Lateral Gene Transfer between Bacteria and *Trichoderma reesei*

By

Charles Worley

November, 2011

Director: Jinling Huang

DEPARTMENT OF BIOLOGY

Lateral gene transfer (LGT) has played an important role in shaping the 'tree of life'. It has promoted evolution by generating gene redundancies, which allow for more freedom in terms of allowed mutations in a genome, and by introducing new metabolic abilities to organisms. An example, with practical implications, is the cellulase activity of the fungus *Trichoderma reesei* which degrades cellulose in biofuel production, environmental waste cleanup, and paper processing.

While partial optimization of conditions for *T. reesei* to express enzymes might be achieved through controlling experimental conditions, potentially the most effective approach would be to bring the power of genetic engineering to bear. If the digestive enzymes in *T. reesei* are the result of LGT, inserting bacterial copies of these genes or their regulators into the *T. reesei* genome may increase enzymatic activity. Furthermore, an appropriate examination of the entire genome for LGT could lead to insights into the evolution of *T. reesei*.

This study performed phylogenomic analyses to identify LGT in the genome of *T. reesei*, specifically those genes transferred from bacteria. Three computational programs were used to identify LGT candidates. All identified LGT candidates were subject to further rigorous phylogenetic analyses to assess their evolutionary origins.

We detected several cases of LGT between fungi and bacteria, only one of which was recently transferred to *T. reesei*. This leads us to conclude that LGT probably did not play a significant role in the recent evolution of *T. reesei*.

Lateral Gene Transfer between Bacteria and *Trichoderma reesei*

A Dissertation

Presented To

The Faculty of the Department of Biology

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

M.S. Biology

By

Charles Worley

January, 2012

Lateral Gene Transfer Between Bacteria and *Trichoderma reesei*

by

Charles Worley

APPROVED BY:
DIRECTOR OF
DISSERTATION/THESIS:_____
Jinling Huang, PhD

COMMITTEE MEMBER:
_____
Matt Schrenk, PhD

COMMITTEE MEMBER:
_____
Edmund Stellwag, PhD

COMMITTEE MEMBER:
_____
Qin Ding, PhD

CHAIR OF THE DEPARTMENT OF BIOLOGY
_____
Terry West, PhD

DEAN OF THE GRADUATE
SCHOOL_____
Paul J. Gemperline, PhD

ACKNOWLEDGEMENTS

I would like to thank my primary advisor Dr. Jinling Huang for his patience and advice. I would also like to thank my thesis committee for their vital assistance and aid in helping me complete this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER1: INTRODUCTION

1. Biology of the Soft Rot Fungus Trichoderma reesei

Fungi are organisms which (with a few exceptions) live by the degradation and absorption of nutrients present in their environment. Fungi degrade molecules too large to be taken up readily through the excretion of digestive exoenzymes, such as cellulases, into the environment. *T. reesei* is a multicelluar fungus whose body is composed of tubular strands called hyphae. All of the hyphae of an individual are together referred to as mycelium and constitute nearly the entire body of the fungus. Hyphae are responsible for nutrient uptake; the tubular form maximizes the ratio of surface area to volume and thus helps to optimize both secretion and uptake, and serves as an indicator of the evolutionary drive to optimize these processes (Freeman, 2002).

Trichoderma reesei is a non-parasitic fungus which obtains nutrients by degrading decaying plant matter, including cellulose and lignin, and absorbing the byproducts of that process. Crucially, it is capable of completely degrading cellulose to glucose (Dashtban, 2009). A byproduct of glucose fermentation is ethanol. Industrial ethanol production relies on this process (Mach-Ainger, 2008). *T. reesei* stands out among organisms which can degrade cellulose at rates sufficient for industrial use. This metabolic efficiency does not appear to be due to a unique variant of an enzyme but rather the ability of the fungus to secrete digestive enzymes at a prodigious rate (Mach-Ainger, 2008). Because of this ability *T. reesei* is an organism of interest in second generation biofuel production (Dashtban, 2009).

While fermentation has been used to produce alcohol since ancient times, efficiency is a critical issue when discussing industrial usage. Refinement of current processes is needed in

order to make ethanol production economically viable candidate as an alternative to fossil fuels. Even under ideal conditions *T. reesei* produces ethanol at a cost inefficient rate (Maki, 2009). It is theoretically possible to improve the efficiency of ethanol production through genetic engineering. Therefore, genetic studies are needed to evaluate which naturally occurring strains are best and which genes are potential targets for genetic modification. Some of these genes may be acquired from other organisms. By identifying the donor(s) of these cellulose degrading genes in *T. reesei,* parts of the regulatory pathway from the donor organism might be used to improve the rate of enzyme secretion.

The genome of *T. reesei* has been fully sequenced and an analysis of proteins related to the digestion of plant cell wall material has been completed. Surprisingly, given its known efficiency at degrading plant matter, *T. reesei* has fewer proteins involved in the degradation of plant cell walls than many other fungi. In addition, it appears that *T. reesei* lacks entire families of esterases involved in the degradation of hemicellulose, indicating a 'severe handicap' in this area (Mach-Ainger, 2008).

2. T. reesei and Lateral Gene Transfer (LGT)

Genes, DNA coding for proteins (or RNA end products), are under selective pressure which shapes their evolution. Regulatory elements that control the expression of coding genes are also under selective pressure. This pressure may lead to more adaptive regulatory genes. This improved adaptivity might take the form of improved efficiency at recruiting transcription factors or binding repressors. This in turn influences the degree to which a gene is expressed; when it's expressed; and in multi cellular organisms, where it is expressed. Finally, genes don't evolve in isolation but rather are subject to synergistic affects across a genome (Carrol, 2009).

LGT creates a situation where the transferred gene, if retained, will be under the influence of selective pressures just like any other gene. The transferred gene will be regulated by mechanisms available in its new host. If the transferred gene is only part of a pathway it might be possible to complement expression by genetically modifying the organism with other parts of that pathway. The first order of business in carrying out such a procedure is determining whether or not any relevant genes were the result of gene transfer. If a lone gene is laterally transferred regulatory mechanisms which might aid in its expression are left behind. It is possible however, that these regulatory genes may be of use if they are later introduced to the recipient organism.

Genetic modification of an organism to improve its utility is a well established practice. For example, bacteria don't naturally produce insulin. However, they can be genetically modified so that they do (Williams, 1982), which is how insulin for diabetics is now produced. *T. reesei* produces a number of exoenzymes which are of use to humans. Therefore, modification of *T. reesei* in order to enhance its effectiveness as a cellulose degrader is already being examined (Dashtban, 2009).

3. LGT in evolution

Speciation is, in the strictest Darwinian sense, descent with modification driven by natural selection. In such a system, a mutation may lead to a change in a protein, gene regulation, or other biological processes. This leads to a different phenotype which can be passed on to descendants who then undergo cycles of mutation and selection. While natural selection and descent with modification are potent and mostly satisfactory explanations of how the diversity of life arose and is maintained, there are corollaries. For instance, sexual selection can promote a trait which is detrimental to survival but increases the chances of producing offspring.

Whereas sexual selection might be viewed as a corollary to natural selection, LGT might be viewed as a corollary to mutation as a source of novel genetic changes. LGT is simply put the nonsexual transfer of genetic material from one organism to another. This is as opposed to vertical gene transfer where genetic material is inherited from a progenitor. The genetic material received via LGT may then be passed onto to future generations. LGT has most likely played a significant role in the shaping of the evolution of life on earth. It has promoted evolution by inducing gene redundancies allowing for more freedom in terms of mutation and by introducing new metabolic abilities to organisms (Hilario, 1993).

A classic example of LGT is antibiotic resistance in bacteria. Some bacteria possess a mechanism called conjugation allowing them to transfer genetic information from one individual to another. Therefore, if a bacterium obtains a gene for antibiotic resistance through evolution it may then able to transfer that resistance to other bacteria. These transfer recipients can in turn pass the information on to their descendants (Koenig, 1999). The transfer of genetic material can occur through a number of different mechanisms and is not limited to bacteria. In addition to conjugation, LGT may occur through the mispackaging of genetic material into viral capsids, uptake from the environment, or through the consumption of another organism. (König, 1999)

Detecting potential LGT between organisms may begin either with the results of phylogenetic analysis or with an observation that two unrelated taxa share a metabolic ability absent from other more closely related organisms. The observation is only suggestive; such a pattern could be caused by convergent evolution or by gene loss from the intermediate taxa. The observation indicates the possibility of LGT that can be elucidated by further analysis (Marcet-Houben, 2010).

There are a number of barriers to LGT in eukaryotes, in particular multicelluar eukaryotes. Universal barriers include the lack of a conjugation system to facilitate the transfer of genetic material among organisms. Furthermore, the existence of a dedicated germ line also renders transfer to non-germ line cells uninheritable. Finally, the complex regulation of eukaryotic genes could potentially prevent a transferred gene from being expressed in a way that would be beneficial to the organism (Keeling, 2008).

Nonetheless, there is considerable evidence for LGT across a wide range of organisms. The degree to which LGT has shaped bacterial evolution is difficult to exaggerate. Although vertical forces still dominate, serious questions have arisen as to whether the traditional species concept can be applied to bacteria and whether a 'tree of life' is a fitting metaphor (Hilario, 1993).

CHAPTER 2: MATERIALS AND METHODS

1. Phylogenetics and LGT

The tools of phylogenetic analysis can be used to detect LGT. Phylogenetic analysis consists of constructing phylogenetic trees and using the trees to infer relationships between organisms or genes. Phylogenetic analysis as it applies to the detection of LGT involves reconstructing the phylogeny of a gene. This gene phylogeny is then compared to a known taxonomic phylogeny. If the query gene forms a strongly supported clade with homologs from distantly related taxa, it may be horizontally transferred (Ricard, 2006).

Detecting LGT using phylogenetic methods can be difficult and time consuming however, and a number of approaches are available (Graur, 2000). Often the approaches entail automated steps carried out by computer programs in order to save time. Three such programs used during the course of this research project were Phylogenie, DarkHorse, and AlienG (discussed below).

2. Homologue Selection and Automated Methods

T. reesei's genome has been sequenced (Martinez, 2008). Putative genes were obtained from the *T. reesei* website (http://genome.jgi-psf.org/Trire2/Trire2.home.html). These translated gene sequences were downloaded from a link on the *T.reesei w*ebsite: TreeseiV2_FrozenGene Catalog20081022.proteins.fasta.gz. In order to determine homologues the National Center for Biotechnology Information (NCBI) provides documented resources to facilitate the detection of homologues for query genes. To identify gene homologues the program BLASTp and NCBI's non-redundant (nr) protein database were used unless otherwise noted. Both BLASTp and the nr

database are available as web based applications which run more quickly for individual genes but are not suitable for running an entire genome at one time.

The initial step, the detection of homologues, is crucial. The purpose is to identify known genes that are actually related to the gene of interest. This is generally done by carrying out an automated search of a database using the gene of interest as a query. Those genes which are deemed sufficiently similar are putative homologues. However, similarity does not correspond perfectly with relatedness so simply ranking genes by degree of similarity does not reveal their relationships. While both amino acid and nucleotide sequences can be used as queries, amino acid sequences are often used to identify homologues in distantly related groups (Ricard, 2006).

In order to reduce the number of genes to be analyzed, three automated programs were used to screen for candidates. These programs were DarkHorse, PhyloGenie, and AlienG. These programs are automated routines for identifying laterally transferred genes. These automated methods differ in their algorithmic detail and implementation. Since these three programs employ different techniques, they each produce a different set of candidates increasing the chance of detecting true transfers.

PhyloGenie is a phylogenomic program package commonly used for detecting gene transfer candidates. PhyloGenie is available as a free download at http://www.eb.tuebingen.mpg.de/departments/1-protein-evolution/software/phylogenie. The program accepts a complete gene set from a whole genome as input and constructs trees for individual genes. LGT candidates can be identified based on subsequent analyses of tree topologies. If a gene is likely LGT-derived, further analysis and manual tree construction should be undertaken to verify its evolutionary origin (Frickery, 2004). Preliminary analysis of the *T. reesei* genome using PhyloGenie identified over 200 LGT candidates.

In order to facilitate detection of transfers, PhyloGenie only displays trees for genes that have BLAST identified homologues in certain lineages. For our purposes the clades chosen were prokaryotes and fungi. The command entered was (Fungi &(Bacteria | Archaea)). PhyloGenie was instructed to ignore genes that that contained certain NCBI terms. These were unknown taxonomy, other sequences, Viruses, Viroids, and unclassified. BLAST was carried out using stand alone protein BLAST (BLASTp) with default settings, except that the maximum number of returned hits was increased to 1000. The configure file was set to draw tree branches regardless of their support value. Customarily PhyloGenie would be run against the nr database. However, attempts to BLAST the *T. reesei* genome against large databases resulted in overflow issues causing the program to crash. For that reason we created our own database comprised of 259 genomes with balanced representatives from throughout the tree of life.

DarkHorse takes an alternative approach to phylogenetic analysis that does not involve tree construction. Instead it calculates a lineage probability index (LPI) of high scoring matches, a value which describes the lineages of high scoring BLAST matches and which expresses the likelihood a gene is a possible case of LGT. Calculation of LPI requires a filter value (Podell, 2007). The threshold value we determined to be suitable for *T. reesei* is 20 percent. Analyses of the *T. reesei* genome using DarkHorse determined 129 LGT candidates. For DarkHorse in general default settings were used. However, the filter threshold used was varied across multiple runs. After comparing the average candidate set size (see Results), produced by each filter setting, it was determined that a threshold of 20 percent was optimal. The exclusion list terms were: cloning, vector, plasmid, cosmid, expression, environmental, synthetic, construct, contaminant, unclassified, unidentified, unknown, untyped, unspecified, clone, *Trichoderma*,

Ascomycota, and *Hypocrea*. DarkHorse was obtained as a downloaded from http://darkhorse.ucsd.edu.

AlienG is a newly developed program which like DarkHorse avoids tree generation. A value called the Alien index is used to score potential transfers. AlienG is relatively fast and it can overcome the difficulties caused by simply checking the top bit score when evaluating possible LGT. AlienG allows the user to tailor their search by inputting exclusion terms, a donor group, and other taxa deemed related. Like PhyloGenie, AlienG carries out BLASTp and attempts to correct for the imperfect association between BLAST rankings and relatedness. The settings used for BLASTp differed from default in two ways. The top 1000 hits and a minimum Evalue of 1 e-5 all other settings were used. The exclusion list included Fungi, contaminant, other sequences, crystal structure, ternary complex, |pdb|, artificial sequences, synthetic construct, viruses, plasmids, environmental samples, and uncultured. The donor group was defined as either bacteria or archaea. All other taxa (group 2) were deemed to be 'related'. AlienG was developed at ECU and a copy of the software was supplied by Dr. Huang's lab.

3. Manual Analysis

Each query gene file was uploaded individually to the web based BLASTp platform and queried against a web accessible nr database. The default settings for web based BLASTp were adhered to except that the maximum number of resulting search hits was increased from 100 to 5000. The number of hits detected by BLASTp was increased in order to improve the diversity of lineages detected and to provide a more complete list of homologues for analysis.

Not all genes homologues detected by BLASTp are true homologues or similar enough to be used for phylogenetic analysis (Xiong, 2006). BLASTp output provides a measure of similarity called the Evalue. The Evalue takes into account the probability (p-score) of two

sequences of given length having a certain degree of similarity by chance and then factors in the size of the relevant database (the larger the database the greater the chance of finding two sequences with any degree of similarity by chance).

