	YP_001734418	A2	548
Synechococcus sp. WH 7803	YP_001224437	A2	562
Synechococcus sp. WH 7803	YP_001225595	A2	556
Thermosynechococcus vulcanus	BAA41041	A2	552
Mycobacterium abscessus	YP_001703352	A1	564
Mycobacterium abscessus	YP_001704119	A1	564
Mycobacterium abscessus	YP_001701786	A1	520
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	NP_871151	A1	654
Paulinella chromatophora	ACB43160	A2	537
Cyanothece sp. PCC 8801	YP_002370814	A2	555
Salmonella enterica subsp. arizonae serovar 62:z4,z23:--	YP_001571493	A1	663
Cyanothece sp. ATCC 51142	YP_001806228	A2	574
Cyanothece sp. ATCC 51142	YP_001805524	A2	594
Cyanothece sp. ATCC 51142	YP_001803392	A2	572
Sulfitobacter sp. EE-36	ZP_00956065	A1	855
Sulfitobacter sp. EE-36	ZP_00955289	A1	557
Sulfitobacter sp. EE-36	ZP_00956569	C	536
Moritella sp. PE36	ZP_01900412	A1	540
Moritella sp. PE36	ZP_01897896	A1	580
Rickettsiella grylli	ZP_02062292	A1	528
Prochlorococcus marinus subsp. pastoris str. CCMP1986	NP_892563	A2	541
Prochlorococcus marinus str. NATL2A	YP_292967	A2	546
Prochlorococcus marinus str. MIT 9303	YP_001016658	A2	558
Shewanella violacea	BAA94865	A1	661
Shewanella sp. MR-4	YP_735912	A1	530
Shewanella sp. MR-4	YP_735944	A1	659
Shewanella sp. MR-4	YP_734136	C	478
Shewanella sp. MR-4	YP_735217	NOR	761
Shewanella sp. MR-7	YP_739899	A1	530
Shewanella sp. MR-7	YP_739949	A1	659
Shewanella sp. MR-7	YP_738015	C	478
Shewanella sp. MR-7	YP_736939	NOR	761
Azoarcus sp. BH72	YP_933261	A1	860
Azoarcus sp. BH72	YP_934805	A1	525
Azoarcus sp. BH72	YP_932845	C	474
Azoarcus sp. BH72	YP_934592	NOR	459

Acinetobacter sp. ADP1	YP_047021	A1	662
Nostoc punctiforme PCC 73102	YP_001865393	A2	558
Nostoc punctiforme PCC 73102	YP_001864063	A2	562
Nostoc punctiforme PCC 73102	ZP_00109104	A2	558
Nostoc punctiforme PCC 73102	YP_001866883	A2	553
Nostoc punctiforme PCC 73102	ZP_00110624	A2	580
Nostoc punctiforme PCC 73102	YP_001868782	A2	582
Halobacterium sp. NRC-1	NP_279674	A1	593
Halobacterium sp. NRC-1	NP_444237	B	565
Synechococcus sp. CC9311	YP_731131	A2	564
Synechococcus sp. CC9311	YP_731328	A2	554
Cyanotheca sp. PCC 7424	YP_002378410	A2	565
Cyanotheca sp. PCC 7424	YP_002379313	A2	560
Bacillus clausii KSM-K16	YP_175632	A1	647
Bacillus clausii KSM-K16	YP_175888	A1	624
Pseudomonas sp. G-179	AAC79449	NOR	472
Prochlorococcus marinus str. MIT 9312	YP_396941	A2	541
Prochlorococcus marinus str. MIT 9313	NP_895169	A2	558
Aromatoleum aromaticum EbN1	YP_157109	A1	835
Aromatoleum aromaticum EbN1	YP_159601	A1	607
Aromatoleum aromaticum EbN1	YP_159414	A1	525
Aromatoleum aromaticum EbN1	YP_159094	B	495
Aromatoleum aromaticum EbN1	YP_159939	C	474
Aromatoleum aromaticum EbN1	YP_158648	C	542
Aromatoleum aromaticum EbN1	YP_157126	NOR	460
Pseudomonas putida GB-1	YP_001666370	A1	529
Pseudomonas putida GB-1	YP_001667096	A1	672
Pseudomonas putida GB-1	YP_001670042	C	474
Pseudomonas putida GB-1	YP_001670046	C	480
Wolbachia endosymbiont of Drosophila simulans	ZP_00372535	A1	516
Xanthobacter autotrophicus Py2	YP_001419524	A1	540
Xanthobacter autotrophicus Py2	YP_001415374	C	555
Xanthobacter autotrophicus Py2	YP_001417931	NOR	767
Mycobacterium tuberculosis CDC1551	NP_337644	A1	573
Mycobacterium tuberculosis H37Rv	NP_217559	A1	573
Synechococcus sp. WH 8102	NP_897622	A2	562
Synechococcus sp. WH 8102	NP_897953	A2	574
Thauera sp. MZ1T	ZP_02844227	A1	536
Thauera sp. MZ1T	ZP_02843808	C	476
Thauera sp. MZ1T	ZP_02841424	NOR	460
Helicobacter pylori 26695	NP_206943	C	488
Helicobacter pylori J99	NP_222853	C	488
Pseudoalteromonas tunicata D2	ZP_01135737	A1	525
Pseudoalteromonas tunicata D2	ZP_01134282	C	477
Roseovarius nubinihibens ISM	ZP_00959150	A1	563
Roseovarius nubinihibens ISM	ZP_00959081	A1	818
Moritella marina	BAA89393	C	465
Prochlorococcus marinus str. MIT 9211	YP_001550327	A2	552
Prochlorococcus marinus str. MIT 9215	YP_001483726	A2	541
Prochlorococcus marinus str. MIT 9215	YP_001484863	A2	563
Staphylococcus aureus subsp. aureus NCTC 8325	YP_499553	A1	662
Staphylococcus aureus subsp. aureus COL	YP_185933	A1	662
Shewanella sp. ANA-3	YP_871655	A1	659
Shewanella sp. ANA-3	YP_871616	A1	530
Shewanella sp. ANA-3	YP_869743	C	478
Shewanella sp. ANA-3	YP_868490	NOR	761
Vitreoscilla sp. C1	AAM20915	A1	666
Pyrobaculum oguniense	BAC56132	A1	803
Pyrobaculum oguniense	BAC56140	B	557
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	NP_459438	A1	663
Streptomyces coelicolor A3(2)	NP_631290	A1	573
Streptomyces coelicolor A3(2)	NP_626411	A1	578
Rhodococcus jostii RHA1	YP_707487	A1	582
Rhodococcus jostii RHA1	YP_706379	A1	596
Rhodococcus jostii RHA1	YP_704173	A1	579
Nostoc sp. PCC 7120	NP_486555	A2	559
Nostoc sp. PCC 7120	NP_486772	A2	555
Nostoc sp. PCC 7120	NP_484994	A2	575
Frankia sp. CcI3	YP_481112	A1	571
Frankia sp. CcI3	YP_482900	A1	609
Frankia sp. CcI3	YP_481589	A1	605
Frankia sp. CcI3	YP_482202	A1	560
Buchnera aphidicola str. APS (Acyrthosiphon pisum)	NP_240282	A1	662
Synechococcus sp. CC9605	YP_380936	A2	555
Synechococcus sp. CC9605	YP_381293	A2	562
Bradyrhizobium sp. ORS278	YP_001203834	A1	669
Bradyrhizobium sp. ORS278	YP_001206809	A1	666
Bradyrhizobium sp. ORS278	YP_001208524	A1	537
Bradyrhizobium sp. ORS278	YP_001204332	B	541
Bradyrhizobium sp. ORS278	YP_001204500	C	564
Bradyrhizobium sp. ORS278	YP_001203888	NOR	448
Halomonas maura	AAY40171	C	474
Microcoleus chthonoplastes PCC 7420	EDX74433	A2	561
Francisella tularensis subsp. holarctica	YP_512987	A1	676
Photobacterium sp. SKA34	ZP_01158716	A1	678
Photobacterium sp. SKA34	ZP_01162438	A1	539
Photobacterium sp. SKA34	ZP_01160483	A1	696
Neisseria meningitidis MC58	NP_274728	C	481

Neisseria meningitidis MC58	NP_274628	NOR	751	Prochlorococcus marinus str. MIT 9515	YP_001011578	A2	563
Neisseria meningitidis Z2491	YP_002343249	C	481	Prochlorococcus marinus str. MIT 9301	YP_001090693	A2	541
Neisseria meningitidis Z2491	YP_002343169	NOR	751	Prochlorococcus marinus str. NATL1A	YP_001014326	A2	546
Desulfotobacterium hafniense Y51	YP_517398	NOR	442	Colwellia psychrerythraea 34H	YP_267002	A1	540
Prochlorococcus marinus str. AS9601	YP_001008895	A2	541	Colwellia psychrerythraea 34H	YP_268725	C	474
Vibrio sp. Ex25	YP_002075077	A1	543	Colwellia psychrerythraea 34H	YP_267351	NOR	463
Vibrio sp. Ex25	YP_002074576	A1	678	Listeria monocytogenes EGD-e	NP_463547	A1	659
Escherichia coli O157:H7 EDL933	NP_286173	A1	663	Candidatus Kuenenia stuttgartiensis	CAJ711174	C	710
Xylella fastidiosa Ann-1	ZP_00682022	A1	622	Staphylococcus epidermidis RP62A	YP_188230	A1	662
Alteromonadales bacterium TW-7	ZP_01613838	A1	530	Staphylococcus epidermidis ATCC 12228	NP_764313	A1	662
Alteromonadales bacterium TW-7	ZP_01614363	C	477	Agrobacterium tumefaciens str. C58	NP_353177	A1	667
Magnetococcus sp. MC-1	YP_866259	C	472	Agrobacterium tumefaciens str. C58	NP_353792	A1	569
Magnetococcus sp. MC-1	YP_864053	NOR	460	Agrobacterium tumefaciens str. C58	NP_354541	C	540
Staphylococcus aureus subsp. aureus Mu50	NP_371584	A1	662	Agrobacterium tumefaciens str. C58	NP_356267	NOR	448
Staphylococcus aureus subsp. aureus N315	NP_374179	A1	662	Francisella tularensis subsp. tularensis SCHU S4	YP_169332	A1	676
Gordonia westfalica	NP_954783	A1	596	Pyrobaculum aerophilum str. IM2	NP_559235	A1	800
Dechloromonas aromatica RCB	YP_284689	A1	531	Pyrobaculum aerophilum str. IM2	NP_559246	B	557
Dechloromonas aromatica RCB	YP_283939	C	475	Pyrobaculum aerophilum str. IM2	NP_560853	NOR	721
Dechloromonas aromatica RCB	YP_286391	NOR	458	Cyanobium sp. PCC 7001	EDY39380	A2	540
Pseudomonas putida KT2440	NP_742974	A1	672	Cyanobium sp. PCC 7001	EDY38696	A2	574
Pseudomonas putida KT2440	NP_742274	A1	529	Xylella fastidiosa Temecula1	NP_778846	A1	665
Pseudomonas putida KT2440	NP_746366	C	474	Alkalilimnicola ehrlichei MLHE-1	YP_741139	A1	526
Pseudomonas putida KT2440	NP_746371	C	480	Alkalilimnicola ehrlichei MLHE-1	YP_742716	C	474
Xylella fastidiosa 9a5c	NP_298678	A1	672	Alkalilimnicola ehrlichei MLHE-1	YP_742950	NOR	447
Erythrobacter sp. SD-21	ZP_01864317	A1	572	Yersinia pestis KIM	NP_668350	A1	663
Erythrobacter sp. SD-21	ZP_01865010	A1	845	Leptospira interrogans serovar Lai str. 56601	NP_710424	A2	534
Wolbachia endosymbiont of Drosophila melanogaster	NP_966102	A1	516	Mycobacterium sp. KMS	YP_937889	A1	576
Wolbachia endosymbiont of Drosophila melanogaster	NP_966115	B	419	Mycobacterium sp. KMS	YP_939485	A1	582
Cupriavidus taiwanensis	YP_002005088	A1	657	Mycobacterium sp. KMS	YP_937591	NOR	783
Cupriavidus taiwanensis	YP_002008421	A1	585	Xanthomonas campestris pv. campestris str. ATCC 33913	NP_639177	A1	534
Cupriavidus taiwanensis	YP_002008864	A1	667	Xanthomonas campestris pv. campestris str. ATCC 33913	NP_636534	A1	666
Cupriavidus taiwanensis	YP_001795680	A1	535	Xanthomonas axonopodis pv. citri str. 306	NP_644193	A1	534
Cupriavidus taiwanensis	YP_002005870	C	482	Xanthomonas axonopodis pv. citri str. 306	NP_641594	A1	666
Cupriavidus taiwanensis	YP_002008716	NOR	762	Caulobacter crescentus CB15	NP_422200	A1	552
Mycobacterium sp. MCS	YP_640601	A1	582	Caulobacter crescentus CB15	NP_420580	A1	665
Mycobacterium sp. MCS	YP_639017	A1	576	Caulobacter crescentus CB15	NP_420214	C	548
Mycobacterium sp. MCS	YP_638736	NOR	783	Campylobacter jejuni subsp. jejuni NCTC 11168	YP_002344870	C	488
Mycobacterium sp. JLS	YP_001070112	A1	576	Campylobacter jejuni RM1221	YP_179644	C	488
Mycobacterium sp. JLS	YP_001071718	A1	582				
Mycobacterium sp. JLS	YP_001069828	NOR	783				
Crocospaera watsonii WH 8501	ZP_00514441	A2	560				
Crocospaera watsonii WH 8501	ZP_00515880	A2	555				
Prochlorococcus marinus subsp. marinus str. CCMP1375	NP_874835	A2	552				
Prochlorococcus marinus str. MIT 9515	YP_001010823	A2	539				

Nocardioides sp. JS614	YP_922823	A1	561
Nocardioides sp. JS614	YP_919310	A1	567
Nocardioides sp. JS614	YP_924329	A1	582
Corynebacterium efficiens YS-314	NP_739028	A1	580
Vibrio vulnificus YJ016	NP_937171	A1	543
Vibrio vulnificus YJ016	NP_934464	C	477
Staphylococcus aureus subsp. aureus MW2	NP_645760	A1	662
Corynebacterium glutamicum ATCC 13032	YP_226765	A1	584
Corynebacterium glutamicum ATCC 13032	NP_601724	A1	584
Thermosynechococcus elongatus BP-1	NP_682800	A2	552
Bacillus anthracis str. Ames	NP_843229	A1	644
Bacillus anthracis str. Ames	NP_846391	A1	611
Buchnera aphidicola str. Sg (Schizaphis graminum)	NP_660787	A1	659
Escherichia coli CFT073	NP_752476	A1	663
Saccharophagus degradans 2-40	YP_525513	A1	516
Saccharophagus degradans 2-40	YP_527894	C	490
Trichodesmium erythraeum IMS101	YP_720232	A2	562
Trichodesmium erythraeum IMS101	YP_721510	A2	590
Tropheryma whippelii str. Twist	NP_787372	A1	543
Candidatus Blochmannia floridanus	NP_878541	A1	654
Acidobacteria bacterium Ellin345	YP_589517	A1	595
Acidobacteria bacterium Ellin345	YP_592070	A2	544
Acidobacteria bacterium Ellin345	YP_589443	NOR	761
Brucella suis 1330	NP_697496	A1	552
Brucella suis 1330	NP_697085	A1	659
Brucella suis 1330	NP_697395	C	541
Brucella suis 1330	NP_699446	NOR	449
Hermiimonas arsenicoxydans	YP_001100944	A1	667
Hermiimonas arsenicoxydans	YP_001101152	A1	533
Hermiimonas arsenicoxydans	YP_001099422	A1	586
Hermiimonas arsenicoxydans	CAL61802	C	483
Hermiimonas arsenicoxydans	YP_001099927	C	475
Hermiimonas arsenicoxydans	YP_001099949	NOR	758
Hermiimonas arsenicoxydans	YP_001098413	NOR	757
Pseudomonas syringae pv. syringae B728a	YP_234234	A1	670
Pseudomonas syringae pv. syringae B728a	YP_236483	C	480
Ehrlichia chaffeensis str. Arkansas	YP_507790	A1	518
Ehrlichia chaffeensis str. Arkansas	YP_507864	A1	432
Pseudomonas fluorescens Pf0-1	YP_345811	A1	530
Pseudomonas fluorescens Pf0-1	YP_350375	A1	676
Pseudomonas fluorescens Pf0-1	YP_347555	C	480
Pseudomonas fluorescens Pf0-1	YP_347559	C	474
Pseudomonas fluorescens Pf0-1	YP_348243	C	475
Desulfovibrio desulfuricans subsp. desulfuricans str. G20	YP_388315	A2	543
Bermanella marisrubri	ZP_01306902	A1	522
Bermanella marisrubri	ZP_01305640	C	476
Neptuniibacter caesariensis	ZP_01166403	A1	524
Neptuniibacter caesariensis	ZP_01166126	C	476
Pseudomonas aeruginosa UCBPP-PA14	YP_788264	A1	530
Pseudomonas aeruginosa UCBPP-PA14	YP_791927	A1	658
Pseudomonas aeruginosa UCBPP-PA14	YP_791385	C	480
Pseudomonas aeruginosa UCBPP-PA14	YP_788989	C	475
Pseudomonas aeruginosa UCBPP-PA14	YP_791695	C	475
Pseudomonas aeruginosa UCBPP-PA14	YP_791698	C	475
Pseudomonas aeruginosa UCBPP-PA14	YP_788694	NOR	465
Pseudomonas aeruginosa PAO1	NP_248796	A1	530
Pseudomonas aeruginosa PAO1	NP_250009	A1	658
Pseudomonas aeruginosa PAO1	NP_250248	C	475
Pseudomonas aeruginosa PAO1	NP_250547	C	480
Pseudomonas aeruginosa PAO1	NP_252822	C	475
Pseudomonas aeruginosa PAO1	NP_250245	C	475
Pseudomonas aeruginosa PAO1	NP_249215	NOR	466
Salmonella enterica subsp. enterica serovar Typhi str. Ty2	NP_806148	A1	663
Shewanella oneidensis MR-1	NP_720123	A1	530
Shewanella oneidensis MR-1	NP_717954	C	478
Anaplasma phagocytophilum HZ	YP_505633	A1	518
Anaplasma phagocytophilum HZ	YP_505736	A1	420
Yersinia pestis CO92	NP_406640	A1	663
Yersinia pestis CO92	YP_002348070	A1	663
Gemmata obscuriglobus UQM 2246	ZP_02733610	A2	586
Gemmata obscuriglobus UQM 2246	ZP_02737089	C	484
Croceibacter atlanticus HTCC2559	ZP_00949206	A1	608
Burkholderia cenocepacia J2315	YP_002229907	A1	535
Burkholderia cenocepacia J2315	YP_002231271	A1	668
Burkholderia cenocepacia J2315	YP_002153874	B	879
Mycobacterium marinum M	YP_001849956	A1	573
Mycobacterium marinum M	YP_001851577	A1	577
Rhizobium leguminosarum bv. viciae 3841	YP_766633	A1	566
Rhizobium leguminosarum bv. viciae 3841	YP_768626	A1	592
Rhizobium leguminosarum bv. viciae 3841	YP_765315	C	540
Rhizobium leguminosarum bv. viciae 3841	YP_766146	C	547

Vibrio vulnificus CMCP6	NP_762523	A1	543
Vibrio vulnificus CMCP6	NP_761442	C	477
Pectobacterium atrosepticum SCRI1043	YP_049248	A1	663
Tropheryma whipplei TW08/27	NP_789455	A1	543
Salmonella enterica subsp. enterica serovar Typhi str. CT18	NP_455039	A1	663
Pseudomonas fluorescens Pf-5	YP_261816	A1	657
Pseudomonas fluorescens Pf-5	YP_262115	A1	672
Pseudomonas fluorescens Pf-5	YP_257214	A1	531
Pseudomonas fluorescens Pf-5	YP_259044	C	474
Pseudomonas fluorescens Pf-5	YP_259940	C	475
Pseudomonas fluorescens Pf-5	YP_259040	C	480
Oceanobacillus iheyensis HTE831	NP_692359	A1	617
Oceanobacillus iheyensis HTE831	NP_693175	A1	650
Oceanobacillus iheyensis HTE831	NP_692667	B	562
Mannheimia succiniciproducens MBEL55E	YP_087486	NOR	624
Bacillus cereus ATCC 10987	NP_977096	A1	644
Bacillus cereus ATCC 10987	NP_980285	A1	611
Neorickettsia sennetsu str. Miyayama	YP_506501	A1	519
Neorickettsia sennetsu str. Miyayama	YP_506021	A1	416
Pseudomonas syringae pv. tomato str. DC3000	NP_791153	A1	670
Pseudomonas syringae pv. tomato str. DC3000	NP_791827	C	480
Vibrio parahaemolyticus RIMD 2210633	NP_800047	A1	545
Vibrio parahaemolyticus RIMD 2210633	NP_800138	A1	678
Vibrio parahaemolyticus RIMD 2210633	NP_797923	C	477
Bacillus subtilis subsp. subtilis str. 168	NP_389373	A1	622
Bacillus subtilis subsp. subtilis str. 168	NP_391695	A1	649
Aquifex aeolicus VF5	NP_213105	A1	386
Aquifex aeolicus VF5	NP_214504	A2	485
Aquifex aeolicus VF5	NP_214506	B	592
Bradyrhizobium japonicum USDA 110	NP_766790	A1	665
Bradyrhizobium japonicum USDA 110	NP_770424	A1	591
Bradyrhizobium japonicum USDA 110	NP_767811	A1	541
Bradyrhizobium japonicum USDA 110	NP_769355	A1	666
Bradyrhizobium japonicum USDA 110	NP_771120	B	540
Bradyrhizobium japonicum USDA 110	NP_769403	C	549
Bradyrhizobium japonicum USDA 110	NP_769855	NOR	448
Brucella melitensis 16M	NP_540382	A1	552
Brucella melitensis 16M	NP_540817	A1	659
Brucella melitensis 16M	NP_540481	C	541
Brucella melitensis 16M	NP_541976	NOR	449
Buchnera aphidicola str. Bp (Baizongia pistaciae)	NP_778021	A1	659
Shewanella piezotolerans WP3	YP_002310796	A1	661
Shewanella piezotolerans WP3	YP_002314036	A1	533
Shewanella piezotolerans WP3	YP_002313394	C	469
Shewanella piezotolerans WP3	YP_002311898	C	478
Shewanella piezotolerans WP3	YP_002310231	NOR	763
Bacillus cereus ATCC 14579	NP_830510	A1	647
Bacillus cereus ATCC 14579	NP_833662	A1	620
Coxiella burnetii RSA 493	NP_820041	A1	668
Streptomyces avermitilis MA-4680	NP_827713	A1	563
Streptomyces avermitilis MA-4680	NP_827224	A1	579
Hyphomonas neptunium ATCC 15444	YP_759761	A1	562
Hyphomonas neptunium ATCC 15444	YP_759950	C	559
Nitrosomonas europaea ATCC 19718	NP_841089	A1	525
Nitrosomonas europaea ATCC 19718	NP_840762	B	492
Nitrosomonas europaea ATCC 19718	NP_842019	NOR	448
Rhizobium arachis	AAP22949	A1	665
Yersinia pestis biovar Microtus str. 91001	NP_992154	A1	663
Acidovorax sp. JS42	YP_986993	A1	646
Acidovorax sp. JS42	YP_987725	A1	544
Acidovorax sp. JS42	YP_985395	C	479
Acidovorax sp. JS42	YP_987627	NOR	457
Acidovorax sp. JS42	YP_984423	NOR	768
Mycobacterium bovis AF2122/97	NP_856714	A1	573
Solibacter usitatus Ellin6076	YP_826822	A1	591
Solibacter usitatus Ellin6076	YP_828106	A1	631
Solibacter usitatus Ellin6076	YP_828774	A1	558
Solibacter usitatus Ellin6076	YP_821793	A2	541
Solibacter usitatus Ellin6076	YP_824765	C	465
Solibacter usitatus Ellin6076	YP_824194	NOR	449
Solibacter usitatus Ellin6076	YP_824206	NOR	752
Rhodococcus erythropolis PR4	YP_345557	A1	583
Anaplasma marginale str. St. Maries	YP_154146	A1	521
Anaplasma marginale str. St. Maries	YP_154212	A1	434
Helicobacter hepaticus ATCC 51449	NP_860787	C	487
Geobacillus kaustophilus HTA426	YP_146936	A1	623
Geobacillus kaustophilus HTA426	YP_149311	A1	648
Geobacillus kaustophilus HTA426	YP_147524	B	387
Geobacillus kaustophilus HTA426	YP_147399	B	549

Geobacillus kaustophilus HTA426	YP_146607	C	413
Geobacillus kaustophilus HTA426	YP_146611	NOR	790
Alcanivorax sp. DG881	EDX89574	A1	666
Alcanivorax sp. DG881	EDX89879	A1	535
Alcanivorax sp. DG881	EDX89384	C	476
Erythrobacter sp. NAP1	ZP_01041429	C	556
Verrucomicrobium spinosum DSM 4136	ZP_02927147	A1	617
Verrucomicrobium spinosum DSM 4136	ZP_02925765	C	800
Anabaena variabilis ATCC 29413	YP_320968	A2	559
Anabaena variabilis ATCC 29413	YP_324793	A2	555
Anabaena variabilis ATCC 29413	YP_324077	A2	553
Anabaena variabilis ATCC 29413	YP_321048	A2	575
Anabaena variabilis ATCC 29413	YP_320079	A2	559
Neisseria gonorrhoeae FA 1090	YP_208434	C	481
Neisseria gonorrhoeae FA 1090	YP_208344	NOR	703
Rhodopirellula baltica SH 1	NP_866086	A1	833
Rhodopirellula baltica SH 1	NP_870646	A2	580
Rhodopirellula baltica SH 1	NP_867199	C	832
Burkholderia mallei ATCC 23344	YP_105022	A1	668
Burkholderia mallei ATCC 23344	YP_104673	A1	535
Burkholderia mallei ATCC 23344	YP_105512	NOR	772
Burkholderia mallei ATCC 23344	YP_102415	NOR	762
Deinococcus radiodurans R1	NP_296339	A2	820
Geobacter sulfurreducens PCA	NP_951280	A2	541
Methylococcus capsulatus str. Bath	YP_113374	A1	541
Methylococcus capsulatus str. Bath	YP_114811	B	537
Methylococcus capsulatus str. Bath	YP_114815	NOR	481
Mycobacterium avium 104	YP_883070	A1	543
Mycobacterium avium 104	YP_883164	NOR	747
Photorhabdus luminescens subsp. laumondii TTO1	NP_931081	A1	663
Vibrio cholerae O1 biovar El Tor str. N16961	NP_231085	C	477
Chromobacterium violaceum ATCC 12472	NP_900270	A1	545
Chromobacterium violaceum ATCC 12472	NP_903665	A1	680
Chromobacterium violaceum ATCC 12472	NP_900844	C	474
Chromobacterium violaceum ATCC 12472	NP_903164	NOR	755
Mycobacterium smegmatis str. MC2 155	YP_888712	A1	583
Myxococcus xanthus DK 1622	YP_632048	A2	545
Myxococcus xanthus DK 1622	YP_634222	A2	556
Myxococcus xanthus DK 1622	YP_633680	C	793
Ruegeria pomeroyi DSS-3	YP_166625	A1	552
Ruegeria pomeroyi DSS-3	YP_168721	C	535
Ruegeria pomeroyi DSS-3	YP_165019	C	540
Ruegeria pomeroyi DSS-3	YP_165045	NOR	455
Nocardia farcinica IFM 10152	YP_120507	A1	581
Nocardia farcinica IFM 10152	YP_117403	A1	588
marine gamma proteobacterium HTCC2143	ZP_01617056	A1	589
marine gamma proteobacterium HTCC2143	ZP_01615637	A1	519
marine gamma proteobacterium HTCC2143	ZP_01616294	C	515
Gloeobacter violaceus PCC 7421	NP_923686	A2	554
Gloeobacter violaceus PCC 7421	NP_925109	A2	560
Oceanicola batsensis HTCC2597	ZP_01001558	A1	854
Oceanicola batsensis HTCC2597	ZP_00997928	A1	556
Oceanicola batsensis HTCC2597	ZP_01000584	C	541
Ehrlichia ruminantium str. Welgevonden	YP_180637	A1	518
Ehrlichia ruminantium str. Welgevonden	YP_197694	A1	518
Ehrlichia ruminantium str. Welgevonden	YP_180694	A1	433
Ehrlichia ruminantium str. Welgevonden	YP_197756	A1	433
Corynebacterium diphtheriae NCTC 13129	NP_940201	A1	564
Corynebacterium diphtheriae NCTC 13129	NP_940552	NOR	763
Bordetella bronchiseptica RB50	NP_887830	A1	660
Bordetella bronchiseptica RB50	NP_891206	A1	531
Bordetella bronchiseptica RB50	NP_891363	A1	535
Bordetella bronchiseptica RB50	NP_887855	A1	673
Bordetella bronchiseptica RB50	NP_889864	C	495
Bordetella parapertussis 12822	NP_886335	A1	531
Bordetella parapertussis 12822	NP_883388	A1	660
Bordetella parapertussis 12822	NP_886371	A1	535
Bordetella parapertussis 12822	NP_883412	A1	673
Bordetella parapertussis 12822	NP_884049	C	523
Bordetella pertussis Tohama I	NP_881514	A1	659
Bordetella pertussis Tohama I	NP_882246	A1	535
Bordetella pertussis Tohama I	NP_881338	A1	666
Rhodopseudomonas palustris CGA009	NP_946185	A1	541
Rhodopseudomonas palustris CGA009	NP_945375	C	548
Rhodopseudomonas palustris CGA009	NP_946806	NOR	449
Psychrobacter arcticus 273-4	YP_264065	C	493
Bacillus anthracis str. Sterne	YP_026945	A1	647
Bacillus anthracis str. Sterne	YP_030105	A1	620
Bacillus anthracis str. 'Ames Ancestor'	YP_017332	A1	644
Bacillus anthracis str. 'Ames Ancestor'	YP_020800	A1	611
Mycobacterium avium subsp. paratuberculosis K-10	NP_962025	A1	581
Mycobacterium avium subsp. paratuberculosis K-10	NP_962115	NOR	551