The p-score is calculated by first aligning two sequences and then generating a raw alignment score. Then random sequences are generated by shuffling residues in one of the two sequences and then realigning. This process is repeated many times generating a random distribution (this is the Gumble extreme value distribution). The score of the initial alignment is then compared to the distribution of scores it falls in the extreme margin of the distribution it means the alignment it is very unlikely to be due to chance. Next, this value is multiplied by the number of residues in the query and in the database. (Xiong, 2006)

Generally, gene sequences with an Evalue of 1e -50 are considered to have a good to almost certain chance of being homologues. Sequences with an Evalue in the range of 1e -50 to .01 are considered to potentially have a distant homology. Sequences with an Evalue below 10 are considered to either be unrelated or so distantly related that homology can't be detected. (Xiong, 2006) For manual analysis generally only sequences with an Evalue below 1e 10-4 were used in subsequent alignment and phylogenetic tree construction. In some cases, sequences with Evalues greater than 1e -4 were included when they were highest ranking representatives of a major lineage. Typically sequences with relatively high Evalues could not be aligned and had to be removed during sequence alignment. Query sequences that were short, contained large regions of low complexity, or with had large portions of the sequence containing repeats were dropped from analysis as such sequences are not generally suitable for alignment in most cases (Xiong, 2006).

The downloaded homologues were more numerous than is suitable for alignment and tree construction and had to be reduced in number. A number of criteria were considered when removing homologues. One organism might have several homologues of a query gene and thus be represented numerous times in BLAST output. These extra homologues were removed to avoid a single organism appearing multiple times in the final tree. Some lineages were often over represented and other lineages underrepresented among the homologues. If disparate representation was carried over to the alignment it could potentially skew the final alignment. Lower scoring matches might be too divergent to be aligned. For these reasons it was necessary to select a representative subset of the BLAST output for use in alignment.

The subset was determined by following a few simple rules. Firstly, the output was reorganized so that sequences from the same organism were grouped together in order of Evalue and that sequences from related organisms were grouped together in order of Evalue. Next one to three representative organisms from a given lineage were kept whereas the remaining members of a lineage were deleted. Only the top scoring homologue from a given organism was kept. A partial exception to this rule was ascomycota (the fungal class to which *T. reesei* belongs). More than three representatives of ascomycota were at times taken. *T. reesei'*s closest relatives would be expected to form a monophyletic grouping with *T. reesei* if there was no LGT was involved in the genes evolution.

A potential short coming of the nr database is the uneven representation of taxa. In order to compensate for the under representation for some lineages in the nr database a second locally constructed database composed of selected eukaryotic genomes was used. A downloadable local version of BLASTp is available from NCBI. This version of BLASTp is functionally identical in the identification of homologues to the web based version. However, it does lack some

secondary features such as repeat detection. The same settings for the local run of BLASTp were used and the same parsing and editing of output took place. Generally, there were fewer hits and these hits were of lower quality than those detected using the nr database. If only a small number of homologues were detected against the local database they were typically included, as the best representatives of their lineages, in subsequent analysis as long as they were above BLASTp's own Evalue cut off.

ClustalX provides a graphical representation of the sequence to be aligned using a color coding scheme to show non-identical residues with similar physiochemical properties (which are more likely to be substituted for one another than less similar residues) which helps facilitate analysis of alignment quality. In addition because it provides a graphical representation of gene sequences, prior to alignment, it is easy to discern which sequences are too short or too long to align properly and these can be removed. ClustalX is available as a free download at www.clustal.org.

Typically, 30-60 sequences per query were loaded into ClustalX for alignment. Following the first run of the program those sequences which aligned poorly or appeared to be disrupting the alignment were removed and the remaining sequences realigned. This was typically repeated until the remaining sequences were adequately aligned. Following alignment regions of low homology where one or more gene sequences contain deletions or insertions remain. Regions of the alignment with low homology must be removed from all the genes in the alignment.

Following construction of an alignment, the appropriate amino acid substitution matrix (a matrix reflecting the probability of any residue be substituted for any other residue) must be

determined. This was performed by the program ModelGenerator (Keene, 2004). For all examined sequences, either the WAG model, BLOSSUM62 model, or the Rtrev substitution model was used. Typically the invariable WAG model was used though in some cases the gamma WAG model was used. After the optimal substitution matrix is determined the alignment can be used for tree construction. ModelGenerator is available at http://bioinf.nuim.ie/modelgenerator.

Two different methods were employed to generate phylogenetic trees. These were neighbor joining and Bayesian analysis. Neighbor joining was carried out by using several programs from the PHYLIP (PHYLogeny Inference Package) program suite (Felsenstein, 1993). Within the PHYLIP suite the following three programs will be used to carry out the neighbor joining and subsequent bootstrapping; they are SEQBOOT, NEIGHBOR, and CONSENSE. Bootstrapping was carried out for 100 pseudoreplicates. The default settings for all programs were used. PHYLIP is available at evolution.genetics.washington.edu/phylip.html.

SEQBOOT is a program which generates multiple resampled datasets. The input for this file was the alignments generated by ClustalX. Default settings (see Table) were used. NEIGHBOR is a tree-constructing program which employs the neighbor joining method. In addition to the original alignment each resampled data set is also converted to a phylogenetic tree. Thus for an original alignment 100 trees are generate. CONSENSE converts the 100 trees into one tree with bootstrap values. The consensus trees produced by CONSENSE are distinct in that branch lengths correspond to bootstrap values rather than evolutionary distances. For the purposes of this project the appropriate bootstrap values were manually entered in the consensus tree so that the tree could be rooted (necessary to show evolutionary direction), a process which

can alter branch lengths. Thus the branch lengths on a PHYLIP output tree do not correspond to evolutionary distance and should be ignored when interpreting the phylogenetic trees.

Bayesian analysis was carried out using the program MrBayes 3.1 (Huelsenbeck, 2001). MrBayes was obtained at http://mrbayes.sourceforge.net/download.php. Default settings were employed where applicable. MrBayes employs Bayesian analysis and was run either for 10000 iterations or a multiple of 10000 iterations until convergence dropped below .01. Typically, trees were sampled every 50 or 100 iterations. Four rate categories were used to approximate the gamma distribution. The following settings were entered upon opening MrBayes: 1. Lset = rates(invgamma), except in a few cases where ModelGenerator indicated that gamma substitution was superior to an invariable gamma rates. 2. Prset aamode = fixedl(x), where x was the amino acid substitution matrix determined by ModelGenerator. 3. Mcmc samplefreq=y ngen=z, where y was typically set to either 50 or 100 depending on an estimation of the number of generations necessary to reach adequate convergence. The estimation was based on input file size. Z varied in value depending on file size but was always a multiple of 10000.

Following completion of a run, two more commands were entered to produce the final output. These were sump burnin=n and sumt burnin=n. Where n is equal 25 percent of the number of generations divided by the sample frequency. If convergence was not achieved in the set number of generations, analysis was continued until it did. In a few cases convergence was never achieved in over 1 – 1.5 million generations. In such cases tree construction was deemed a failure and no output was available from MrBayes.

The trees generated had branch tips which were labeled in, an at times, unclear manner. Furthermore the organisms were identified by different levels of nomenclature by each approach.

The trees were edited so that the organism classification was more consistent and clear between approaches. Graphical editing of the trees was carried out by a usage of either NJ Plot and Adobe Illustrator or NJ Plot and PowerPoint. The phylogenetic trees from both approaches were evaluated for indications of gene transfer. In this study, genes determined to be transfers by both programs are considered to be likely cases of LGT.

Analysis of phylogenetic trees for LGT involves identifying a difference between an organismal phylogeny and a gene phylogeny which can't be reasonably explained by another hypothesis. A phylogenetic tree consists of a series of branches joined together at nodes. The tips of the branches represent extant taxa or genes and the underlying pattern of branches and nodes is a model of their relatedness. An organismal phylogeny is a phylogeny which shows the evolutionary relationships between organisms. A gene phylogeny shows the apparent evolutionary relationship between genes. Besides gene transfer there are a number of reasons a gene phylogeny may differ from an organismal phylogeny. Genes can evolve at different rates from one another and their rate of evolution can change over time, so that two sets of genes in two different organisms may appear to differ greatly in terms of divergence from a third gene. Another factor is gene duplication and gene families, in such cases related genes are found in the same organism which can potentially lead to confusion. Other problems include gene loss, 'noise' and divergence which can obscure real genetic signals.

Gene transfer can be deduced as reasonably certain when a gene is present in two widely divergent lineages and absent in all intermediary lineages or when a strong stable alliance is suggested between two widely diverse lineages while intermediary lineages represent a separate clade or clades.

CHAPTER III: RESULTS

1. General Overview

Following the full analysis of the genome, a total of 13 genes were identified as derived from LGT events to *T. reesei* or its recent ancestors. Of these 13 genes, 5 have been deemed as strong transfer candidates. The remaining 7 genes while strong candidates have complex phylogenies that complicate definitive identification. An additional 9 genes were identified as possible cases of more ancient LGT. Finally, a number of genes were deemed to be of indeterminate status, either due to a lack of strong BLAST matches, short length of the alignment, or repeat richness.

Analysis of the initial PhyloGenie output resulted in the detection of 289 putative candidates. DarkHorse identified 129 candidates (this included significant overlap with PhyloGenie output) and AlienG identified 14 transfers at the specified settings. PhyloGenie detected of 8 genuine LGTs over 289 putative recent transfers giving it a high false positive rate. DarkHorse fared better with 11 over 129 transfers candidates being genuine LGTs, though that is still less than 10 percent. AlienG had the lowest false positive rate with 5 over 14 resulting in a false positive rate of less than 60 percent.

DarkHorse, AlienG and PhyloGenie were all responsible for detecting at least one case of recent transfer. PhyloGenie detected 8 of the 13 transfer candidates. DarkHorse detected 11 of the 13 LGTs and AlienG detected 5. Only in 2 cases did all 3 programs detect the same transfer candidate. DarkHorse and PhyloGenie detected 6 of the same transfers (counting those which were detected by AlienG as well). DarkHorse and AlienG detected 5 of the same transfers

meaning that all of the recent transfers detected by AlienG were also detected by DarkHorse and that only two detected by PhyloGenie were also detected by AlienG.

PhyloGenie had a relatively high false positive rate. The 289 putative candidates represent only about a third of the 690 output trees which needed to be examined (PhyloGenie outputs trees containing certain preselected taxa) – however it is up to the user to ascertain whether or not the tree indicates possible transfer. In order to find the highest number of transfers possible, even trees of ambiguous status underwent manual analysis. Thus a PhyloGenie output tree that did not have strong bootstrap analysis (a frequent issue) was still considered a potential candidate because the tree did not rule transfer out.

Another factor contributing to PhyloGenie's high false positive rate was the fact that it makes no allowance for multiple orthologous genes coming from a single organism. Thus a single organism could appear ten times on a tree once for every related gene in its genome. The complex interaction between different sets of orthologous genes being drawn on a single phylogeny complicates attempts to interpret the resulting tree as a true phylogeny. A final drawback to PhyloGenie is that the program did not regard the possibility of an alignment containing too many or too few sequences for meaningful analysis. For example, in at least one case PhyloGenie drew a phylogenetic 'tree' containing only one branch connecting two organisms.

AlienG was far more efficient than PhyloGenie at detecting transfers, being markedly quicker to run taking around 24 hours, compared to PhyloGenie which took weeks to complete a single run. AlienG false positives were generally caused by gene sequences which were too low in quality to be amendable to phylogenetic analysis. In other words some of these false positives

might represent transfers however, it is not possible to reliably demonstrate this using phylogenetic methods.

DarkHorse, like AlienG was quicker and had a lower false positive rate than PhyloGenie. Some DarkHorse false positives were also due in part to low quality gene sequences that could not be analyzed, though it also detected false positives that were demonstrated to not be transfers only following manual analysis. DarkHorse, with a filter threshold of twenty percent, failed to detect only two cases of transfer, the fewest false negatives of any approach. Increasing the filter percentage to thirty percent allowed DarkHorse to detect one of the remaining transfers. Further increases did not result in the additional detection of the remaining transferred genes and were accompanied by increasing numbers of false positives (see Table 1.)

Table 1. DarkHorse Filter Threshold. The table includes the number of putative transfers, the number of these transfers that were verified by manual analyses (true positives) and the percentage of putative transfers that were true positives at various filter threshold settings.

DarkHorse Filter Threshold	**Putative Transfers**	**True Positives**	**Total Hits/True Positives**
0.02	79	10	0.13
0.03	79	10	0.13
0.1	85	10	0.12
0.2	129	11	0.09
0.3	185	12	0.06
0.4	309	12	0.04
0.5	531	12	0.02

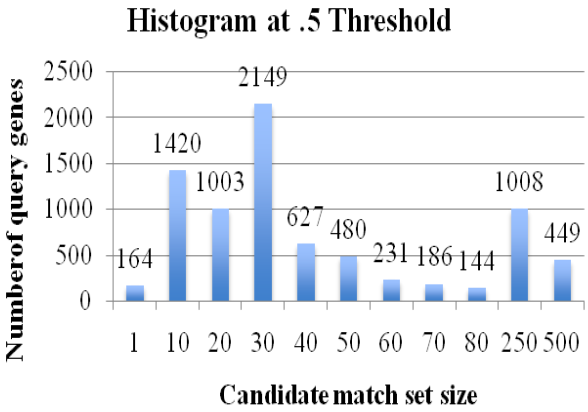
The authors of the DarkHorse paper (Podell, 2007.) did not describe a mathematically rigorous method of determining the size of the filter threshold but did construct a plot of the gene, in their study, with the largest number of matches below 500 versus the threshold percentage. Based on the shape of that curve, they estimated where the filtering threshold should be set. This approach was not feasible for the current study, given the number of genes in our data set. Until the filter was set close to zero there was always a few outliers with large candidate set sizes. However, examining the average candidate set size we were able to select a reasonable threshold value that optimized what was considered to be an ideal candidate set size. (see Figure 1.) Essentially, as the typical candidate set size begins to run off towards 500 it follows that the filter threshold is filtering out fewer sequences and thus is less effective. If the candidate set size is typically one or close to one it follows that the filter may be too effective.

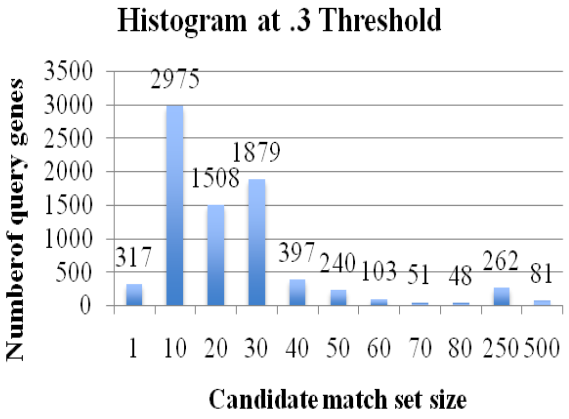
Table 2 below is a list of the transferred genes detected in this study. The table includes the label the gene was given during the project, the loci/gene name annotated on the *T. reesei* genome website, the most prominent domain/or super family (when this is known) associated with the gene and finally, the program or programs which detected it. The trees following the table are the phylogenetic trees generated for each these genes.

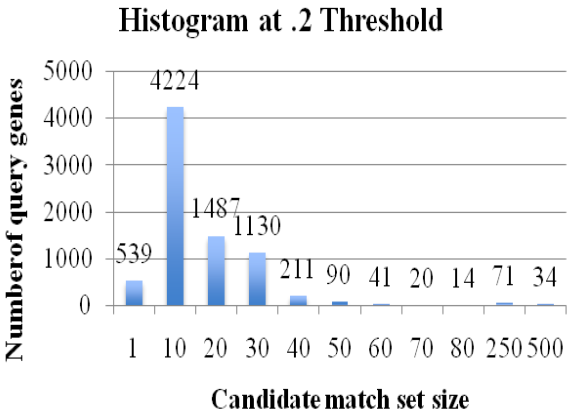
2. Summary of Transferred Genes

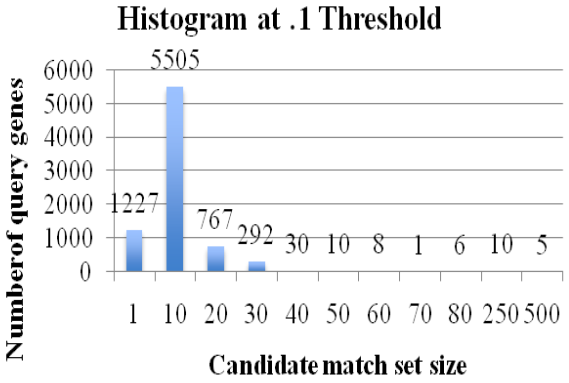
Triree1049, Triree7365, and Triree8375 lack known functions. Furthermore, Triree7365 and Triree8375 lack domains or membership in gene families rendering them enigmas. Triree1049 contains one recognizable domain DUF3626 (Martinez, 2008). While the functions of these protein are unknown this domain family is considered to be bacterial and therefore its presence in a *T. reesei* gene further strengthens the case for gene transfer. There was no other

Figure 1. Query gene's candidate set size distribution at various filter thresholds The scale on the Y axis has been varied from graph to graph in order to facilitate visual comparison between small and large candidate match set sizes.









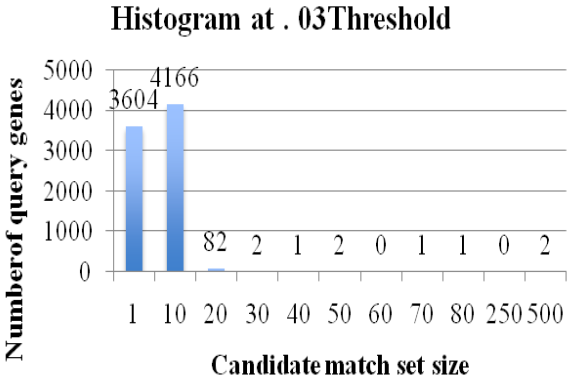


Table 2. LGT-derived genes detected in *T. reesei*

*Duff: Domain of Unknown Function

Gene Index	Official loci/gene name	Putative Function/ Major Domain(s)	Detected by:
Triree1049	Jgi|Trire2|103442 Scaffold_2000030	Duff3626*	Phylogenie, AlienG and DarkHorse
Triree1708b	Jgi|Trire2|104197 Scaffold_3000152	metallo dependent hydrolase	AlienG, DarkHorse
Triree1630	Jgi|Trire2|70025 Scaffold_3000071	Zinc_dependant_alcohol_dehydrogenase	DarkHorse
Triree3136	Jgi|Trire2|121136 Scaffold_5000073	Periplasmic_Binding_Protein_Type_1	AlienG, DarkHorse
Triree4966	Jgi|Trire2|108007 Scaffold_10000397	COG1037, other hydrolases	AlienG, DarkHorse, PhyloGenie
Triree5571	Jgi|Trire2|108671 Scaffold_13000017	Glyco_hydro_3_superfamily	DarkHorse, Phylogenie
Triree6978	Jgi|Trire2|110271 Scaffold_19000156	NADB rossman	DarkHorse
Triree7021	Jgi|Trire2|66598	MDR superfamily	DarkHorse
Triree7365	Jgi|Trire2|110709 Scaffold_22000022	ADP_ribosyl_ GH	PhyloGenie
Triree7540	Jgi|Trire2|68064	Glycho_hydro_43	PhyloGenie, DarkHorse
Triree8067	Jgi|Trire2|72488 Scaffold_27000020	D-aminoacylase	PhyloGenie
Triree8375	Jgi|Trire2|111865 Scaffold 30000004	No Putative domains	AlienG, DarkHorse, PhyloGenie
Triree9071	Jgi|Trire2|71101 Scaffold_53	Crotonase-like superfamily	DarkHorse, PhyloGenie

information available for this protein family and Triree1049 also lacks any other domains or conserved regions.

A number of transferred genes have known domains of functions about which little can at present be said. Triree6978 and Triree7921 are proteins with dihydroflavonol-4-reductase activity (Martinez, 2008). This type of reductase has a variety of functions. It is involved in lignin biosynthesis in plants, and it has been suggested to provide tolerance for cell death in some other organisms. Triree4966 is a potential esterase, lipase, or thioesterase (Martinez,

2008). While the specific function of this enzyme has not been elucidated, NCBI web based analysis indicates that it possesses a putative dienelactone hydrolase.

Triree1630 is a zinc dependent alcohol dehydrogenase (Martinez, 2008). Alcohol dehydrogenases are involved in a variety of pathways in various organisms. An alcohol dehydrogenase is responsible for the reduction of acetyl aldehyde to ethanol during fermentation (and the coupled oxidation of NADH or NADPH to NAD+ or NADP respectively). Thus this particular enzyme may be implicated in alcoholic fermentation which is carried out by *T. reesei*.

Triree1708 contains regions with high similarity to NagA and Metallo-dependent hydrolases, subgroup A (Martinez, 2008). These domains contain metal binding sites and are involved in amino acid synthesis. This is in keeping with top matching BLAST hits which include proteins with amido acid synthesis functions. Specifically it also contains a D-amino-acyl hydrolase domain. This protein contributes to D amino acid synthesis by hydrolyzing N-acyl D-amino acids. Amino acid synthesis is a crucial part of maintaining homeostasis. An inability to synthesize amino acids would restrict any organism to a life style where amino acids were all available from the environment.

Triree3136 is a hypothetical transmembrane protein (Martinez, 2008). Transmembrane proteins have a variety of functions including transporting substance in and out of a cell and detecting environmental stimuli. The specific function of this protein is not designated by the JGI *Trichoderma reesei* website. A search of this gene using NCBI resources indicated that it possesses a ligand binding site, possibly for monosaccharides. This suggests that Triree3136 may be a sugar transporter.

Triree5571 and Triree7540 are enzymes involved in hydrolyzing O-glycosyl compounds and thus implicated in carbohydrate metabolism (Martinez, 2008). The specific targets of these

proteins are unclear. However, they are probably involved in hydrolyzing the bond between a carbohydrate and a non-carbohydrate so they are not likely to be directly involved in cellulose degradation. Triree7540 belongs to glycoside hydrolase family 43 (Martinez, 2008). Enzymes in this family have a variety of functions. However, it has been noted that organisms which degrade plant cell walls have been found to have an expanded complement of enzymes in this class.

Triree8067 is a member of glycoside hydrolase family 95 (Martinez, 2008). These glycoside hydrolases metabolize a variety of polysaccharides including some of the blood group antigens. This protein's involvement in the degradation of carbohydrates is noteworthy.

Triree9071 is a putative member of the crotonase superfamily and is also a class one amino acyl hydratase/isomerase (Martinez, 2008). This enzyme contains tRNA and ATP binding sites and functions as an amino acid activator. The activation of amino acids is necessary for translation and is a core or 'house keeping' biological process.

CHAPTER IV: DISCUSSION

The enzymes primarily responsible for the degradation of cellulose are a class of hydrolases referred to as cellulases. These enzymes work by hydrolyzing the bonds between glucose monomers in cellulose. None of these genes was found to be a transfer candidate. However, not all of the genes involved in cellulose degradation are known. Potentially Triree1049 might have some roll. Furthermore there is more to cell wall degradation than simply breaking up cellulose. Several of the candidate genes, such as Triree7540 and Triree5571, help breakdown oligoproteins and by doing so contribute to cell wall degradation. Nonetheless, it is unlikely that these genes play even an indirect role in cellulose degradation.

Potentially, of more interest is Triree1630 which is an alcohol dehydrogenase. A primary function of alcohol dehydrogenases is the terminal step in alcoholic fermentation, the reduction of acetylaldehyde to ethanol (a process coupled to the oxidation of NADH). Triree6978 is involved in lignin degradation. Lignin degradation is the breakdown of wood. This is of relevance to ethanol production and paper processing. For that reason this gene is a good, probably the best, potential target for further research along the lines of this project.

It's worth noting that most of the transfers detected in the course of this project are hydrolases, many of which are involved in carbohydrate processing. This suggests that many of the genes identified in this study are involved in facilitating the absorptive life style of *T. reesei*. It must be noted that while these genes are examples LGT most were not so recent as to be transfers to *T. reesei*, suggesting that they are unlikely to represent a source of novel genetic change unique to it.

Of the three methods used to select genes for manual analysis, DarkHorse and AlienG proved effective at detecting gene transfer candidates, with DarkHorse detecting the largest number of transfers and AlienG having the lowest ratio of false positives. By evaluating input genes sequences for quality prior to using these programs, it might be possible to reduce the number of false positives. How these programs might be improved upon in terms of avoiding false negatives is less clear but using a generous filter threshold for DarkHorse would make it more inclusive. PhyloGenie is a much slower approach taking weeks or longer to complete a single run. It produces more false positives than any other approach and fails to detect a high percentage of transfers. One issue may have been the use of a customized database but using the nr database led to overflow problems causing the program to crash.

Trees generated by PHYLIP and MrBayes were generally consistent - With one exception, T. reesei was drawn in the same clade on both trees. In the exceptional case, it was drawn equally distant between that clade and another, and this difference did not change any conclusions about direction of transfer. On many trees included here only fungal and bacterial lineages appear. Whenever possible various eukaryotic lineages were included. However, in some cases, no homologues in non-fungal eukaryotes were detected by BLAST, or some were detected right at the edge of significance. Attempting to include these highly divergent sequences in an alignment was often not possible, as these sequences aligned poorly.

In conclusion, with one possible exception there were no transfers from bacteria directly to *T. reesei*. As such it is unlikely that LGT has played a role in *T. reesei'*s ability to degrade cellulose efficiently. Those transfers that were detected were typically older and present in many fungal lineages. Given the documentation of transferred genes between bacteria and eukaryotes to date, it is not surprising that a lineage as ancient as fungi or ascomycota contains evidence

gene transfer events. Finally, the evolutionary cause(s) of *T. reesei*'s cellulase activity does not seem attributable to LGT and most likely arose from vertical evolutionary forces.

REFERENCES

Anderson, J.O. 2005. *Lateral gene transfer in eukaryotes*. CMLS, Cellular and Molecular Life Sciences 62 1182–1197

Baxevanis, A., Ouellette, B. F. 2005. *Bioinformatics: A Practical Guide To The Analysis of Genes and Proteins*. Hoboken, New Jersey. John Wiley & Sons, Inc.

Dashtban, M., Schraft, H., Qin, W. editors. 2009. *Fungal Bioconversion of Lignocellulosic Residues; Opportunities & Perspectives.* International Journal of Biological Sciences. 5(6):578-595

Druzhinina et al. 2010. *Evolution and Ecophysiology of the Industrial Producer Hypocrea jecorina (Anamorph Trichoderma reesei) and a New Sympatric Agamospecies Related to It.* PLoS ONE 5(2):

Freeman, S. editor. *Biological Science.* 2003. Prentice Hall. Upper Saddle River, NJ 2002

Frickery, T. Lupas, A. 2004. *PhyloGenie: automated phylome generation and analysis.* Nucleic Acids Research. Vol. 32, No. 17 5231–5238

Garland, T. Bennett, A. and Rezendel, E. 2005. *Phylogenetic approaches in comparative physiology.* The Journal of Experimental Biology. 208, 3015-3035

Graur, D. Wen-Hsiung L. 2000. *Fundamentals of Molecular Evolution.* Sinauer Associates Inc. Sunderland Massachusetts

Guarro, J. Gene´, J., and Stchigel, A. 1999. *Developments in Fungal Taxonomy*. Clinical Microbiology Reviews. July p. 454–500

Hilario E, Gogarten J.P. 1993. *Lateral transfer of ATPase genes--the tree of life becomes a net of life.* Biosystems. Vol 31, Issues 2-3

König, J. et al. 1999. *Conjugate export pumps of the multidrug resistance protein (MRP) family: localization, substrate specificity, and MRP2-mediated drug resistance.* Biochimica et Biophysica Acta (BBA) – Biomembranes Vol. 1461, Issue 2, Pages 377-394

Kubicek, .C P. et al. 2009. *Fungal genus Hypocrea/Trichoderma: from barcodes to biodiversity.* Journal of Zhejiang University Science. October; 9(10): 753–763

Mach-Aigner, A. R. 2008. *Transcriptional Regulation of xyr1, Encoding the Main Regulator of the Xylanolytic and Cellulolytic Enzyme System in Hypocrea jecorina.* Applied and environmental microbiology. Nov. 6554–6562

Maki, M., Leung, K. T., and Qin, W. 2009. *The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass*. International Journal of Biological Sciences. 5(5): 500–516.

Marcet-Houben, M., Gabaldón, T. 2010 *Acquisition of prokaryotic genes by fungal genomes*. Trends in Genetics Volume 26, Issue 1.

Martinez, Diego et al. 2008. *Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina.* Nature biotechnology. volume 26 number 5 553-560.

Pavlopoulos, Georgios et al. 2010. *A reference guide for tree analysis and visualization.* BioData Mining. 3:1

Podell, S. Gaasterland, T. 2007. *DarkHorse: a method for genome-wide prediction of Lateral gene transfer*. Genome Biology. 8(2): R16

Rosenburg, A., Carroll, S. B. 2010. *Endless forms: the evolution of gene regulation and morphological diversity*. Philosophy of Biology: An Anthology. Blackwell Publishing Ltd.

Salzberg, S. et al. 2001. *Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss?* Science 292. 1903

Seidl, V. 2009. *Sexual development in the industrial workhorse Trichoderma reesei.* PNAS August. vol. 106

Podell, S et al. 2008. *A database of phylogenetically atypical genes in archaeal and bacterial genomes, identified using the DarkHorse algorithm* BMC Bioinformatics. 9: 419. 7

Ricard, G. 2006. *Lateral gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics 10 Feb*

Willams, D.C. 1982. *Cytoplasmic inclusion bodies in Escherichia coli producing biosynthetic human insulin proteins*. Science. Vol. 215 no. 4533 687-689

Xiong, J. 2006. *Essential Bioinformatics.* Cambridge University Press, Cambridge, New York

APPENDIX: PHYLOGENETIC TREES

Figure 2. The molecular phylogeny of Triree1049, Jgi|Trire2|103442. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node. Following completion of research an additional fungal homologue for Triree1049 was added to the nr database.