delta proteobacterium MLMS-1	ZP_01287602	C	466
Exiguobacterium sibiricum 255-15	YP_001814456	A1	627
Exiguobacterium sibiricum 255-15	YP_001813096	B	563
Brucella abortus bv. 1 str. 9-941	YP_221246	A1	552
Brucella abortus bv. 1 str. 9-941	YP_221148	C	541
Brucella abortus bv. 1 str. 9-941	YP_223677	NOR	449
Thermus thermophilus HB27	YP_005640	A2	791
Thermus thermophilus HB27	YP_004743	B	562
Picrophilus torridus DSM 9790	YP_024082	A1	841
Ralstonia eutropha JMP134	YP_294541	A1	534
Ralstonia eutropha JMP134	YP_295206	A1	657
Ralstonia eutropha JMP134	YP_297835	A1	585
Ralstonia eutropha JMP134	YP_298772	A1	659
Ralstonia eutropha JMP134	YP_299083	A1	667
Ralstonia eutropha JMP134	YP_298631	A2	545
Ralstonia eutropha JMP134	YP_296249	C	482
Ralstonia eutropha JMP134	YP_299247	NOR	761
Candidatus Protochlamydia amoebophila UWE25	YP_008189	A1	653
Candidatus Protochlamydia amoebophila UWE25	YP_008977	C	471
Bdellovibrio bacteriovorus HD100	NP_967290	A2	535
Bdellovibrio bacteriovorus HD100	NP_969411	C	706
Bdellovibrio bacteriovorus HD100	NP_969401	NOR	441
Pseudomonas syringae pv. phaseolicola 1448A	YP_273477	A1	670
Pseudomonas syringae pv. phaseolicola 1448A	YP_275495	C	480
Methylobacillus flagellatus KT	YP_545403	A1	659
Methylobacillus flagellatus KT	YP_545557	A1	631
Methylobacillus flagellatus KT	YP_545543	C	494
Methylobacillus flagellatus KT	YP_544740	C	485
Listeria monocytogenes str. 4b F2365	YP_012629	A1	659
Ralstonia metallidurans CH34	YP_583104	A1	657
Ralstonia metallidurans CH34	YP_587920	A1	658
Ralstonia metallidurans CH34	YP_587737	A1	667
Ralstonia metallidurans CH34	YP_582417	A1	536
Ralstonia metallidurans CH34	YP_586093	A1	585
Ralstonia metallidurans CH34	YP_584521	C	518
Ralstonia metallidurans CH34	YP_586456	C	529
Ralstonia metallidurans CH34	YP_584192	C	482
Ralstonia metallidurans CH34	YP_585308	NOR	759
Burkholderia xenovorans LB400	YP_553679	A1	662
Burkholderia xenovorans LB400	YP_555443	A1	877
Burkholderia xenovorans LB400	YP_556882	A1	537
Burkholderia xenovorans LB400	YP_556973	B	541
Burkholderia xenovorans LB400	YP_558837	B	679
Burkholderia xenovorans LB400	YP_555356	B	680
Mesorhizobium sp. BNC1	YP_675332	A1	845
Mesorhizobium sp. BNC1	YP_676071	A1	676
Mesorhizobium sp. BNC1	YP_673307	A1	550
Mesorhizobium sp. BNC1	YP_675024	A1	591
Mesorhizobium sp. BNC1	YP_674805	C	547
Sinorhizobium meliloti 1021	NP_386623	A1	593
Sinorhizobium meliloti 1021	NP_437856	A1	669
Sinorhizobium meliloti 1021	NP_385011	A1	562
Sinorhizobium meliloti 1021	NP_437488	A1	841
Sinorhizobium meliloti 1021	NP_435656	C	539
Sinorhizobium meliloti 1021	NP_435566	C	551
Sinorhizobium meliloti 1021	NP_435940	NOR	448
Mesorhizobium loti MAFF303099	NP_109481	A1	623
Mesorhizobium loti MAFF303099	NP_102720	A1	589
Mesorhizobium loti MAFF303099	NP_107796	A1	550
Mesorhizobium loti MAFF303099	NP_107274	A1	665
Mesorhizobium loti MAFF303099	NP_106928	C	536
Mesorhizobium loti MAFF303099	NP_107099	C	536
Kineococcus radiotolerans SRS30216	YP_001362986	A1	588
Ralstonia solanacearum GMI1000	NP_523124	A1	586
Ralstonia solanacearum GMI1000	NP_521037	A1	659
Ralstonia solanacearum GMI1000	NP_518484	A1	534
Ralstonia solanacearum GMI1000	NP_519980	A1	658
Ralstonia solanacearum GMI1000	NP_519397	C	485
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	YP_000200	A2	534
Propionibacterium acnes KPA171202	YP_055417	A1	572
Propionibacterium acnes KPA171202	YP_056649	NOR	748
Synechococcus elongatus PCC 6301	YP_172217	A2	541
Synechococcus elongatus PCC 6301	YP_172019	C	439
Burkholderia vietnamiensis G4	YP_001120772	A1	535
Burkholderia vietnamiensis G4	YP_001119810	A1	667
Burkholderia vietnamiensis G4	YP_001120648	A1	670
Burkholderia vietnamiensis G4	YP_001115325	B	541
Burkholderia vietnamiensis G4	YP_001115017	B	879
Burkholderia sp. 383	YP_367049	A1	586
Burkholderia sp. 383	YP_369613	A1	668
Burkholderia sp. 383	YP_370413	A1	535
Burkholderia sp. 383	YP_372161	B	541
Ehrlichia canis str. Jake	YP_303439	A1	518
Ehrlichia canis str. Jake	YP_303497	B	436
Rhodospirillum rubrum ATCC 11170	YP_428413	C	496
Cytophaga hutchinsonii ATCC 33406	YP_678811	A1	609
Cytophaga hutchinsonii ATCC 33406	YP_677753	C	712
Geobacter metallireducens GS-15	YP_383219	A2	541

Geobacter metallireducens GS-15	YP_386423	NOR	756
Thermobifida fusca YX	YP_288942	A1	560
Thermobifida fusca YX	YP_289077	A1	563
Thermobifida fusca YX	YP_289753	A1	570
Bacillus cereus G9241	ZP_00236857	A1	611
Bacillus cereus G9241	ZP_00239526	A1	647
Marinobacter sp. ELB17	ZP_01736607	A1	535
Marinobacter sp. ELB17	ZP_01737388	C	538
Marinobacter sp. ELB17	ZP_01738263	C	456
Marinobacter sp. ELB17	ZP_01736782	NOR	471
Burkholderia thailandensis E264	YP_438680	A1	668
Burkholderia thailandensis E264	YP_440985	A1	535
Burkholderia thailandensis E264	YP_442319	A1	670
Burkholderia thailandensis E264	YP_443386	B	534
Burkholderia thailandensis E264	YP_439812	C	535
Burkholderia thailandensis E264	YP_442345	NOR	762
Burkholderia thailandensis E264	YP_439142	NOR	773
Aeropyrum pernix K1	NP_147500	A1	815
Aeropyrum pernix K1	NP_148062	B	548
Bacillus halodurans C-125	NP_243480	A1	620
Bacillus halodurans C-125	NP_242931	A1	647
Bacillus halodurans C-125	NP_241605	B	564
Burkholderia pseudomallei K96243	YP_107079	A1	535
Burkholderia pseudomallei K96243	YP_108972	A1	668
Burkholderia pseudomallei K96243	YP_111902	A1	670
Burkholderia pseudomallei K96243	YP_110109	A1	531
Burkholderia pseudomallei K96243	YP_107348	B	941
Burkholderia pseudomallei K96243	YP_107881	B	541
Burkholderia pseudomallei K96243	YP_108946	NOR	762
Desulfotobacterium hafniense DCB-2	YP_002458720	NOR	442
Gluconacetobacter diazotrophicus PAI 5	YP_001602279	A1	664
Gluconacetobacter diazotrophicus PAI 5	YP_001602346	A1	470
Gluconacetobacter diazotrophicus PAI 5	YP_002274731	A1	454
Gluconacetobacter diazotrophicus PAI 5	YP_002274670	A1	664
Gluconacetobacter diazotrophicus PAI 5	YP_001602839	A2	540
Gluconacetobacter diazotrophicus PAI 5	YP_002275232	A2	540
Haloarcula marismortui ATCC 43049	YP_135564	A1	760
Haloarcula marismortui ATCC 43049	YP_135922	A1	594
Haloarcula marismortui ATCC 43049	YP_136841	A1	854
43049			
Haloarcula marismortui ATCC 43049	YP_135811	B	562
Haloarcula marismortui ATCC 43049	YP_136806	NOR	760
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	YP_001334075	A1	663
Legionella pneumophila subsp. pneumophila str. Philadelphia 1	YP_096890	A1	538
Legionella pneumophila subsp. pneumophila str. Philadelphia 1	YP_096248	NOR	802
Listeria innocua Clip11262	NP_469361	A1	659
Magnetospirillum magnetotacticum MS-1	ZP_00050953	A1	542
Magnetospirillum magnetotacticum MS-1	ZP_00054051	A1	534
Magnetospirillum magnetotacticum MS-1	ZP_00055725	C	484
Mycobacterium leprae TN	NP_302190	A1	574
Neisseria meningitidis FAM18	YP_975604	C	481
Neisseria meningitidis FAM18	YP_975519	NOR	750
Rhodobacter sphaeroides 2.4.1	YP_351928	A1	566
Rhodobacter sphaeroides 2.4.1	YP_354614	A1	825
Rhodobacter sphaeroides 2.4.1	YP_353192	A1	857
Rhodobacter sphaeroides 2.4.1	YP_353773	C	535
Rhodobacter sphaeroides 2.4.1	YP_353400	NOR	446
Rickettsia conorii str. Malish 7	NP_360190	A1	532
Rickettsia conorii str. Malish 7	NP_360163	C	507
Rickettsia prowazekii str. Madrid E	NP_220786	A1	534
Rickettsia prowazekii str. Madrid E	NP_220763	B	510
Rickettsia sibirica 246	ZP_00142009	A1	532
Salmonella enterica subsp. enterica serovar Paratyphi B str. SPB7	YP_001589333	A1	663
Staphylococcus aureus RF122	YP_416411	A1	662
Sulfolobus solfataricus P2	NP_344288	A1	790
Sulfolobus solfataricus P2	NP_341619	B	593
Sulfolobus solfataricus P2	NP_343985	B	511
Sulfolobus solfataricus P2	NP_342994	NOR	534
Sulfolobus tokodaii str. 7	NP_375951	A1	815
Sulfolobus tokodaii str. 7	NP_375983	B	524
Sulfolobus tokodaii str. 7	NP_376500	B	569
Sulfolobus tokodaii str. 7	NP_376567	B	522
Sulfolobus tokodaii str. 7	NP_378599	B	590
Sulfolobus tokodaii str. 7	NP_378042	B	593
Sulfolobus tokodaii str. 7	NP_378396	B	524
Wolinella succinogenes DSM 1740	NP_906440	C	488
Yersinia pseudotuberculosis IP 32953	YP_069492	A1	663
Opitutaceae bacterium TAV2	ZP_02013954	C	505
Opitutaceae bacterium TAV2	ZP_02011608	C	796
Streptomyces ambofaciens ATCC 23877	CAJ89853	A1	570
Bacillus licheniformis ATCC	YP_081116	A1	649

14580				Anaeromyxobacter dehalogenans 2CP-C	YP_464382	C	473
Bacillus licheniformis ATCC 14580	YP_078884	A1	622	Anaeromyxobacter dehalogenans 2CP-C	YP_466923	NOR	763
Bacillus licheniformis ATCC 14580	YP_091296	A1	622	Anaeromyxobacter dehalogenans 2CP-C	YP_464614	NOR	762
Bacillus licheniformis ATCC 14580	YP_093545	A1	649	Anaeromyxobacter dehalogenans 2CP-C	YP_466383	NOR	769
Bacillus licheniformis ATCC 14580	YP_091667	NOR	784	Chromohalobacter salexigens DSM 3043	YP_573011	A1	659
Bacillus licheniformis ATCC 14580	YP_079248	NOR	784	Arthrobacter sp. FB24	YP_831696	A1	575
Novosphingobium aromaticivorans DSM 12444	YP_496193	A1	562	Jannaschia sp. CCS1	YP_509192	A1	628
Novosphingobium aromaticivorans DSM 12444	YP_497848	C	548	Jannaschia sp. CCS1	YP_511799	C	532
Staphylococcus haemolyticus JCSC1435	YP_253817	A1	662	Gluconobacter oxydans 621H	YP_192257	A1	388
Leifsonia xyli subsp. xyli str. CTCB07	YP_062417	A1	557	Gluconobacter oxydans 621H	YP_192306	A1	662
Bacillus thuringiensis serovar konkukian str. 97-27	YP_034958	A1	647	Candidatus Blochmannia pennsylvanicus str. BPEN	YP_277755	A1	653
Bacillus thuringiensis serovar konkukian str. 97-27	YP_038005	A1	620	Xanthomonas oryzae pv. oryzae KACC10331	YP_200269	A1	673
Staphylococcus aureus subsp. aureus MRSA252	YP_040448	A1	662	Xanthomonas oryzae pv. oryzae KACC10331	YP_202779	A1	534
Staphylococcus aureus subsp. aureus MRSA252	YP_039724	NOR	763	Ruegeria sp. TM1040	YP_614285	A1	552
Staphylococcus aureus subsp. aureus MSSA476	YP_043120	A1	662	Ruegeria sp. TM1040	YP_614279	A1	678
Bartonella quintana str. Toulouse	YP_031835	A1	672	Ruegeria sp. TM1040	YP_614539	C	534
Bartonella henselae str. Houston-1	YP_032982	A1	672	Thiobacillus denitrificans ATCC 25259	YP_314084	A1	526
Idiomarina loihiensis L2TR	YP_154651	A1	528	Thiobacillus denitrificans ATCC 25259	YP_314096	C	484
Idiomarina loihiensis L2TR	YP_155686	C	478	Thiobacillus denitrificans ATCC 25259	YP_314401	C	473
Idiomarina loihiensis L2TR	YP_154578	NOR	746	Thiobacillus denitrificans ATCC 25259	YP_314581	NOR	456
Aurantimonas sp. SI85-9A1	ZP_01228023	A1	533	Thiobacillus denitrificans ATCC 25259	YP_314319	NOR	467
Aurantimonas sp. SI85-9A1	ZP_01228584	A1	853	Symbiobacterium thermophilum IAM 14863	YP_075926	A1	628
Aurantimonas sp. SI85-9A1	ZP_01225850	B	647	Symbiobacterium thermophilum IAM 14863	YP_074758	B	436
Aurantimonas sp. SI85-9A1	ZP_01225540	C	547	Symbiobacterium thermophilum IAM 14863	YP_076979	C	467
Bradyrhizobium sp. BTAi1	YP_001236997	A1	537	Wolbachia endosymbiont strain TRS of Brugia malayi	YP_198138	A1	516
Bradyrhizobium sp. BTAi1	YP_001238127	A1	669	Rickettsia canadensis str. McKiel	YP_001492392	A1	532
Bradyrhizobium sp. BTAi1	YP_001239191	A1	666	Rickettsia akari str. Hartford	YP_001493401	A1	540
Bradyrhizobium sp. BTAi1	YP_001238831	C	539	Rickettsia akari str. Hartford	YP_001493371	B	516
Bradyrhizobium sp. BTAi1	YP_001238182	NOR	448	Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150	YP_151477	A1	663
Bacillus cereus E33L	YP_082219	A1	647	Polaromonas sp. JS666	YP_547581	A1	588
Bacillus cereus E33L	YP_085284	A1	620	Polaromonas sp. JS666	YP_548095	A1	544
Renibacterium salmoninarum ATCC 33209	YP_001624733	A1	585	Polaromonas sp. JS666	YP_551095	C	477
Chlorobium limicola DSM 245	YP_001943824	C	461	Polaromonas sp. JS666	YP_548116	C	547
Citrobacter koseri ATCC BAA-895	YP_001454272	A1	663	Legionella pneumophila str. Lens	YP_128137	A1	538
Cronobacter sakazakii ATCC BAA-894	YP_001438936	A1	663	Legionella pneumophila str. Lens	YP_127497	NOR	780
Arthrobacter aurescens TC1	YP_947955	A1	574	Legionella pneumophila str. Paris	YP_122635	A1	664
Anaeromyxobacter dehalogenans 2CP-C	YP_465481	A2	554	Legionella pneumophila str. Paris	YP_125263	A1	538
Anaeromyxobacter dehalogenans 2CP-C	YP_464016	A2	554				

Legionella pneumophila str. Paris	YP_124501	NOR	780
Photobacterium profundum SS9	YP_128409	A1	554
Photobacterium profundum SS9	YP_132060	A1	600
Photobacterium profundum SS9	YP_130037	C	475
Photobacterium profundum SS9	YP_132726	NOR	462
Frankia sp. EAN1pec	YP_001505257	A1	561
Frankia sp. EAN1pec	YP_001506146	A1	568
Shigella boydii Sb227	YP_406866	A1	663
Shigella sonnei Ss046	YP_309425	A1	663
Thermus thermophilus HB8	YP_143578	A2	791
Thermus thermophilus HB8	YP_144401	B	562
Ehrlichia ruminantium str. Gardel	YP_196734	A1	518
Ehrlichia ruminantium str. Gardel	YP_196797	B	433
Campylobacter coli RM2228	ZP_00368201	C	488
Campylobacter lari RM2100	ZP_00368853	C	488
Campylobacter upsaliensis RM3195	ZP_00370506	C	488
Corynebacterium jeikeium K411	YP_250256	A1	565
Salinibacter ruber DSM 13855	YP_446205	A2	581
Salinibacter ruber DSM 13855	YP_444460	B	557
Salinibacter ruber DSM 13855	YP_444468	C	504
Polynucleobacter necessarius subsp. asymbioticus QLW-PIDMWA-1	YP_001156715	A1	543
Polynucleobacter necessarius subsp. asymbioticus QLW-PIDMWA-1	YP_001155237	C	480
marine actinobacterium PHSC20C1	ZP_01131187	A1	559
Vibrio fischeri ES114	YP_204682	C	476
Dokdonia donghaensis MED134	ZP_01050582	A1	607
Dokdonia donghaensis MED134	ZP_01050474	C	734
Polaribacter irgensii 23-P	ZP_01117638	A1	593
Psychroflexus torquis ATCC 700755	ZP_01253939	A1	606
Psychroflexus torquis ATCC 700755	ZP_01254673	C	734
Robiginitalea biformata HTCC2501	ZP_01119437	A1	572
Robiginitalea biformata HTCC2501	ZP_01119531	C	733
Polaribacter sp. MED152	ZP_01053676	A1	593
Polaribacter sp. MED152	ZP_01052258	C	731
Flavobacteriales bacterium HTCC2170	ZP_01106280	NOR	446
Microscilla marina ATCC 23134	ZP_01690072	A1	635
Microscilla marina ATCC 23134	ZP_01690773	C	715
Lyngbya sp. PCC 8106	ZP_01623743	A2	564
Lyngbya sp. PCC 8106	ZP_01621649	A2	574
Nodularia spumigena CCY 9414	ZP_01628240	A2	502
Nodularia spumigena CCY 9414	ZP_01628371	A2	549
Nodularia spumigena CCY 9414	ZP_01628175	A2	564
Lentisphaera araneosa HTCC2155	ZP_01877020	A2	539
Erythrobacter litoralis HTCC2594	YP_457435	A1	567
Blastopirellula marina DSM 3645	ZP_01094316	A1	554
Blastopirellula marina DSM 3645	ZP_01091121	A2	595
Blastopirellula marina DSM 3645	ZP_01093641	A2	566
Fulvimarina pelagi HTCC2506	ZP_01440869	A1	667
Fulvimarina pelagi HTCC2506	ZP_01438693	A1	875
Fulvimarina pelagi HTCC2506	ZP_01440921	A1	533
Loktanella vestfoldensis SKA53	ZP_01004262	A1	552
Loktanella vestfoldensis SKA53	ZP_01004045	C	537
Nitrobacter sp. Nb-311A	ZP_01045192	A1	538
Nitrobacter sp. Nb-311A	ZP_01048077	B	657
Oceanicaulis alexandrii HTCC2633	ZP_00953043	A1	538
Oceanicaulis alexandrii HTCC2633	ZP_00953505	C	552
Oceanicola granulosus HTCC2516	ZP_01155740	A1	818
Oceanicola granulosus HTCC2516	ZP_01158274	A1	864
Oceanicola granulosus HTCC2516	ZP_01157859	A1	557
Parvularcula bermudensis HTCC2503	ZP_01016500	A1	558
Candidatus Pelagibacter ubique HTCC1002	ZP_01264700	A1	544
Roseovarius sp. 217	ZP_01038050	NOR	433
Roseovarius sp. HTCC2601	ZP_01446254	A1	461
Roseovarius sp. HTCC2601	ZP_01445618	A1	557
Roseovarius sp. HTCC2601	ZP_01445338	A1	852
Roseovarius sp. HTCC2601	ZP_01444008	C	537
Sphingomonas sp. SKA58	ZP_01301620	A1	687
Sphingomonas sp. SKA58	ZP_01302641	A1	556
Sphingomonas sp. SKA58	ZP_01301694	A1	840
Sphingomonas sp. SKA58	ZP_01305192	A1	657
Sphingomonas sp. SKA58	ZP_01302542	C	552
Rhodobacterales bacterium HTCC2654	ZP_01012401	NOR	448
Alteromonas macleodii 'Deep ecotype'	YP_002125779	A1	835
Alteromonas macleodii 'Deep ecotype'	YP_002128296	A1	537
Alteromonas macleodii 'Deep ecotype'	ZP_01108658	A1	537
Alteromonas macleodii 'Deep ecotype'	ZP_01109066	A1	835
Alteromonas macleodii 'Deep ecotype'	YP_002128195	A1	662
Alteromonas macleodii 'Deep ecotype'	YP_002126350	C	477
Idiomarina baltica OS145	ZP_01041859	A1	528
Idiomarina baltica OS145	ZP_01043949	C	478
Marinomonas sp. MED121	ZP_01077262	A1	522
Marinomonas sp. MED121	ZP_01075802	C	484
Nitrococcus mobilis Nb-231	ZP_01127177	A1	536
Photobacterium profundum 3TCK	ZP_01217907	A1	679
Photobacterium profundum 3TCK	ZP_01219385	A1	600

Photobacterium profundum 3TCK	ZP_01220712	A1	554
Photobacterium profundum 3TCK	ZP_01223000	C	475
Photobacterium profundum 3TCK	ZP_01219400	C	470
Photobacterium profundum 3TCK	ZP_01220308	NOR	462
Psychromonas sp. CNPT3	ZP_01215617	C	477
Reinekea sp. MED297	ZP_01114101	A1	528
Reinekea sp. MED297	ZP_01113912	C	478
Congregibacter litoralis KT71	ZP_01103079	A1	523
Congregibacter litoralis KT71	ZP_01101285	C	478
Congregibacter litoralis KT71	ZP_01104031	C	552
gamma proteobacterium HTCC2207	ZP_01223355	A1	524
Vibrio alginolyticus 12G01	ZP_01262713	A1	543
Vibrio alginolyticus 12G01	ZP_01260779	A1	678
Vibrio alginolyticus 12G01	ZP_01258960	C	477
Vibrio campbellii AND4	ZP_02194303	A1	678
Vibrio campbellii AND4	ZP_02197718	A1	546
Vibrio campbellii AND4	ZP_02194286	C	469
Vibrio campbellii AND4	ZP_02194798	C	475
Vibrio sp. MED222	ZP_01063507	C	474
Vibrio sp. MED222	ZP_01064301	C	477
Vibrio splendidus 12B01	ZP_00990597	A1	553
Vibrio splendidus 12B01	ZP_00989662	A1	695
Vibrio splendidus 12B01	ZP_00988490	C	477
Vibrio splendidus 12B01	ZP_00988835	C	474
Vibrio angustum S14	ZP_01237208	A1	539
Vibrio angustum S14	ZP_01235351	A1	696
Vibrio angustum S14	ZP_01234682	C	477
Mariprofundus ferrooxydans PV-1	ZP_01453733	C	488
Xanthomonas campestris pv. campestris str. 8004	YP_244149	A1	666
Xanthomonas campestris pv. campestris str. 8004	YP_244964	A1	534
beta proteobacterium KB13	EDZ65002	A1	529
Shewanella benthica KT99	ZP_02158631	A1	584
Shewanella benthica KT99	ZP_02156063	A1	541
Shewanella benthica KT99	ZP_02156284	C	474
Rickettsia felis URRWXC2	YP_246643	A1	532
Rickettsia felis URRWXC2	YP_246614	B	526
Bacillus weihenstephanensis KBAB4	YP_001646566	A1	620
Bacillus weihenstephanensis KBAB4	YP_001643510	A1	644
Bacillus cereus subsp. cytotoxis NVH 391-98	YP_001373933	A1	644
Bacillus cereus subsp. cytotoxis NVH 391-98	YP_001375874	A1	620
Bacillus pumilus SAFR-032	YP_001486625	A1	623
Bacillus pumilus SAFR-032	YP_001488673	A1	652
Rhodopseudomonas palustris BisA53	YP_782773	A1	841
Rhodopseudomonas palustris BisA53	YP_783651	A1	546
Rhodopseudomonas palustris BisA53	YP_778959	C	548
Rhodopseudomonas palustris BisA53	YP_779559	NOR	448
Rhodopseudomonas palustris BisB18	YP_533702	A1	840
Rhodopseudomonas palustris BisB18	YP_534629	A1	540
Rhodopseudomonas palustris BisB18	YP_529913	C	550
Rhodopseudomonas palustris BisB5	YP_567735	A1	669
Rhodopseudomonas palustris BisB5	YP_567954	A1	541
Rhodopseudomonas palustris BisB5	YP_567869	C	548
Rhodopseudomonas palustris HaA2	YP_483778	A1	841
Rhodopseudomonas palustris HaA2	YP_483863	A1	669
Rhodopseudomonas palustris HaA2	YP_488181	A1	541
Rhodopseudomonas palustris HaA2	YP_483640	C	548
Xanthomonas campestris pv. vesicatoria str. 85-10	YP_363039	A1	666
Xanthomonas campestris pv. vesicatoria str. 85-10	YP_365736	A1	534
Herpetosiphon aurantiacus ATCC 23779	YP_001542885	A1	615
Herpetosiphon aurantiacus ATCC 23779	YP_001544075	C	420
Aliivibrio salmonicida LF11238	YP_002262995	C	476
Synechococcus sp. RCC307	YP_001226927	A2	556
Synechococcus sp. RCC307	YP_001228191	A2	468
Synechococcus sp. CC9902	YP_377756	A2	556
Escherichia coli str. K-12 substr. DH10B	YP_001729337	A1	663
Escherichia coli str. K-12 substr. W3110	AP_001081	A1	663
Thiomicrospira crunogena XCL-2	YP_392229	C	476
Sphingopyxis alaskensis RB2256	YP_616451	A1	556
Sphingopyxis alaskensis RB2256	YP_616553	A1	672
Sphingopyxis alaskensis RB2256	YP_616722	C	552
Shewanella denitrificans OS217	YP_564516	A1	535
Shewanella denitrificans OS217	YP_562858	C	477
Shewanella denitrificans OS217	YP_562986	NOR	471
Shewanella frigidimarina NCIMB 400	YP_748924	A1	660
Shewanella frigidimarina NCIMB 400	YP_748958	A1	531
Shewanella frigidimarina NCIMB 400	YP_750694	C	475
Shewanella frigidimarina NCIMB 400	YP_751928	NOR	762
Paracoccus denitrificans PD1222	YP_915727	A1	558

Paracoccus denitrificans PD1222	YP_916807	A1	556
Paracoccus denitrificans PD1222	YP_915641	C	539
Paracoccus denitrificans PD1222	ZP_00629589	C	539
Paracoccus denitrificans PD1222	YP_916266	NOR	462
Paracoccus denitrificans PD1222	ZP_00632126	NOR	462
Shewanella putrefaciens CN-32	YP_001185289	A1	530
Shewanella putrefaciens CN-32	YP_001185364	A1	659
Shewanella putrefaciens CN-32	YP_001183481	C	478
Shewanella putrefaciens CN-32	YP_001184584	NOR	761
Deinococcus geothermalis DSM 11300	YP_603878	A2	818
Deinococcus geothermalis DSM 11300	YP_603490	B	587
Burkholderia pseudomallei 1710b	YP_332087	A1	535
Burkholderia pseudomallei 1710b	YP_334218	A1	668
Burkholderia pseudomallei 1710b	YP_336150	A1	668
Burkholderia pseudomallei 1710b	YP_336750	A1	531
Burkholderia pseudomallei 1710b	YP_332357	B	950
Burkholderia pseudomallei 1710b	YP_332896	B	534
Burkholderia pseudomallei 1710b	YP_334186	NOR	762
Burkholderia pseudomallei 1710b	YP_335633	NOR	683
Burkholderia pseudomallei 668	YP_001059739	A1	668
Burkholderia pseudomallei 668	YP_001061143	A1	531
Burkholderia pseudomallei 668	YP_001063707	A1	668
Burkholderia pseudomallei 668	YP_001057543	A1	535
Burkholderia pseudomallei 668	YP_001057822	B	962
Burkholderia pseudomallei 668	YP_001058396	B	541
Burkholderia pseudomallei 668	YP_001059704	NOR	762
Burkholderia pseudomallei 668	YP_001063055	NOR	773
Burkholderia mallei SAVP1	YP_990966	A1	668
Burkholderia mallei SAVP1	YP_991520	A1	535
Burkholderia mallei SAVP1	YP_993718	A1	560
Burkholderia mallei SAVP1	YP_990155	NOR	772
Burkholderia mallei SAVP1	YP_993685	NOR	762
Burkholderia mallei NCTC 10247	YP_001077431	A1	668
Burkholderia mallei NCTC 10247	YP_001082371	A1	535
Burkholderia mallei NCTC 10247	YP_001081265	A1	560
Burkholderia mallei NCTC 10247	YP_001078795	NOR	772
Burkholderia mallei NCTC 10247	YP_001081235	NOR	762
Burkholderia mallei GB8 horse 4	ZP_00440513	A1	664
bacterium Ellin514	ZP_02965874	A1	745
bacterium Ellin514	ZP_02964949	A2	551
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	YP_215471	A1	663
Synechococcus sp. JA-3-3Ab	YP_474942	A2	569
Synechococcus sp. JA-2-3B'a(2-13)	YP_478886	A2	568
Brevibacterium linens BL2	ZP_00378590	A1	565
Nitrobacter hamburgensis X14	YP_578651	A1	537
Nitrobacter hamburgensis X14	YP_575616	A1	537
Nitrobacter hamburgensis X14	YP_578831	A1	534
Nitrobacter hamburgensis X14	YP_571634	A2	557
Nitrobacter hamburgensis X14	YP_577950	B	484
Nitrobacter hamburgensis X14	YP_578641	B	658
Nitrobacter winogradskyi Nb-255	YP_317379	A1	539
Nitrobacter winogradskyi Nb-255	YP_316844	A1	539
Nitrobacter winogradskyi Nb-255	YP_318920	A1	839
Nitrobacter winogradskyi Nb-255	YP_317387	B	657
Nitrobacter winogradskyi Nb-255	YP_317176	C	402
Nitrosococcus oceani ATCC 19707	YP_345017	A1	532
Nitrosococcus oceani ATCC 19707	YP_343768	A1	853
Nitrosococcus oceani ATCC 19707	YP_343276	A2	535
Nitrosococcus oceani ATCC 19707	YP_344942	B	499
Nitrosococcus oceani ATCC 19707	YP_343846	NOR	448
Nitrosospira multiformis ATCC 25196	YP_410884	A1	526
Nitrosospira multiformis ATCC 25196	YP_412465	A1	529
Nitrosospira multiformis ATCC 25196	YP_413345	B	495
Nitrosospira multiformis ATCC 25196	YP_411159	B	671
Nitrosospira multiformis ATCC 25196	YP_411950	NOR	463
Shewanella loihica PV-4	YP_001092498	A1	662
Shewanella loihica PV-4	YP_001092255	A1	536
Shewanella loihica PV-4	YP_001094110	C	475
Shewanella loihica PV-4	YP_001092788	NOR	763
Paenibacillus sp. JDR-2	ZP_02845309	A1	627
Paenibacillus sp. JDR-2	ZP_02850431	A1	658
Chloroflexus aurantiacus J-10-fl	YP_001635743	A2	552
Chloroflexus aurantiacus J-10-fl	YP_001636024	B	559
Pelodictyon phaeoclathratiforme BU-1	YP_002018898	C	463
Shewanella baltica OS155	YP_001048556	A1	530
Shewanella baltica OS155	YP_001050566	C	478
Shewanella baltica OS155	YP_001041605	C	478
Shewanella amazonensis SB2B	YP_929356	A1	531
Shewanella amazonensis SB2B	YP_927666	C	478
Shewanella amazonensis SB2B	YP_926606	NOR	758
Sulfurimonas denitrificans DSM 1251	YP_392620	A2	545
Sulfurimonas denitrificans DSM 1251	YP_392597	C	497
Sulfurimonas denitrificans DSM 1251	YP_394493	NOR	474
Bacillus amyloliquefaciens FZB42	YP_001421070	A1	622
Bacillus amyloliquefaciens FZB42	YP_001423102	A1	649
Frankia alni ACN14a	YP_714342	A1	572
Frankia alni ACN14a	YP_715300	A1	557
Frankia alni ACN14a	YP_716216	A1	580

Chloroflexus aggregans DSM 9485	YP_002462860	A2	552
Chloroflexus aggregans DSM 9485	ZP_01514910	A2	552
Chloroflexus aggregans DSM 9485	YP_002464262	B	559
Chloroflexus aggregans DSM 9485	ZP_01516700	B	559
Pseudoalteromonas haloplanktis TAC125	YP_341349	A1	530
Pseudoalteromonas haloplanktis TAC125	YP_340728	A1	667
Pseudoalteromonas haloplanktis TAC125	YP_340357	C	477
Pseudoalteromonas haloplanktis TAC125	YP_340908	NOR	752
Acaryochloris marina MBIC11017	YP_001518912	A2	554
Acaryochloris marina MBIC11017	YP_001514849	NOR	764
Acaryochloris marina MBIC11017	YP_001520798	NOR	770
Sulfolobus acidocaldarius DSM 639	YP_256842	A1	788
Sulfolobus acidocaldarius DSM 639	YP_254815	B	588
Sulfolobus acidocaldarius DSM 639	YP_256675	B	508
Escherichia coli E24377A	YP_001461617	A1	663
Escherichia coli HS	YP_001457274	A1	663
Burkholderia cenocepacia AU 1054	YP_622106	A1	535
Burkholderia cenocepacia AU 1054	YP_625845	A1	668
Burkholderia cenocepacia AU 1054	YP_621172	B	879
Burkholderia cenocepacia HI2424	YP_835714	A1	668
Burkholderia cenocepacia HI2424	YP_836489	A1	535
Burkholderia cenocepacia HI2424	YP_840161	B	879
Chlorobium phaeobacteroides BS1	YP_001959988	C	468
alpha proteobacterium BAL199	ZP_02186257	A1	536
Sphingobium indicum	AAR05968	C	552
Orientia tsutsugamushi str. Ikeda	YP_001937888	A1	528
Lactobacillus fermentum IFO 3956	YP_001844459	NOR	780
Nitrosomonas eutropha C91	YP_748573	A1	525
Nitrosomonas eutropha C91	YP_746933	A1	703
Nitrosomonas eutropha C91	YP_748072	B	492
Nitrosomonas eutropha C91	YP_747791	C	552
Nitrosomonas eutropha C91	YP_746761	NOR	448
Psychrobacter cryohalolentis K5	YP_580051	C	493
Syntrophobacter fumaroxidans MPOB	YP_847042	NOR	758
Candidatus Pelagibacter ubique HTCC1062	YP_265561	A1	544
Rickettsia bellii RML369-C	YP_538061	A1	532
Rickettsia bellii RML369-C	YP_537972	B	528
Mycobacterium tuberculosis F11	YP_001288988	A1	573
Vibrio harveyi ATCC BAA-1116	YP_001447623	A1	678
Vibrio harveyi ATCC BAA-1116	YP_001448390	A1	546
Vibrio harveyi ATCC BAA-1116	YP_001447597	C	469
Vibrio harveyi ATCC BAA-1116	YP_001445492	C	477
Pelobacter carbinolicus DSM 2380	YP_357934	A2	543
Rhodoferrax ferrireducens T118	YP_522942	A1	539
Rhodoferrax ferrireducens T118	YP_525356	C	550
Rhodoferrax ferrireducens T118	YP_523187	C	478
Rhodoferrax ferrireducens T118	YP_523143	NOR	774
Burkholderia ambifaria AMMD	YP_778088	A1	586
Burkholderia ambifaria AMMD	YP_774791	A1	535
Burkholderia ambifaria AMMD	YP_773996	A1	668
Burkholderia ambifaria AMMD	YP_778451	B	882
Actinobacillus succinogenes 130Z	YP_001344891	NOR	753
Bacillus thuringiensis serovar israelensis ATCC 35646	ZP_00742474	A1	552
Bacillus thuringiensis serovar israelensis ATCC 35646	ZP_00741993	A1	620
Bordetella petrii DSM 12804	YP_001631310	A1	658
Bordetella petrii DSM 12804	YP_001628771	A1	535
Bordetella petrii DSM 12804	YP_001630971	A1	556
Bordetella petrii DSM 12804	YP_001632007	A1	667
Bordetella petrii DSM 12804	YP_001631825	A1	844
Bordetella petrii DSM 12804	YP_001631855	C	491
Bordetella petrii DSM 12804	YP_001632665	NOR	473
Bordetella petrii DSM 12804	YP_001632171	NOR	762
Pyrobaculum arsenaticum DSM 13514	YP_001152749	NOR	734
Chlorobium chlorochromatii CaD3	YP_379659	C	462
Corynebacterium glutamicum R	YP_001139341	A1	584
Magnetospirillum magneticum AMB-1	YP_421532	A1	532
Magnetospirillum magneticum AMB-1	YP_421585	B	560
Magnetospirillum magneticum AMB-1	YP_422307	NOR	439
Xanthomonas oryzae pv. oryzae MAFF 311018	YP_450545	A1	666
Xanthomonas oryzae pv. oryzae MAFF 311018	YP_452948	A1	534
Ralstonia solanacearum UW551	ZP_00944709	C	529
Ralstonia solanacearum UW551	ZP_00945695	C	485
Ralstonia solanacearum UW551	ZP_00946507	NOR	758
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	YP_301821	A1	662
Pseudoalteromonas atlantica T6c	YP_660871	A1	664
Pseudoalteromonas atlantica T6c	YP_663798	A1	532
Pseudoalteromonas atlantica T6c	YP_661645	C	479
Sodalis glossinidius str. 'morsitans'	YP_454346	A1	664
Shigella boydii CDC 3083-94	YP_001879141	A1	663
Planctomyces maris DSM 8797	ZP_01854217	A1	754
Planctomyces maris DSM 8797	ZP_01857356	A1	846
Planctomyces maris DSM 8797	ZP_01852349	A2	606
Planctomyces maris DSM 8797	ZP_01855645	A2	572
Vibrio cholerae O395	YP_001216999	C	477

Bacillus coagulans 36D1	ZP_01696322	A1	623	Mycobacterium vanbaalenii PYR-1	YP_954495	A1	581
Bacillus coagulans 36D1	ZP_01697439	A1	649	Mycobacterium vanbaalenii PYR-1	YP_952903	A1	581
Rickettsia africae ESF-5	ZP_02336393	A1	532	Mycobacterium vanbaalenii PYR-1	YP_956381	A1	584
Rhizobium etli CFN 42	YP_470115	A1	592	Burkholderia dolosa AUO158	YP_002098023	A1	668
Rhizobium etli CFN 42	YP_468491	A1	566	Burkholderia dolosa AUO158	YP_002097350	A1	535
Rhizobium etli CFN 42	YP_472195	A1	667	Burkholderia dolosa AUO158	YP_002101429	B	694
Rhizobium etli CFN 42	YP_471835	A1	637	Roseobacter sp. AzwK-3b	ZP_01904953	C	534
Rhizobium etli CFN 42	YP_468520	B	511	Marinobacter aquaeolei VT8	YP_957357	A1	527
Rhizobium etli CFN 42	YP_473132	NOR	446	Marinobacter aquaeolei VT8	YP_959053	C	476
Natronomonas pharaonis DSM 2160	YP_326879	A1	587	Marinobacter aquaeolei VT8	YP_960376	NOR	471
Natronomonas pharaonis DSM 2160	YP_330877	B	575	Geobacter uraniireducens Rf4	YP_001229197	A2	542
Natronomonas pharaonis DSM 2160	YP_327130	B	581	Acidothermus cellulolyticus 11B	YP_872716	A1	570
Natronomonas pharaonis DSM 2160	YP_327124	NOR	448	Acidothermus cellulolyticus 11B	YP_872249	A1	589
Rhodobacter sphaeroides ATCC 17029	YP_001042415	A1	566	Shewanella sp. W3-18-1	YP_961501	A1	659
Rhodobacter sphaeroides ATCC 17029	YP_001043635	A1	857	Shewanella sp. W3-18-1	YP_961539	A1	530
Rhodobacter sphaeroides ATCC 17029	YP_001045676	A1	825	Shewanella sp. W3-18-1	YP_963429	C	478
Rhodobacter sphaeroides ATCC 17029	YP_001044225	C	535	Shewanella sp. W3-18-1	YP_962277	NOR	761
Rhodobacter sphaeroides ATCC 17029	YP_001043844	NOR	446	Pseudomonas putida F1	YP_001265479	A1	529
Rhodobacter sphaeroides ATCC 17025	YP_001166875	A1	566	Pseudomonas putida F1	YP_001266182	A1	672
Rhodobacter sphaeroides ATCC 17025	YP_001167902	A1	878	Pseudomonas putida F1	YP_001266953	C	480
Rhodobacter sphaeroides ATCC 17025	YP_001166745	C	535	Pseudomonas putida F1	YP_001266958	C	474
Rhodobacter sphaeroides ATCC 17025	YP_001167177	NOR	446	Campylobacter jejuni subsp. jejuni 81-176	YP_001001138	C	488
Psychrobacter sp. PRwf-1	YP_001280634	C	496	Leptospira borgpetersenii serovar Hardjo-bovis L550	YP_796822	A2	534
Psychrobacter sp. PRwf-1	YP_001280413	NOR	759	Leptospira borgpetersenii serovar Hardjo-bovis L550	YP_798081	C	476
Acidiphilium cryptum JF-5	YP_001234229	A1	689	Leptospira borgpetersenii serovar Hardjo-bovis JB197	YP_801959	A2	534
Acidiphilium cryptum JF-5	YP_001235284	A1	532	Leptospira borgpetersenii serovar Hardjo-bovis JB197	YP_800805	C	476
Hahella chejuensis KCTC 2396	YP_431395	A1	526	Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ames)'	YP_001963663	A2	537
Hahella chejuensis KCTC 2396	YP_433526	C	481	Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ames)'	ABZ95888	C	476
Hahella chejuensis KCTC 2396	YP_435539	NOR	461	Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ames)'	YP_001964801	C	476
Hahella chejuensis KCTC 2396	YP_432036	NOR	756	Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ames)'	YP_001963907	NOR	449
Yersinia pestis Angola	YP_001607427	A1	663	Orientia tsutsugamushi str. Boryong	YP_001249069	A1	528
Yersinia pseudotuberculosis IP 31758	YP_001402059	A1	663	Burkholderia pseudomallei 1106a	YP_001076601	A1	670
Yersinia intermedia ATCC 29909	ZP_00834817	A1	663	Burkholderia pseudomallei 1106a	YP_001074155	A1	524
Yersinia frederiksenii ATCC 33641	ZP_00829851	A1	663	Burkholderia pseudomallei 1106a	YP_001067024	A1	668
Yersinia bercovieri ATCC 43970	ZP_00821016	A1	663	Burkholderia pseudomallei 1106a	YP_001064791	A1	535
Mycobacterium gilvum PYR-GCK	YP_001135534	A1	580	Burkholderia pseudomallei 1106a	YP_001065632	B	541
Mycobacterium gilvum PYR-GCK	YP_001136626	A1	581	Burkholderia pseudomallei 1106a	YP_001065064	B	974
Mycobacterium gilvum PYR-GCK	YP_001134083	A1	581	Burkholderia pseudomallei 1106a	YP_001066989	NOR	762
Mycobacterium gilvum PYR-GCK	YP_001132466	A1	592	Burkholderia pseudomallei 1106a	YP_001075999	NOR	518
				Helicobacter pylori HPAG1	YP_626883	C	488
				Psychromonas ingrahamii 37	YP_944025	A1	534
				Psychromonas ingrahamii 37	YP_942380	C	482

Roseiflexus sp. RS-1	YP_001276593	A1	641	Rhodobacterales bacterium HTCC2255	ZP_01450302	A1	532
Roseiflexus sp. RS-1	YP_001275293	B	561	Rhodobacterales bacterium HTCC2255	ZP_01449141	A1	557
Brucella melitensis biovar Abortus 2308	YP_413953	A1	552	Rhodobacterales bacterium HTCC2255	ZP_01449753	C	477
Brucella melitensis biovar Abortus 2308	YP_413854	C	541	Arcobacter butzleri RM4018	YP_001490968	C	489
Brucella melitensis biovar Abortus 2308	YP_419097	NOR	449	Arcobacter butzleri RM4018	YP_001489383	C	523
Staphylococcus aureus subsp. aureus JH9	YP_001246495	A1	662	Arcobacter butzleri RM4018	YP_001490279	NOR	744
Staphylococcus aureus subsp. aureus JH1	YP_001316283	A1	662	Staphylococcus aureus subsp. aureus USA300	YP_493660	A1	662
Xanthomonas oryzae pv. oryzae PXO99A	YP_001911827	A1	534	Salinispora tropica CNB-440	YP_001157943	A1	644
Xanthomonas oryzae pv. oryzae PXO99A	YP_001912706	A1	666	Salinispora tropica CNB-440	YP_001158972	A1	665
Bartonella bacilliformis KC583	YP_989523	A1	672	Buchnera aphidicola str. Cc (Cinara cedri)	YP_802843	A1	657
Yersinia pestis Antiqua	YP_652566	A1	663	Baumannia cicadellinica str. Hc (Homalodisca coagulata)	YP_588718	A1	665
Campylobacter concisus 13826	ABW74731	C	496	Neisseria meningitidis 053442	YP_001599744	C	452
Campylobacter concisus 13826	YP_001466413	NOR	743	Neisseria meningitidis 053442	YP_001599647	NOR	751
Campylobacter curvus 525.92	YP_001407573	C	494	Janthinobacterium sp. Marseille	YP_001354841	A1	533
Campylobacter curvus 525.92	YP_001407973	NOR	746	Janthinobacterium sp. Marseille	YP_001354329	A1	667
Campylobacter fetus subsp. fetus 82-40	YP_891580	C	489	Janthinobacterium sp. Marseille	YP_001353321	C	474
Campylobacter hominis ATCC BAA-381	YP_001406948	C	488	Roseobacter denitrificans OCh 114	YP_682261	A1	556
Campylobacter jejuni subsp. doylei 269.97	YP_001398790	C	488	Roseobacter denitrificans OCh 114	YP_681602	C	537
Coxiella burnetii RSA 331	YP_001596689	A1	668	Roseobacter denitrificans OCh 114	YP_681875	NOR	456
Bordetella avium 197N	YP_785316	A1	659	Flavobacterium johnsoniae UW101	YP_001193992	A1	600
Exiguobacterium sp. AT1b	ZP_02991613	A1	628	Flavobacterium johnsoniae UW101	YP_001194883	C	727
Exiguobacterium sp. AT1b	ZP_02992279	B	561	Flavobacterium johnsoniae UW101	YP_001194761	NOR	416
Mycobacterium ulcerans Agy99	YP_905820	A1	573	Flavobacterium johnsoniae UW101	YP_001194759	NOR	744
Mycobacterium ulcerans Agy99	YP_905358	A1	578	Chlorobium ferrooxidans DSM 13031	ZP_01385091	C	464
Escherichia coli 536	YP_668419	A1	663	Yersinia pestis Nepal516	YP_646858	A1	663
Haloquadratum walsbyi DSM 16790	YP_657381	A1	634	Kocuria rhizophila DC2201	YP_001855138	A1	573
Escherichia coli UTI89	YP_539487	A1	663	Pseudomonas stutzeri A1501	YP_001174628	A1	508
Polaromonas naphthalenivorans CJ2	YP_981117	A1	544	Pseudomonas stutzeri A1501	YP_001172366	A1	842
Polaromonas naphthalenivorans CJ2	YP_983999	C	478	Pseudomonas stutzeri A1501	YP_001172356	C	475
Polaromonas naphthalenivorans CJ2	YP_981564	NOR	758	Pseudomonas stutzeri A1501	YP_001172359	C	474
Sinorhizobium medicae WSM419	YP_001313492	A1	840	Pseudomonas stutzeri A1501	YP_001173991	NOR	474
Sinorhizobium medicae WSM419	YP_001326210	A1	562	Acidithiobacillus ferrooxidans ATCC 53993	YP_002219239	A1	716
Sinorhizobium medicae WSM419	YP_001328093	A1	592	Aeromonas hydrophila subsp. hydrophila ATCC 7966	YP_857449	A1	658
Sinorhizobium medicae WSM419	YP_001314880	A1	635	Aeromonas hydrophila subsp. hydrophila ATCC 7966	YP_856818	C	496
Sinorhizobium medicae WSM419	YP_001313108	A1	669	Hydrogenobaculum sp. Y04AAS1	YP_002121488	C	472
Sinorhizobium medicae WSM419	YP_001314560	C	539	Hydrogenobaculum sp. Y04AAS1	YP_002121019	NOR	748
Sinorhizobium medicae WSM419	YP_001314551	C	557	Ralstonia eutropha H16	YP_841573	A1	585
Sinorhizobium medicae WSM419	YP_001314819	NOR	448	Ralstonia eutropha H16	YP_726142	A1	658
Caulobacter sp. K31	YP_001686123	A1	555	Ralstonia eutropha H16	YP_725581	A1	658
Caulobacter sp. K31	YP_001684062	C	547	Ralstonia eutropha H16	YP_724862	A1	535
				Ralstonia eutropha H16	YP_729182	A1	667
				Ralstonia eutropha H16	YP_726783	C	482

Ralstonia eutropha H16	AAC45803	NOR	762	Sagittula stellata E-37	ZP_01746009	A1	692
Ralstonia eutropha H16	YP_841835	NOR	762	Sagittula stellata E-37	ZP_01746085	C	544
Pseudomonas aeruginosa PA7	YP_001349426	A1	658	Bacillus sp. B14905	ZP_01723962	A1	649
Pseudomonas aeruginosa PA7	YP_001345577	A1	530	Bacillus sp. B14905	ZP_01721998	A1	627
Pseudomonas aeruginosa PA7	YP_001346344	C	475	Bacillus sp. B14905	ZP_01725697	A1	647
Pseudomonas aeruginosa PA7	YP_001348796	C	480	Bacillus sp. B14905	ZP_01723349	B	553
Pseudomonas aeruginosa PA7	YP_001349127	C	475	Algoriphagus sp. PR1	ZP_01720126	A1	628
Pseudomonas aeruginosa PA7	YP_001349131	C	475	Algoriphagus sp. PR1	ZP_01720277	C	710
Pseudomonas aeruginosa PA7	YP_001346020	NOR	465	Roseobacter sp. SK209-2-6	ZP_01754686	NOR	447
Aeromonas salmonicida subsp. salmonicida A449	YP_001142711	A1	658	Pseudomonas putida W619	YP_001746998	A1	529
Aeromonas salmonicida subsp. salmonicida A449	YP_001141804	C	474	Pseudomonas putida W619	YP_001751224	A1	672
Verrucomicrobiae bacterium DG1235	EDY82944	A1	574	Pseudomonas putida W619	YP_001750420	C	474
Verrucomicrobiae bacterium DG1235	EDY83083	C	547	Pseudomonas putida W619	YP_001750424	C	480
Helicobacter acinonychis str. Sheeba	YP_664165	C	495	Stenotrophomonas maltophilia R551-3	YP_002026699	A1	535
Bartonella tribocorum CIP 105476	YP_001608617	A1	672	Stenotrophomonas maltophilia R551-3	YP_002027529	A1	665
Roseiflexus castenholzii DSM 13941	YP_001433624	A1	641	Stenotrophomonas maltophilia R551-3	YP_002030176	A1	658
Roseiflexus castenholzii DSM 13941	YP_001431686	B	561	Salinispora arenicola CNS-205	YP_001535885	A1	585
Phaeobacter gallaeciensis 2.10	ZP_02150241	C	519	Salinispora arenicola CNS-205	YP_001537144	A1	666
Phaeobacter gallaeciensis 2.10	ZP_02147755	C	545	Burkholderia phymatum STM815	YP_001860947	A1	874
Phaeobacter gallaeciensis 2.10	ZP_02151373	NOR	452	Burkholderia phymatum STM815	YP_001861937	A1	666
Methylophilales bacterium HTCC2181	ZP_01552610	A1	531	Burkholderia phymatum STM815	YP_001856516	A1	537
Pseudomonas entomophila L48	YP_606676	A1	673	Burkholderia phymatum STM815	YP_001859829	A1	670
Pseudomonas entomophila L48	YP_605868	A1	529	Burkholderia phymatum STM815	YP_001860138	A1	668
Pseudomonas entomophila L48	YP_607451	C	480	Burkholderia phymatum STM815	YP_001856586	B	541
Pseudomonas entomophila L48	YP_607455	C	474	Burkholderia phymatum STM815	YP_001861613	B	669
Labrenzia aggregata IAM 12614	ZP_01548128	A1	531	Granulibacter thesedensis CGDNIH1	YP_745096	A1	663
Labrenzia aggregata IAM 12614	ZP_01548610	A1	679	Granulibacter thesedensis CGDNIH1	YP_746079	A1	538
Labrenzia aggregata IAM 12614	ZP_01549668	NOR	456	Vibionales bacterium SWAT-3	ZP_01815045	A1	685
Listeria welshimeri serovar 6b str. SLCC5334	YP_848218	A1	659	Vibionales bacterium SWAT-3	ZP_01812538	A1	553
Escherichia coli O157:H7 str. Sakai	NP_308512	A1	663	Vibionales bacterium SWAT-3	ZP_01815176	C	477
Yersinia pestis Pestoides F	YP_001164135	A1	663	Vibionales bacterium SWAT-3	ZP_01812411	C	474
Nitratiruptor sp. SB155-2	YP_001357081	C	487	Kordia algicida OT-1	ZP_02161004	A1	599
Nitratiruptor sp. SB155-2	YP_001357257	NOR	486	Vibrio shilonii AK1	ZP_01865541	A1	547
Sulfurovum sp. NBC37-1	YP_001357604	B	560	Vibrio shilonii AK1	ZP_01867969	A1	678
Sulfurovum sp. NBC37-1	YP_001358387	C	527	Vibrio shilonii AK1	ZP_01865726	C	469
Sulfurovum sp. NBC37-1	YP_001357495	C	488	Vibrio shilonii AK1	ZP_01869570	C	475
Sulfurovum sp. NBC37-1	YP_001357561	NOR	481	Roseobacter sp. CCS2	ZP_01751185	A1	557
Candidatus Carsonella ruddii PV	YP_802566	A1	606	Roseobacter litoralis Och 149	ZP_02140791	A1	534
Pseudomonas aeruginosa PACS2	ZP_01363428	NOR	465	Roseobacter litoralis Och 149	ZP_02141397	C	537
Vibrio fischeri MJ11	YP_002156096	C	476	Roseobacter litoralis Och 149	ZP_02141655	NOR	456
Sagittula stellata E-37	ZP_01747947	A1	851	Pedobacter sp. BAL39	ZP_01883290	A1	593
Sagittula stellata E-37	ZP_01744693	A1	557	Pedobacter sp. BAL39	ZP_01883863	C	703
				Limnobacter sp. MED105	ZP_01915177	A1	640
				Limnobacter sp. MED105	ZP_01913894	A1	534
				Flavobacteria bacterium BAL38	ZP_01734088	A1	588

Flavobacteria bacterium BAL38	ZP_01734935	C	723
Brevundimonas sp. BAL3	EDX81437	A1	559
Brevundimonas sp. BAL3	EDX79523	A1	825
Brevundimonas sp. BAL3	EDX80508	A1	647
Flavobacteriales bacterium ALC-1	ZP_02180989	A1	597
Octadecabacter antarcticus 238	EDY92197	C	432
Phaeobacter gallaeciensis BS107	ZP_02145844	A1	557
Oceanibulbus indolifex HEL-45	ZP_02154310	A1	557
Oceanibulbus indolifex HEL-45	ZP_02154201	A1	854
Oceanibulbus indolifex HEL-45	ZP_02153683	C	536
Plesiocystis pacifica SIR-1	ZP_01911748	A2	542
Verminephrobacter eiseniae EF01-2	YP_996163	A1	543
Verminephrobacter eiseniae EF01-2	YP_995960	C	486
Desulfovibrio vulgaris DP4	YP_966790	A2	541
Rickettsia bellii OSU 85-389	YP_001495898	A1	532
Rickettsia bellii OSU 85-389	YP_001496394	B	533
Rickettsia rickettsii str. 'Sheila Smith'	YP_001494671	A1	533
Rickettsia rickettsii str. 'Sheila Smith'	YP_001494642	B	507
Hydrogenivirga sp. 128-5-R1-1	ZP_02177583	A2	537
Hydrogenivirga sp. 128-5-R1-1	ZP_02176553	A2	537
Hydrogenivirga sp. 128-5-R1-1	ZP_02176982	NOR	467
Sphingomonas wittichii RW1	YP_001264359	A1	562
Sphingomonas wittichii RW1	YP_001262300	C	546
Sphingomonas wittichii RW1	YP_001260078	NOR	759
Sphingomonas wittichii RW1	YP_001265090	NOR	756
Shewanella woodyi ATCC 51908	YP_001758573	A1	540
Shewanella woodyi ATCC 51908	YP_001758825	A1	661
Shewanella woodyi ATCC 51908	YP_001760798	C	474
Shewanella woodyi ATCC 51908	YP_001759151	NOR	762
Paenibacillus larvae subsp. larvae BRL-230010	ZP_02330430	A1	667
Paenibacillus larvae subsp. larvae BRL-230010	ZP_02328499	A1	645
Francisella tularensis subsp. holarctica OSU18	YP_762853	A1	676
Francisella tularensis subsp. tularensis FSC198	YP_666464	A1	676
Yersinia enterocolitica subsp. enterocolitica 8081	YP_001007324	A1	663
Alcanivorax borkumensis SK2	YP_693619	A1	535
Alcanivorax borkumensis SK2	YP_693756	A1	669
Alcanivorax borkumensis SK2	YP_693071	C	476
Maricaulis maris MCS10	YP_757463	A1	544
Maricaulis maris MCS10	YP_756380	C	550
Burkholderia multivorans ATCC 17616	YP_001946490	A1	668
Burkholderia multivorans ATCC 17616	YP_001947222	A1	535
Burkholderia multivorans ATCC 17616	YP_001579395	A1	668
Burkholderia multivorans ATCC 17616	YP_001578649	A1	535
Burkholderia multivorans ATCC 17616	YP_001941788	B	878
Burkholderia multivorans ATCC 17616	YP_001941565	B	541
Burkholderia multivorans ATCC 17616	YP_001585463	B	876
Burkholderia multivorans ATCC 17616	YP_001585717	B	541
Rhizobium leguminosarum bv. trifolii WSM1325	ZP_02292511	A1	670
Rhizobium leguminosarum bv. trifolii WSM1325	ZP_02295284	A1	667
Rhizobium leguminosarum bv. trifolii WSM1325	ZP_02293248	A1	567
Rhizobium leguminosarum bv. trifolii WSM1325	ZP_02295062	A1	592
Rhizobium leguminosarum bv. trifolii WSM1325	ZP_02293874	A2	554
Rhizobium leguminosarum bv. trifolii WSM2304	YP_002280126	A1	565
Rhizobium leguminosarum bv. trifolii WSM2304	YP_002281771	A1	589
Leptothrix cholodnii SP-6	YP_001789293	A1	588
Leptothrix cholodnii SP-6	YP_001792838	A1	544
Leptothrix cholodnii SP-6	YP_001791809	C	479
Leptothrix cholodnii SP-6	YP_001790501	NOR	483
Rhodopseudomonas palustris TIE-1	YP_001989933	A1	541
Rhodopseudomonas palustris TIE-1	YP_001989058	C	548
Rhodopseudomonas palustris TIE-1	YP_001990646	NOR	448
Beijerinckia indica subsp. indica ATCC 9039	YP_001831868	A1	555
Methylocella silvestris BL2	YP_002361569	A1	554
Methylocella silvestris BL2	ZP_02957295	A1	554
Methylocella silvestris BL2	YP_002363645	B	650
Methylocella silvestris BL2	ZP_02956138	B	650
Methylocella silvestris BL2	YP_002363906	C	557
Methylocella silvestris BL2	YP_002361828	NOR	451
Methylocella silvestris BL2	ZP_02947292	NOR	451
Thioalkalivibrio sp. HL-EbGR7	ZP_03278562	A1	533
Thioalkalivibrio sp. HL-EbGR7	ZP_03278292	C	472
Thioalkalivibrio sp. HL-EbGR7	ZP_03278572	C	496
Burkholderia graminis C4D1M	ZP_02882963	A1	646
Burkholderia graminis C4D1M	ZP_02883997	A1	649
Burkholderia graminis C4D1M	ZP_02884057	A1	668
Burkholderia graminis C4D1M	ZP_02882701	A1	537
Burkholderia graminis C4D1M	ZP_02887806	B	684
Acidovorax avenae subsp. citrulli AAC00-1	YP_971882	A1	569
Acidovorax avenae subsp. citrulli AAC00-1	YP_970574	A1	667