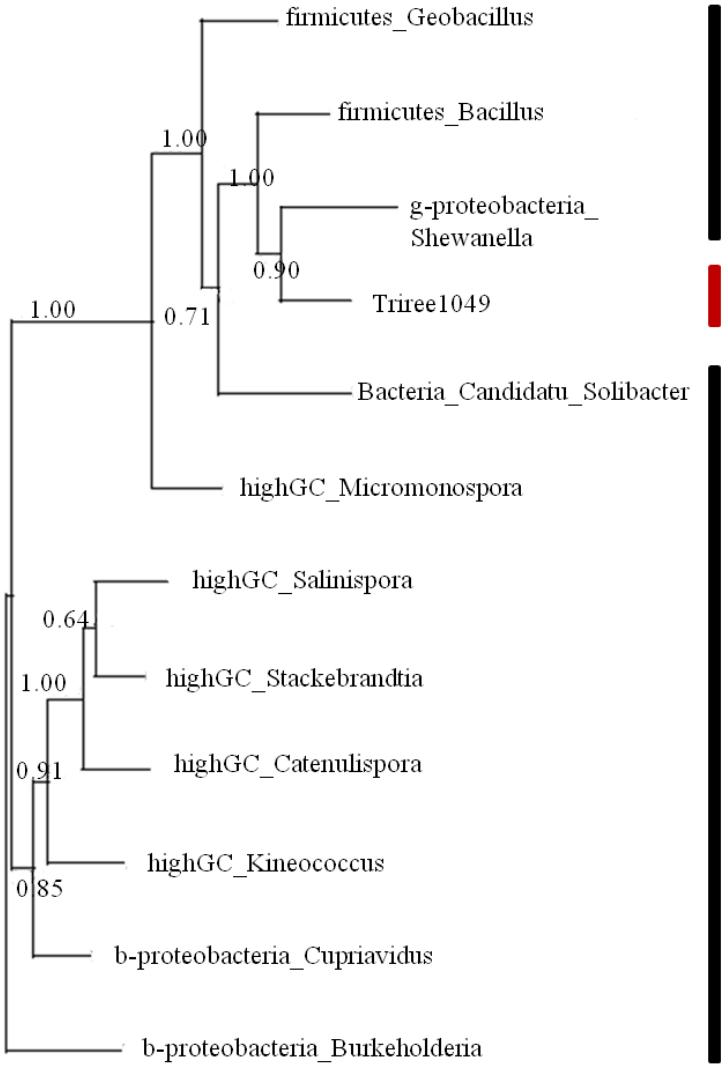


Figure 3. The molecular phylogeny of Triree1049, Jgi|Trire2|103442. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

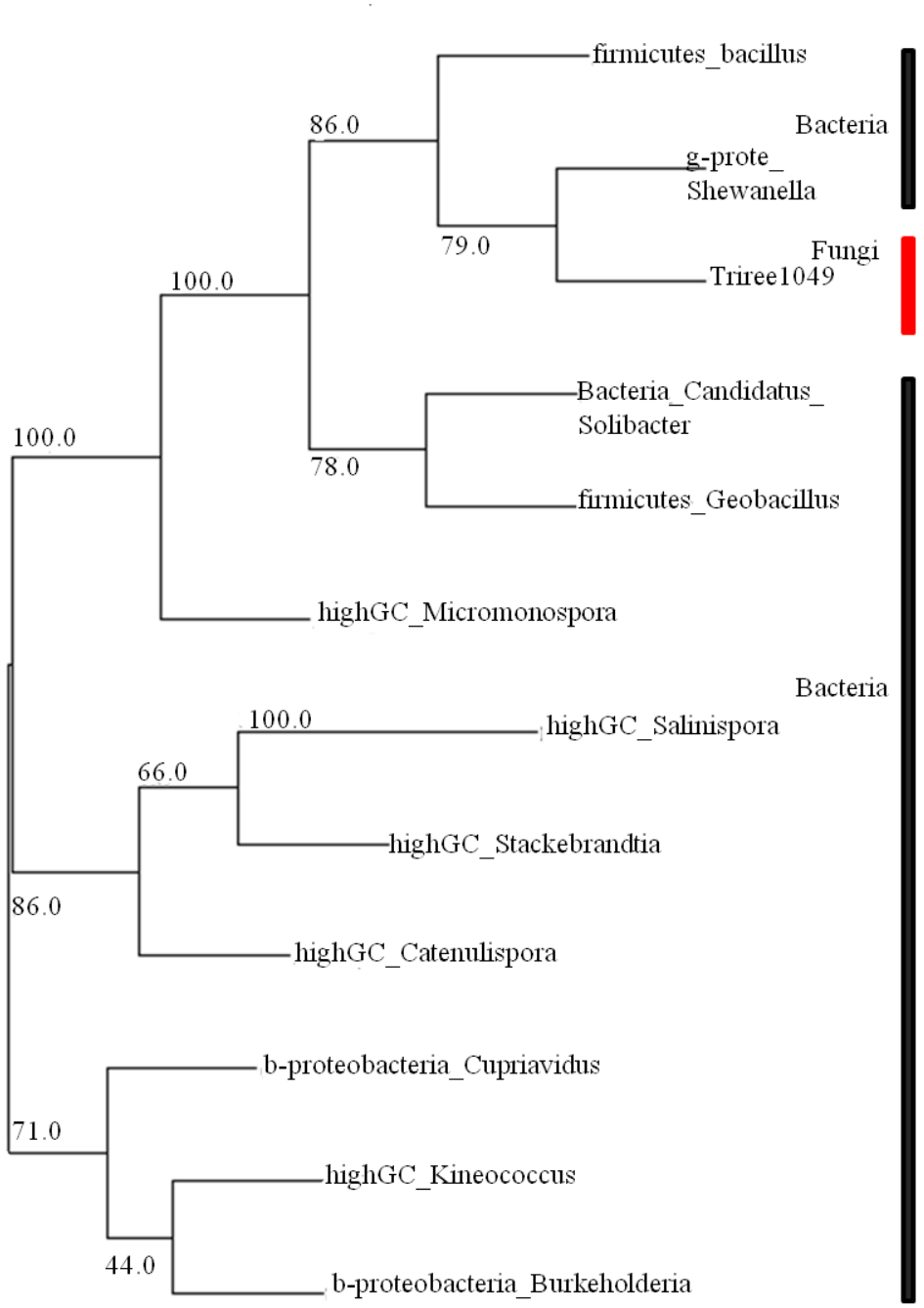


Figure 4. The molecular phylogeny of Triree1630, Jgi|Trire2|70025. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

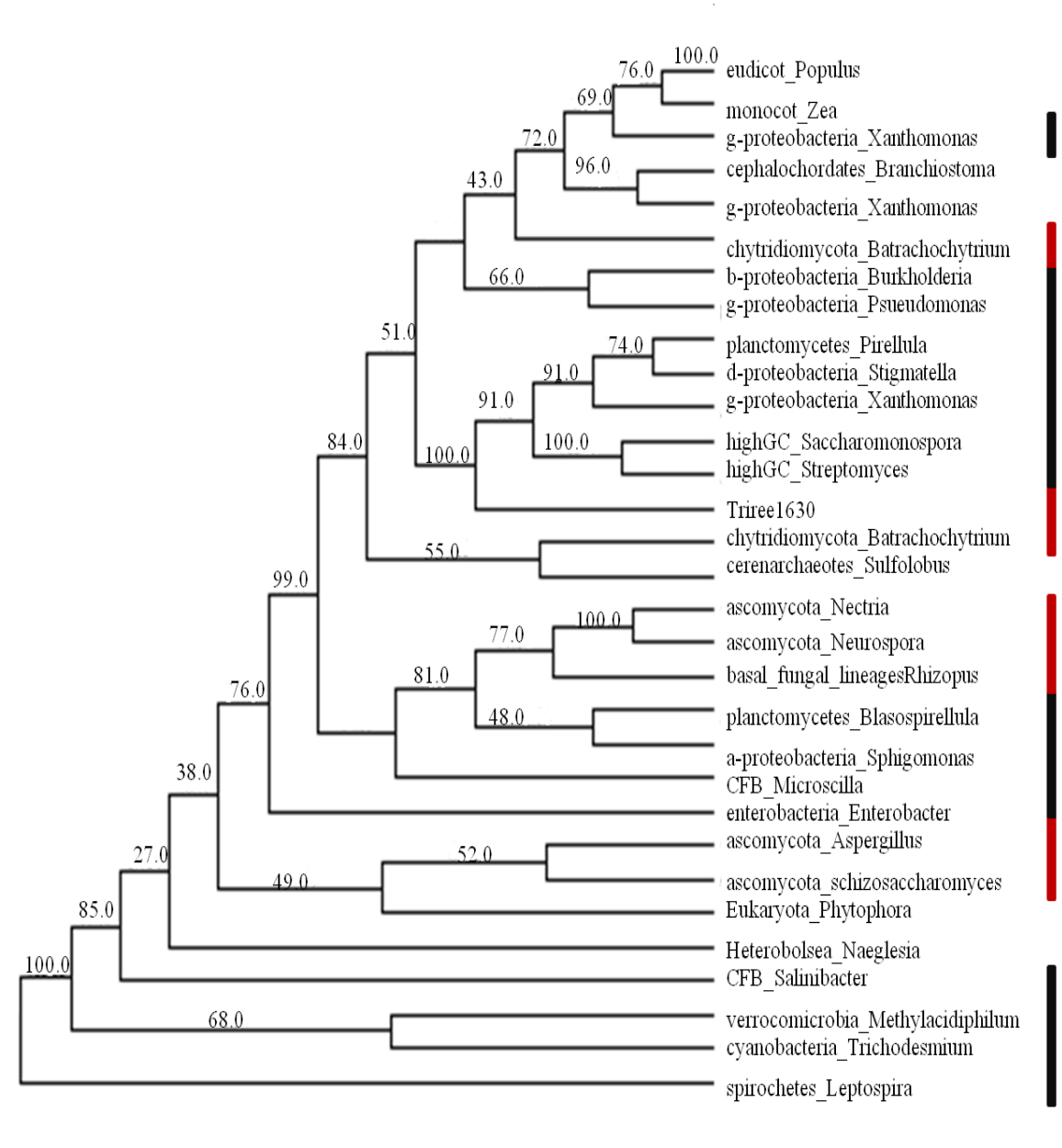


Figure 5. The molecular phylogeny of Triree1708, Jgi|Trire2|104197. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

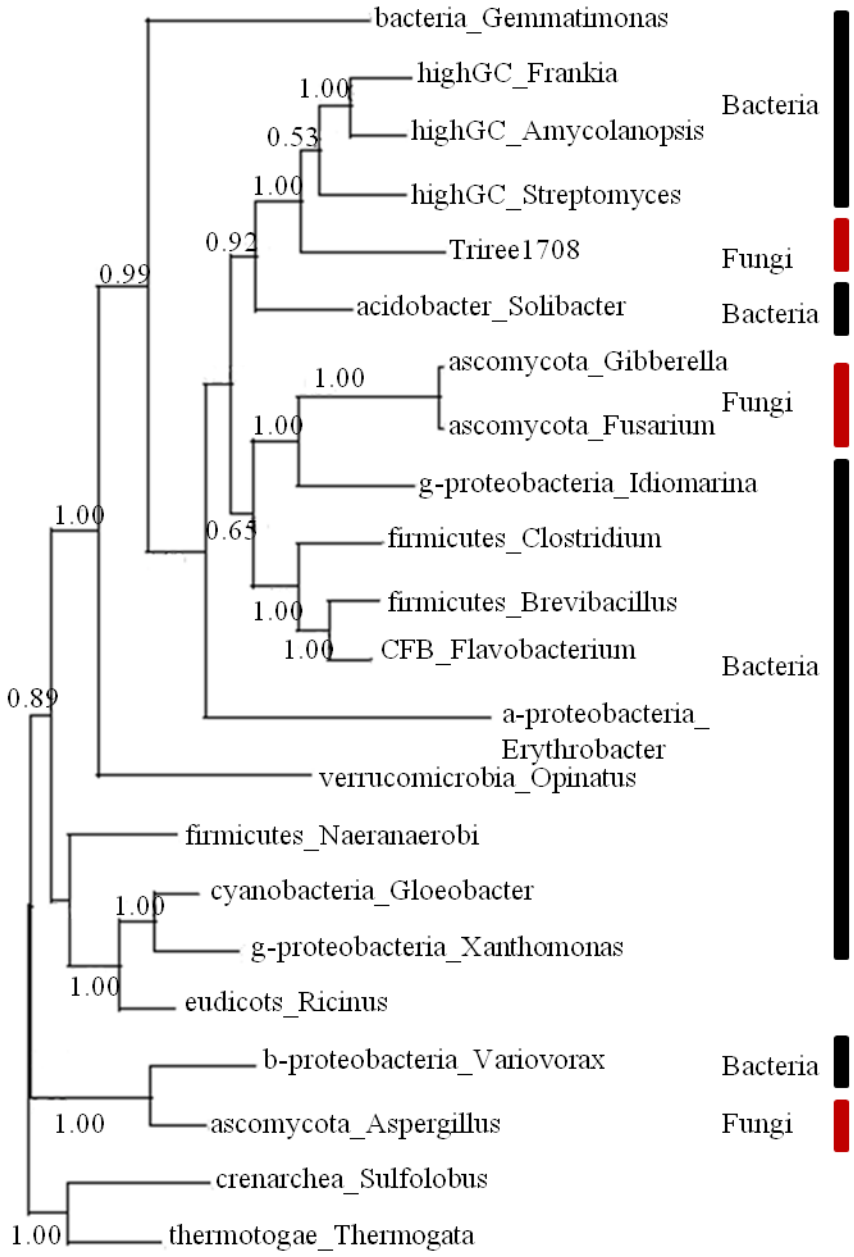


Figure 6. The molecular phylogeny of Triree1708, Jgi|Trire2|104197. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

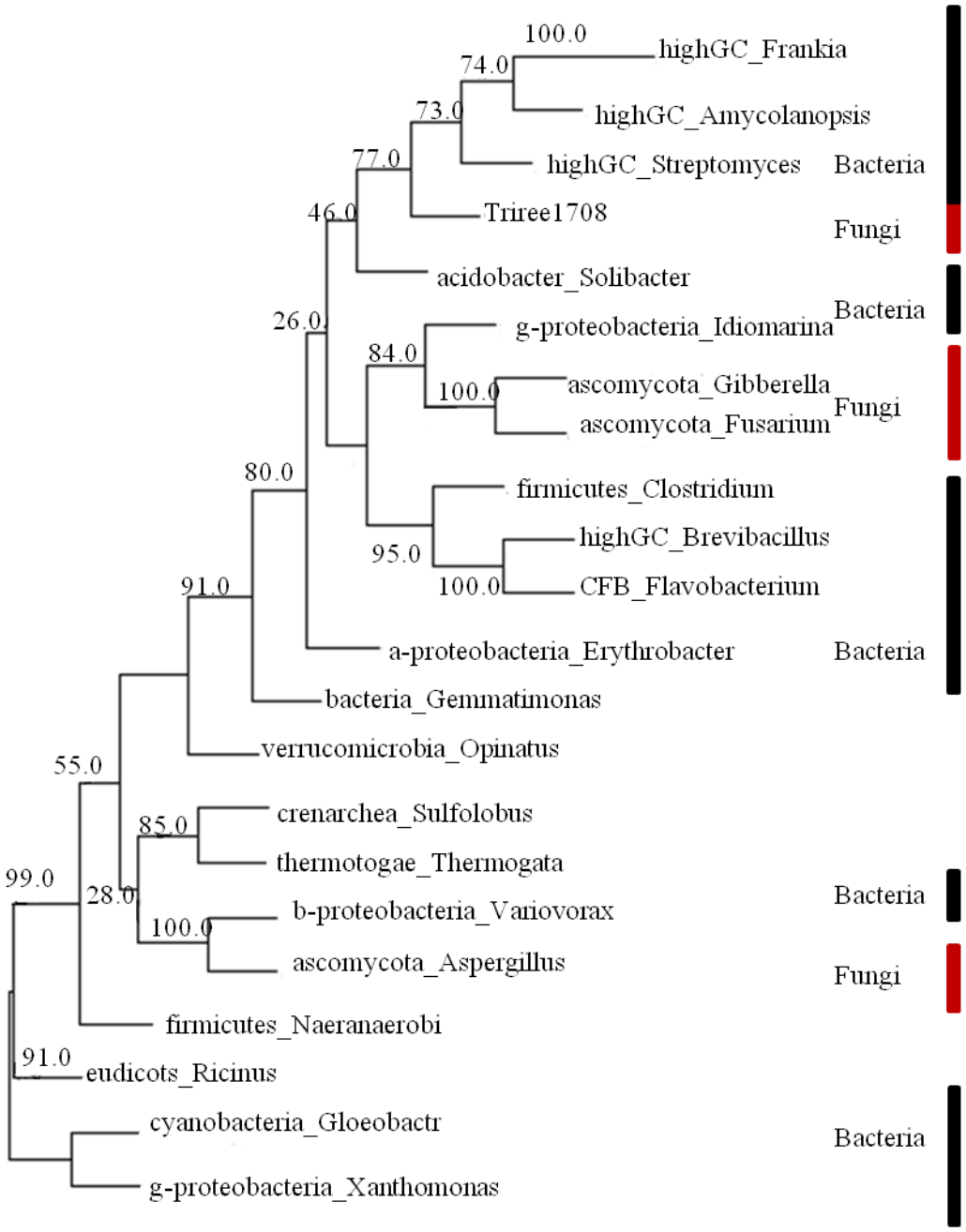


Figure 7. The molecular phylogeny of Triree3136, Jgi|Trire2|121136. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