Acidovorax avenae subsp. citrulli AAC00-1	YP_972223	A1	547
Acidovorax avenae subsp. citrulli AAC00-1	YP_969886	C	479
Calddivirga maquilensis IC-167	YP_001541744	B	529
Calddivirga maquilensis IC-167	YP_001540124	NOR	731
Burkholderia phytodimans PsJN	YP_001888744	A1	662
Burkholderia phytodimans PsJN	YP_001894194	A1	537
Burkholderia phytodimans PsJN	YP_001894271	B	541
Burkholderia phytodimans PsJN	YP_001895658	B	678
Burkholderia ambifaria MC40-6	YP_001808673	A1	668
Burkholderia ambifaria MC40-6	YP_001809456	A1	535
Burkholderia ambifaria MC40-6	YP_001815927	A1	586
Burkholderia ambifaria MC40-6	YP_001816273	B	879
Burkholderia ambifaria MC40-6	YP_001811293	C	554
Delftia acidovorans SPH-1	YP_001562703	A1	556
Delftia acidovorans SPH-1	YP_001565074	A1	645
Delftia acidovorans SPH-1	YP_001563798	C	478
Shewanella pealeana ATCC 700345	YP_001503845	A1	534
Shewanella pealeana ATCC 700345	YP_001500817	C	468
Shewanella pealeana ATCC 700345	YP_001502069	C	478
Dinoroseobacter shibae DFL 12	YP_001533720	A1	558
Dinoroseobacter shibae DFL 12	YP_001542100	A1	853
Dinoroseobacter shibae DFL 12	YP_001532008	C	533
Dinoroseobacter shibae DFL 12	YP_001534518	NOR	459
Leeuwenhoekella blandensis MED217	ZP_01059311	A1	608
Leeuwenhoekella blandensis MED217	ZP_01060293	C	734
Metallosphaera sedula DSM 5348	YP_001190425	A1	801
Metallosphaera sedula DSM 5348	YP_001190669	B	546
Metallosphaera sedula DSM 5348	YP_001192096	B	585
Metallosphaera sedula DSM 5348	YP_001190585	B	570
Metallosphaera sedula DSM 5348	YP_001190586	B	585
Metallosphaera sedula DSM 5348	YP_001190391	B	550
Shewanella baltica OS195	YP_001556730	A1	530
Shewanella baltica OS195	YP_001556769	A1	659
Shewanella baltica OS195	YP_001554649	C	478
Pseudomonas mendocina ymp	YP_001185649	A1	527
Pseudomonas mendocina ymp	YP_001187757	A1	843
Pseudomonas mendocina ymp	YP_001188092	C	480
Pseudomonas mendocina ymp	YP_001188087	C	479
Serratia proteamaculans 568	YP_001477321	A1	663
Enterobacter sp. 638	YP_001175633	A1	663
Comamonas testosteroni KF-1	ZP_01518706	A1	646
Comamonas testosteroni KF-1	ZP_01517839	A1	547
Comamonas testosteroni KF-1	ZP_01519090	C	478
Shewanella putrefaciens 200	ZP_01706816	A1	659
Shewanella putrefaciens 200	ZP_01706832	A1	530
Shewanella putrefaciens 200	ZP_01705821	C	478
Shewanella putrefaciens 200	ZP_01705890	NOR	761
Acinetobacter baumannii ATCC 17978	YP_001085196	A1	641
Marinomonas sp. MWYL1	YP_001338967	A1	659
Legionella pneumophila str. Corby	YP_001252428	A1	538
Legionella pneumophila str. Corby	YP_001250986	NOR	780
Francisella novicida U112	YP_897860	A1	676
Flavobacterium psychrophilum JIP02/86	YP_001295310	A1	601
Flavobacterium psychrophilum JIP02/86	YP_001296054	C	726
Flavobacterium psychrophilum JIP02/86	YP_001295831	C	722
Parvibaculum lavamentivorans DS-1	YP_001411950	A1	535
Parvibaculum lavamentivorans DS-1	YP_001411844	A1	840
Parvibaculum lavamentivorans DS-1	YP_001412808	C	563
Parvibaculum lavamentivorans DS-1	YP_001412284	NOR	769
Shewanella baltica OS185	YP_001368360	A1	530
Shewanella baltica OS185	YP_001368398	A1	659
Shewanella baltica OS185	YP_001366375	C	478
Geobacter bemidjensis Bem	YP_002136870	A2	539
Geobacter bemidjensis Bem	YP_002136948	C	463
Geobacter bemidjensis Bem	YP_002138052	C	467
Geobacter bemidjensis Bem	YP_002138063	NOR	457
Geobacter bemidjensis Bem	YP_002140689	NOR	763
Anaeromyxobacter sp. Fw109-5	YP_001378044	A2	555
Anaeromyxobacter sp. Fw109-5	YP_001380045	A2	596
Anaeromyxobacter sp. Fw109-5	YP_001378410	C	477
Anaeromyxobacter sp. Fw109-5	YP_001379612	NOR	760
Anaeromyxobacter sp. Fw109-5	YP_001379942	NOR	452
Anaeromyxobacter sp. Fw109-5	YP_001381011	NOR	764
Acinetobacter baumannii ACICU	YP_001847029	A1	663
Xylella fastidiosa M12	YP_001775363	A1	665
Xylella fastidiosa M23	YP_001829371	A1	665
Bacillus cereus G9842	YP_002444166	A1	644
Bacillus cereus G9842	YP_002447510	A1	620
Bacillus cereus AH187	YP_002336832	A1	644
Bacillus cereus AH187	YP_002339996	A1	620
Bacillus cereus AH820	YP_002449730	A1	644
Bacillus cereus AH820	YP_002452908	A1	620
Saccharopolyspora erythraea NRRL 2338	YP_001103438	A1	557
Saccharopolyspora erythraea NRRL 2338	YP_001109127	A1	566
Escherichia coli APEC O1	YP_851604	A1	663
Burkholderia cenocepacia MC0-3	YP_001765372	A1	668

Burkholderia cenocepacia MC0-3	YP_001766138	A1	535
Burkholderia cenocepacia MC0-3	YP_001773760	B	1004
Campylobacter jejuni subsp. jejuni 81116	YP_001482970	C	488
Shewanella baltica OS223	YP_002360007	A1	531
Shewanella baltica OS223	YP_002360041	A1	659
Shewanella baltica OS223	YP_002358136	C	478
Escherichia coli SE11	YP_002291731	A1	663
Mycobacterium bovis BCG str. Pasteur 1173P2	YP_979151	A1	573
Pyrobaculum calidifontis JCM 11548	YP_001056825	A1	798
Pyrobaculum calidifontis JCM 11548	YP_001056831	B	560
Pyrobaculum calidifontis JCM 11548	YP_001056790	NOR	720
Gramella forsetii KT0803	YP_860231	A1	607
Gramella forsetii KT0803	YP_861474	C	731
Gramella forsetii KT0803	YP_861439	C	602
Hoeflea phototrophica DFL-43	ZP_02167414	A1	550
Hoeflea phototrophica DFL-43	ZP_02167623	C	539
Burkholderia mallei NCTC 10229	YP_001025367	A1	668
Burkholderia mallei NCTC 10229	YP_001027406	A1	535
Burkholderia mallei NCTC 10229	YP_001028824	A1	560
Burkholderia mallei NCTC 10229	YP_001024473	NOR	772
Burkholderia mallei NCTC 10229	YP_001028854	NOR	762
Bacillus thuringiensis str. Al Hakam	YP_893531	A1	647
Bacillus thuringiensis str. Al Hakam	YP_896311	A1	620
Candidatus Vesicomysocius okutanii HA	YP_001218898	A1	533
Candidatus Vesicomysocius okutanii HA	YP_001218903	C	472
Vibrio cholerae MAK 757	ZP_01953281	C	475
Candidatus Ruthia magnifica str. Cm (Calypotgena magnifica)	YP_903309	A1	538
Candidatus Ruthia magnifica str. Cm (Calypotgena magnifica)	YP_903313	C	472
Cenarchaeum symbiosum A	YP_876890	A1	508
Rhodospirillum centenum SW	YP_002299792	A1	535
Rhodospirillum centenum SW	YP_002297025	A1	542
Rhodospirillum centenum SW	YP_002297500	C	494
Rhodospirillum centenum SW	YP_002300129	NOR	752
Rickettsia massiliae MTU5	YP_001499310	A1	540
Rickettsia massiliae MTU5	YP_001499284	C	528
Halorubrum lacusprofundi ATCC 49239	ZP_02014739	B	597
Halorubrum lacusprofundi ATCC 49239	ZP_02015245	NOR	763
Staphylococcus aureus subsp. aureus Mu3	YP_001441642	A1	662
Francisella tularensis subsp. tularensis WY96-3418	YP_001122595	A1	676
Vibrio parahaemolyticus AQ3810	ZP_01990224	A1	678
Vibrio parahaemolyticus AQ3810	ZP_01990060	A1	545
Leptospirillum sp. Group II UBA	EAY56843	C	479
Methylobacterium extorquens PA1	YP_001637594	A1	638
Methylobacterium extorquens PA1	YP_001640720	A1	542
Methylobacterium extorquens PA1	YP_001641776	A1	667
Methylobacterium extorquens PA1	YP_001637720	A2	563
Mycobacterium tuberculosis H37Ra	YP_001284419	A1	573
Geobacillus thermodenitrificans NG80-2	YP_001127472	A1	648
Geobacillus thermodenitrificans NG80-2	YP_001125071	A1	622
Geobacillus thermodenitrificans NG80-2	YP_001125633	B	482
Geobacillus thermodenitrificans NG80-2	YP_001125509	B	549
Geobacillus thermodenitrificans NG80-2	YP_001124764	C	413
Geobacillus thermodenitrificans NG80-2	YP_001124768	NOR	790
Methylibium petroleiphilum PM1	YP_001022368	A1	539
Methylibium petroleiphilum PM1	YP_001020937	A1	668
Methylibium petroleiphilum PM1	YP_001021668	C	480
Beggiatoa sp. PS	ZP_01999494	A2	525
Beggiatoa sp. PS	ZP_02000496	NOR	534
Beggiatoa sp. PS	ZP_02001268	NOR	477
Salmonella enterica subsp. enterica serovar Newport str. SL254	YP_002039684	A1	663
Shewanella sediminis HAW-EB3	YP_001471938	A1	534
Shewanella sediminis HAW-EB3	YP_001473948	C	475
Shewanella sediminis HAW-EB3	YP_001475554	NOR	761
Methylobacterium sp. 4-46	YP_001769836	A1	638
Methylobacterium sp. 4-46	YP_001770700	A1	589
Methylobacterium sp. 4-46	YP_001769241	A1	542
Methylobacterium sp. 4-46	YP_001772457	A1	667
Methylobacterium sp. 4-46	YP_001772537	A2	558
Methylobacterium sp. 4-46	YP_001770109	C	552
Methylobacterium radiotolerans JCM 2831	YP_001754410	A1	666
Methylobacterium radiotolerans JCM 2831	YP_001752742	A1	542
Methylobacterium radiotolerans JCM 2831	YP_001766956	A2	565
Methylobacterium radiotolerans JCM 2831	YP_001753729	A2	558
Staphylococcus aureus subsp. aureus str. Newman	YP_001331963	A1	662
Ralstonia pickettii 12D	ZP_02005338	A1	586
Ralstonia pickettii 12D	ZP_02008323	A1	659
Ralstonia pickettii 12D	ZP_02005784	A1	534
Ralstonia pickettii 12D	ZP_02009629	A1	667
Ralstonia pickettii 12D	ZP_02007918	C	485

Ralstonia pickettii 12D	ZP_02008815	C	550	Methylobacterium chloromethanicum CM4	YP_002423407	A1	667
Ralstonia pickettii 12D	ZP_02005927	NOR	758	Methylobacterium chloromethanicum CM4	ZP_02057820	A1	667
Brucella abortus S19	YP_001934469	A1	552	Methylobacterium chloromethanicum CM4	ZP_02059811	A1	542
Brucella abortus S19	YP_001934376	C	541	Methylobacterium chloromethanicum CM4	YP_002419042	A2	563
Brucella abortus S19	YP_001932818	NOR	449	Methylobacterium chloromethanicum CM4	ZP_02056159	A2	563
Francisella tularensis subsp. holarctica FSC022	YP_002167604	A1	676	Burkholderia pseudomallei 7894	ZP_02482537	NOR	762
Magnetospirillum gryphiswaldense MSR-1	CAM74645	A1	526	Burkholderia oklahomensis C6786	ZP_02363473	NOR	762
Magnetospirillum gryphiswaldense MSR-1	CAM77358	C	532	Burkholderia oklahomensis EO147	ZP_02354324	A1	535
Magnetospirillum gryphiswaldense MSR-1	CAM75374	NOR	445	Burkholderia oklahomensis EO147	ZP_02356343	NOR	746
Coxiella burnetii Dugway 5J108-111	YP_001424381	A1	668	Burkholderia ubonensis Bu	ZP_02377758	A1	535
Coxiella burnetii CbuG_Q212	YP_002303478	A1	668	Burkholderia ubonensis Bu	ZP_02381572	A1	669
Coxiella burnetii CbuK_Q154	YP_002305192	A1	668	Burkholderia ubonensis Bu	ZP_02377878	A1	670
Sulfurihydrogenibium sp. YO3AOP1	YP_001930662	C	410	Burkholderia ubonensis Bu	ZP_02382486	A1	587
Sulfurihydrogenibium sp. YO3AOP1	YP_001930359	C	469	Methylobacterium populi BJ001	YP_001927484	A1	667
Sulfurihydrogenibium sp. YO3AOP1	YP_001930667	NOR	739	Methylobacterium populi BJ001	YP_001926144	A1	542
Nitrosopumilus maritimus SCM1	YP_001581518	A1	508	Methylobacterium populi BJ001	YP_001922756	A2	563
Azorhizobium caulinodans ORS 571	YP_001525015	A1	670	Bacillus coahuilensis m4-4	ZP_03226521	B	548
Azorhizobium caulinodans ORS 571	YP_001525875	A1	568	Francisella tularensis subsp. mediasiatica FSC147	YP_001892168	A1	676
Azorhizobium caulinodans ORS 571	YP_001527439	C	551	Geobacter sp. M21	ZP_03025377	A2	542
Bacillus selenitireducens MLS10	ZP_02171841	A1	613	Geobacter sp. M21	ZP_03025828	C	463
Ochrobactrum anthropi ATCC 49188	YP_001368609	A1	659	Geobacter sp. M21	ZP_03025602	NOR	762
Ochrobactrum anthropi ATCC 49188	YP_001369142	A1	552	Geobacter sp. M21	ZP_03024466	NOR	457
Ochrobactrum anthropi ATCC 49188	YP_001369029	C	541	Marinobacter algicola DG893	ZP_01893361	A1	528
Ochrobactrum anthropi ATCC 49188	YP_001372916	NOR	448	Marinobacter algicola DG893	ZP_01894120	C	476
Campylobacteriales bacterium GD 1	EDZ63164	C	497	Streptomyces clavuligerus ATCC 27064	ZP_03183291	A1	579
Campylobacteriales bacterium GD 1	EDZ63129	NOR	477	Streptomyces clavuligerus ATCC 27064	ZP_03185674	A1	569
Pseudovibrio sp. JE062	EEA93213	A1	665	Clavibacter michiganensis subsp. michiganensis NCPPB 382	YP_001222583	A1	581
Pseudovibrio sp. JE062	EEA94034	A1	530	Lysinibacillus sphaericus C3-41	YP_001698546	A1	649
Pseudovibrio sp. JE062	EEA94158	C	551	Lysinibacillus sphaericus C3-41	YP_001699583	A1	399
Pseudovibrio sp. JE062	EEA92005	NOR	461	Lysinibacillus sphaericus C3-41	YP_001697134	A1	627
Pseudovibrio sp. JE062	EEA93383	NOR	768	Brucella ovis ATCC 25840	YP_001258090	A1	659
Rhodobacterales bacterium Y4I	EDZ45255	C	536	Brucella ovis ATCC 25840	YP_001258486	A1	552
Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633	YP_002113473	A1	663	Brucella ovis ATCC 25840	YP_001258397	C	541
Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853	YP_002214396	A1	663	Candidatus Sulcia muelleri GWSS	YP_001597952	C	736
Escherichia coli SMS-3-5	YP_001742572	A1	663	Escherichia coli O157:H7 str. EC4115	YP_002269077	A1	663
Methylobacterium chloromethanicum CM4	YP_002422334	A1	542	Anaeromyxobacter sp. K	YP_002133217	A2	554
				Anaeromyxobacter sp. K	YP_002133959	A2	555
				Anaeromyxobacter sp. K	ZP_02173620	A2	555
				Anaeromyxobacter sp. K	YP_002133592	C	473
				Anaeromyxobacter sp. K	YP_002135648	NOR	769
				Anaeromyxobacter sp. K	YP_002134809	NOR	762
				Anaeromyxobacter sp. K	YP_002136117	NOR	763

Sorangium cellulosum 'So ce 56'	YP_001618242	A2	545	Methylobacterium nodulans ORS 2060	ZP_02123893	A1	542
Sorangium cellulosum 'So ce 56'	YP_001615167	A2	552	Methylobacterium nodulans ORS 2060	ZP_02118651	A1	591
Sorangium cellulosum 'So ce 56'	YP_001613920	NOR	772	Methylobacterium nodulans ORS 2060	ZP_02124708	A1	667
Sorangium cellulosum 'So ce 56'	YP_001619234	NOR	834	Methylobacterium nodulans ORS 2060	ZP_02118662	A1	638
Microcystis aeruginosa NIES-843	YP_001657308	A2	558	Methylobacterium nodulans ORS 2060	ZP_02117357	A2	558
Phenylobacterium zucineum HLK1	YP_002132014	A1	551	Methylobacterium nodulans ORS 2060	ZP_02124900	C	550
Phenylobacterium zucineum HLK1	YP_002129386	A1	549	Streptomyces sviveus ATCC 29083	YP_002204028	A1	580
Phenylobacterium zucineum HLK1	YP_002130425	A1	844	Streptomyces sviveus ATCC 29083	ZP_03196979	A1	545
Phenylobacterium zucineum HLK1	YP_002131652	C	567	Streptomyces sviveus ATCC 29083	YP_002204627	A1	567
Phenylobacterium zucineum HLK1	YP_002129462	NOR	760	Micrococcus luteus NCTC 2665	ZP_02944669	A1	569
Staphylococcus aureus subsp. aureus USA300_TCH1516	YP_001574903	A1	662	Streptomyces sp. Mg1	ZP_03171263	A1	576
Opitutus terrae PB90-1	YP_001820418	A1	743	Erwinia tasmaniensis Et1/99	YP_001908438	A1	663
Opitutus terrae PB90-1	YP_001818954	A1	661	Brucella suis ATCC 23445	YP_001626723	A1	659
Opitutus terrae PB90-1	YP_001817463	C	427	Brucella suis ATCC 23445	YP_001627153	A1	552
Opitutus terrae PB90-1	YP_001820802	C	468	Brucella suis ATCC 23445	YP_001627054	C	541
Opitutus terrae PB90-1	YP_001819752	C	572	Brucella suis ATCC 23445	YP_001622088	NOR	449
Opitutus terrae PB90-1	YP_001820951	C	772	Geobacillus sp. WCH70	ZP_02913585	B	549
Polynucleobacter necessarius subsp. necessarius STIR1	YP_001798332	A1	535	Providencia stuartii ATCC 25827	ZP_02962546	A1	663
Polynucleobacter necessarius subsp. necessarius STIR1	YP_001798088	C	491	Halobacterium salinarum R1	YP_001688771	A1	593
Rickettsia rickettsii str. Iowa	YP_001649930	A1	533	Halobacterium salinarum R1	YP_001689972	B	565
Rickettsia rickettsii str. Iowa	YP_001649899	C	507	Acinetobacter baumannii AB0057	YP_002319850	A1	663
Salmonella enterica subsp. enterica serovar Agona str. SL483	YP_002145423	A1	663	Methylacidiphilum infernorum V4	YP_001940602	A1	615
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476	YP_002044477	A1	663	Methylacidiphilum infernorum V4	YP_001938924	A1	630
Anaeromyxobacter dehalogenans 2CP-1	ZP_02323500	A2	555	Methylacidiphilum infernorum V4	YP_001939709	B	561
Anaeromyxobacter dehalogenans 2CP-1	ZP_02325394	C	473	Methylacidiphilum infernorum V4	YP_001940605	C	429
Streptomyces griseus subsp. griseus NBRC 13350	YP_001826865	A1	578	Methylacidiphilum infernorum V4	YP_001938701	NOR	474
Streptomyces griseus subsp. griseus NBRC 13350	YP_001821828	A1	602	Methylacidiphilum infernorum V4	ABX56670	NOR	471
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	YP_001840037	A2	537	Geobacillus sp. Y412MC10	ZP_03036781	A1	632
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	YP_001964466	C	476	Geobacillus sp. Y412MC10	ZP_03037388	A1	647
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	YP_001840298	NOR	449	Escherichia coli ATCC 8739	YP_001726151	A1	663
Streptomyces pristinaespiralis ATCC 25486	YP_002196331	A1	576	Brucella canis ATCC 23365	YP_001591920	A1	659
Streptomyces pristinaespiralis ATCC 25486	YP_002196596	A1	577	Brucella canis ATCC 23365	YP_001592327	A1	552
Streptomyces pristinaespiralis ATCC 25486	YP_002198657	A1	579	Brucella canis ATCC 23365	YP_001592230	C	541
Francisella tularensis subsp. holarctica FTNF002-00	YP_001427640	A1	645	Brucella canis ATCC 23365	YP_001594211	NOR	449
Shewanella halifaxensis HAW-EB4	YP_001672492	A1	534	Anoxybacillus flavithermus WK1	YP_002314642	A1	665
Shewanella halifaxensis HAW-EB4	YP_001674415	C	478	Anoxybacillus flavithermus WK1	YP_002316213	A1	622
Shewanella halifaxensis HAW-EB4	YP_001675756	NOR	764	Anoxybacillus flavithermus WK1	YP_002315712	B	549
				Rhizobium etli CIAT 652	YP_001977211	A1	568
				Rhizobium etli CIAT 652	YP_001978875	A1	589
				Geobacillus sp. G11MC16	ZP_03146778	NOR	787
				Chthoniobacter flavus Ellin428	ZP_03129065	A1	623
				Chthoniobacter flavus Ellin428	ZP_03130087	C	775
				Cellvibrio japonicus Ueda107	YP_001981374	A1	532
				Cellvibrio japonicus Ueda107	YP_001983094	C	485

Cellvibrio japonicus Ueda107	YP_001982028	NOR	745
Thermus aquaticus Y51MC23	ZP_03134051	A2	791
Enterobacter cancerogenus ATCC 35316	ZP_03281366	A1	663
Escherichia albertii TW07627	ZP_02902141	A1	663
Yersinia pseudotuberculosis YPIII	YP_001721964	A1	663
Yersinia pseudotuberculosis PB1/+	YP_001871430	A1	663
Corynebacterium urealyticum DSM 7109	YP_001800872	A1	565
Oligotropha carboxidovorans OM5	EDT29063	A1	541
Oligotropha carboxidovorans OM5	YP_002210494	A1	541
Oligotropha carboxidovorans OM5	YP_002287693	A1	541
Oligotropha carboxidovorans OM5	YP_002213116	C	540
Oligotropha carboxidovorans OM5	YP_002212912	C	550
Oligotropha carboxidovorans OM5	YP_002290108	C	550
Oligotropha carboxidovorans OM5	YP_002290312	C	540
Klebsiella pneumoniae 342	YP_002240090	A1	663
Xanthomonas campestris pv. campestris str. B100	YP_001904584	A1	666
Xanthomonas campestris pv. campestris str. B100	YP_001905423	A1	534
Acinetobacter baumannii SDF	YP_001706978	A1	663
Acinetobacter baumannii AYE	YP_001713301	A1	663
Escherichia coli str. K-12 substr. MG1655	NP_414965	A1	663
Helicobacter pylori Shi470	YP_001909635	C	488
Arthrosira maxima CS-328	ZP_03271753	A2	585
Arthrosira maxima CS-328	ZP_03273012	NOR	521
Burkholderia sp. H160	ZP_03268378	A1	586
Burkholderia sp. H160	ZP_03268598	A1	537
Burkholderia sp. H160	ZP_03267566	A1	673
Burkholderia sp. H160	ZP_03268689	B	845
Burkholderia sp. H160	ZP_03264540	NOR	775
Chlorobaculum parvum NCIB 8327	YP_001998491	A1	663
Chlorobaculum parvum NCIB 8327	YP_001999252	C	466
Neisseria gonorrhoeae NCCP11945	YP_002002235	C	477
Neisseria gonorrhoeae NCCP11945	YP_002002114	NOR	742
Stenotrophomonas maltophilia K279a	YP_001971217	A1	667
Stenotrophomonas maltophilia K279a	YP_001970342	A1	535
Stenotrophomonas maltophilia K279a	YP_001974059	A1	658
Sulfolobus metallicus DSM 6482	ABG91823	B	595
Proteus mirabilis HI4320	YP_002149891	A1	660
Candidatus Liberibacter asiaticus str. psy62	ZP_03287165	A1	671
Salmonella enterica subsp. enterica serovar Enteritidis str. P125109	YP_002242573	A1	663
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	YP_002225554	A1	663
Listeria monocytogenes HCC23	YP_002351598	A1	659
Acinetobacter baumannii AB307-0294	YP_002325200	A1	663
Pseudomonas aeruginosa LESB58	YP_002437715	A1	530
Pseudomonas aeruginosa LESB58	YP_002441446	A1	658
Pseudomonas aeruginosa LESB58	YP_002438446	C	475
Pseudomonas aeruginosa LESB58	YP_002441054	C	480
Pseudomonas aeruginosa LESB58	YP_002441354	C	475
Pseudomonas aeruginosa LESB58	YP_002441357	C	475
Pseudomonas aeruginosa LESB58	YP_002438129	NOR	465
Buchnera aphidicola str. Tuc7 (Acyrtosiphon pisum)	YP_002468760	A1	662
Helicobacter pylori G27	YP_002265766	C	488
Buchnera aphidicola str. 5A (Acyrtosiphon pisum)	YP_002468207	A1	662
Francisella novicida FTG	ZP_03247914	A1	645
Wolbachia endosymbiont of Culex quinquefasciatus Pel	YP_001974910	A1	516
Wolbachia endosymbiont of Culex quinquefasciatus Pel	YP_001974920	B	430
Helicobacter pylori P12	YP_002300783	C	488
Escherichia coli O127:H6 str. E2348/69	YP_002327943	A1	663
Escherichia coli IA11	YP_002385910	A1	663
Escherichia coli S88	YP_002390253	A1	663
Escherichia fergusonii ATCC 35469	YP_002383692	A1	663
Escherichia coli 55989	YP_002401561	A1	663
Escherichia coli UMN026	YP_002411233	A1	663
Escherichia coli IA139	YP_002406286	A1	663
Escherichia coli ED1a	YP_002396506	A1	663

Table 0.8 Detailed information on mitochondrial sequences.