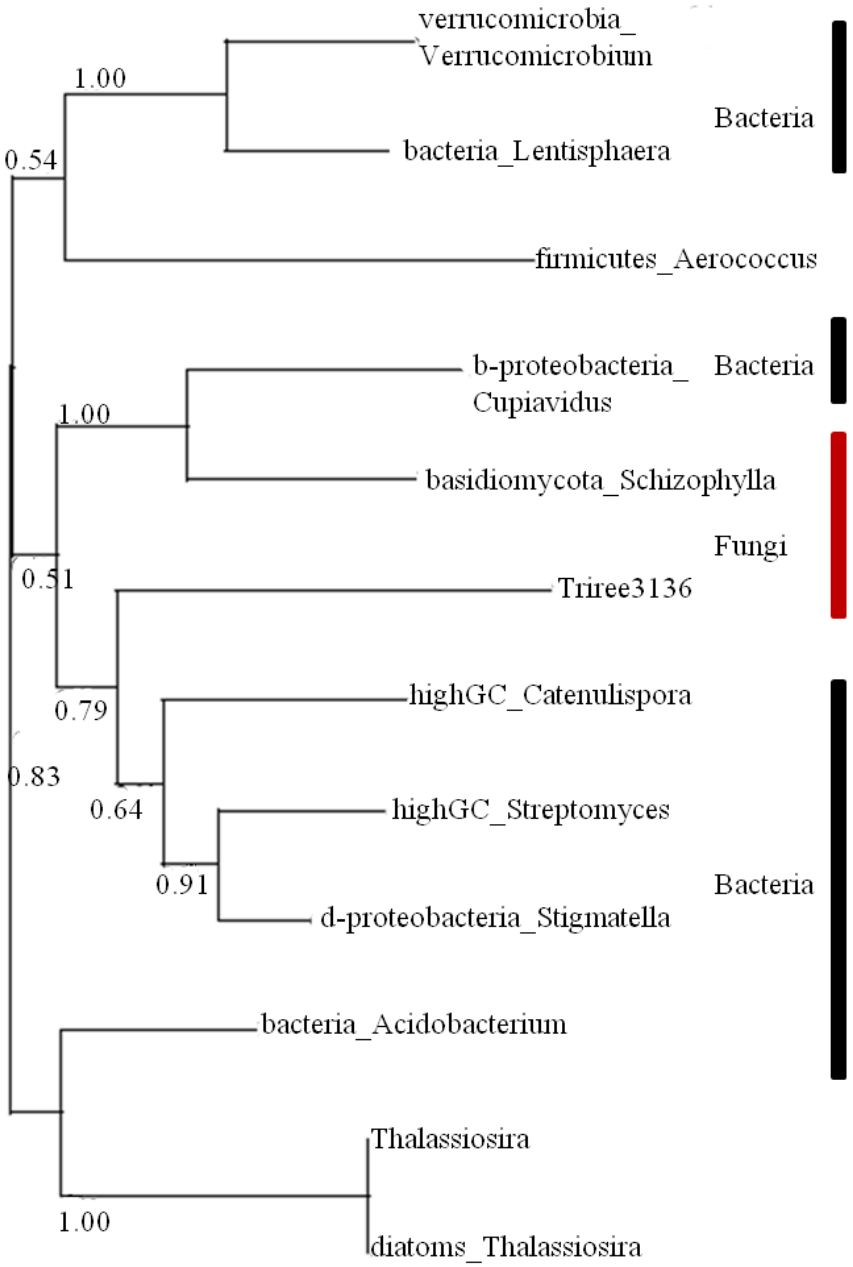


Figure 8. The molecular phylogeny of Triree3136, Jgi|Trire2|121136. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

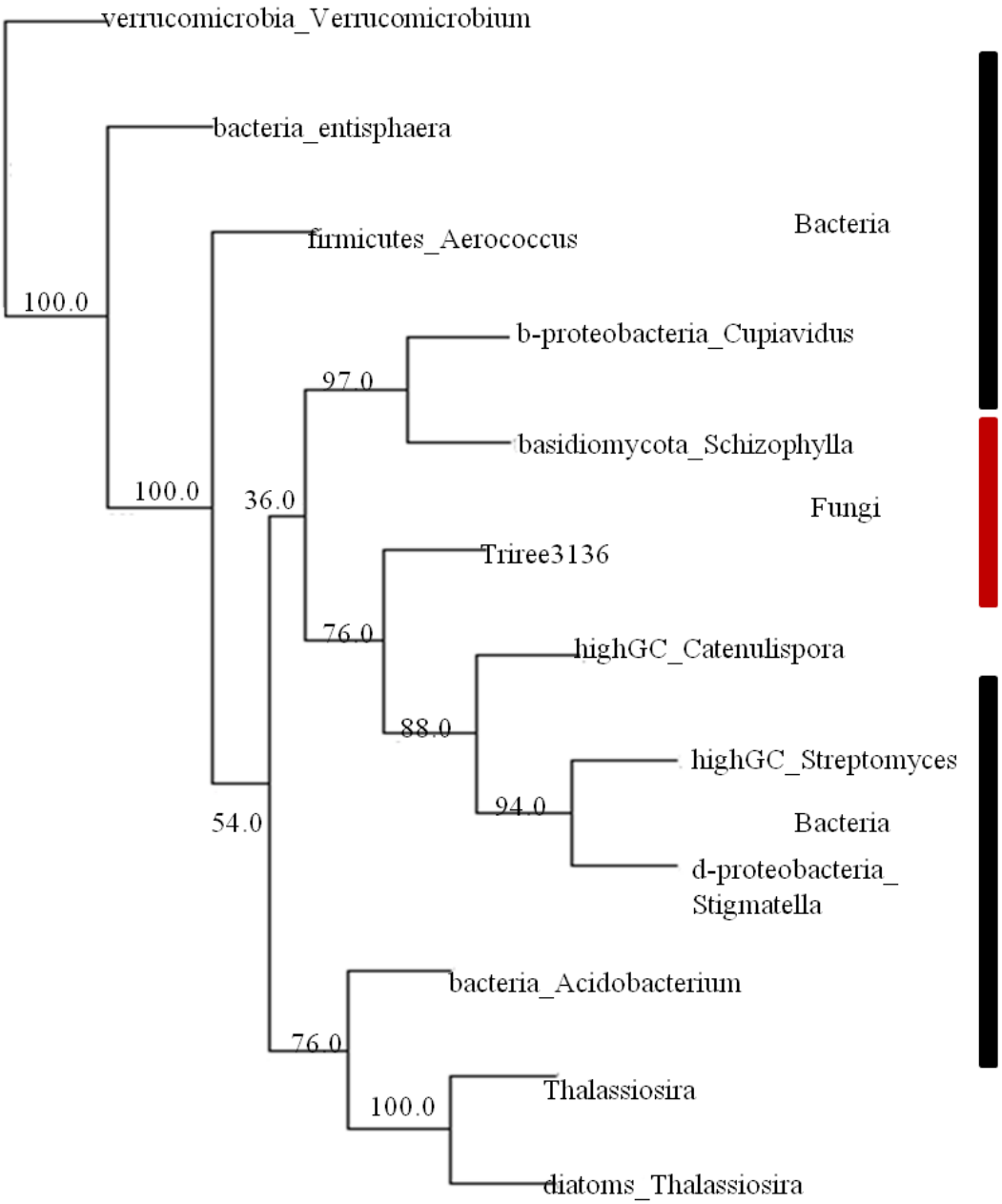


Figure 9. The molecular phylogeny of Triree4966, Jgi|Trire2|108007. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

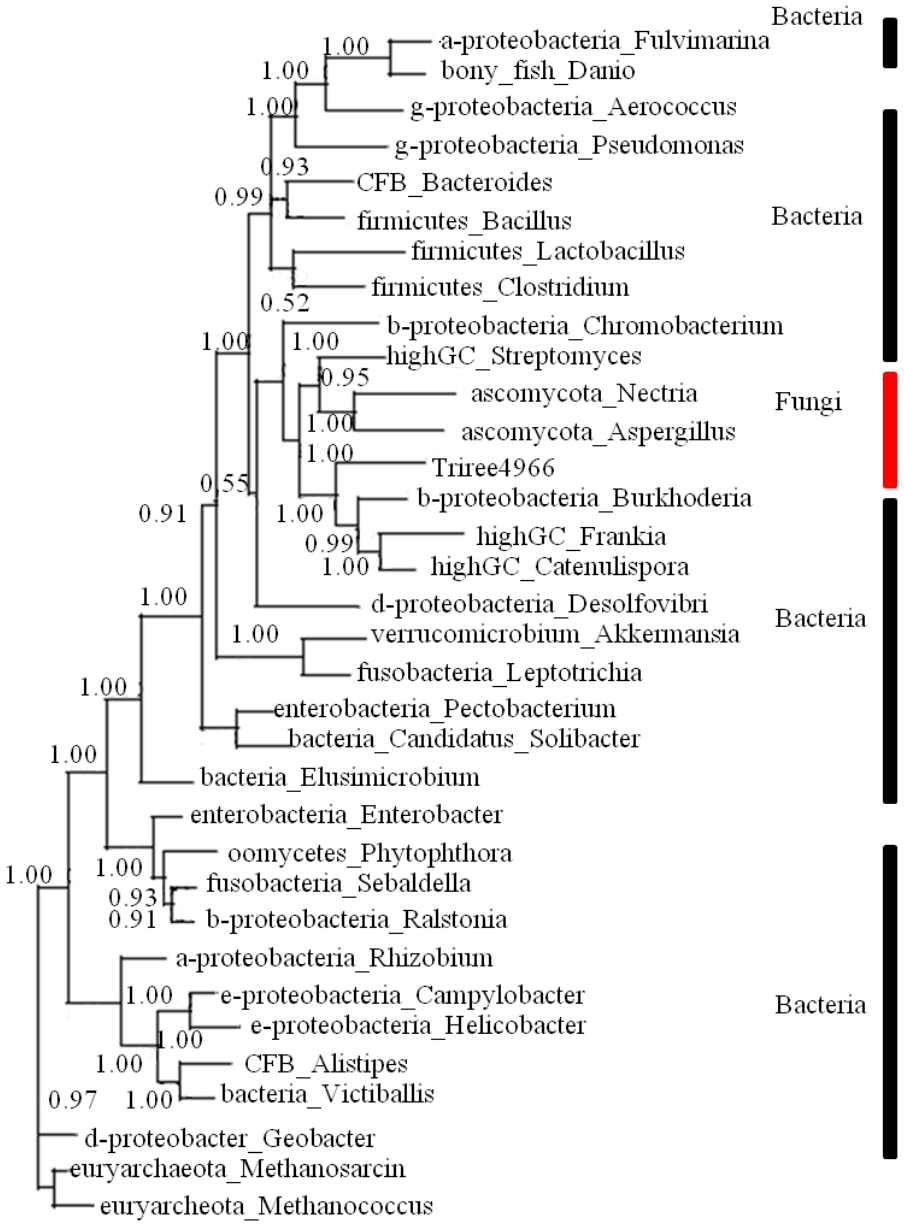


Figure 10. The molecular phylogeny of Triree4966, Jgi|Trire2|108007. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

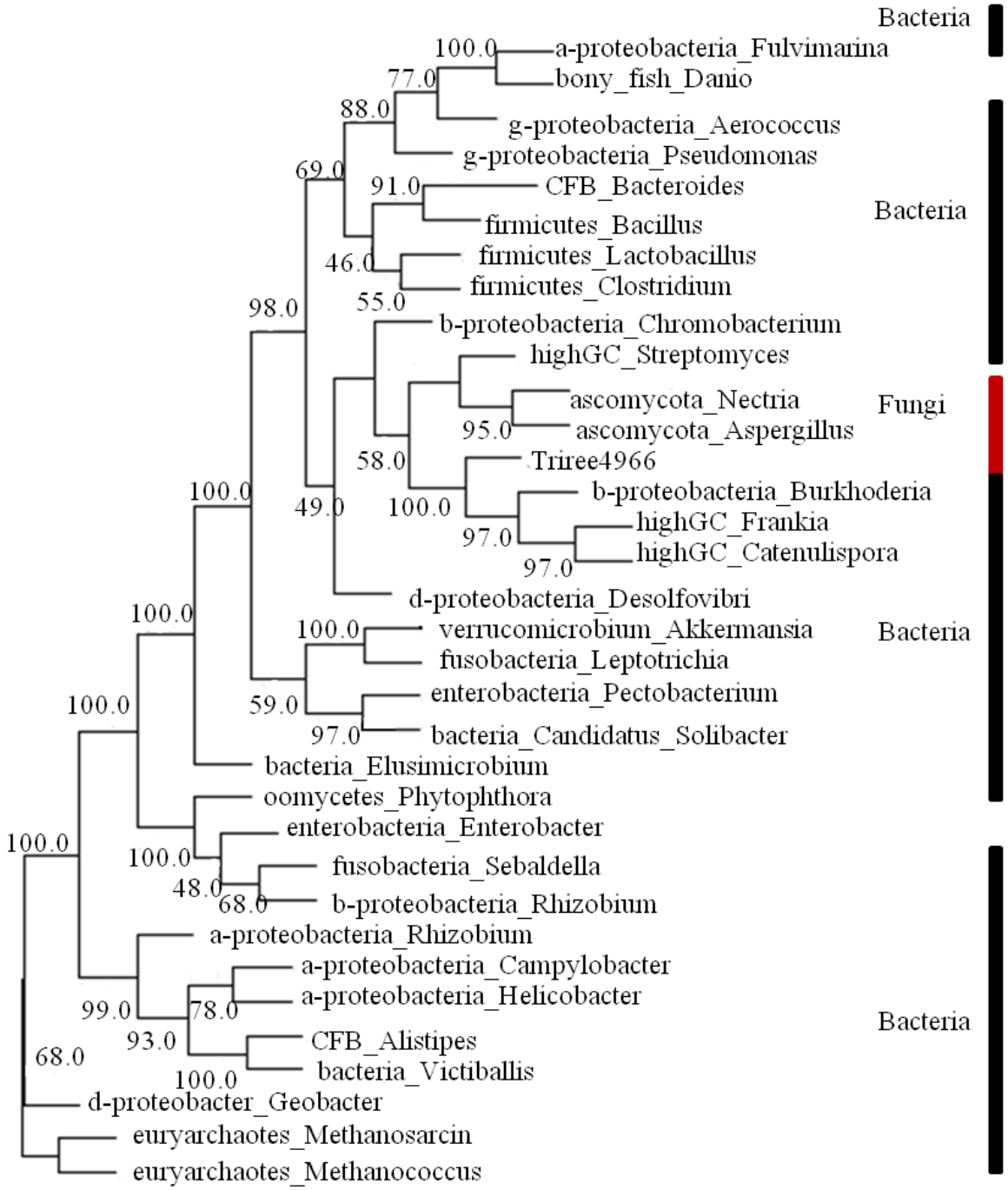


Figure 11. The molecular phylogeny of Triree5571, Jgi|Trire2|108671. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