Genome	Accession	Length		
Mammalia (mammals)			<i>Tarsipes rostratus</i>	NC_006518 15,744
Prototheria			Vombatidae	
Monotremata (egg-laying mammals)			<i>Vombatus ursinus</i>	NC_003322 16,996
Ornithorhynchidae			Microbiotheria	
<i>Ornithorhynchus anatinus</i>	NC_000891	17,019	<i>Dromiciops gliroides</i>	NC_005826 15,350
Tachyglossidae			Notoryctemorphia	
<i>Tachyglossus aculeatus</i>	NC_003321	16,360	<i>Notoryctes typhlops</i>	NC_006522 15,923
<i>Zaglossus bruijni</i>	NC_006364	16,367	Paucituberculata	
Theria			Caenolestidae	
Metatheria (marsupials)			<i>Caenolestes fuliginosus</i>	NC_005828 16,629
Dasyuromorphia			<i>Rhyncholestes raphanurus</i>	NC_005829 16,727
Dasyuridae			Peramelemorphia	
<i>Dasyurus hallucatus</i>	NC_007630	17,159	Peramelidae	
<i>Phascogale tapoatafa</i>	NC_006523	17,405	<i>Isoodon macrourus</i>	NC_002746 16,852
<i>Sarcophilus harrisii</i>	NC_018788	17,117	<i>Macrotis lagotis</i>	NC_006520 15,289
<i>Sminthopsis crassicaudata</i>	NC_007631	17,363	<i>Perameles gunnii</i>	NC_006521 15,360
<i>Sminthopsis douglasi</i>	NC_006517	16,538	Peroryctidae	
Myrmecobiidae			<i>Echymipera rufescens australis</i>	NC_007632 16,642
<i>Myrmecobius fasciatus</i>	NC_011949	16,813	Eutheria (placentals)	
Thylacinidae			Afrotheria	
<i>Thylacinus cynocephalus</i>	NC_011944	17,411	Chrysochloridae	
Didelphimorphia			<i>Chrysochloris asiatica</i>	NC_004920 16,652
Didelphidae			<i>Eremitalpa granti</i>	NC_010304 16,602
<i>Didelphis virginiana</i>	NC_001610	17,084	Hyracoidea	
<i>Metachirus nudicaudatus</i>	NC_006516	15,882	<i>Dendrohyrax dorsalis</i>	NC_010301 16,590
<i>Monodelphis domestica</i>	NC_006299	17,079	<i>Procavia capensis</i>	NC_004919 16,743
<i>Thylamys elegans</i>	NC_005825	15,452	Macroscelidea	
Diprotodontia			<i>Elephantulus sp. VB001</i>	NC_004921 16,677
Acrobatidae			<i>Macroscelides proboscideus</i>	NC_004026 16,641
<i>Distoechurus pennatus</i>	NC_008145	16,775	Proboscidea	
Macropodidae			<i>Elephas maximus</i>	NC_005129 16,902
<i>Lagorchestes hirsutus</i>	NC_008136	16,889	<i>Loxodonta africana</i>	NC_000934 16,866
<i>Lagostrophus fasciatus</i>	NC_008447	16,866	<i>Mammut americanum</i>	NC_009574 16,469
<i>Macropus robustus</i>	NC_001794	16,896	<i>Mammuthus columbi</i>	NC_015529 16,459
Petauridae			<i>Mammuthus primigenius</i>	NC_007596 16,770
<i>Dactylopsila trivirgata</i>	NC_008134	16,058	Sirenia	
<i>Petaurus breviceps</i>	NC_008135	16,103	<i>Dugong dugon</i>	NC_003314 16,850
Phalangeridae			<i>Trichechus manatus</i>	NC_010302 16,882
<i>Phalanger vestitus</i>	NC_008137	15,318	Tenrecidae	
<i>Trichosurus vulpecula</i>	NC_003039	17,191	<i>Echinops telfairi</i>	NC_002631 16,549
Phascolarctidae			Tubulidentata	
<i>Phascolarctos cinereus</i>	NC_008133	16,357	<i>Orycteropus afer</i>	NC_002078 16,816
Potoroidae			Euarchontoglires	
<i>Potorous tridactylus</i>	NC_006524	16,922	Dermoptera	
Pseudocheiridae			<i>Galeopterus variegatus</i>	NC_004031 16,748
<i>Pseudocheirus peregrinus</i>	NC_006519	16,022	Glires	
Tarsipedidae			<i>Anomalurus sp. GP-2005</i>	NC_009056 16,923
			<i>Apodemus agrarius</i>	NC_016428 16,263

<i>Apodemus chejuensis</i>	NC_016662	16,261
<i>Apodemus chevrieri</i>	NC_017599	16,298
<i>Apodemus draco</i>	NC_019584	16,282
<i>Apodemus latronum</i>	NC_019585	16,288
<i>Apodemus peninsulae</i>	NC_016060	16,269
<i>Castor canadensis</i>	NC_015108	16,701
<i>Castor fiber</i>	NC_015072	16,730
<i>Cavia porcellus</i>	NC_000884	16,801
<i>Cricetulus griseus</i>	NC_007936	16,284
<i>Eospalax baileyi</i>	NC_018098	16,351
<i>Eospalax rothschildi</i>	NC_018535	16,377
<i>Eothenomys chinensis</i>	NC_013571	16,362
<i>Glis glis</i>	NC_001892	16,602
<i>Heterocephalus glaber</i>	NC_015112	16,386
<i>Jaculus jaculus</i>	NC_005314	16,546
<i>Leggadina lakedownensis</i>	NC_014696	16,262
<i>Lepus capensis</i>	NC_015841	17,722
<i>Lepus europaeus</i>	NC_004028	17,734
<i>Marmota himalayana</i>		32,886
<i>Mesocricetus auratus</i>	NC_013276	16,264
<i>Microtus fortis</i>		
<i>Microtus fortis calamorum</i>	NC_015243	16,310
<i>Microtus fortis fortis</i>	NC_015241	16,310
<i>Microtus kikuchii</i>	NC_003041	16,312
<i>Microtus levis</i>	NC_008064	16,283
<i>Mus musculus</i>		
<i>Mus musculus</i>	NC_005089	16,299
<i>Mus musculus castaneus</i>	NC_012387	16,299
<i>Mus musculus domesticus</i>	NC_006914	16,300
<i>Mus musculus molossinus</i>	NC_006915	16,300
<i>Mus musculus musculus</i>	NC_010339	16,300
<i>Mus terricolor</i>	NC_010650	16,310
<i>Myodes regulus</i>	NC_016427	16,379
<i>Nannospalax ehrenbergi</i>	NC_005315	16,408
<i>Neodon irene</i>	NC_016055	16,367
<i>Niviventer excelsior</i>	NC_019617	16,298
<i>Ochotona collaris</i>	NC_003033	16,968
<i>Ochotona curzoniae</i>	NC_011029	17,131
<i>Ochotona princeps</i>	NC_005358	16,481
<i>Oryctolagus cuniculus</i>	NC_001913	17,245
<i>Proedromys liangshanensis</i>	NC_013563	16,296
<i>Pseudomys chapmani</i>	NC_014698	16,249
<i>Pteromys volans</i>	NC_019612	16,513
<i>Rattus exulans</i>	NC_012389	16,307
<i>Rattus fuscipes</i>	NC_014867	16,301
<i>Rattus leucopus</i>	NC_014855	16,297
<i>Rattus lutreolus</i>	NC_014858	16,295
<i>Rattus norvegicus strain BN/SsNHsdMCW</i>	NC_001665	16,313
<i>Rattus praetor</i>	NC_012461	16,293

<i>Rattus rattus</i>	NC_012374	16,305
<i>Rattus sordidus</i>	NC_014871	16,309
<i>Rattus tanezumi</i>	NC_011638	16,306
<i>Rattus tunneyi</i>	NC_014861	16,292
<i>Rattus villosissimus</i>	NC_014864	16,303
<i>Sciurus vulgaris</i>	NC_002369	16,507
<i>Thryonomys swinderianus</i>	NC_002658	16,626
<i>Tscherskia triton</i>	NC_013068	16,488
Primates		
<i>Aotus azarai azarai</i>	NC_018115	16,585
<i>Aotus lemurinus</i>	NC_019799	16,580
<i>Aotus nancymae</i>	NC_018116	16,469
<i>Ateles belzebuth</i>	NC_019800	16,560
<i>Callicebus donacophilus</i>	NC_019801	16,563
<i>Cebus albifrons</i>	NC_002763	16,554
<i>Cebus apella</i>	NC_016666	16,538
<i>Chlorocebus aethiops</i>	NC_007009	16,389
<i>Chlorocebus pygerythrus</i>	NC_009747	16,343
<i>Chlorocebus sabaeus</i>	NC_008066	16,550
<i>Chlorocebus tantalus</i>	NC_009748	16,368
<i>Colobus guereza</i>	NC_006901	16,648
<i>Daubentonia madagascariensis</i>	NC_010299	16,620
<i>Eulemur fulvus</i>		
<i>Eulemur fulvus fulvus</i>	NC_012766	17,126
<i>Eulemur fulvus mayottensis</i>	NC_012769	17,118
<i>Eulemur macaco macaco</i>	NC_012771	17,149
<i>Eulemur mongoz</i>	NC_010300	16,686
<i>Galago senegalensis</i>	NC_012761	16,795
<i>Gorilla gorilla</i>		
<i>Gorilla gorilla</i>	NC_001645	16,364
<i>Gorilla gorilla gorilla</i>	NC_011120	16,412
<i>Homo sapiens</i>		
<i>Homo sapiens</i>	NC_012920	16,569
<i>Homo sapiens neanderthalensis</i>	NC_011137	16,565
<i>Homo sp. Altai</i>	NC_013993	16,570
<i>Hylobates agilis</i>	NC_014042	16,500
<i>Hylobates lar</i>	NC_002082	16,472
<i>Hylobates pileatus</i>	NC_014045	16,502
<i>Lemur catta</i>	NC_004025	17,036
<i>Lepilemur hubbardorum</i>	NC_014453	16,854
<i>Loris tardigradus</i>	NC_012763	16,776
<i>Macaca fascicularis</i>	NC_012670	16,575
<i>Macaca mulatta</i>	NC_005943	16,564
<i>Macaca sylvanus</i>	NC_002764	16,586
<i>Macaca thibetana</i>	NC_011519	16,540
<i>Nasalis larvatus</i>	NC_008216	16,570
<i>Nomascus gabriellae</i>	NC_018753	16,478
<i>Nomascus siki</i>	NC_014051	16,476
<i>Nycticebus coucang</i>	NC_002765	16,764

<i>Otolemur crassicaudatus</i>	NC_012762	16,827
<i>Pan paniscus</i>	NC_001644	16,563
<i>Pan troglodytes</i>	NC_001643	16,554
<i>Papio anubis isolate east</i>	NC_020006	16,450
<i>Papio cynocephalus isolate north</i>	NC_020007	16,450
<i>Papio hamadryas</i>	NC_001992	16,521
<i>Papio kindae</i>	NC_020008	16,450
<i>Papio papio</i>	NC_020009	16,450
<i>Papio ursinus isolate north</i>	NC_020010	16,450
<i>Perodicticus potto</i>	NC_012764	16,638
<i>Ptilocolobus badius</i>	NC_008219	16,595
<i>Pongo abelii</i>	NC_002083	16,499
<i>Pongo pygmaeus</i>	NC_001646	16,389
<i>Presbytis melalophos</i>	NC_008217	16,543
<i>Propithecus coquereli</i>	NC_011053	17,104
<i>Pygathrix cinerea 1 RL-2012</i>	NC_018062	16,535
<i>Pygathrix cinerea 2 RL-2012</i>	NC_018063	16,535
<i>Pygathrix nemaeus</i>	NC_008220	15,467
<i>Pygathrix nigripes</i>	NC_018061	16,536
<i>Rhinopithecus avunculus</i>	NC_015485	16,548
<i>Rhinopithecus bieti</i>	NC_015486	16,551
<i>Rhinopithecus bieti 1 RL-2012</i>	NC_018058	16,550
<i>Rhinopithecus bieti 2 RL-2012</i>	NC_018060	16,550
<i>Rhinopithecus brelichi</i>	NC_018057	16,553
<i>Rhinopithecus roxellana</i>	NC_008218	16,549
<i>Rhinopithecus strykeri</i>	NC_018059	16,550
<i>Saimiri boliviensis boliviensis</i>	NC_018096	16,473
<i>Saimiri sciureus</i>	NC_012775	16,474
<i>Semnopithecus entellus</i>	NC_008215	16,532
<i>Symphalangus syndactylus</i>	NC_014047	16,516
<i>Tarsius bancanus</i>	NC_002811	16,927
<i>Tarsius syrichta</i>	NC_012774	17,004
<i>Theropithecus gelada</i>	NC_019802	16,546
<i>Trachypithecus germaini</i>	NC_019580	16,552
<i>Trachypithecus hatinhensis</i>	NC_019579	16,547
<i>Trachypithecus johnii</i>	NC_019583	16,544
<i>Trachypithecus obscurus</i>	NC_006900	16,560
<i>Trachypithecus shortridgei</i>	NC_019581	16,529
<i>Trachypithecus vetulus</i>	NC_019582	16,554
<i>Varecia variegata variegata</i>	NC_012773	16,965
Scandentia		
<i>Tupaia belangeri</i>	NC_002521	16,754
Laurasiatheria		
Carnivora		
<i>Acinonyx jubatus</i>	NC_005212	17,047
<i>Ailuropoda melanoleuca</i>	NC_009492	16,805
<i>Ailurus fulgens</i>		
<i>Ailurus fulgens</i>	NC_011124	16,493
<i>Ailurus fulgens styani</i>	NC_009691	16,374

<i>Arctocephalus forsteri</i>	NC_004023	15,413
<i>Arctocephalus pusillus</i>	NC_008417	16,652
<i>Arctocephalus townsendi</i>	NC_008420	16,571
<i>Arctodus simus</i>	NC_011116	16,754
<i>Callorhinus ursinus</i>	NC_008415	16,668
<i>Canis latrans</i>	NC_008093	16,724
<i>Canis lupus</i>		
<i>Canis lupus</i>	NC_008092	16,729
<i>Canis lupus chanco</i>	NC_010340	16,774
<i>Canis lupus familiaris</i>	NC_002008	16,727
<i>Canis lupus laniger</i>	NC_011218	16,723
<i>Canis lupus lupus</i>	NC_009686	16,757
<i>Cuon alpinus</i>	NC_013445	16,672
<i>Cystophora cristata</i>	NC_008427	16,952
<i>Enhydra lutris</i>	NC_009692	16,431
<i>Erignathus barbatus</i>	NC_008426	16,705
<i>Eumetopias jubatus</i>	NC_004030	16,638
<i>Felis catus</i>	NC_001700	17,009
<i>Gulo gulo</i>	NC_009685	16,541
<i>Halichoerus grypus</i>	NC_001602	16,797
<i>Helarctos malayanus</i>	NC_009968	16,783
<i>Herpestes javanicus</i>	NC_006835	16,763
<i>Hydrurga leptonyx</i>	NC_008425	16,602
<i>Leptonychotes weddellii</i>	NC_008424	16,603
<i>Lobodon carcinophaga</i>	NC_008423	16,759
<i>Lutra lutra</i>	NC_011358	16,536
<i>Lynx rufus</i>	NC_014456	17,056
<i>Martes flavigula</i>	NC_012141	16,549
<i>Martes melampus</i>	NC_009678	16,498
<i>Martes zibellina</i>	NC_011579	16,523
<i>Meles anakuma</i>	NC_009677	16,492
<i>Meles meles</i>	NC_011125	16,442
<i>Melursus ursinus</i>	NC_009970	16,817
<i>Mirounga leonina</i>	NC_008422	16,970
<i>Monachus schauinslandi</i>	NC_008421	16,754
<i>Neofelis nebulosa</i>	NC_008450	16,844
<i>Neophoca cinerea</i>	NC_008419	16,736
<i>Nyctereutes procyonoides</i>	NC_013700	16,713
<i>Odobenus rosmarus rosmarus</i>	NC_004029	16,565
<i>Panthera leo persica</i>	NC_018053	16,818
<i>Panthera pardus</i>	NC_010641	16,964
<i>Panthera tigris</i>		
<i>Panthera tigris</i>	NC_010642	16,990
<i>Panthera tigris amoyensis</i>	NC_014770	17,001
<i>Phoca fasciata</i>	NC_008428	16,694
<i>Phoca groenlandica</i>	NC_008429	16,551
<i>Phoca largha</i>	NC_008430	16,701
<i>Phoca vitulina</i>	NC_001325	16,826
<i>Phocarcotus hookeri</i>	NC_008418	16,722

<i>Prionailurus bengalensis</i>	NC_016189	16,990
<i>euphilurus</i>		
<i>Procyon lotor</i>	NC_009126	16,623
<i>Puma concolor</i>	NC_016470	17,153
<i>Pusa caspica</i>	NC_008431	16,737
<i>Pusa hispida</i>	NC_008433	16,754
<i>Pusa sibirica</i>	NC_008432	16,773
<i>Spilogale putorius</i>	NC_010497	16,597
<i>Tremarctos ornatus</i>	NC_009969	16,766
<i>Uncia uncia</i>	NC_010638	16,773
<i>Ursus americanus</i>	NC_003426	16,841
<i>Ursus arctos</i>	NC_003427	17,020
<i>Ursus maritimus</i>	NC_003428	17,017
<i>Ursus spelaeus</i>	NC_011112	16,780
<i>Ursus thibetanus</i>		
<i>Ursus thibetanus</i>	NC_009971	16,795
<i>Ursus thibetanus formosanus</i>	NC_009331	17,044
<i>Ursus thibetanus mupinensis</i>	NC_008753	16,868
<i>Ursus thibetanus thibetanus</i>	NC_011118	17,034
<i>Ursus thibetanus ussuricus</i>	NC_011117	16,824
<i>Vulpes vulpes</i>	NC_008434	16,813
<i>Zalophus californianus</i>	NC_008416	16,677
<i>Cetartiodactyla</i>		
<i>Ammotragus lervia</i>	NC_009510	16,530
<i>Antilope cervicapra</i>	NC_012098	16,431
<i>Balaena mysticetus</i>	NC_005268	16,390
<i>Balaenoptera acutorostrata</i>	NC_005271	16,417
<i>Balaenoptera bonaerensis</i>	NC_006926	16,421
<i>Balaenoptera borealis</i>	NC_006929	16,410
<i>Balaenoptera brydei</i>	NC_006928	16,408
<i>Balaenoptera edeni</i>	NC_007938	16,409
<i>Balaenoptera musculus</i>	NC_001601	16,402
<i>Balaenoptera omurai</i>	NC_007937	16,404
<i>Balaenoptera physalus</i>	NC_001321	16,398
<i>Berardius bairdii</i>	NC_005274	16,346
<i>Bison bison</i>	NC_012346	16,319
<i>Bison bonasus</i>	NC_014044	16,326
<i>Bos grunniens</i>	NC_006380	16,323
<i>Bos indicus</i>	NC_005971	16,341
<i>Bos javanicus</i>	NC_012706	16,339
<i>Bos primigenius</i>	NC_013996	16,338
<i>Bos taurus</i>	NC_006853	16,338
<i>Bubalus bubalis</i>	NC_006295	16,359
<i>Budorcas taxicolor</i>	NC_013069	16,667
<i>Camelus bactrianus</i>	NC_009628	16,659
<i>Camelus dromedarius</i>	NC_009849	16,643
<i>Camelus ferus</i>	NC_009629	16,680
<i>Caperea marginata</i>	NC_005269	16,383
<i>Capra hircus</i>	NC_005044	16,643
<i>Capricornis crispus</i>	NC_012096	16,453

<i>Capricornis swinhoei</i>	NC_010640	16,524
<i>Cervus elaphus</i>		
<i>Cervus elaphus</i>	NC_007704	16,357
<i>Cervus elaphus songaricus</i>	NC_014703	16,419
<i>Cervus elaphus xanthopygus</i>	NC_013836	16,416
<i>Cervus elaphus yarkandensis</i>	NC_013840	16,351
<i>Cervus nippon</i>		
<i>Cervus nippon centralis</i>	NC_006993	16,663
<i>Cervus nippon kopschi</i>	NC_016178	16,429
<i>Cervus nippon sichuanicus</i>	NC_018595	16,429
<i>Cervus nippon yakushimae</i>	NC_007179	16,440
<i>Cervus nippon hortulorum</i>	NC_013834	16,434
<i>Cervus nippon taiouanus</i>	NC_008462	16,436
<i>Cervus nippon yesoensis</i>	NC_006973	16,543
<i>Delphinus capensis</i>	NC_012061	16,385
<i>Elaphodus cephalophus</i>	NC_008749	16,347
<i>Elaphurus davidianus</i>		32,710
<i>Eschrichtius robustus</i>	NC_005270	16,412
<i>Eubalaena australis</i>	NC_006930	16,385
<i>Eubalaena japonica</i>	NC_006931	16,386
<i>Feresa attenuata</i>	NC_019588	16,387
<i>Giraffa camelopardalis angolensis</i>	NC_012100	16,433
<i>Globicephala macrorhynchus</i>	NC_019578	16,387
<i>Globicephala melas</i>	NC_019441	16,390
<i>Grampus griseus</i>	NC_012062	16,386
<i>Hippopotamus amphibius</i>	NC_000889	16,407
<i>Hydropotes inermis</i>		
<i>Hydropotes inermis</i>	NC_011821	16,355
<i>Hydropotes inermis argyropus</i>	NC_018032	16,356
<i>Hyperoodon ampullatus</i>	NC_005273	16,346
<i>Inia geoffrensis</i>	NC_005276	16,588
<i>Kobus leche</i>	NC_018603	16,426
<i>Kogia breviceps</i>	NC_005272	16,406
<i>Lagenorhynchus albirostris</i>	NC_005278	16,393
<i>Lama glama</i>	NC_012102	16,597
<i>Lama guanicoe</i>	NC_011822	16,649
<i>Lama pacos</i>	NC_002504	16,652
<i>Lipotes vexillifer</i>	NC_007629	16,392
<i>Megaptera novaeangliae</i>	NC_006927	16,398
<i>Monodon monoceros</i>	NC_005279	16,383
<i>Moschus anhuiensis</i>	NC_020017	16,351
<i>Moschus berezovskii</i>	NC_012694	16,354
<i>Moschus chrysogaster</i>	NC_020093	16,353
<i>Moschus moschiferus</i>	NC_013753	16,351
<i>Muntiacus crinifrons</i>	NC_004577	16,357
<i>Muntiacus muntjak</i>	NC_004563	16,354
<i>Muntiacus reevesi</i>		
<i>Muntiacus reevesi</i>	NC_004069	16,354
<i>Muntiacus reevesi micrurus</i>	NC_008491	16,353

<i>Muntiacus vuquangensis</i>	NC_016920	16,361	<i>Plecotus rafinesquii</i>	NC_016872	16,438
<i>Naemorhedus caudatus</i>	NC_013751	16,519	<i>Pteropus dasymallus</i>	NC_002612	16,705
<i>Odocoileus virginianus</i>	NC_015247	16,477	<i>Pteropus scapulatus</i>	NC_002619	16,741
<i>Orcaella brevirostris</i>	NC_019590	16,383	<i>Rhinolophus ferrumequinum korai</i>	NC_016191	16,839
<i>Orcaella heinsohni</i>	NC_019591	16,389	<i>Rhinolophus formosae</i>	NC_011304	16,915
<i>Orcinus orca</i>	NC_014682	16,388	<i>Rhinolophus luctus</i>	NC_018539	16,775
<i>Oryx dammah</i>	NC_016421	16,756	<i>Rhinolophus monoceros</i>	NC_005433	16,851
<i>Oryx gazella</i>	NC_016422	16,661	<i>Rhinolophus pumilus</i>	NC_005434	16,869
<i>Ovis aries</i>	NC_001941	16,616	<i>Rousettus aegyptiacus</i>	NC_007393	16,706
<i>Ovis canadensis</i>	NC_015889	16,463	Insectivora		
<i>Pantholops hodgsonii</i>	NC_007441	16,498	<i>Crocidura russula</i>	NC_006893	17,202
<i>Pecari tajacu</i>	NC_012103	16,839	<i>Echinosorex gymnura</i>	NC_002808	17,088
<i>Peponocephala electra</i>	NC_019589	16,388	<i>Episoriculus fumidus</i>	NC_003040	17,488
<i>Phacochoerus africanus</i>	NC_008830	16,719	<i>Erinaceus europaeus</i>	NC_002080	17,447
<i>Phocoena phocoena</i>	NC_005280	16,382	<i>Galemys pyrenaicus</i>	NC_008156	16,510
<i>Physeter catodon</i>	NC_002503	16,428	<i>Hemiechinus auritus</i>	NC_005033	17,283
<i>Platanista minor</i>	NC_005275	16,324	<i>Hylomys suillus</i>	NC_010298	15,353
<i>Pontoporia blainvillei</i>	NC_005277	16,593	<i>Mogera wogura</i>	NC_005035	16,814
<i>Procapra przewalskii</i>	NC_014875	16,548	<i>Neotetracus sinensis</i>	NC_019626	16,982
<i>Przewalskium albirostris</i>	NC_016707	16,478	<i>Sorex unguiculatus</i>	NC_005435	17,105
<i>Pseudois schaeferi</i>	NC_016689	16,663	<i>Talpa europaea</i>	NC_002391	16,884
<i>Pseudorca crassidens</i>	NC_019577	16,392	<i>Uropsilus gracilis</i>	NC_018598	16,530
<i>Rangifer tarandus</i>	NC_007703	16,362	<i>Urotrichus talpoides</i>	NC_005034	16,904
<i>Rucervus eldi</i>	NC_014701	16,357	Perissodactyla		
<i>Rusa unicolor swinhoei</i>	NC_008414	16,505	<i>Ceratotherium simum</i>	NC_001808	16,832
<i>Sousa chinensis</i>	NC_012057	16,388	<i>Coelodonta antiquitatis</i>	NC_012681	16,436
<i>Stenella attenuata</i>	NC_012051	16,386	<i>Dicerorhinus sumatrensis</i>	NC_012684	16,466
<i>Stenella coeruleoalba</i>	NC_012053	16,384	<i>Diceros bicornis</i>	NC_012682	16,411
<i>Sus scrofa</i>			<i>Equus asinus</i>	NC_001788	16,670
<i>Sus scrofa</i>	NC_000845	16,613	<i>Equus caballus</i>	NC_001640	16,660
<i>Sus scrofa domesticus</i>	NC_012095	16,770	<i>Equus hemionus</i>	NC_016061	16,634
<i>Sus scrofa taiwanensis</i>	NC_014692	16,480	<i>Rhinoceros sondaicus</i>	NC_012683	16,417
<i>Tursiops aduncus</i>	NC_012058	16,385	<i>Rhinoceros unicornis</i>	NC_001779	16,829
<i>Tursiops truncatus</i>	NC_012059	16,388	Pholidota		
<i>Vicugna vicugna</i>	NC_013558	16,084	<i>Manis pentadactyla</i>	NC_016008	16,578
Chiroptera			<i>Manis tetradactyla</i>	NC_004027	16,571
<i>Artibeus jamaicensis</i>	NC_002009	16,651	Xenarthra		
<i>Artibeus lituratus</i>	NC_016871	16,709	Cingulata		
<i>Chalinolobus tuberculatus</i>	NC_002626	16,818	<i>Dasypus novemcinctus</i>	NC_001821	17,056
<i>Hipposideros armiger</i>	NC_018540	16,784	Pilosa		
<i>Lasiurus borealis</i>	NC_016873	17,048	<i>Bradypus tridactylus</i>	NC_006923	16,975
<i>Myotis formosus</i>	NC_015828	17,159	<i>Choloepus didactylus</i>	NC_006924	16,543
<i>Mystacina tuberculata</i>	NC_006925	16,784	<i>Tamandua tetradactyla</i>	NC_004032	16,395
<i>Pipistrellus abramus</i>	NC_005436	16,976			
<i>Plecotus auritus</i>	NC_015484	16,900			

Table 0.9 Optimization results of HCO sequence data.

alignment position	Common Alignments (Mode)					Frequency of the residue					Frequency within solution subsets
	A1	A2	B	C	NOR	A1	A2	B	C	NOR	
1	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
2	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
3	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
4	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
5	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
6	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
7	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
8	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
9	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
10	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
11	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
12	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
13	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
14	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
15	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
16	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
17	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
18	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
19	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
20	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
21	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
22	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
23	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
24	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
25	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
26	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
27	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
28	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
29	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
30	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
31	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
32	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
33	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
34	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
35	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
36	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
37	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
38	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
39	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
40	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
41	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
42	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
43	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
44	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
45	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
46	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
47	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
48	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
49	-	-	-	-	-	1.00	1.00	1.00	1.00	0.99	0.00
50	-	-	-	-	-	1.00	1.00	1.00	1.00	0.97	0.00
51	-	-	-	-	-	1.00	1.00	1.00	1.00	0.97	0.00
52	-	-	-	-	-	1.00	1.00	1.00	1.00	0.94	0.00
53	-	-	-	-	-	1.00	1.00	1.00	1.00	0.92	0.00
54	-	-	-	-	-	1.00	1.00	1.00	1.00	0.91	0.00
55	-	-	-	-	-	1.00	1.00	1.00	1.00	0.88	0.00
56	-	-	-	-	-	1.00	1.00	1.00	1.00	0.88	0.00
57	-	-	-	-	-	1.00	1.00	1.00	1.00	0.88	0.00
58	-	-	-	-	-	1.00	1.00	1.00	1.00	0.88	0.00
59	-	-	-	-	-	1.00	1.00	1.00	1.00	0.87	0.00
60	-	-	-	-	-	1.00	1.00	1.00	1.00	0.86	0.00
61	-	-	-	-	-	1.00	1.00	1.00	1.00	0.85	0.00
62	-	-	-	-	-	1.00	1.00	1.00	1.00	0.85	0.00
63	-	-	-	-	-	1.00	1.00	1.00	1.00	0.84	0.00
64	-	-	-	-	-	1.00	1.00	1.00	1.00	0.84	0.00
65	-	-	-	-	-	1.00	1.00	1.00	1.00	0.82	0.00
66	-	-	-	-	-	1.00	1.00	1.00	1.00	0.82	0.00
67	-	-	-	-	-	1.00	1.00	1.00	1.00	0.80	0.00
68	-	-	-	-	-	1.00	1.00	1.00	1.00	0.54	0.00
69	-	-	-	-	-	1.00	1.00	1.00	1.00	0.52	0.00
70	-	-	-	-	-	1.00	1.00	1.00	1.00	0.49	0.00
71	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
72	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
73	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
74	-	-	-	-	-	1.00	1.00	1.00	1.00	0.49	0.00
75	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
76	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
77	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
78	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
79	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
80	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
81	-	-	-	-	-	1.00	1.00	1.00	1.00	0.48	0.00
82	-	-	-	-	-	1.00	1.00	1.00	1.00	0.47	0.00
83	-	-	-	-	-	1.00	1.00	1.00	1.00	0.47	0.00

- 140 -

- 141 -

268	-	-	-	-	-	1.00	1.00	1.00	0.91	0.46	0.00	314	G	T	-	-	-	0.30	0.45	0.78	0.72	0.45	0.04
269	-	-	-	-	-	1.00	1.00	1.00	0.91	0.46	0.00	315	-	I	-	-	P	0.27	0.15	0.78	0.95	0.47	0.03
270	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	316	L	T	-	-	-	0.33	0.23	0.83	0.95	0.45	0.03
271	-	-	-	-	-	1.00	1.00	1.00	0.94	0.46	0.00	317	-	A	-	-	-	0.28	0.17	0.92	0.96	0.45	0.01
272	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	318	-	A	-	-	-	0.30	0.15	0.97	0.96	0.45	0.01
273	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	319	-	-	-	-	N	0.32	0.26	0.97	0.97	0.52	0.03
274	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	320	-	-	-	-	-	0.32	0.34	0.94	0.96	0.45	0.00
275	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	321	-	-	-	-	-	0.70	0.36	0.96	0.97	0.45	0.00
276	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	322	-	-	-	-	-	0.77	0.41	0.95	0.98	0.45	0.00
277	-	-	-	-	-	1.00	1.00	1.00	0.95	0.46	0.00	323	-	-	-	-	-	0.79	0.42	0.95	0.99	0.45	0.00
278	-	-	-	-	-	1.00	0.99	1.00	0.93	0.46	0.00	324	-	-	-	-	-	0.88	0.55	0.95	0.99	0.45	0.00
279	-	-	-	-	-	1.00	0.99	1.00	0.93	0.46	0.00	325	-	-	-	-	-	0.92	0.53	0.95	0.99	0.44	0.00
280	-	-	-	-	-	1.00	0.99	1.00	0.93	0.46	0.00	326	-	-	-	-	-	0.91	0.63	0.95	0.99	0.44	0.00
281	-	-	-	-	-	1.00	0.99	1.00	0.92	0.46	0.00	327	-	-	-	-	-	0.93	0.73	0.95	0.99	0.44	0.00
282	-	-	-	-	-	1.00	0.99	1.00	0.92	0.46	0.00	328	-	-	-	-	W	0.93	0.74	0.97	0.99	0.53	0.03
283	-	-	-	-	-	1.00	0.99	1.00	0.89	0.46	0.00	329	-	-	-	-	S	0.65	0.91	0.94	0.99	0.46	0.02
284	-	-	-	-	F	1.00	0.99	1.00	0.85	0.53	0.03	330	-	-	-	-	-	0.64	0.98	0.94	0.99	0.44	0.00
285	-	-	-	-	-	1.00	0.99	1.00	0.85	0.46	0.00	331	-	-	-	-	-	0.64	0.87	0.94	0.79	0.44	0.00
286	-	-	-	-	-	1.00	0.99	1.00	0.84	0.46	0.00	332	-	-	-	-	S	0.66	0.86	0.93	0.79	0.50	0.03
287	-	-	-	-	W	1.00	0.99	1.00	0.85	0.49	0.02	333	-	-	-	-	-	0.89	1.00	0.97	0.79	0.44	0.00
288	-	-	-	-	-	1.00	0.99	1.00	0.84	0.46	0.00	334	-	-	-	-	-	0.89	1.00	0.97	0.79	0.44	0.00
289	-	-	-	-	-	0.99	0.99	1.00	0.82	0.46	0.00	335	-	-	-	-	-	0.89	1.00	0.95	0.79	0.44	0.00
290	-	-	-	-	W	0.99	0.99	1.00	0.71	0.46	0.02	336	-	-	-	-	L	0.89	1.00	0.89	0.79	0.46	0.02
291	-	-	-	-	-	0.99	0.99	1.00	0.71	0.45	0.00	337	-	-	-	-	-	0.88	1.00	0.89	0.79	0.44	0.00
292	-	-	-	-	-	0.99	0.99	1.00	0.71	0.45	0.00	338	-	-	-	-	-	0.51	1.00	0.78	0.76	0.44	0.00
293	-	-	-	-	-	0.98	0.99	1.00	0.71	0.45	0.00	339	-	-	-	-	-	0.50	1.00	0.76	0.77	0.44	0.00
294	-	-	-	-	-	0.99	0.99	1.00	0.70	0.45	0.00	340	-	-	-	-	-	0.49	1.00	0.76	0.77	0.44	0.00
295	-	-	-	-	-	0.99	0.99	1.00	0.70	0.45	0.00	341	-	-	-	-	-	0.49	1.00	0.74	0.76	0.44	0.00
296	-	-	-	-	R	0.99	0.97	1.00	0.70	0.52	0.03	342	-	-	-	-	-	0.52	1.00	0.71	0.75	0.44	0.00
297	-	-	-	-	P	0.98	0.97	1.00	0.70	0.49	0.02	343	-	-	-	-	-	0.52	1.00	0.71	0.75	0.44	0.00
298	-	-	-	-	-	0.97	0.97	1.00	0.69	1.00	0.00	344	-	-	-	-	-	0.51	1.00	0.69	0.73	0.44	0.00
299	-	-	-	-	-	0.97	0.97	1.00	0.69	1.00	0.00	345	-	-	-	-	-	0.53	1.00	0.67	0.73	0.44	0.00
300	-	-	-	-	-	0.97	0.97	1.00	0.69	0.98	0.00	346	-	-	-	-	-	0.52	1.00	0.66	0.73	0.43	0.00
301	-	-	-	-	-	0.97	0.97	1.00	0.69	0.77	0.00	347	-	-	-	-	-	0.51	1.00	0.66	0.72	0.43	0.00
302	-	-	-	-	-	0.93	0.97	1.00	0.68	0.47	0.00	348	-	-	-	-	-	0.50	1.00	0.66	0.72	0.43	0.00
303	-	-	-	-	-	0.93	0.97	1.00	0.64	0.48	0.00	349	-	-	-	-	-	0.44	0.92	0.57	0.72	0.43	0.00
304	-	-	-	-	-	0.91	0.97	1.00	0.55	0.61	0.00	350	-	-	-	-	-	0.43	0.91	0.54	0.72	0.44	0.00
305	-	-	-	M	-	0.90	0.97	1.00	0.51	0.47	0.03	351	-	-	-	-	-	0.35	0.94	0.54	0.72	0.46	0.00
306	-	-	-	S	-	0.90	0.96	1.00	0.21	0.47	0.01	352	-	-	-	-	-	0.31	0.94	0.57	0.74	0.51	0.00
307	-	-	-	T	Y	0.88	0.96	0.98	0.16	0.48	0.03	353	-	-	-	-	-	0.30	0.88	0.56	0.74	0.61	0.00
308	-	-	-	A	T	0.87	0.96	0.98	0.16	0.52	0.03	354	A	-	-	-	-	0.30	0.78	0.53	0.74	0.76	0.01
309	-	-	-	A	-	0.84	0.94	0.80	0.14	0.45	0.01	355	-	-	-	-	-	0.21	0.77	0.49	0.76	0.72	0.00
310	-	-	-	A	N	0.82	0.92	0.80	0.25	0.52	0.04	356	H	-	-	-	-	0.16	0.53	0.47	0.78	0.56	0.01
311	-	-	-	-	W	0.78	0.90	0.80	0.25	0.50	0.02	357	G	-	-	-	-	0.22	0.47	0.43	0.77	0.52	0.01
312	-	-	-	-	P	0.56	0.89	0.78	0.56	0.55	0.03	358	A	-	-	-	-	0.22	0.45	0.42	0.77	0.54	0.01
313	M	M	-	-	-	0.29	0.82	0.78	0.65	0.45	0.08	359	I	-	-	-	-	0.23	0.44	0.36	0.77	0.49	0.01

360	T	-	-	-	-	0.39	0.44	0.35	0.79	0.59	0.02	406	Q	E	Q	Q	Y	0.56	0.81	0.35	0.78	0.52	0.52
361	Y	-	-	-	-	0.27	0.20	0.32	0.79	0.56	0.01	407	L	L	R	-	Y	0.73	0.97	0.43	0.96	0.51	0.60
362	-	Y	-	-	-	0.22	0.27	0.34	0.76	0.79	0.01	408	A	A	A	L	V	0.53	0.29	0.34	0.84	0.69	0.45
363	-	L	-	-	-	0.36	0.42	0.36	0.76	0.75	0.02	409	L	T	-	A	E	0.22	0.59	0.49	0.38	0.47	0.08
364	-	Q	-	-	-	0.34	0.20	0.54	0.76	0.55	0.01	410	P	P	-	W	-	0.64	0.92	0.53	0.47	0.40	0.53
365	W	P	-	-	-	0.37	0.24	0.53	0.75	0.46	0.03	411	G	G	-	P	-	0.44	0.20	0.44	0.94	0.42	0.10
366	-	G	-	-	-	0.26	0.20	0.56	0.75	0.46	0.01	412	L	-	-	-	P	0.26	0.52	0.44	0.39	0.29	0.03
367	Y	K	-	-	-	0.22	0.19	0.58	0.72	0.40	0.02	413	-	-	-	-	-	0.35	0.52	0.61	0.96	0.99	0.00
368	L	W	-	-	-	0.39	0.51	0.53	0.76	0.39	0.05	414	-	-	-	-	-	0.45	0.52	0.62	0.97	0.99	0.00
369	W	L	-	-	P	0.53	0.32	0.52	0.73	0.36	0.06	415	Q	-	-	-	-	0.25	0.53	0.65	0.98	0.99	0.01
370	R	R	-	-	-	0.19	0.32	0.43	0.81	0.34	0.04	416	-	-	-	-	-	0.49	0.58	0.65	1.00	1.00	0.00
371	R	Y	-	-	-	0.38	0.41	0.44	0.74	0.33	0.04	417	-	-	-	-	-	0.50	0.97	0.65	1.00	1.00	0.00
372	W	F	-	-	-	0.75	0.45	0.34	0.72	0.33	0.06	418	-	-	-	-	-	0.65	0.99	0.89	1.00	1.00	0.00
373	L	L	L	-	-	0.50	0.31	0.32	0.71	0.36	0.11	419	-	-	-	-	-	0.70	0.99	0.89	1.00	1.00	0.00
374	T	F	-	-	M	0.53	0.68	0.29	0.57	0.30	0.38	420	-	-	-	-	-	0.90	1.00	0.89	1.00	1.00	0.00
375	S	S	T	-	L	0.51	0.41	0.22	0.44	0.30	0.09	421	-	-	-	-	-	0.91	1.00	0.99	0.93	1.00	0.00
376	T	T	V	-	T	0.45	0.48	0.22	0.34	0.53	0.11	422	-	-	-	L	-	0.93	1.00	1.00	0.50	1.00	0.03
377	D	D	D	Y	P	0.59	0.97	0.27	0.84	0.41	0.65	423	-	-	-	N	-	0.94	1.00	1.00	0.48	1.00	0.02
378	H	H	H	N	S	0.93	0.92	0.23	0.49	0.81	0.74	424	-	-	-	-	-	0.94	0.99	1.00	0.71	1.00	0.00
379	K	K	K	Y	Q	0.93	0.97	0.32	0.46	0.84	0.76	425	-	-	-	-	-	0.97	0.99	1.00	0.72	1.00	0.00
380	D	V	R	K	K	0.35	0.50	0.41	0.28	0.40	0.11	426	-	-	-	-	-	0.97	0.99	1.00	0.99	1.00	0.00
381	I	I	L	V	A	0.81	0.95	0.40	0.56	0.39	0.64	427	-	-	-	-	-	0.97	0.99	1.00	1.00	1.00	0.00
382	G	G	V	V	A	0.88	0.85	0.28	0.69	0.45	0.66	428	-	-	-	-	-	0.98	0.99	1.00	1.00	1.00	0.00
383	I	I	L	R	G	0.37	0.59	0.45	0.62	0.27	0.14	429	-	-	-	-	-	0.98	0.98	1.00	1.00	1.00	0.00
384	M	Q	A	Q	K	0.56	0.52	0.28	0.39	0.44	0.38	430	-	-	-	-	-	0.97	0.98	1.00	1.00	1.00	0.00
385	Y	Y	Y	F	Y	0.96	0.90	0.32	0.70	0.86	0.76	431	-	-	-	-	-	0.94	0.98	1.00	0.99	1.00	0.00
386	L	L	L	A	F	0.46	0.67	0.34	0.33	0.80	0.21	432	-	-	-	-	-	0.94	0.98	1.00	0.97	1.00	0.00
387	I	V	V	I	Y	0.46	0.53	0.35	0.41	0.19	0.12	433	-	-	-	-	-	0.94	0.98	1.00	0.97	1.00	0.00
388	F	T	L	A	V	0.29	0.41	0.28	0.37	0.41	0.09	434	-	-	-	-	-	1.00	0.98	1.00	1.00	1.00	0.00
389	A	I	A	T	A	0.53	0.24	0.59	0.64	0.56	0.38	435	-	-	-	-	-	1.00	0.99	1.00	1.00	1.00	0.00
390	F	F	F	V	L	0.31	0.50	0.57	0.45	0.26	0.18	436	-	-	-	-	-	1.00	0.99	1.00	1.00	1.00	0.00
391	V	L	A	V	A	0.32	0.25	0.28	0.33	0.31	0.07	437	-	-	-	-	-	1.00	0.99	1.00	1.00	1.00	0.00
392	M	F	A	W	L	0.56	0.80	0.32	0.85	0.88	0.51	438	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
393	L	F	L	G	F	0.39	0.55	0.70	0.78	0.75	0.48	439	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
394	L	L	L	I	L	0.52	0.61	0.34	0.54	0.38	0.48	440	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
395	R	I	L	V	A	0.38	0.46	0.38	0.58	0.23	0.10	441	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
396	G	G	G	G	Q	0.88	0.90	0.65	0.92	0.98	0.82	442	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
397	G	G	G	M	V	0.57	0.87	0.58	0.62	0.41	0.62	443	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
398	A	L	L	L	L	0.41	0.27	0.43	0.40	0.49	0.16	444	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
399	D	L	E	V	F	0.38	0.40	0.21	0.53	0.33	0.09	445	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
400	A	A	G	G	G	0.74	0.83	0.53	0.95	0.99	0.73	446	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
401	L	M	L	V	L	0.36	0.23	0.53	0.65	0.30	0.10	447	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
402	M	L	L	L	I	0.36	0.37	0.22	0.36	0.35	0.11	448	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
403	M	I	Q	I	T	0.55	0.49	0.41	0.67	0.38	0.39	449	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
404	R	R	R	A	A	0.98	0.99	0.46	0.89	0.52	0.84	450	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
405	A	L	L	A	H	0.32	0.28	0.42	0.43	0.57	0.12	451	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00

452	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	498	-	-	-	S	P	0.94	0.99	0.47	0.30	0.32	0.03
453	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	499	-	-	-	-	-	0.94	0.99	0.47	0.80	0.50	0.00
454	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	500	-	-	-	-	-	0.94	0.99	0.47	0.97	0.96	0.00
455	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	501	-	-	-	-	-	0.93	0.99	0.65	0.99	0.96	0.00
456	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	502	-	-	-	-	-	0.93	1.00	0.89	0.99	0.96	0.00
457	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	503	-	-	-	-	-	0.92	0.49	1.00	0.99	0.96	0.00
458	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	504	-	-	-	-	-	0.56	0.48	0.97	1.00	0.96	0.00
459	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	505	-	-	-	-	-	0.45	0.41	0.97	1.00	0.45	0.00
460	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	506	-	-	-	-	-	0.45	0.41	0.91	0.99	0.43	0.00
461	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	507	-	-	-	-	-	0.41	0.41	0.77	0.91	0.43	0.00
462	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	508	-	P	-	-	-	0.37	0.30	0.78	0.86	0.43	0.02
463	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	509	E	D	-	-	L	0.31	0.38	0.42	0.80	0.62	0.07
464	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	510	H	T	Q	-	P	0.39	0.31	0.22	0.79	0.94	0.09
465	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	511	Y	Y	Y	F	F	0.65	0.99	0.79	0.74	0.58	0.76
466	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	512	N	N	Y	G	N	0.60	0.97	0.58	0.76	0.51	0.60
467	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	513	Q	Q	Q	R	V	0.70	0.27	0.34	0.96	0.37	0.45
468	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	514	I	L	L	L	A	0.35	0.36	0.24	0.82	0.28	0.14
469	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	515	F	F	L	R	R	0.54	0.58	0.45	0.97	0.92	0.60
470	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	516	T	T	T	P	T	0.90	0.93	0.84	0.90	0.41	0.79
471	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	517	A	L	M	L	W	0.37	0.50	0.22	0.76	0.35	0.11
472	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	518	H	H	H	H	H	1.00	0.99	0.99	0.99	1.00	0.99
473	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	519	G	G	G	T	T	0.91	0.87	0.78	0.90	0.57	0.81
474	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	520	V	T	V	N	Q	0.34	0.51	0.47	0.64	0.52	0.37
475	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	521	I	I	L	A	L	0.70	0.51	0.29	0.70	0.45	0.46
476	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	522	M	M	M	V	L	0.99	0.97	0.46	0.85	0.38	0.79
477	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	523	I	I	I	I	I	0.53	0.55	0.27	0.83	0.52	0.54
478	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	524	F	F	F	F	V	0.75	0.85	0.38	0.83	0.42	0.69
479	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	525	F	L	V	A	W	0.58	0.73	0.40	0.69	0.97	0.51
480	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	526	M	W	F	F	I	0.30	0.48	0.38	0.88	0.49	0.41
481	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	527	A	I	A	G	A	0.57	0.55	0.36	0.69	0.50	0.41
482	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	528	M	I	T	G	T	0.70	0.45	0.34	0.59	0.44	0.13
483	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	529	P	P	F	N	G	0.88	0.94	0.25	0.35	0.52	0.65
484	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	530	A	S	F	A	F	0.33	0.31	0.28	0.66	0.54	0.11
485	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	531	F	I	I	L	L	0.28	0.36	0.30	0.70	0.48	0.15
486	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00	532	I	P	S	F	G	0.22	0.39	0.25	0.52	0.58	0.10
487	-	-	-	-	-	0.99	1.00	1.00	1.00	1.00	0.00	533	G	G	G	A	A	0.88	0.55	0.62	0.75	0.51	0.65
488	-	-	-	-	-	0.99	1.00	1.00	1.00	1.00	0.00	534	-	A	-	T	G	0.79	0.29	0.46	0.70	0.54	0.08
489	-	-	-	-	-	0.99	1.00	1.00	1.00	1.00	0.00	535	L	F	F	S	L	0.50	0.64	0.27	0.71	0.45	0.43
490	-	-	-	-	-	1.00	1.00	1.00	0.95	1.00	0.00	536	A	G	Y	Y	Y	0.36	0.71	0.17	0.58	0.50	0.16
491	-	-	-	-	-	1.00	1.00	1.00	0.95	1.00	0.00	537	N	N	A	Y	L	0.93	0.98	0.25	0.84	0.50	0.70
492	-	-	-	F	-	0.99	1.00	0.51	0.43	1.00	0.02	538	W	Y	Y	V	A	0.30	0.44	0.36	0.54	0.38	0.12
493	-	-	-	D	-	0.99	0.99	0.47	0.36	1.00	0.02	539	V	L	L	V	P	0.31	0.74	0.26	0.72	0.96	0.47
494	-	-	-	L	F	0.99	0.99	0.47	0.31	0.48	0.04	540	V	I	W	Q	E	0.53	0.41	0.19	0.87	0.37	0.12
495	-	-	-	P	F	0.99	1.00	0.47	0.50	0.42	0.05	541	P	P	R	R	E	0.99	0.99	0.28	0.93	0.41	0.76
496	-	-	-	W	G	0.94	1.00	0.47	0.57	0.55	0.06	542	L	L	L	T	S	0.83	0.62	0.22	0.70	0.40	0.58
497	-	-	-	L	I	0.94	0.99	0.47	0.42	0.38	0.04	543	Q	M	L	C	G	0.64	0.68	0.45	0.55	0.51	0.47

544	I	I	L	Q	-	0.88	0.62	0.26	0.38	0.85	0.52
545	G	G	G	T	-	0.98	0.99	0.29	0.35	0.45	0.74
546	A	A	S	R	E	0.94	0.97	0.27	0.67	0.41	0.69
547	R	R	L	L	P	0.54	0.59	0.31	0.80	0.47	0.47
548	D	D	D	F	K	0.98	0.99	0.28	0.42	0.39	0.76
549	V	M	M	G	Y	0.52	0.55	0.21	0.32	0.32	0.38
550	A	A	-	-	Q	0.89	0.95	0.31	0.59	0.53	0.62
551	F	F	F	-	P	0.81	0.96	0.29	0.89	0.23	0.66
552	P	P	P	-	-	0.90	0.99	0.30	0.90	0.99	0.69
553	R	R	R	-	-	0.52	0.69	0.22	0.89	0.99	0.45
554	L	L	L	-	-	0.62	0.88	0.28	0.29	0.99	0.56
555	N	N	W	K	L	0.95	0.99	0.39	0.20	0.38	0.68
556	N	A	A	L	L	0.50	0.56	0.37	0.72	0.54	0.46
557	L	V	G	A	V	0.48	0.29	0.24	0.57	0.47	0.10
558	S	S	-	A	D	0.83	0.50	0.55	0.33	0.26	0.46
559	F	F	-	F	I	0.82	0.62	0.49	0.67	0.30	0.52
560	W	W	-	T	L	0.97	0.70	0.49	0.39	0.63	0.59
561	L	L	-	F	F	0.56	0.61	0.49	0.79	0.58	0.48
562	F	I	-	W	W	0.31	0.22	0.49	0.79	0.44	0.12
563	V	P	-	G	A	0.26	0.51	0.51	0.75	0.51	0.10
564	-	-	-	-	-	1.00	1.00	0.94	0.99	1.00	0.00
565	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
566	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
567	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
568	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
569	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
570	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
571	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
572	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
573	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
574	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
575	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
576	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
577	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
578	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
579	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
580	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
581	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00
582	-	-	-	W	L	1.00	1.00	0.51	0.60	0.54	0.06
583	-	-	-	Q	V	1.00	0.99	0.50	0.65	0.38	0.05
584	V	P	L	L	V	0.26	0.36	0.35	0.44	0.42	0.11
585	G	G	A	V	V	0.56	0.43	0.37	0.33	0.34	0.14
586	A	G	G	I	V	0.35	0.59	0.26	0.76	0.55	0.40
587	L	L	L	V	V	0.24	0.38	0.28	0.57	0.37	0.16
588	L	L	G	L	G	0.63	0.49	0.47	0.37	0.66	0.44
589	L	L	L	A	S	0.41	0.45	0.28	0.81	0.43	0.42
590	N	L	V	A	V	0.33	0.46	0.21	0.55	0.27	0.09
591	I	A	A	I	V	0.22	0.30	0.21	0.48	0.30	0.09
592	S	S	S	T	G	0.78	0.67	0.34	0.41	0.73	0.60
593	L	F	A	L	Y	0.50	0.34	0.29	0.63	0.37	0.11
594	F	F	V	P	L	0.48	0.50	0.20	0.53	0.36	0.37
595	V	V	P	L	L	0.38	0.26	0.23	0.57	0.48	0.14
596	G	G	G	G	G	0.48	0.49	0.28	0.97	0.43	0.53
597	G	G	L	Y	I	0.31	0.61	0.26	0.24	0.33	0.11
598	F	-	-	T	H	0.33	0.66	0.25	0.81	0.19	0.07
599	-	-	G	Q	G	0.70	0.99	0.37	0.37	0.30	0.05
600	-	-	-	-	-	0.70	0.99	0.78	0.98	0.24	0.00
601	-	-	-	-	L	0.71	1.00	0.82	0.99	0.36	0.02
602	-	-	-	-	-	0.89	1.00	0.94	0.99	0.29	0.00
603	-	-	-	-	-	0.99	1.00	0.95	0.99	0.33	0.00
604	-	-	-	-	-	1.00	1.00	0.95	0.99	0.74	0.00
605	-	-	-	-	-	1.00	1.00	0.95	1.00	0.95	0.00
606	-	-	-	-	-	1.00	1.00	0.95	1.00	1.00	0.00
607	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
608	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
609	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
610	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
611	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
612	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
613	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
614	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
615	-	-	-	-	-	1.00	1.00	0.95	1.00	0.59	0.00
616	-	-	-	-	-	0.99	1.00	0.99	1.00	0.28	0.00
617	-	-	-	-	-	0.99	1.00	0.99	1.00	0.25	0.00
618	-	-	-	-	F	0.98	1.00	0.99	1.00	0.47	0.02
619	-	-	-	-	W	0.62	0.33	0.77	1.00	0.38	0.02
620	A	A	P	-	-	0.47	0.41	0.21	1.00	0.29	0.08
621	D	D	A	-	G	0.19	0.42	0.47	0.99	0.54	0.10
622	T	T	S	-	-	0.35	0.41	0.21	0.99	0.33	0.07
623	G	G	V	-	Q	0.97	0.97	0.44	0.99	0.54	0.68
624	W	W	L	-	G	0.98	0.99	0.41	0.99	0.54	0.69
625	-	-	Y	S	G	0.98	0.99	0.36	0.51	0.41	0.06
626	T	T	T	K	E	0.51	0.83	0.53	0.78	0.42	0.59
627	A	F	F	E	-	0.43	0.40	0.41	0.96	0.58	0.13
628	Y	Y	Y	Y	Y	0.98	1.00	0.82	0.93	0.59	0.93
629	P	P	P	A	L	0.54	0.59	0.40	0.87	0.54	0.52
630	P	P	P	E	E	0.96	0.99	0.77	0.95	0.56	0.87
631	L	L	L	L	L	0.90	0.43	0.58	0.60	0.45	0.58
632	S	S	S	E	G	0.75	0.97	0.17	0.83	0.52	0.69
633	G	L	A	-	R	0.26	0.33	0.28	0.95	0.59	0.07
634	L	T	L	-	-	0.23	0.21	0.22	0.98	1.00	0.03
635	-	T	-	-	-	0.94	0.29	0.77	1.00	1.00	0.01

636	-	-	-	-	-	1.00	0.99	0.77	1.00	1.00	0.00	682	-	-	-	-	-	0.99	0.99	0.39	0.99	0.92	0.00
637	-	-	-	-	-	1.00	0.99	0.86	1.00	1.00	0.00	683	-	-	-	-	-	1.00	0.99	0.49	0.99	0.92	0.00
638	E	P	-	-	-	0.19	0.31	0.62	1.00	1.00	0.02	684	-	-	-	-	-	1.00	0.99	0.97	1.00	0.93	0.00
639	Y	T	-	-	-	0.27	0.23	0.62	1.00	1.00	0.02	685	-	-	-	-	-	1.00	0.99	0.97	1.00	0.93	0.00
640	S	G	-	-	-	0.49	0.27	0.62	1.00	1.00	0.04	686	-	-	-	-	-	1.00	0.99	0.98	1.00	0.93	0.00
641	P	-	-	-	-	0.61	0.34	0.62	1.00	1.00	0.03	687	-	-	-	-	-	1.00	0.99	0.98	1.00	0.93	0.00
642	G	-	-	-	-	0.63	0.52	0.61	1.00	1.00	0.03	688	-	-	-	-	-	1.00	0.99	0.98	1.00	0.93	0.00
643	-	-	-	-	-	0.97	0.81	0.98	1.00	1.00	0.00	689	-	-	-	-	-	1.00	0.99	0.98	1.00	0.93	0.00
644	-	-	-	-	-	0.99	0.81	0.99	1.00	1.00	0.00	690	-	-	-	-	-	1.00	0.99	0.98	1.00	1.00	0.00
645	-	-	-	-	-	0.99	0.98	0.99	1.00	1.00	0.00	691	-	-	-	-	-	1.00	0.99	0.98	1.00	1.00	0.00
646	-	-	-	-	-	0.94	0.97	0.99	1.00	1.00	0.00	692	-	-	-	-	-	1.00	0.99	0.98	1.00	1.00	0.00
647	-	-	-	-	-	0.94	0.97	0.99	1.00	1.00	0.00	693	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
648	V	-	-	-	-	0.38	0.96	0.65	1.00	1.00	0.02	694	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
649	G	G	P	-	-	0.70	0.47	0.26	0.99	1.00	0.10	695	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
650	V	V	I	-	-	0.48	0.41	0.24	0.99	1.00	0.08	696	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
651	D	I	F	-	-	0.76	0.23	0.40	0.99	1.00	0.07	697	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
652	Y	L	Y	-	F	0.38	0.36	0.44	0.97	0.21	0.07	698	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
653	W	W	L	W	W	0.31	0.52	0.23	0.92	0.56	0.44	699	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
654	I	I	G	P	Q	0.58	0.45	0.52	0.64	0.53	0.44	700	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
655	W	L	L	I	I	0.33	0.57	0.55	0.50	0.49	0.42	701	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
656	A	A	G	D	G	0.45	0.33	0.39	0.82	0.47	0.39	702	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
657	L	V	L	I	L	0.65	0.62	0.47	0.62	0.44	0.46	703	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
658	Q	F	V	L	V	0.41	0.37	0.30	0.44	0.36	0.09	704	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
659	I	L	L	I	V	0.49	0.54	0.39	0.48	0.49	0.40	705	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
660	S	L	V	T	G	0.43	0.38	0.36	0.47	0.55	0.11	706	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
661	G	G	G	I	L	0.87	1.00	0.67	0.39	0.41	0.77	707	-	-	-	-	-	1.00	0.99	0.99	1.00	1.00	0.00
662	I	F	S	V	L	0.32	0.42	0.27	0.74	0.65	0.36	708	-	-	-	-	-	1.00	0.99	0.98	1.00	1.00	0.00
663	G	S	W	W	I	0.55	0.93	0.41	0.88	0.30	0.55	709	A	A	-	-	-	0.79	0.57	0.53	1.00	1.00	0.44
664	T	S	I	V	W	0.49	0.92	0.25	0.79	0.54	0.50	710	P	P	-	-	-	0.88	0.90	0.53	1.00	1.00	0.58
665	I	I	-	-	L	0.44	0.93	0.50	0.97	0.69	0.14	711	G	G	-	-	-	0.93	0.85	0.53	1.00	1.00	0.58
666	L	L	L	A	F	0.55	0.49	0.17	0.34	0.40	0.42	712	M	M	-	-	-	0.94	0.73	0.53	1.00	0.99	0.54
667	G	G	G	Y	L	0.52	0.55	0.35	0.70	0.44	0.47	713	T	T	-	-	-	0.68	0.45	0.53	1.00	0.99	0.08
668	G	G	A	A	V	0.48	0.56	0.42	0.49	0.37	0.43	714	L	W	-	-	-	0.54	0.44	0.53	0.99	0.99	0.05
669	I	V	A	I	T	0.62	0.39	0.28	0.33	0.21	0.09	715	M	F	-	K	-	0.54	0.36	0.53	0.53	0.95	0.07
670	N	N	N	V	R	0.93	0.99	0.28	0.56	0.49	0.76	716	K	R	-	E	-	0.55	0.45	0.68	0.41	0.94	0.07
671	F	F	F	F	T	0.52	0.92	0.37	0.71	0.44	0.63	717	M	M	-	P	-	0.82	0.45	0.67	0.25	0.94	0.11
672	I	I	V	F	L	0.51	0.61	0.36	0.58	0.28	0.42	718	P	P	P	H	-	0.97	0.97	0.34	0.79	0.73	0.75
673	V	T	V	G	L	0.53	0.36	0.23	0.75	0.37	0.11	719	L	L	I	I	-	0.36	0.66	0.19	0.75	0.52	0.15
674	T	T	T	T	P	0.96	0.97	0.26	0.87	0.40	0.79	720	F	F	I	Y	-	0.90	0.69	0.25	0.91	0.46	0.58
675	I	I	L	I	A	0.74	0.69	0.22	0.53	0.40	0.52	721	T	V	P	V	-	0.47	0.40	0.45	0.88	0.46	0.11
676	L	H	L	V	L	0.51	0.38	0.22	0.19	0.43	0.09	722	W	W	L	A	L	0.97	0.97	0.44	0.65	0.42	0.71
677	N	K	W	K	K	0.38	0.52	0.44	0.60	0.37	0.14	723	T	A	G	N	T	0.46	0.59	0.24	0.71	0.20	0.11
678	M	M	K	R	R	0.75	0.59	0.33	0.95	0.38	0.55	724	I	M	T	W	N	0.19	0.34	0.41	0.97	0.20	0.31
679	R	R	R	-	K	0.97	0.94	0.34	0.96	0.43	0.72	725	L	Y	F	F	V	0.73	0.40	0.34	0.76	0.32	0.15
680	-	-	-	-	T	0.99	0.99	0.39	0.98	0.38	0.02	726	I	A	M	Y	L	0.31	0.64	0.31	0.46	0.61	0.12
681	-	-	-	-	T	0.99	0.99	0.39	0.98	0.21	0.01	727	T	T	A	G	L	0.71	0.73	0.33	0.42	0.33	0.52