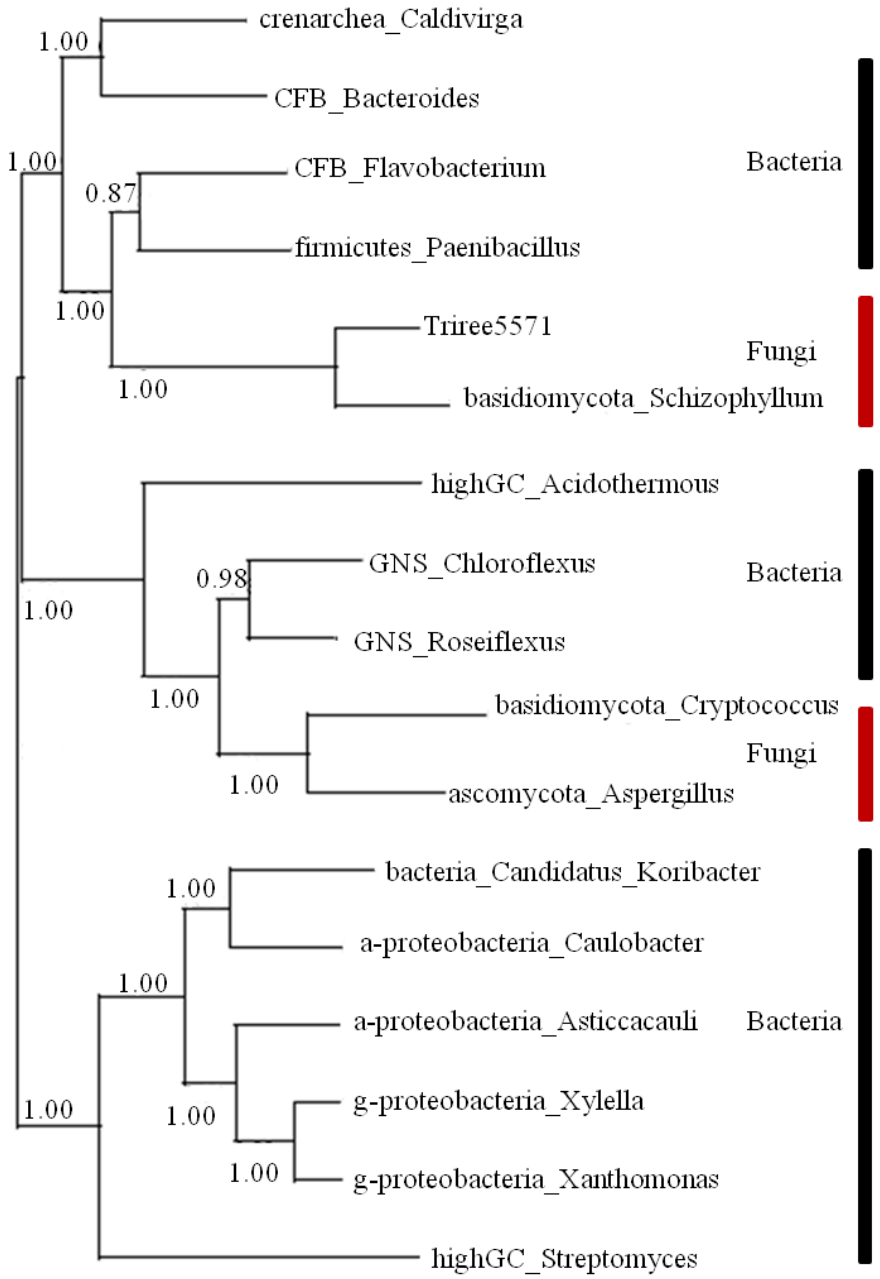


Figure 12. The molecular phylogeny of Triree5571, Jgi|Trire2|108671. This tree was generated by neighbor joining using PHYLIP. The numbers represent bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

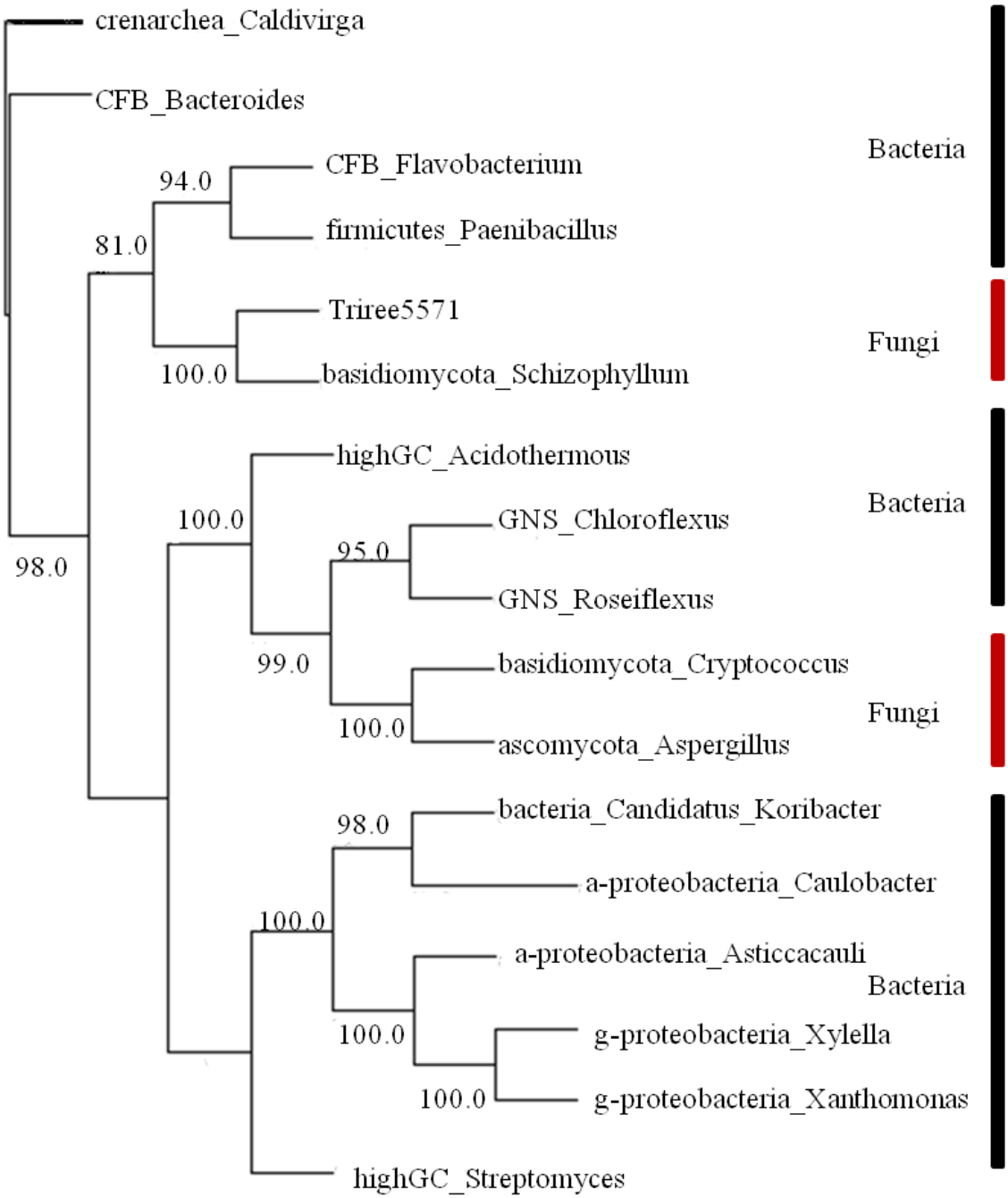


Figure 13. The molecular phylogeny of Triree6978, Jgi|Trire2|110271. This tree was generated by neighbor joining using PHYLIP. The numbers represent bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

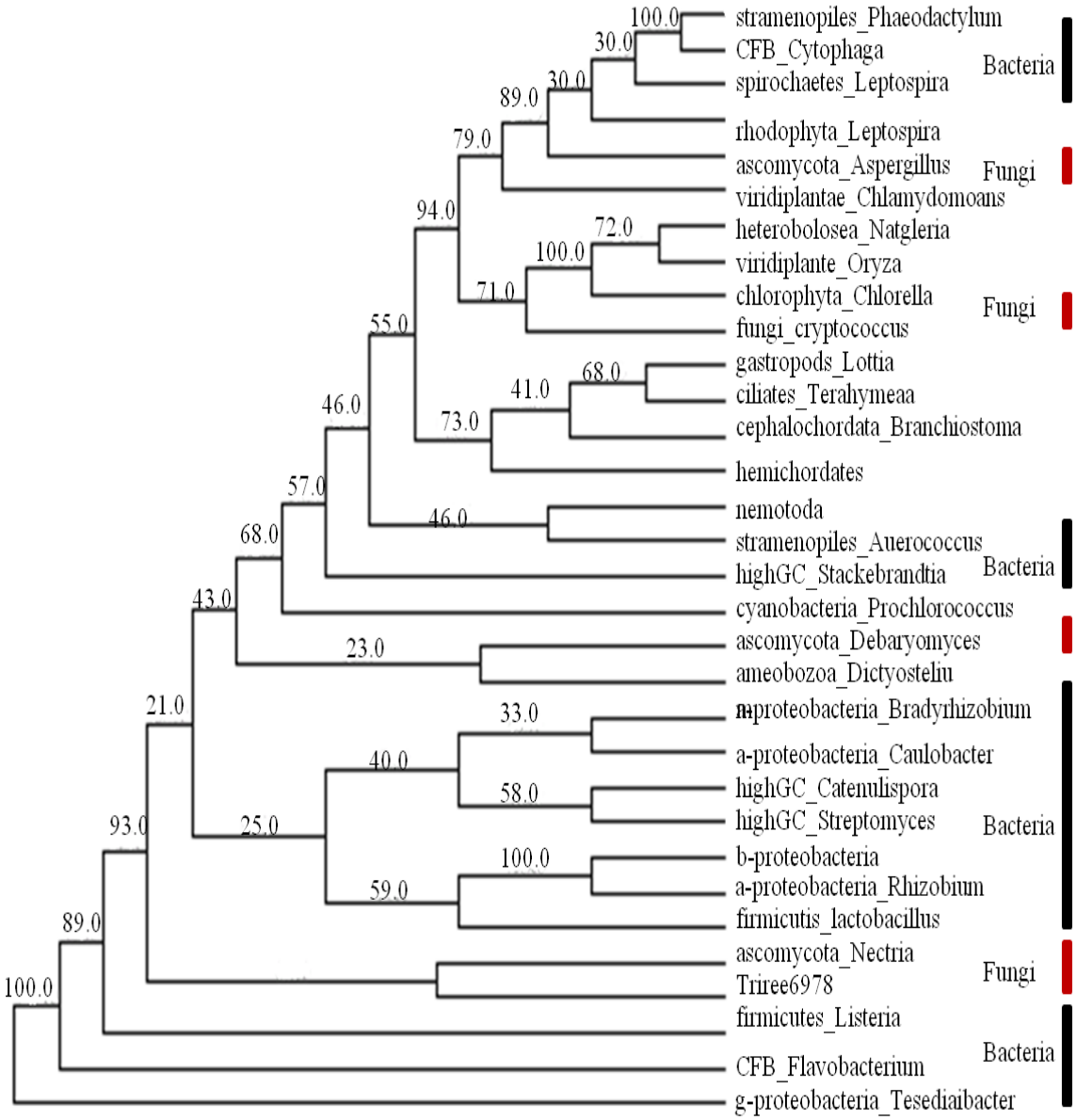


Figure 14. The molecular phylogeny of Triree7021, Jgi|Trire2|66598 .This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

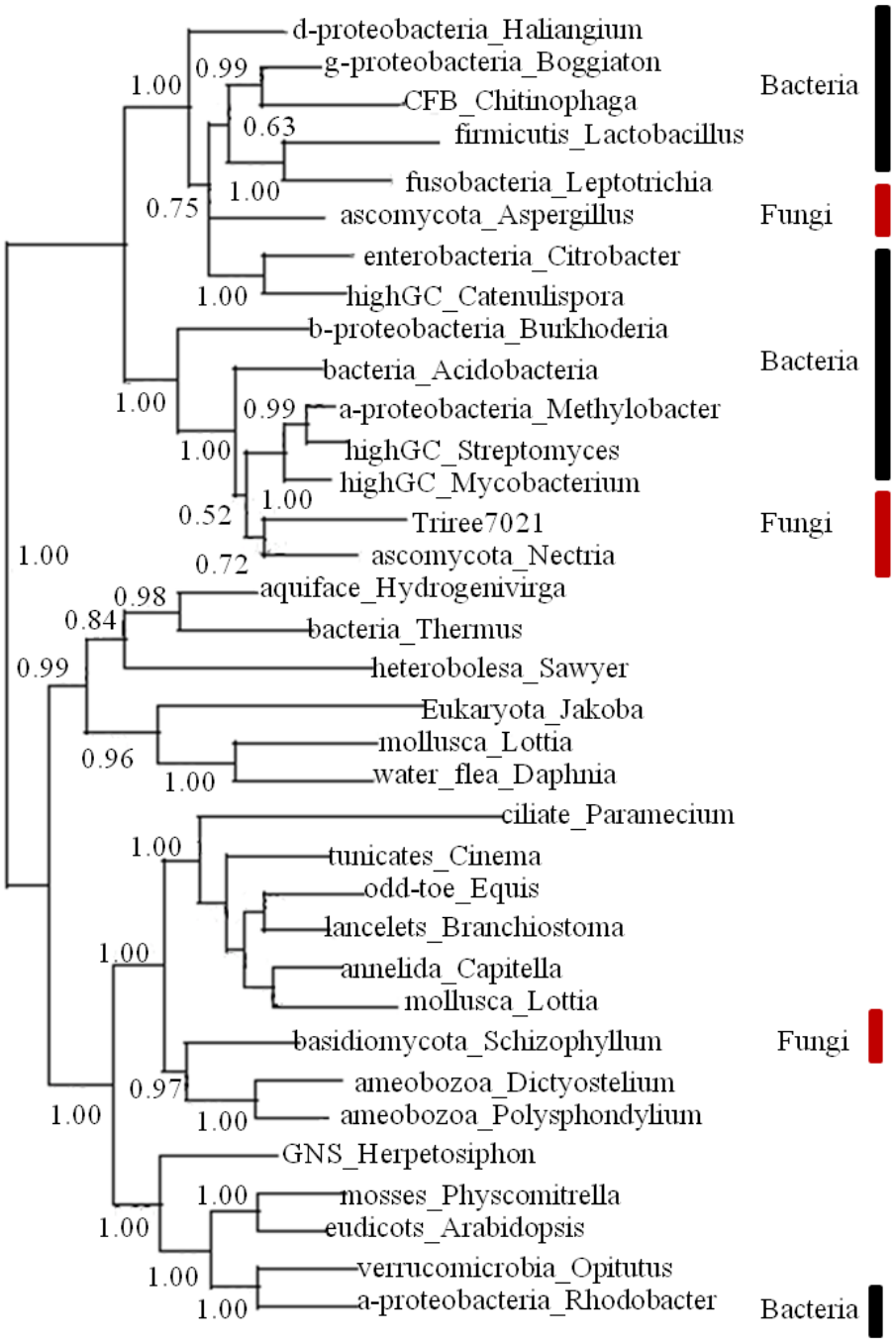


Figure 15. The molecular phylogeny of Triree7021, Jgi|Trire2|66598. This tree was generated by neighbor joining using PHYLIP. The numbers represent bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