728	A	S	V	A	L	0.36	0.66	0.24	0.73	0.24	0.11
729	F	I	V	F	S	0.26	0.41	0.25	0.64	0.46	0.10
730	L	I	T	I	L	0.80	0.48	0.21	0.72	0.40	0.45
731	I	Q	M	I	W	0.50	0.51	0.22	0.35	0.24	0.34
732	I	L	L	T	A	0.39	0.46	0.34	0.58	0.34	0.13
733	A	L	L	V	I	0.43	0.68	0.34	0.47	0.48	0.15
734	A	A	W	A	A	0.44	0.40	0.42	0.85	0.45	0.44
735	F	T	V	M	L	0.63	0.73	0.23	0.35	0.56	0.47
736	P	P	P	L	F	0.98	0.96	0.28	0.88	0.64	0.78
737	V	V	A	H	Y	0.52	0.73	0.28	0.71	0.63	0.49
738	L	L	V	I	L	0.95	0.93	0.28	0.61	0.38	0.67
739	T	A	S	V	F	0.46	0.51	0.22	0.63	0.41	0.11
740	V	G	G	N	G	0.32	0.30	0.41	0.88	0.50	0.38
741	A	T	A	N	L	0.53	0.46	0.47	0.58	0.42	0.12
742	L	L	A	-	Y	0.43	0.63	0.44	0.96	0.28	0.12
743	T	V	L	-	-	0.34	0.33	0.31	0.96	0.47	0.05
744	M	L	E	-	-	0.43	0.77	0.28	0.95	0.89	0.07
745	L	L	L	-	-	0.69	0.68	0.31	0.96	0.99	0.51
746	L	L	L	-	-	0.39	0.28	0.41	0.96	0.99	0.11
747	L	F	L	-	-	0.38	0.47	0.27	0.96	0.99	0.06
748	-	-	-	-	-	1.00	1.00	0.91	1.00	1.00	0.00
749	D	D	-	-	-	0.93	0.57	0.54	1.00	1.00	0.49
750	R	R	R	L	-	0.87	0.38	0.27	0.39	1.00	0.16
751	Y	I	L	A	-	0.35	0.20	0.25	0.43	1.00	0.06
752	F	F	L	I	-	0.54	0.29	0.37	0.32	1.00	0.10
753	G	G	P	P	-	0.82	0.68	0.42	0.92	1.00	0.59
754	T	T	W	V	-	0.62	0.57	0.24	0.65	1.00	0.43
755	H	G	S	-	-	0.42	0.50	0.32	0.48	1.00	0.06
756	F	F	L	S	-	0.86	0.57	0.36	0.30	1.00	0.50
757	F	F	G	L	-	0.92	0.96	0.41	0.35	1.00	0.65
758	T	D	-	G	-	0.39	0.45	0.52	0.28	1.00	0.06
759	N	P	-	-	-	0.29	0.95	0.53	0.64	1.00	0.06
760	-	-	-	-	-	0.98	0.98	0.91	0.98	1.00	0.00
761	-	-	-	-	-	1.00	1.00	0.99	0.99	1.00	0.00
762	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
763	-	-	-	-	-	0.93	0.99	0.93	0.99	1.00	0.00
764	-	-	-	-	-	0.93	0.99	0.93	0.99	1.00	0.00
765	-	-	-	-	-	0.94	0.99	1.00	0.99	1.00	0.00
766	-	-	-	-	-	0.94	0.99	1.00	0.93	1.00	0.00
767	-	-	-	-	-	0.94	0.98	1.00	0.93	1.00	0.00
768	-	-	-	-	-	0.94	0.98	1.00	0.92	1.00	0.00
769	-	-	-	-	-	0.94	0.98	0.99	0.92	1.00	0.00
770	-	-	-	-	-	0.94	0.98	0.99	0.89	1.00	0.00
771	-	-	-	K	-	0.94	0.98	0.99	0.69	1.00	0.03
772	-	-	-	S	-	0.94	0.97	0.57	0.84	0.43	0.04
773	-	-	-	Y	-	0.99	0.99	0.54	0.74	0.43	0.04
774	A	A	A	S	-	0.36	0.30	0.18	0.68	0.43	0.12
775	G	L	T	A	P	0.40	0.28	0.17	0.27	0.32	0.07
776	G	G	I	Y	T	0.87	0.96	0.22	0.57	0.35	0.65
777	G	G	D	A	N	0.98	0.99	0.34	0.44	0.55	0.71
778	D	D	R	G	L	0.40	0.68	0.16	0.93	0.51	0.43
779	P	P	P	V	A	0.56	0.77	0.24	0.47	0.33	0.53
780	M	V	L	Q	V	0.39	0.62	0.61	0.44	0.27	0.12
781	L	L	L	D	D	0.54	0.53	0.28	0.86	0.44	0.50
782	Y	Y	R	A	E	0.48	0.52	0.42	0.93	0.40	0.44
783	Q	Q	Q	M	Y	0.55	0.99	0.28	0.58	0.50	0.63
784	H	H	L	V	W	0.60	0.92	0.38	0.56	0.58	0.57
785	L	L	F	Q	R	0.75	0.53	0.52	0.87	0.60	0.52
786	F	F	W	W	W	0.69	0.99	0.40	0.96	0.92	0.74
787	W	W	W	W	W	0.99	1.00	0.56	0.94	0.74	0.92
788	F	F	F	Y	V	0.52	0.92	0.33	0.84	0.77	0.62
789	F	Y	G	G	V	0.59	0.99	0.39	0.93	0.82	0.61
790	G	S	H	H	H	0.99	0.99	0.53	0.98	1.00	0.80
791	H	H	H	N	L	0.99	0.99	0.46	0.94	0.93	0.83
792	P	P	P	A	W	0.99	0.99	0.32	0.88	0.96	0.82
793	E	A	V	V	V	0.99	0.95	0.52	0.91	0.97	0.79
794	V	V	Y	G	E	0.99	0.99	0.42	0.76	0.99	0.75
795	Y	Y	-	F	G	0.99	0.99	0.53	0.88	0.77	0.73
796	I	I	F	F	F	0.92	0.71	0.44	0.56	0.42	0.65
797	L	M	W	L	F	0.48	0.87	0.42	0.81	0.50	0.49
798	I	I	L	T	E	0.61	0.56	0.44	0.95	0.96	0.56
799	L	L	L	A	L	0.95	0.99	0.50	0.55	0.44	0.78
800	P	P	P	G	F	0.99	0.99	0.77	0.65	0.56	0.83
801	A	A	A	F	A	0.56	0.50	0.74	0.72	0.48	0.54
802	F	F	Y	L	T	0.88	0.55	0.39	0.84	0.42	0.55
803	G	G	G	G	A	0.97	0.96	0.34	0.75	0.42	0.79
804	I	V	A	M	V	0.63	0.39	0.38	0.67	0.33	0.12
805	F	I	W	M	L	0.35	0.41	0.33	0.73	0.46	0.34
806	S	S	Y	Y	A	0.92	0.92	0.64	0.97	0.83	0.76
807	E	E	T	Y	F	0.58	0.92	0.28	0.96	0.50	0.57
808	I	I	L	F	L	0.56	0.47	0.32	0.85	0.48	0.43
809	I	L	L	V	L	0.44	0.33	0.41	0.56	0.43	0.13
810	P	P	P	P	V	0.39	0.63	0.84	0.97	0.28	0.61
811	T	V	K	K	K	0.68	0.57	0.26	0.92	0.36	0.51
812	F	F	F	Q	L	0.86	0.43	0.25	0.58	0.65	0.21
813	S	A	A	A	G	0.67	0.51	0.27	0.70	0.51	0.47
814	R	R	G	G	L	0.67	0.91	0.54	0.38	0.43	0.60
815	-	-	-	-	-	1.00	1.00	0.99	0.99	0.99	0.00
816	-	-	-	-	-	1.00	1.00	0.99	0.99	0.99	0.00
817	K	K	R	R	V	0.91	0.85	0.45	0.79	0.59	0.69
818	P	P	P	P	D	0.47	0.71	0.40	0.95	0.35	0.56
819	L	I	L	V	R	0.65	0.54	0.64	0.61	0.58	0.45

820	F	F	F	Y	E	0.92	0.99	0.29	0.79	0.42	0.76
821	G	G	S	S	V	0.98	0.95	0.61	0.96	0.35	0.76
822	Y	Y	D	Y	A	0.94	0.99	0.45	0.89	0.52	0.72
823	T	K	K	R	E	0.24	0.36	0.20	0.72	0.49	0.11
824	S	A	M	L	K	0.48	0.34	0.28	0.94	0.40	0.36
825	M	I	A	S	-	0.69	0.45	0.49	0.92	0.99	0.13
826	V	A	R	-	-	0.81	0.67	0.56	0.97	0.99	0.47
827	Y	Y	V	-	-	0.52	0.46	0.28	0.97	0.99	0.09
828	A	S	S	-	-	0.88	0.78	0.47	0.97	0.99	0.55
829	-	-	-	-	-	0.99	0.99	0.78	0.99	1.00	0.00
830	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
831	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
832	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
833	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
834	-	-	-	-	-	0.99	0.99	0.47	0.99	1.00	0.00
835	-	-	-	-	W	0.99	0.99	0.47	0.99	0.42	0.02
836	-	-	-	I	L	0.99	0.99	0.46	0.72	0.46	0.06
837	T	S	F	V	Y	0.57	0.66	0.41	0.55	0.53	0.44
838	V	L	L	H	V	0.35	0.38	0.52	0.75	0.34	0.37
839	A	A	L	F	I	0.43	0.57	0.27	0.96	0.34	0.41
840	I	I	L	W	V	0.92	0.98	0.25	0.96	0.28	0.69
841	A	A	S	A	I	0.35	0.52	0.33	0.62	0.26	0.13
842	V	V	L	L	L	0.31	0.31	0.21	0.88	0.52	0.41
843	L	L	P	I	F	0.83	0.48	0.42	0.65	0.34	0.50
844	S	G	I	F	L	0.71	0.71	0.24	0.58	0.63	0.50
845	F	L	G	L	F	0.77	0.39	0.46	0.42	0.33	0.12
846	I	L	F	Y	G	0.26	0.31	0.21	0.96	0.44	0.33
847	V	V	-	I	G	0.93	0.90	0.53	0.59	0.99	0.67
848	W	W	-	W	I	0.99	0.99	0.53	0.90	0.54	0.71
849	A	A	-	A	L	0.52	0.38	0.53	0.85	0.49	0.13
850	H	H	-	G	G	0.99	0.99	0.53	0.97	0.99	0.79
851	H	H	-	-	-	0.99	0.99	0.54	0.98	1.00	0.64
852	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
853	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
854	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
855	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
856	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
857	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
858	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
859	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
860	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
861	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
862	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
863	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
864	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
865	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
866	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
867	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
868	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
869	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
870	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
871	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
872	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
873	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
874	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
875	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
876	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
877	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
878	-	-	-	-	-	1.00	1.00	1.00	0.99	1.00	0.00
879	-	-	-	-	T	1.00	1.00	1.00	0.99	0.89	0.04
880	-	-	-	P	G	0.99	0.99	0.47	0.67	0.52	0.06
881	-	-	H	H	H	0.99	0.99	0.50	0.97	0.99	0.45
882	-	-	-	H	H	0.99	0.99	0.46	0.96	0.99	0.34
883	M	M	L	L	Y	0.60	0.95	0.28	0.95	0.42	0.62
884	F	F	F	H	Y	0.82	0.90	0.22	0.63	0.84	0.70
885	T	T	D	Y	W	0.74	0.51	0.51	0.82	0.53	0.50
886	M	S	P	T	I	0.38	0.85	0.44	0.72	0.39	0.42
887	G	G	G	A	G	0.97	0.94	0.53	0.69	0.97	0.81
888	M	T	I	L	T	0.28	0.49	0.47	0.74	0.46	0.37
889	G	P	P	P	P	0.39	0.54	0.31	0.96	0.74	0.54
890	A	P	P	D	T	0.35	0.27	0.25	0.73	0.28	0.11
891	N	W	L	W	Y	0.25	0.57	0.23	0.96	0.40	0.36
892	V	M	A	A	W	0.36	0.38	0.22	0.42	0.56	0.11
893	N	R	K	Q	Q	0.42	0.53	0.51	0.82	0.26	0.41
894	A	L	F	S	A	0.37	0.37	0.33	0.45	0.46	0.10
895	F	F	F	L	I	0.73	0.36	0.27	0.73	0.35	0.18
896	F	F	F	G	G	0.98	0.97	0.27	0.82	0.98	0.84
897	G	S	G	M	A	0.32	0.38	0.21	0.68	0.46	0.10
898	I	I	V	V	V	0.34	0.32	0.35	0.70	0.27	0.17
899	A	L	L	F	F	0.42	0.29	0.28	0.53	0.92	0.42
900	T	T	T	S	S	0.89	0.79	0.54	0.94	0.97	0.78
901	M	M	F	L	A	0.77	0.47	0.23	0.40	0.65	0.16
902	L	L	L	I	L	0.37	0.36	0.41	0.42	0.77	0.41
903	I	V	V	L	E	0.94	0.54	0.42	0.88	0.97	0.59
904	A	A	A	L	V	0.71	0.74	0.35	0.30	0.74	0.60
905	V	V	V	A	V	0.56	0.73	0.35	0.47	0.46	0.54
906	P	P	P	P	P	0.98	0.95	0.79	0.96	0.95	0.94
907	T	T	S	S	L	0.92	0.68	0.31	0.88	0.60	0.61
908	G	G		W	F	0.89	0.61		0.89	0.34	0.55
909	-	-	L	G	L	0.99	0.99	0.44	0.84	0.39	0.25
910	V	I	L	G	M	0.70	0.80	0.35	0.76	0.39	0.53
911	K	K	T	M	G	0.92	0.98	0.64	0.71	0.36	0.70

912	I	V	V	I	L	0.46	0.55	0.35	0.79	0.21	0.14
913	F	F	F	N	E	0.94	0.96	0.72	0.96	0.48	0.81
914	N	N	A	G	A	0.77	0.60	0.24	0.76	0.69	0.53
915	W	W	W	L	F	0.95	0.95	0.23	0.37	0.30	0.72
916	L	V	I	M	E	0.47	0.36	0.18	0.69	0.29	0.10
917	F	A	A	T	M	0.38	0.71	0.69	0.96	0.31	0.49
918	T	T	T	L	Y	0.99	0.98	0.47	0.81	0.24	0.79
919	M	L	V	S	R	0.75	0.62	0.29	0.60	0.32	0.47
920	W	W	E	G	L	0.55	0.57	0.44	0.89	0.29	0.45
921	R	G	T	-	-	0.36	0.52	0.20	0.98	0.97	0.05
922	G	G	G	-	-	0.98	0.97	0.33	0.98	0.97	0.71
923	-	-	-	-	-	1.00	0.99	0.46	0.99	0.98	0.00
924	-	-	-	-	-	0.99	0.99	0.50	0.99	0.99	0.00
925	-	-	-	-	-	0.99	0.99	0.35	0.99	0.99	0.00
926	-	-	-	-	-	0.99	0.99	0.36	1.00	0.99	0.00
927	-	-	G	-	-	0.94	0.99	0.43	1.00	0.99	0.02
928	-	-	-	-	-	1.00	0.99	0.98	1.00	0.99	0.00
929	-	-	-	-	-	1.00	0.99	0.99	1.00	0.99	0.00
930	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
931	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
932	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
933	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
934	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00
935	-	-	-	-	-	1.00	1.00	0.98	1.00	0.59	0.00
936	-	-	-	-	-	1.00	1.00	0.97	1.00	0.58	0.00
937	-	-	-	-	-	1.00	1.00	0.97	1.00	0.58	0.00
938	-	-	-	-	-	1.00	1.00	0.97	1.00	0.94	0.00
939	-	-	-	-	-	1.00	1.00	0.97	1.00	0.43	0.00
940	-	-	-	-	-	1.00	1.00	0.93	1.00	0.43	0.00
941	-	-	-	-	-	1.00	1.00	0.52	1.00	0.43	0.00
942	-	-	-	-	-	1.00	1.00	0.52	1.00	0.43	0.00
943	-	-	-	-	-	1.00	1.00	0.52	1.00	0.44	0.00
944	-	-	-	-	-	1.00	1.00	0.52	1.00	0.94	0.00
945	-	-	-	-	-	1.00	1.00	0.52	1.00	0.97	0.00
946	-	-	-	-	-	1.00	1.00	0.52	1.00	0.99	0.00
947	-	-	-	-	-	1.00	1.00	0.51	0.99	0.99	0.00
948	-	-	-	-	-	1.00	1.00	0.51	0.99	0.94	0.00
949	-	-	-	-	-	1.00	1.00	0.51	0.99	0.94	0.00
950	-	-	-	-	-	1.00	1.00	0.51	0.99	0.94	0.00
951	-	-	-	-	-	1.00	1.00	0.51	0.99	0.94	0.00
952	-	-	-	A	-	1.00	1.00	0.51	0.73	0.58	0.04
953	-	-	-	W	M	0.99	1.00	0.47	0.86	0.16	0.05
954	-	-	-	H	H	0.99	1.00	0.47	0.46	0.32	0.06
955	-	-	-	K	P	0.99	1.00	0.48	0.76	0.44	0.06
956	-	-	-	L	N	0.99	1.00	0.91	0.73	0.40	0.06
957	-	-	-	R	K	1.00	1.00	0.98	0.81	0.49	0.06
958	R	K	-	-	-	0.37	0.45	0.53	0.99	1.00	0.04
959	I	I	-	-	-	0.51	0.65	0.52	0.99	1.00	0.38
960	T	R	-	-	-	0.29	0.35	0.68	0.99	1.00	0.03
961	F	L	P	T	W	0.77	0.55	0.33	0.54	0.40	0.46
962	E	N	R	D	A	0.35	0.35	0.22	0.76	0.61	0.34
963	T	T	F	P	L	0.51	0.47	0.26	0.81	0.36	0.14
964	P	P	P	I	L	0.83	0.62	0.22	0.51	0.43	0.57
965	M	M	A	L	W	0.84	0.79	0.41	0.63	0.41	0.60
966	L	L	L	R	F	0.81	0.80	0.34	0.65	0.37	0.65
967	W	F	F	F	V	0.55	0.61	0.23	0.74	0.42	0.46
968	A	A	F	L	G	0.40	0.57	0.36	0.50	0.44	0.14
969	I	L	A	V	V	0.46	0.44	0.40	0.60	0.34	0.14
970	G	G	G	V	A	0.71	0.80	0.28	0.55	0.47	0.61
971	F	F	F	S	F	0.94	0.73	0.28	0.49	0.48	0.65
972	I	I	V	L	W	0.46	0.43	0.24	0.50	0.53	0.13
973	V	F	F	S	N	0.32	0.29	0.48	0.32	0.58	0.12
974	T	L	G	F	F	0.47	0.34	0.24	0.83	0.53	0.40
975	F	F	F	Y	V	0.98	0.96	0.41	0.97	0.39	0.79
976	T	V	G	G	G	0.42	0.44	0.41	0.77	0.99	0.53
977	I	I	G	M	A	0.42	0.37	0.57	0.72	0.97	0.48
978	G	G	-	S	G	0.99	0.85	0.52	0.68	0.95	0.68
979	G	G	A	T	V	0.99	0.99	0.27	0.88	0.66	0.73
980	M	L	G	F	W	0.30	0.53	0.28	0.81	0.37	0.11
981	T	T	G	E	G	0.83	0.78	0.51	0.85	0.98	0.64
982	G	G	G	G	-	0.99	0.99	0.24	0.96	0.90	0.80
983	V	V	V	P	F	0.69	0.50	0.36	0.81	0.85	0.55
984	M	M	N	M	L	0.39	0.35	0.49	0.64	0.58	0.39
985	L	L	A	M	I	0.83	0.87	0.30	0.64	0.51	0.63
986	A	A	S	S	N	0.78	0.59	0.40	0.50	0.56	0.54
987	V	S	Y	I	L	0.25	0.27	0.21	0.75	0.47	0.10
988	P	V	V	K	P	0.43	0.52	0.26	0.71	0.67	0.41
989	P	P	P	T	I	0.47	0.62	0.24	0.44	0.38	0.45
990	A	F	N	V	V	0.45	0.34	0.30	0.86	0.51	0.40
991	D	D	D	N	N	0.93	0.97	0.32	0.90	0.53	0.80
992	F	I	W	A	Y	0.50	0.50	0.20	0.57	0.75	0.38
993	Q	H	V	L	Y	0.38	0.85	0.33	0.78	0.84	0.48
994	L	V	H	S	T	0.48	0.71	0.46	0.68	0.36	0.43
995	H	H	N	H	H	0.78	0.67	0.47	0.96	0.56	0.61
996	D	D	T	Y	G	0.53	0.77	0.45	0.69	0.98	0.55
997	T	T	T	T	T	0.51	0.92	0.45	0.93	0.51	0.70
998	Y	Y	W	D	Q	0.63	0.93	0.41	0.86	0.42	0.59
999	F	F	I	W	T	0.79	0.90	0.28	0.91	0.26	0.62
1000	V	V	V	T	T	0.58	0.94	0.39	0.73	0.92	0.70
1001	V	V	G	I	A	0.64	0.96	0.45	0.61	0.61	0.60

1059	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	7											
1060	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1088	F	R	R	S	-	0.29	0.73	0.23	0.91	0.77	0.13
1061	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1089	M	M	S	L	-	0.44	0.79	0.21	0.18	0.29	0.11
1062	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1090	L	Y	M	K	Q	0.56	0.57	0.18	0.30	0.32	0.38
1063	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1091	N	N	S	L	L	0.53	0.42	0.22	0.80	0.37	0.17
1064	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1092	E	E	-	I	L	0.81	0.87	0.33	0.41	0.53	0.59
1065	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1093	T	G	-	N	K	0.29	0.24	0.33	0.53	0.29	0.07
1066	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1094	L	L	L	W	M	0.56	0.42	0.54	0.25	0.27	0.18
1067	-	-	-	-	-	1.00	1.00	0.98	1.00	1.00	0.00	1095	G	G	G	-	-	0.85	0.80	0.34	0.96	0.97	0.61
1068	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1096	K	R	R	-	-	0.74	0.35	0.24	0.96	0.97	0.08
1069	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1097	I	L	L	-	-	0.21	0.31	0.27	0.96	0.57	0.05
1070	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1098	H	H	Q	H	A	0.48	0.57	0.37	0.87	0.40	0.43
1071	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1099	F	F	F	F	F	0.96	0.57	0.42	0.95	0.88	0.74
1072	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1100	W	W	W	W	W	0.91	0.22	0.56	0.95	0.91	0.69
1073	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1101	L	L	L	L	L	0.27	0.69	0.34	0.65	0.41	0.24
1074	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1102	W	T	W	A	M	0.32	0.50	0.51	0.67	0.38	0.36
1075	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1103	F	F	F	T	N	0.44	0.76	0.36	0.80	0.56	0.53
1076	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1104	I	I	I	I	I	0.58	0.52	0.36	0.52	0.31	0.50
1077	-	-	-	-	-	1.00	1.00	0.95	0.99	1.00	0.00	1105	G	G	G	G	G	0.81	0.98	0.93	0.98	0.74	0.90
1078	-	-	-	-	-	1.00	1.00	0.96	0.99	1.00	0.00	1106	F	F	M	I	L	0.66	0.59	0.42	0.43	0.55	0.48
1079	-	-	-	-	-	1.00	1.00	0.96	0.99	1.00	0.00	1107	N	N	N	V	A	0.44	0.85	0.21	0.70	0.30	0.21
1080	-	-	-	-	-	1.00	1.00	0.96	0.99	1.00	0.00	1108	V	L	L	L	F	0.36	0.57	0.35	0.70	0.21	0.16
1081	-	-	-	-	-	1.00	1.00	0.96	0.99	1.00	0.00	1109	T	T	F	Y	M	0.35	0.56	0.20	0.91	0.77	0.44
1082	-	-	-	-	A	1.00	1.00	0.96	0.84	0.23	0.01	1110	F	F	S	I	T	0.96	0.85	0.24	0.55	0.38	0.65
1083	-	-	-	-	P	1.00	1.00	0.95	0.44	0.23	0.01	1111	F	L	F	A	-	0.47	0.38	0.29	0.37	0.43	0.08
1084	-	-	-	E	W	1.00	1.00	0.95	0.18	0.33	0.03	1112	P	P	P	S	L	0.81	0.87	0.36	0.51	0.34	0.67
1085	-	-	-	R	N	1.00	1.00	0.95	0.30	0.44	0.04	1113	M	Q	M	M	S	0.34	0.52	0.47	0.82	0.40	0.42
1086	-	-	-	M	-	1.00	1.00	0.32	0.48	0.70	0.02	1114	H	H	H	W	L	0.58	0.52	0.59	0.84	0.79	0.58
1087	G	-	L	Y	-	0.90	0.53	0.29	0.65	0.73	0.09	1115	F	W	I	V	L	0.42	0.20	0.19	0.46	0.39	0.08

111	6	L	L	A	A	P	0.81	0.73	0.33	0.37	0.54	0.58
111	7	G	G	G	G	A	0.99	0.99	1.00	0.97	0.46	0.91
111	8	L	L	-	I	G	0.53	0.57	0.51	0.57	0.95	0.43
111	9	M	L	L	M	I	0.33	0.28	0.51	0.42	0.36	0.11
112	0	G	G	G	Q	L	0.98	0.99	0.40	0.83	0.28	0.79
112	1	M	M	G	G	Q	0.86	0.97	0.47	0.92	0.92	0.74
112	2	P	P	P	L	V	0.62	0.79	0.29	0.73	0.25	0.58
112	3	R	R	R	M	H	0.99	0.99	0.43	0.89	0.24	0.79
112	4	R	R	R	W	A	0.99	1.00	0.98	0.89	0.40	0.86
112	5	-	-	R	R	S	0.93	0.97	0.45	0.73	0.31	0.10
112	6	-	-	-	A	R	0.93	0.97	0.54	0.60	0.42	0.05
112	7	-	-	-	V	V	0.93	0.97	0.55	0.38	0.23	0.05
112	8	-	-	-	N	H	0.93	0.97	0.56	0.49	0.19	0.03
112	9	-	-	-	E	G	0.94	0.97	0.54	0.23	0.82	0.05
113	0	-	-	-	D	L	0.94	0.97	0.55	0.50	0.24	0.04
113	1	Y	Y	Y	G	W	0.30	0.47	0.28	0.88	0.44	0.41
113	2	A	A	-	-	Y	0.20	0.52	0.39	0.93	0.36	0.07
113	3	D	S	-	-	A	0.42	0.20	0.64	0.98	0.57	0.06
113	4	Y	Y	-	-	R	0.72	0.97	0.66	0.98	0.55	0.58
113	5	-	D	-	-	S	0.92	0.47	0.89	0.98	0.51	0.05
113	6	-	-	-	-	-	1.00	1.00	0.97	0.99	0.98	0.00
113	7	-	-	-	-	-	1.00	1.00	1.00	0.99	0.98	0.00
113	8	-	-	-	-	-	1.00	1.00	1.00	1.00	1.00	0.00
113	9	-	-	-	-	-	1.00	1.00	1.00	0.92	1.00	0.00
114	0	-	-	-	T	-	1.00	1.00	1.00	0.43	0.98	0.02
114	1	-	-	-	L	-	1.00	1.00	0.94	0.81	0.98	0.04
114	2	-	-	-	T	-	1.00	1.00	0.76	0.40	0.98	0.02
114	3	-	P	-	Y	-	0.64	0.48	0.42	0.73	0.98	0.06
114		-	-	-	S	-	0.34	0.43	0.40	0.65	0.56	0.03
4												
114	5	N	F	-	-	M	0.20	0.41	0.33	0.93	0.34	0.05
114	6	P	P	-	F	E	0.28	0.31	0.26	0.93	0.46	0.11
114	7	G	E	G	V	V	0.27	0.26	0.22	0.49	0.38	0.10
114	8	-	-	-	E	Q	0.99	0.54	0.34	0.66	0.40	0.05
114	9	-	-	-	T	D	0.99	1.00	0.34	0.37	0.30	0.03
115	0	-	-	-	V	Q	0.99	0.54	0.39	0.62	0.39	0.05
115	1	-	-	-	E	-	0.99	1.00	0.32	0.23	0.41	0.01
115	2	-	-	-	A	-	0.97	0.99	0.42	0.66	0.41	0.03
115	3	F	-	W	S	L	0.50	0.99	0.34	0.31	0.39	0.08
115	4	A	-	A	H	A	0.27	0.99	0.17	0.28	0.21	0.05
115	5	P	-	T	P	L	0.21	0.53	0.22	0.90	0.30	0.08
115	6	W	L	Y	Y	L	0.31	0.47	0.25	0.31	0.38	0.10
115	7	N	N	N	Y	Y	0.62	0.83	0.25	0.68	0.32	0.61
115	8	M	V	L	V	W	0.18	0.38	0.27	0.29	0.79	0.10
115	9	I	I	L	V	L	0.40	0.43	0.28	0.37	0.32	0.11
116	0	A	S	T	R	R	0.44	0.59	0.23	0.96	0.94	0.52
116	1	S	T	A	A	L	0.36	0.53	0.29	0.32	0.25	0.10
116	2	I	I	I	L	P	0.42	0.25	0.35	0.33	0.28	0.13
116	3	G	G	G	G	G	0.89	0.95	0.83	0.92	0.47	0.88
116	4	A	A	G	G	D	0.69	0.67	0.45	0.98	0.56	0.58
116	5	F	F	L	L	V	0.41	0.31	0.24	0.29	0.49	0.12
116	6	L	L	I	L	V	0.35	0.38	0.53	0.43	0.47	0.13
116	7	I	L	L	F	F	0.30	0.78	0.39	0.48	0.69	0.45
116	8	G	A	F	L	A	0.42	0.54	0.22	0.67	0.25	0.11
116	9	L	I	I	S	I	0.33	0.23	0.27	0.24	0.36	0.08
117	0	G	S	G	G	G	0.52	0.46	0.48	0.95	0.83	0.57
117	1	I	T	I	M	A	0.30	0.38	0.20	0.43	0.50	0.09
117	2	L	L	L	L	L	0.50	0.54	0.34	0.53	0.49	0.50

1173	L	P	L	I	L	0.20	0.51	0.45	0.35	0.32	0.09
1174	F	F	F	M	L	0.57	0.54	0.68	0.70	0.29	0.51
1175	L	L	V	A	A	0.47	0.50	0.29	0.63	0.21	0.15
1176	I	I	V	Y	W	0.22	0.31	0.25	0.68	0.31	0.10
1177	Q	N	N	N	F	0.31	0.64	0.29	0.92	0.25	0.17
1178	V	L	L	V	F	0.41	0.31	0.34	0.31	0.21	0.09
1179	F	V	L	Y	L	0.21	0.27	0.24	0.29	0.34	0.07
1180	V	W	V	K	V	0.31	0.26	0.20	0.32	0.31	0.07
1181	S	S	S	T	L	0.54	0.79	0.36	0.88	0.22	0.56
1182	-	L	-	V	G	0.48	0.43	0.25	0.42	0.32	0.06
1183	-	R	-	R	R	0.74	0.20	0.25	0.39	0.25	0.06
1184	-	K	-	A	K	0.79	0.18	0.46	0.31	0.15	0.03
1185	-	G	-	G	A	0.91	0.95	0.53	0.20	0.15	0.07
1186	I	E	-	K	R	0.22	0.28	0.50	0.28	0.16	0.05
1187	R	K	-	-	A	0.37	0.34	0.49	0.34	0.13	0.04
1188	-	A	-	-	-	0.22	0.90	0.46	0.33	0.15	0.04
1189	R	-	-	-	-	0.35	0.53	0.46	0.31	0.25	0.02
1190	-	-	-	-	-	0.37	0.53	0.47	0.20	0.35	0.00
1191	-	-	-	A	-	0.90	0.99	0.50	0.26	0.36	0.01
1192	-	-	-	-	-	0.95	1.00	0.73	0.24	0.67	0.00
1193	-	-	-	A	-	0.66	1.00	0.75	0.18	0.65	0.01
1194	-	-	-	A	-	0.64	1.00	0.73	0.17	0.63	0.01
1195	-	-	-	A	-	0.58	1.00	0.75	0.30	0.68	0.02
1196	-	-	-	A	-	0.41	1.00	0.32	0.31	0.31	0.02
1197	-	-	-	A	-	0.29	1.00	0.32	0.19	0.28	0.01
1198	T	-	-	A	-	0.32	1.00	0.32	0.21	0.28	0.03
1199	-	-	-	-	-	0.57	1.00	0.34	0.30	0.40	0.00
1200	-	-	-	-	-	0.51	1.00	0.37	0.42	0.45	0.00
120	-	-	-	-	-	0.45	0.47	0.38	0.55	0.49	0.00

1201	G	-	-	-	-	0.38	0.47	0.34	0.59	0.52	0.02
1203	D	N	-	-	-	0.55	0.95	0.32	0.78	0.53	0.45
1204	P	P	P	-	-	0.84	0.99	0.35	0.81	0.54	0.68
1205	W	W	-	-	-	0.97	0.99	0.33	0.88	0.59	0.64
1206	G	G	-	-	-	0.42	0.31	0.34	0.90	0.63	0.06
1207	-	-	-	-	-	0.68	0.54	0.58	0.87	0.69	0.00
1208	-	-	-	-	-	0.52	0.54	0.67	0.87	0.74	0.00
1209	-	-	-	-	-	0.98	1.00	0.91	0.86	0.79	0.00
1210	-	-	-	-	-	0.98	1.00	0.97	0.93	0.90	0.00
1211	-	-	-	-	-	0.99	1.00	0.97	0.93	0.91	0.00
1212	-	-	-	-	-	0.99	1.00	0.97	0.96	0.94	0.00
1213	-	-	-	-	-	0.99	1.00	0.95	0.96	0.95	0.00
1214	-	-	-	-	-	0.99	1.00	0.95	0.98	0.91	0.00
1215	-	-	-	-	-	0.99	1.00	0.64	0.98	0.92	0.00
1216	-	-	-	-	-	1.00	1.00	0.59	0.98	0.94	0.00
1217	-	-	-	-	-	0.99	1.00	0.59	0.87	0.94	0.00
1218	-	-	-	-	-	0.99	1.00	0.56	0.88	0.94	0.00
1219	-	-	-	-	-	0.99	1.00	0.56	0.75	0.94	0.00
1220	-	-	-	-	-	0.99	1.00	0.56	0.75	0.94	0.00
1221	-	-	-	-	-	0.99	0.99	0.55	0.76	0.94	0.00
1222	-	-	-	-	-	0.99	0.99	0.55	0.76	0.95	0.00
1223	-	-	-	-	-	0.99	0.99	0.55	0.77	0.95	0.00
1224	-	-	-	-	-	0.99	0.99	0.55	0.96	0.95	0.00
1225	-	-	-	-	-	0.99	0.99	0.55	0.97	0.95	0.00
1226	-	-	-	-	-	0.99	0.99	0.55	0.97	0.95	0.00
1227	-	-	-	-	-	0.99	0.99	0.55	0.99	0.96	0.00
1228	-	-	-	-	-	0.99	0.99	0.55	0.99	0.98	0.00
1229	-	-	-	-	-	0.99	0.99	0.55	0.99	0.99	0.00

123	0	-	-	-	-	-	0.99	0.99	0.55	0.99	0.99	0.00
123	1	-	-	-	-	-	0.99	0.99	0.55	0.99	0.99	0.00
123	2	-	-	-	-	-	0.99	0.99	0.55	0.99	1.00	0.00
123	3	-	-	-	-	-	0.99	0.99	0.55	0.99	1.00	0.00
123	4	-	-	-	-	-	0.99	1.00	0.55	0.99	1.00	0.00
123	5	-	-	-	-	-	0.99	1.00	0.55	0.99	1.00	0.00
123	6	-	-	-	-	-	0.99	1.00	0.56	0.99	1.00	0.00
123	7	-	-	-	-	-	0.99	1.00	0.56	0.99	1.00	0.00
123	8	-	-	-	-	-	0.99	0.99	0.55	0.99	1.00	0.00
123	9	-	-	-	-	-	0.99	0.99	0.55	0.99	1.00	0.00
124	0	-	-	-	-	-	0.99	0.99	0.55	0.99	1.00	0.00
124	1	-	-	-	-	-	0.99	0.99	0.55	0.99	1.00	0.00
124	2	-	-	-	-	-	0.99	0.99	0.55	0.98	1.00	0.00
124	3	-	-	-	-	-	0.99	0.99	0.55	0.98	1.00	0.00
124	4	-	-	-	-	-	0.99	1.00	0.55	0.98	1.00	0.00
124	5	-	-	-	-	-	0.99	1.00	0.55	0.98	1.00	0.00
124	6	-	-	-	-	-	1.00	1.00	0.55	0.98	1.00	0.00
124	7	-	-	-	-	-	1.00	1.00	0.57	0.98	1.00	0.00
124	8	-	-	-	-	-	0.99	1.00	0.56	0.98	1.00	0.00
124	9	-	-	-	-	-	0.99	1.00	0.73	0.98	1.00	0.00
125	0	-	-	-	-	-	0.99	1.00	0.70	0.98	1.00	0.00
125	1	-	-	-	-	-	0.99	1.00	0.75	0.98	1.00	0.00
125	2	-	-	-	-	-	0.99	1.00	0.57	0.98	1.00	0.00
125	3	-	-	-	-	-	0.99	1.00	0.57	0.98	1.00	0.00
125	4	-	-	-	-	-	0.99	1.00	0.58	0.98	1.00	0.00
125	5	-	-	-	-	-	0.99	1.00	0.58	0.98	1.00	0.00
125	6	-	-	-	-	-	0.99	1.00	0.58	0.98	1.00	0.00
125	7	-	-	-	-	-	0.99	1.00	0.99	0.99	1.00	0.00
125		-	-	-	-	-	0.99	1.00	0.99	0.99	1.00	0.00
8												
125	9	-	-	-	-	-	0.99	1.00	0.99	0.98	1.00	0.00
126	0	-	-	-	-	-	0.99	1.00	0.99	0.98	1.00	0.00
126	1	-	-	-	-	-	0.99	1.00	0.99	0.98	1.00	0.00
126	2	-	-	-	-	-	0.99	1.00	0.99	0.98	1.00	0.00
126	3	-	-	-	-	-	0.99	1.00	0.96	0.98	1.00	0.00
126	4	-	-	-	-	-	0.99	1.00	0.58	0.94	1.00	0.00
126	5	-	-	-	-	-	0.99	1.00	0.59	0.94	1.00	0.00
126	6	-	-	-	-	-	1.00	1.00	0.68	0.94	1.00	0.00
126	7	-	-	-	-	-	0.99	1.00	0.72	0.94	1.00	0.00
126	8	-	-	-	-	-	0.99	1.00	0.74	0.94	1.00	0.00
126	9	-	-	-	-	-	0.99	1.00	0.94	0.94	1.00	0.00
127	0	-	-	-	-	-	0.99	1.00	0.94	0.94	1.00	0.00
127	1	-	-	-	-	-	0.99	1.00	0.94	0.95	1.00	0.00
127	2	-	-	-	-	-	0.99	1.00	0.95	0.98	1.00	0.00
127	3	-	-	-	-	-	0.99	1.00	0.95	0.98	1.00	0.00
127	4	-	-	-	-	-	0.99	1.00	0.97	0.95	1.00	0.00
127	5	-	-	-	-	-	0.99	1.00	0.97	0.95	1.00	0.00
127	6	-	-	-	-	-	0.99	1.00	0.97	0.94	1.00	0.00
127	7	-	-	-	-	-	0.99	1.00	0.97	0.94	1.00	0.00
127	8	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
127	9	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	0	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	1	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	2	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	3	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	4	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	5	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00
128	6	-	-	-	-	-	0.99	1.00	0.98	0.94	1.00	0.00

- 155 -

134	4	-	-	-	-	-	0.99	1.00	1.00	0.94	1.00	0.00
134	5	-	-	-	-	-	0.91	1.00	1.00	0.94	1.00	0.00
134	6	G	-	-	-	-	0.48	0.47	0.77	0.94	1.00	0.02
134	7	R	-	-	-	-	0.41	0.47	0.83	0.94	1.00	0.02
134	8	T	T	-	-	-	0.66	0.51	0.62	0.94	1.00	0.38
134	9	L	L	-	-	-	0.95	0.79	0.55	0.94	1.00	0.57
135	0	E	E	-	-	-	0.97	0.99	0.55	0.94	1.00	0.64
135	1	W	W	-	-	-	0.98	0.99	0.55	0.94	1.00	0.64
135	2	A	L	-	-	-	0.35	0.31	0.56	0.94	1.00	0.03
135	3	T	-	-	-	-	0.52	0.45	0.56	0.94	1.00	0.03
135	4	S	T	-	-	-	0.37	0.35	0.56	0.94	1.00	0.04
135	5	S	-	-	-	-	0.49	0.53	0.56	0.94	1.00	0.02
135	6	P	S	-	-	-	0.56	0.93	0.56	0.94	1.00	0.45
135	7	P	P	-	-	-	0.79	0.97	0.58	0.98	1.00	0.57
135	8	P	P	-	-	-	0.67	0.98	0.59	0.99	1.00	0.53
135	9	-	P	-	-	-	0.58	0.64	0.81	0.99	1.00	0.03
136	0	F	V	-	-	-	0.36	0.27	0.59	0.99	1.00	0.03
136	1	Y	E	-	-	-	0.45	0.62	0.59	1.00	1.00	0.05
136	2	N	N	-	-	-	0.63	0.98	0.59	1.00	1.00	0.52
136	3	F	F	-	-	-	0.87	0.76	0.59	0.96	1.00	0.53
136	4	A	E	-	-	-	0.40	0.40	0.59	0.94	1.00	0.04
136	5	T	E	-	-	-	0.20	0.25	0.60	0.94	1.00	0.02
136	6	-	-	-	-	-	0.93	0.57	0.84	0.94	1.00	0.00
136	7	L	-	-	-	-	0.36	0.31	0.60	0.94	1.00	0.02
136	8	P	P	-	-	-	0.96	0.83	0.62	0.94	1.00	0.58
136	9	R	V	-	-	-	0.16	0.38	0.62	0.94	1.00	0.03
137	0	V	V	-	-	-	0.56	0.51	0.62	0.94	1.00	0.35
137	1	R	T	-	-	-	0.21	0.31	0.62	0.94	1.00	0.03
137	-	T	-	-	-	-	0.81	0.14	0.85	0.94	1.00	0.01
2												
137	3	-	G	-	-	-	0.25	0.44	0.65	0.94	1.00	0.02
137	4	-	P	-	-	-	0.25	0.87	0.65	0.94	1.00	0.04
137	5	D	Y	-	-	-	0.42	0.52	0.74	0.94	1.00	0.05
137	6	A	-	-	-	-	0.27	0.43	0.74	0.94	1.00	0.01
137	7	F	Y	-	-	-	0.27	0.51	0.73	0.94	1.00	0.04
137	8	W	G	-	-	-	0.33	0.51	0.73	0.94	1.00	0.04
137	9	-	D	-	-	-	0.27	0.31	0.73	0.94	1.00	0.02
138	0	-	Y	-	-	-	0.31	0.33	0.73	0.94	1.00	0.02
138	1	K	E	-	-	-	0.43	0.14	0.73	0.94	1.00	0.03
138	2	-	P	-	-	-	0.37	0.23	0.72	0.94	1.00	0.01
138	3	-	-	-	-	-	0.38	0.52	0.72	0.94	1.00	0.00
138	4	-	-	-	-	-	0.60	0.54	0.72	0.94	1.00	0.00
138	5	-	-	-	-	-	0.92	0.55	0.85	0.94	1.00	0.00
138	6	-	-	-	-	-	0.93	0.62	0.86	0.94	1.00	0.00
138	7	-	-	-	-	-	0.98	0.64	0.92	0.94	1.00	0.00
138	8	-	-	-	-	-	0.98	0.67	1.00	0.94	1.00	0.00
138	9	-	-	-	-	-	0.98	0.67	1.00	0.94	1.00	0.00
139	0	-	-	-	-	-	0.98	0.67	1.00	0.94	1.00	0.00
139	1	-	-	-	-	-	0.99	0.71	1.00	0.94	1.00	0.00
139	2	-	-	-	-	-	0.99	0.69	1.00	0.94	1.00	0.00
139	3	-	-	-	-	-	0.99	0.70	1.00	1.00	1.00	0.00
139	4	-	-	-	-	-	0.99	0.71	1.00	1.00	1.00	0.00
139	5	-	-	-	-	-	0.99	0.78	1.00	1.00	1.00	0.00
139	6	-	-	-	-	-	0.99	0.79	1.00	1.00	1.00	0.00
139	7	-	-	-	-	-	0.99	0.80	1.00	1.00	1.00	0.00
139	8	-	-	-	-	-	0.99	0.81	0.96	1.00	1.00	0.00
139	9	-	-	-	-	-	0.64	0.85	0.91	1.00	1.00	0.00
140	0	-	-	-	-	-	0.53	0.88	0.89	0.94	1.00	0.00

- 157 -

145	8	-	-	-	-	-	0.98	0.96	0.88	0.94	1.00	0.00
145	9	-	-	-	-	-	0.98	0.97	0.90	0.94	1.00	0.00
146	0	-	-	-	-	-	0.47	0.97	0.66	0.94	1.00	0.00
146	1	-	-	-	-	-	0.48	0.97	0.66	0.94	1.00	0.00
146	2	-	-	-	-	-	0.48	0.97	0.66	0.94	1.00	0.00
146	3	-	-	-	-	-	0.46	0.97	0.66	0.94	1.00	0.00
146	4	-	-	-	-	-	0.47	0.97	0.64	0.94	1.00	0.00
146	5	-	-	-	-	-	0.47	0.97	0.64	0.94	1.00	0.00
146	6	-	-	-	-	-	0.47	0.97	0.63	0.94	1.00	0.00
146	7	-	-	-	-	-	0.48	0.97	0.63	0.94	1.00	0.00
146	8	-	-	-	-	-	0.49	0.97	0.62	0.94	1.00	0.00
146	9	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	0	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	1	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	2	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	3	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	4	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	5	-	-	-	-	-	0.50	0.97	0.62	0.94	1.00	0.00
147	6	-	-	-	-	-	0.52	0.97	0.62	0.94	1.00	0.00
147	7	-	-	-	-	-	0.50	0.97	0.68	0.94	1.00	0.00
147	8	-	-	-	-	-	0.51	0.97	0.70	0.94	1.00	0.00
147	9	-	-	-	-	-	0.50	0.97	0.75	0.94	1.00	0.00
148	0	-	-	-	-	-	0.51	0.97	0.76	0.94	1.00	0.00
148	1	-	-	-	-	-	0.51	0.97	0.76	0.95	1.00	0.00
148	2	-	-	-	-	-	0.51	0.97	0.76	0.97	1.00	0.00
148	3	-	-	-	-	-	0.51	0.97	0.77	0.97	1.00	0.00
148	4	-	-	-	-	-	0.56	0.97	0.97	0.97	1.00	0.00
148	5	-	-	-	-	-	0.98	0.97	1.00	0.97	1.00	0.00
148		-	-	-	-	-	0.56	0.97	0.97	0.97	1.00	0.00
6												
148	7	-	-	-	-	-	0.56	0.97	0.97	0.97	1.00	0.00
148	8	-	-	-	-	-	0.56	0.97	0.97	0.97	1.00	0.00
148	9	-	-	-	-	-	0.56	0.97	0.97	0.97	1.00	0.00
149	0	-	-	-	-	-	0.61	0.97	0.97	0.98	1.00	0.00
149	1	-	-	-	-	-	0.98	0.97	1.00	0.98	1.00	0.00
149	2	-	-	-	-	-	0.98	0.97	1.00	0.98	1.00	0.00
149	3	-	-	-	-	-	0.94	0.97	1.00	0.98	1.00	0.00
149	4	-	-	-	-	-	0.51	0.97	0.77	0.98	1.00	0.00
149	5	-	-	-	-	-	0.51	0.97	0.77	0.98	1.00	0.00
149	6	-	-	-	-	-	0.51	0.97	0.77	0.98	1.00	0.00
149	7	-	-	-	-	-	0.51	0.97	0.77	0.98	1.00	0.00
149	8	-	-	-	-	-	0.51	0.97	0.77	0.98	1.00	0.00
149	9	-	-	-	-	-	0.51	0.97	0.77	0.98	1.00	0.00
150	0	-	-	-	-	-	0.51	0.97	0.77	0.94	1.00	0.00
150	1	-	-	-	-	-	0.52	0.97	0.77	0.94	1.00	0.00
150	2	-	-	-	-	-	0.55	0.97	0.77	0.94	1.00	0.00
150	3	-	-	-	-	-	0.53	0.97	0.77	0.94	1.00	0.00
150	4	-	-	-	-	-	0.61	0.97	0.77	0.94	1.00	0.00
150	5	-	-	-	-	-	0.86	0.97	0.77	0.95	1.00	0.00
150	6	-	-	-	-	-	0.92	0.97	0.77	0.94	1.00	0.00
150	7	-	-	-	-	-	0.89	0.97	0.77	0.94	1.00	0.00
150	8	-	-	-	-	-	0.89	0.97	0.78	0.94	1.00	0.00
150	9	-	-	-	-	-	0.89	0.97	0.79	0.94	1.00	0.00
151	0	-	-	-	-	-	0.90	0.97	0.79	0.94	1.00	0.00
151	1	-	-	-	-	-	0.93	0.97	0.82	0.94	1.00	0.00
151	2	-	-	-	-	-	0.91	0.97	0.82	0.94	1.00	0.00
151	3	-	-	-	-	-	0.93	0.97	0.82	0.94	1.00	0.00
151	4	-	-	-	-	-	0.93	0.97	0.95	0.94	1.00	0.00

151	5	-	-	-	-	-	0.92	0.97	0.96	0.94	1.00	0.00
151	6	-	-	-	-	-	0.94	0.97	0.97	0.94	1.00	0.00
151	7	-	-	-	-	-	0.98	0.97	0.97	0.94	1.00	0.00
151	8	-	-	-	-	-	0.98	0.97	0.97	0.94	1.00	0.00
151	9	-	-	-	-	-	0.91	0.97	0.88	0.94	1.00	0.00
152	0	-	-	-	-	-	0.67	0.97	0.88	0.94	1.00	0.00
152	1	-	-	-	-	-	0.64	0.97	0.85	0.94	1.00	0.00
152	2	-	-	-	-	-	0.64	0.97	0.85	0.94	1.00	0.00
152	3	-	-	-	-	-	0.65	0.97	0.85	0.94	1.00	0.00
152	4	-	-	-	-	-	0.65	0.97	0.85	0.94	1.00	0.00
152	5	-	-	-	-	-	0.67	0.97	0.86	0.94	1.00	0.00
152	6	-	-	-	-	-	0.67	0.97	0.97	0.94	1.00	0.00
152	7	-	-	-	-	-	0.71	0.97	0.97	0.95	1.00	0.00
152	8	-	-	-	-	-	0.76	0.97	0.97	0.95	1.00	0.00
152	9	-	-	-	-	-	0.77	0.97	0.97	0.96	1.00	0.00
153	0	-	-	-	-	-	0.78	0.97	0.97	0.96	1.00	0.00
153	1	-	-	-	-	-	0.79	0.97	0.97	0.96	1.00	0.00
153	2	-	-	-	-	-	0.93	0.97	0.97	0.96	1.00	0.00
153	3	-	-	-	-	-	0.95	0.97	0.97	0.96	1.00	0.00
153	4	-	-	-	-	-	0.95	0.97	0.97	0.96	1.00	0.00
153	5	-	-	-	-	-	0.97	0.97	0.97	0.97	1.00	0.00
153	6	-	-	-	-	-	0.97	0.97	0.97	0.97	1.00	0.00
153	7	-	-	-	-	-	0.97	0.97	0.97	0.98	1.00	0.00
153	8	-	-	-	-	-	0.97	0.97	0.97	0.99	1.00	0.00
153	9	-	-	-	-	-	0.97	0.97	0.97	0.99	1.00	0.00
154	0	-	-	-	-	-	0.97	0.97	0.97	0.99	1.00	0.00
154	1	-	-	-	-	-	0.97	0.97	0.97	0.99	1.00	0.00
154	2	-	-	-	-	-	0.98	0.97	0.97	0.99	1.00	0.00
154		-	-	-	-	-	0.99	0.97	0.97	0.99	1.00	0.00
3												
154	4	-	-	-	-	-	0.99	0.97	0.97	0.99	1.00	0.00
154	5	-	-	-	-	-	0.99	0.97	0.97	0.99	1.00	0.00
154	6	-	-	-	-	-	0.99	0.97	0.97	0.99	1.00	0.00
154	7	-	-	-	-	-	0.99	0.97	0.97	0.99	1.00	0.00
154	8	-	-	-	-	-	0.99	0.97	0.97	0.99	1.00	0.00
154	9	-	-	-	-	-	0.98	0.97	0.97	0.99	1.00	0.00
155	0	-	-	-	-	-	0.98	0.97	0.97	0.99	1.00	0.00
155	1	-	-	-	-	-	0.98	0.97	0.89	0.99	1.00	0.00
155	2	-	-	-	-	-	0.98	0.97	0.89	0.99	1.00	0.00
155	3	-	-	-	-	-	0.98	0.97	0.89	0.99	1.00	0.00
155	4	-	-	-	-	-	0.98	0.97	0.89	0.99	1.00	0.00
155	5	-	-	-	-	-	0.93	0.97	0.88	0.99	1.00	0.00
155	6	-	-	-	-	-	0.93	0.97	0.88	0.99	1.00	0.00
155	7	-	-	-	-	-	0.93	0.97	0.88	0.99	1.00	0.00
155	8	-	-	-	-	-	0.93	0.97	0.88	0.99	1.00	0.00
155	9	-	-	-	-	-	0.94	0.97	0.88	0.99	1.00	0.00
156	0	-	-	-	-	-	0.94	0.97	0.90	0.99	1.00	0.00
156	1	-	-	-	-	-	0.98	0.97	0.91	0.99	1.00	0.00
156	2	-	-	-	-	-	0.98	0.97	0.91	0.99	1.00	0.00
156	3	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
156	4	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
156	5	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
156	6	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
156	7	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
156	8	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
156	9	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
157	0	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00
157	1	-	-	-	-	-	0.99	0.97	0.89	1.00	1.00	0.00

- 160 -

- 161 -

- 162 -

- 163 -

- 164 -

- 165 -

- 166 -

0											
198 1	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 2	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 3	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 4	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 5	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 6	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 7	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00
198 8	-	-	-	-	-	1.00	1.00	0.99	1.00	1.00	0.00

CURRICULUM VITAE

Name : Yasin BAKIŞ

Permanent Address : Abant İzzet Baysal
Üniversitesi, Fen Edebiyat
Fakültesi, Biyoloji Bölümü,
14280 BOLU

Degree and date to be conferred : (Ph.D. 2013)

Place/Date of Birth : AYDIN/1978

Education

Collegiate institutions attended	Dates	Degree	Date of degree
Abant İzzet Baysal Üniversitesi	1996-2001	B.S	2001
Abant İzzet Baysal Üniversitesi	2002-2005	M.S	2005

Positions held

Position	Institution	Dates
Internship	TÜBİTAK, GMBAE, Biyoinformatik Birimi	2001 (3 months)
Research Assistant	Abant İzzet Baysal Üniversitesi, Biyoloji Bölümü	2001-

Publications

[Conference papers]

Bakış Y, Sezerman OU, Babaç MT and Meydan C (2009) Optimization of morphological data in numerical taxonomy analysis using genetic algorithms feature selection method. 11th Genetic and Evolutionary Computation Conference, GECCO 2009, Montreal, Québec, Canada

Yörükoğlu D, Bakış Y and Sezerman OU (2009) An Entropy Based Heuristic Model for Predicting Functional Sub-type Divisions of Protein Families. 11th Genetic and Evolutionary Computation Conference, GECCO 2009, Montreal, Québec, Canada

Bakış Y, Sezerman OU and Otu HH (2009) Hizalama Temelli Olmayan Benzerlik/Farklılık Ölçüm Metodları İçin Yeni Tekrar-Örnekleme Teknikleri. 16. Ulusal Biyoteknoloji Kongresi, Kongre Kitapçığı, Antalya

Bakış Y, Sezerman OU and Otu HH (2010) Physico-chemical Properties of DNA in Phylogeny Construction. 5th International Symposium on Health Informatics and Bioinformatics (HIBIT). IEEE Xplore, Antalya

Bakış Y and Sezerman OU (2010) A Genetic Algorithm Based Model for Predicting Functional Classification of Proteins. Intelligent Systems for Molecular Biology, Proceedings Book, Boston

Bakış Y, Babaç MT and Uslu E (2011) Updates and improvements of Turkish Plants Data Service (TÜBİVES). 6th International Symposium on Health Informatics and Bioinformatics (HIBIT). IEEE Xplore, İzmir

Bakış Y and Sezerman OU (2012) Proteinlerin İşlevsel Gruplamasında Muhtemel Önemli Pozisyonların Tesbiti İçin Biyoenformatik Araç. Ulusal Moleküler Biyoloji ve Biyoteknoloji Kongresi, Kongre Kitapçığı, Antalya

[Regular papers]

Uslu E, Bakış Y and Babaç MT (2011) A study on biogeographical distribution of Turkish oak species and their relations with the Anatolian diagonal. Acta Botanica Hungarica 53: 3-4

Bakış Y, Otu HH and Sezerman OU (2012) Inferring Phylogenies from Physico-Chemical Properties of DNA. American Journal of Bioinformatics Research 2(1): 1-6

Uslu E and Bakış Y (2012) Geographic distribution of Turkish Oaks. Dendrobiology 67: 41-48

Bakış Y, Otu HH, Taşçı N, Meydan C, Bilgin N, Yüzbaşıoğlu S and Sezerman OU (2013) Testing robustness of relative complexity measure method constructing robust phylogenetic trees for *Galanthus* L. Using the relative complexity measure. BMC Bioinformatics, 14:20

Bakış Y and Babaç MT (UNDER REVISION) Morphological variability of acorns and its taxonomic significance in *Quercus* L. genus from Turkey. Botanical Studies

Uslu E and Bakış Y (UNDER REVISION) Morphometric analyses of the leaf variations among the Section *Cerris* Loudon (Fagaceae: *Quercus* L.) in Turkey. Annals of Forest Research

Bakış Y, Babaç MT and Sezerman OU (UNDER REVISION) Optimization of morphological data and extracting informative characters. Current Bioinformatics

Bakış Y (UNDER REVISION) Turkey at the Junction of Three Phytogeographical Regions: A Grid Base Floristic Study. Botanical Studies

Bakış Y and Sezerman OU (UNDER REVISION) A new bio-inspired resampling technique for molecular sequences. Journal of Bioinformatics and Computational Biology