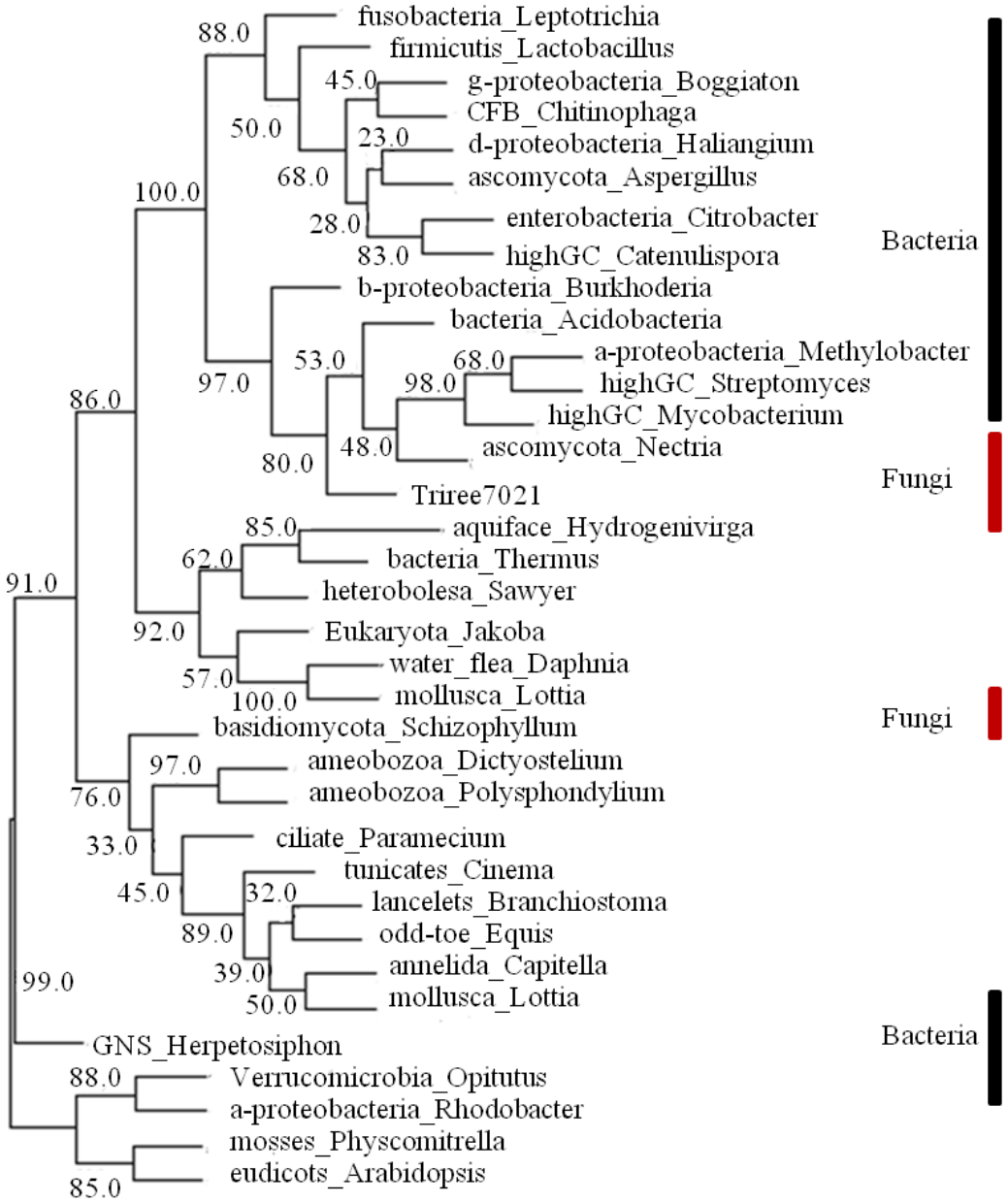


Figure 16. The molecular phylogeny of Triree7365, Jgi|Trire2|110709. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

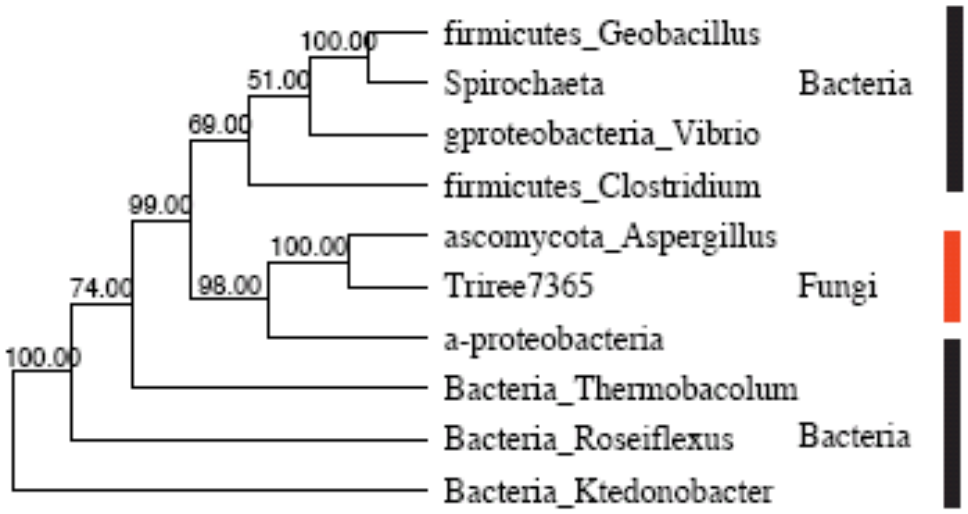


Figure 17. The molecular phylogeny of Triree7540, Jgi|Trire2|68064 . This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

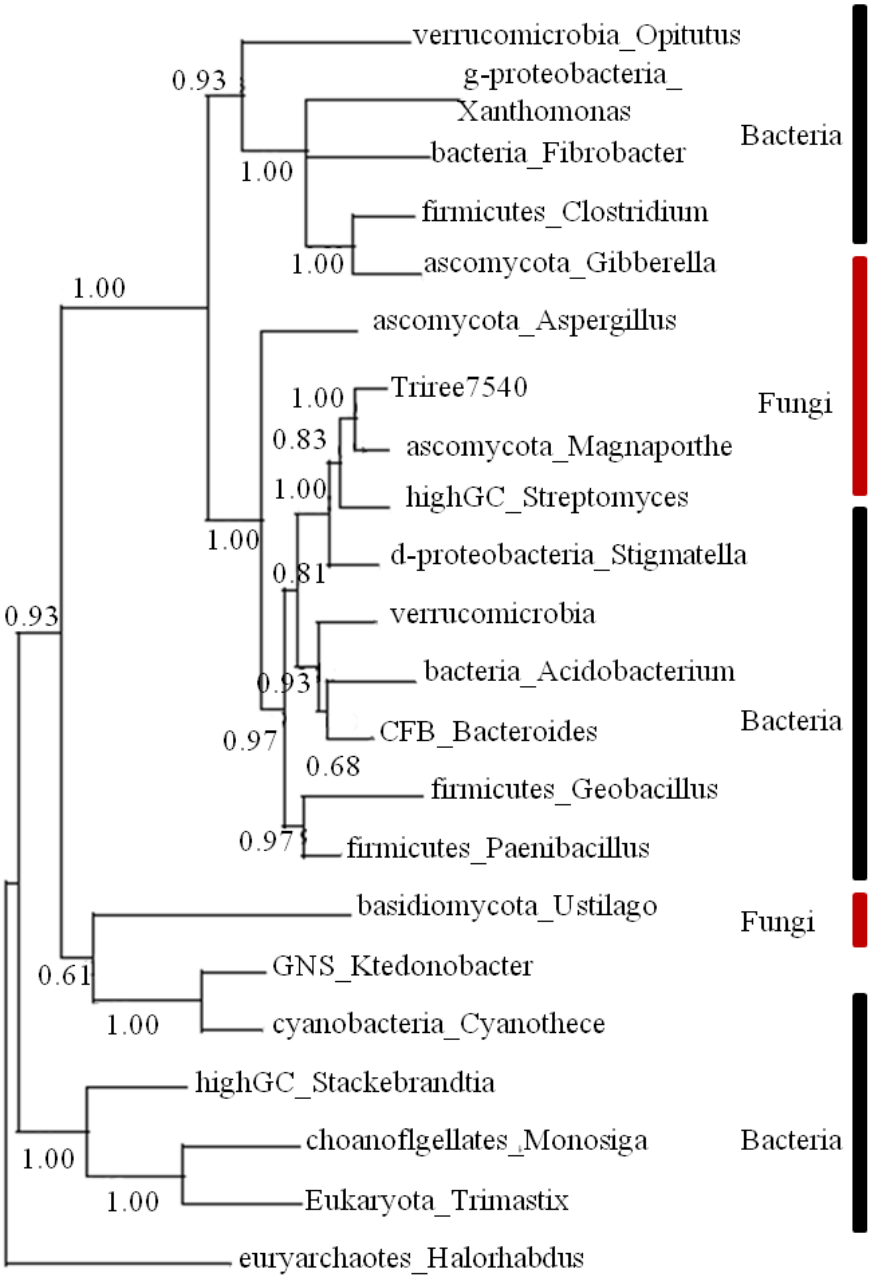


Figure 18. The molecular phylogeny of Triree7540, Jgi|Trire2|68064 . This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

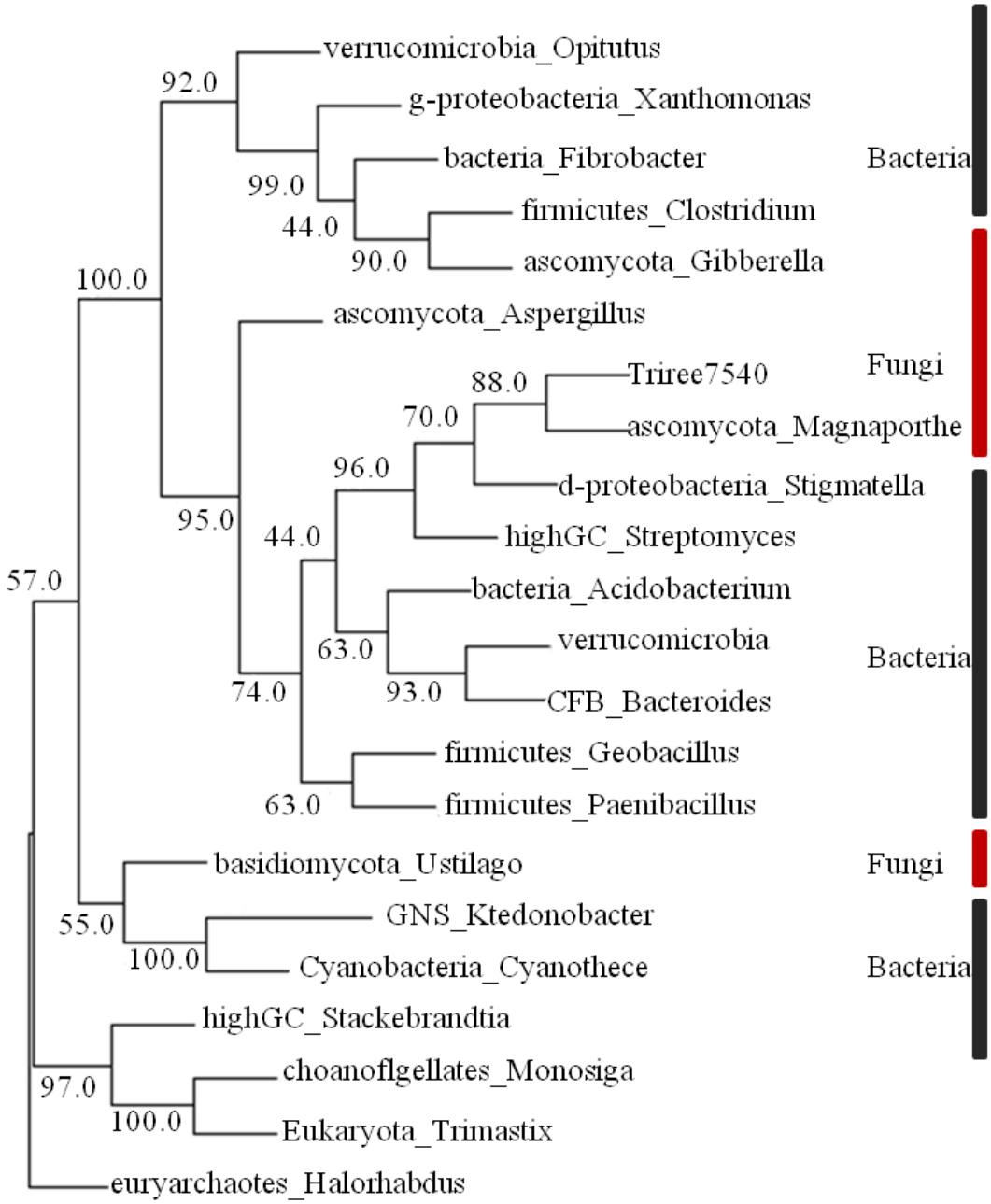


Figure 19. The molecular phylogeny of Triree8067, Jgi|Trire2|72488. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

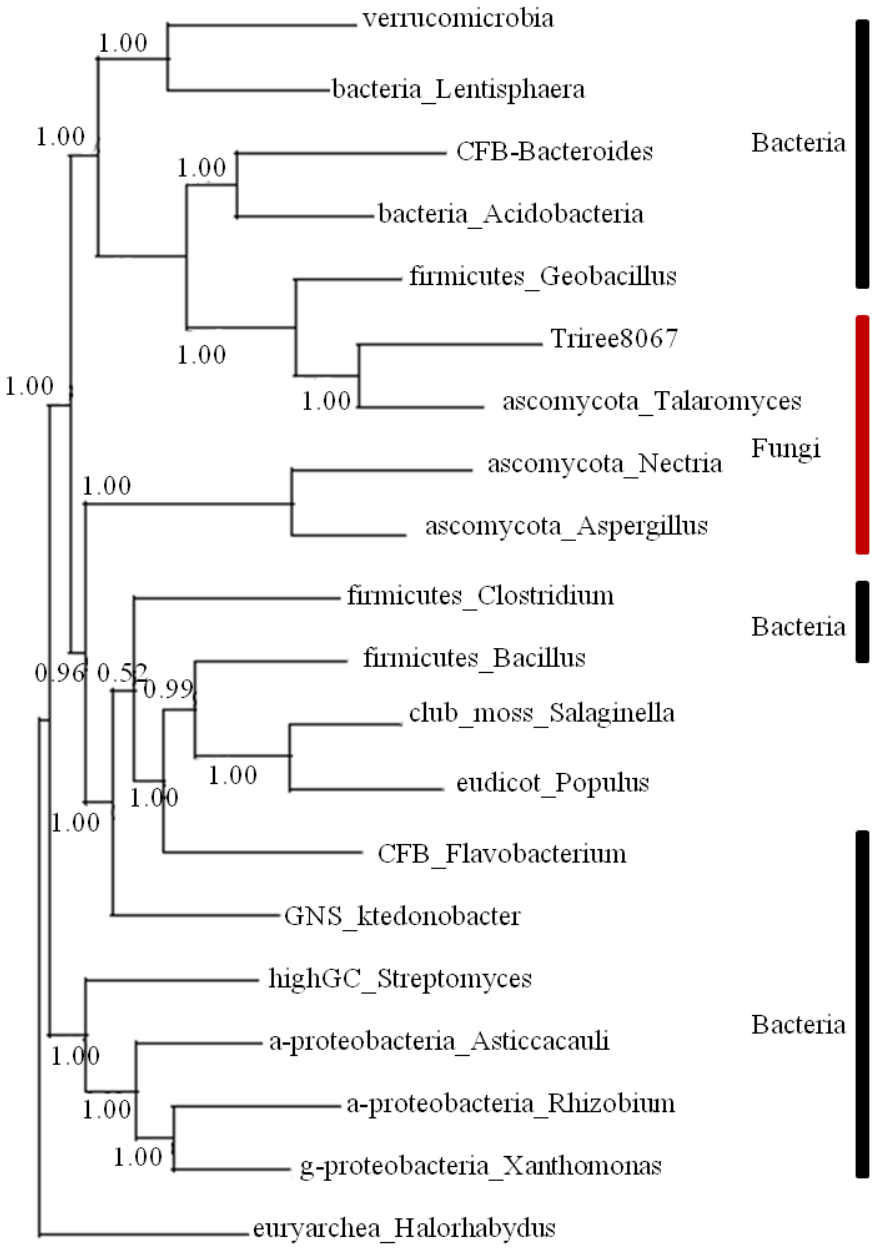


Figure 20. The molecular phylogeny of Triree8067, Jgi|Trire2|72488. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

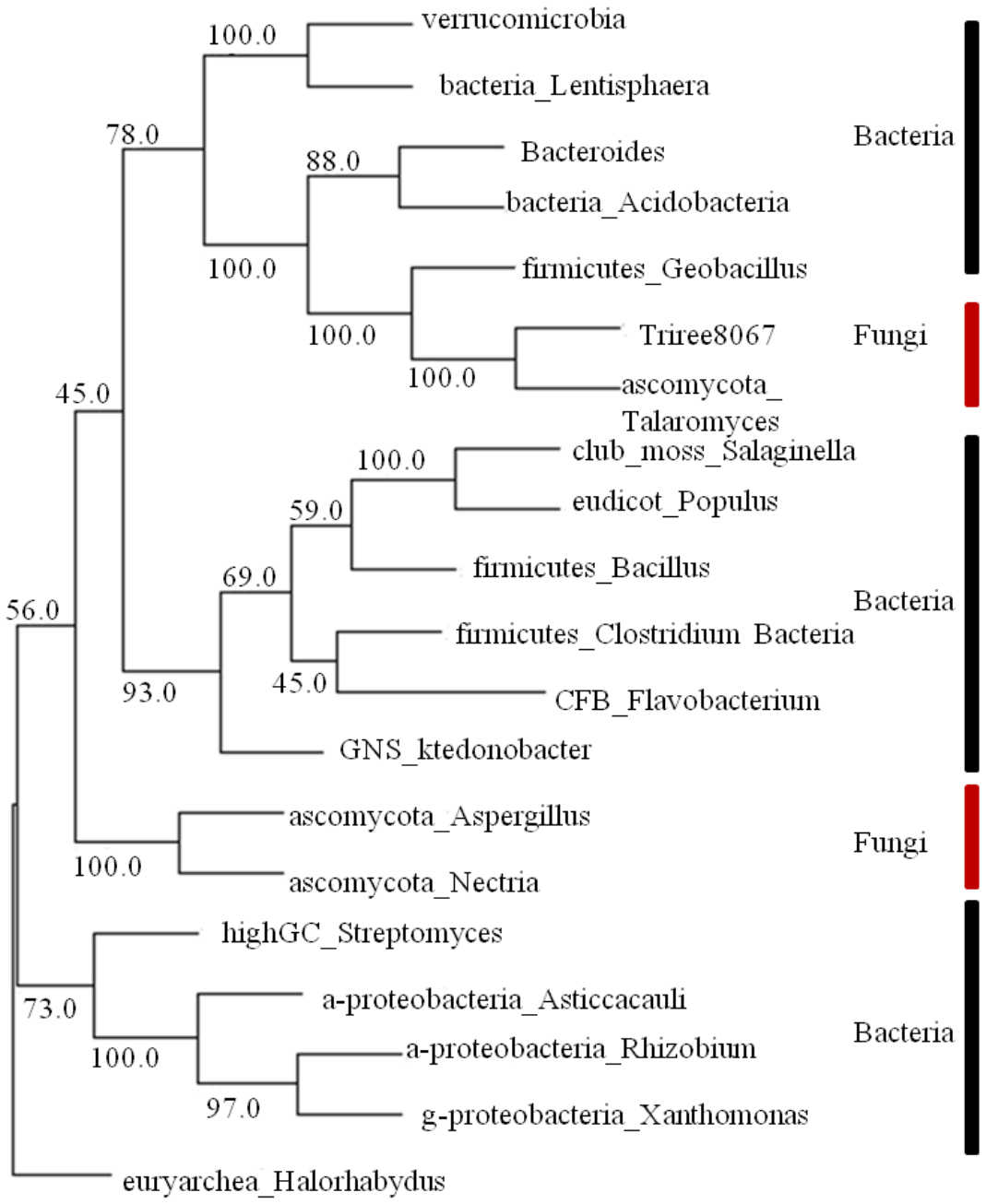


Figure 21. The molecular phylogeny of Triree8375, Jgi|Trire2|111865 . This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities for their respective branches (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

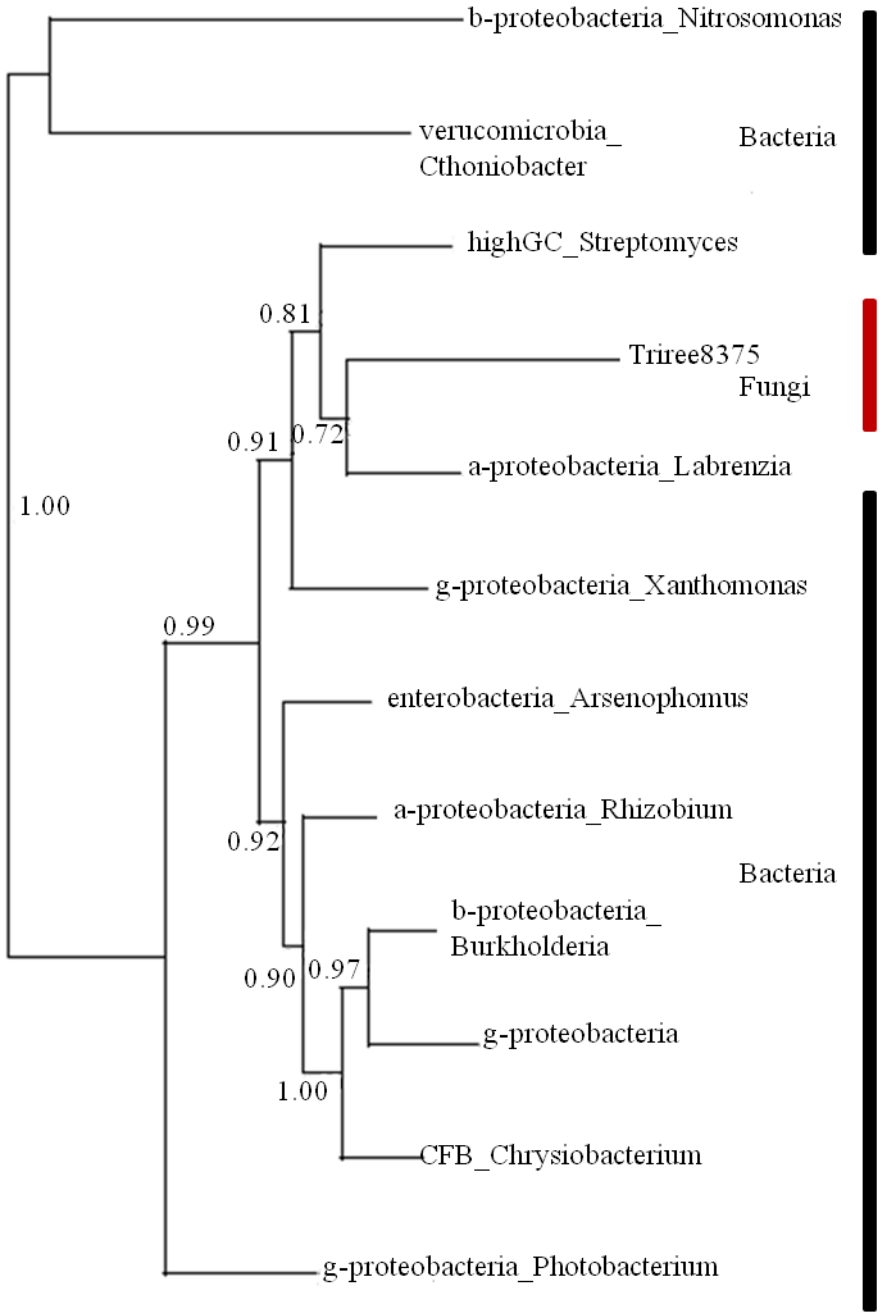


Figure 22. The molecular phylogeny of Triree8375, Jgi|Trire2|111865. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

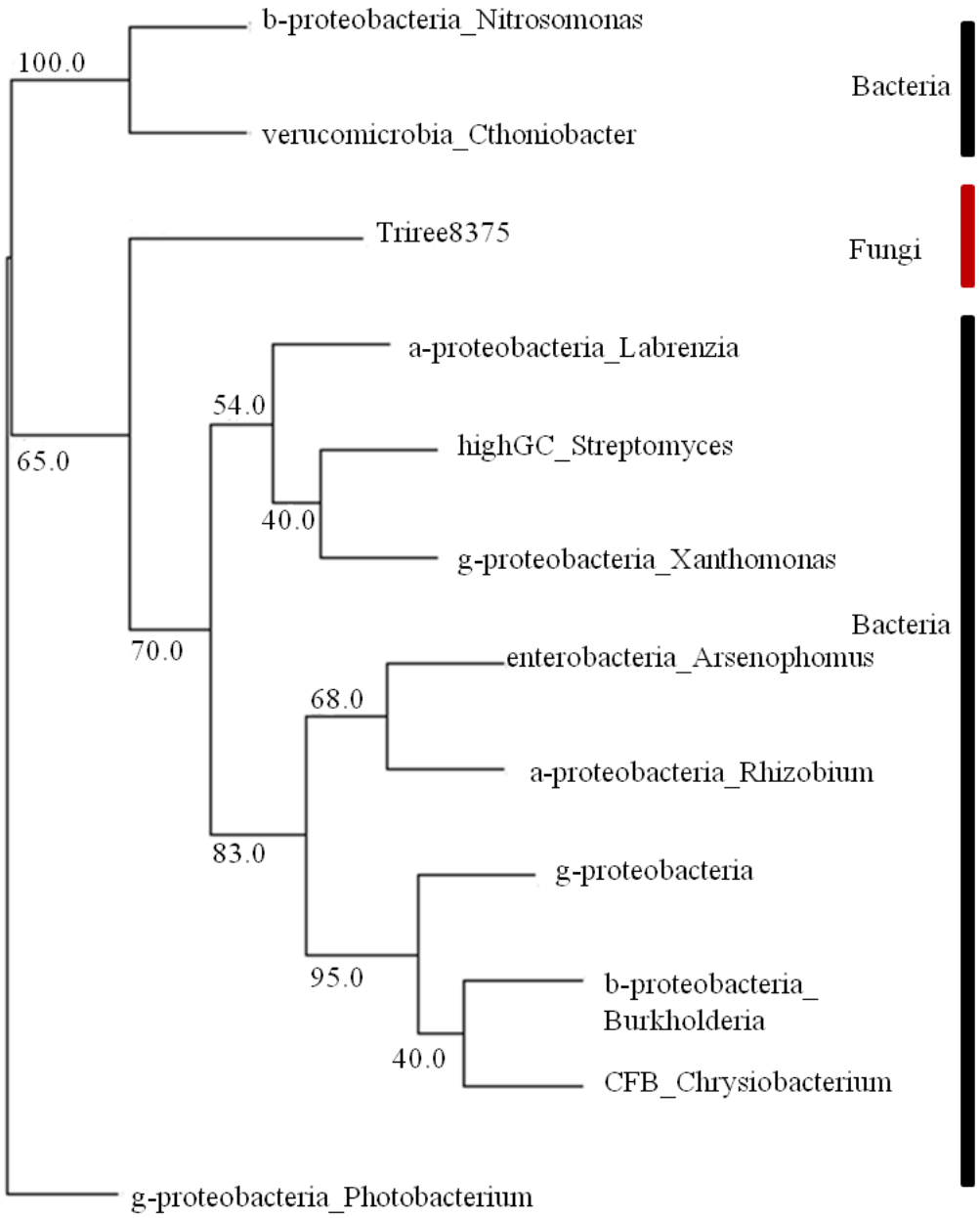


Figure 23. The molecular phylogeny of Triree9071, Jgi|Trire2|71101. This tree was generated by Bayesian analysis using MrBayes. The numbers represent the posterior probabilities (1.00 =100%). The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.

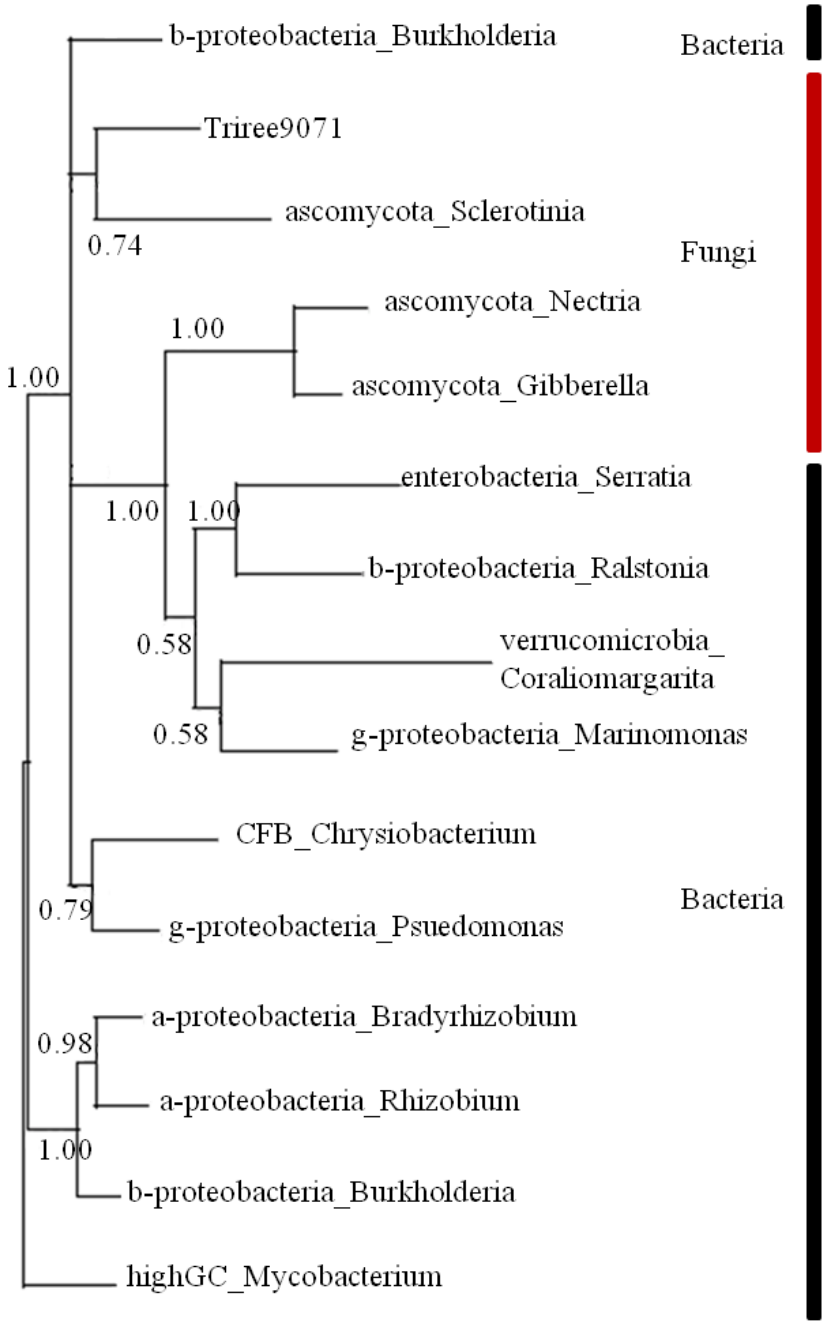


Figure 24. The molecular phylogeny of Triree9071, Jgi|Trire2|71101. This tree was generated by the neighbor joining method using the PHYLIP software package. The numbers represent the bootstrap values for their respective branches. The red bars indicate fungal lineages. The black bars indicate bacterial lineages. Branch length represents relative divergence since the last node.