2012-12-13

A Tree Theory Case Study in *Steinernema*

Camille Eileen Finlinson Porter
Brigham Young University - Provo

A Tree Theory Case Study in *Steinernema*

Camille Eileen Finlinson Porter

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Byron J. Adams, Chair
Keith A. Crandall
Michael F. Whiting

Department of Biology

Brigham Young University

December 2012

ABSTRACT

A Tree Theory Case Study in *Steinernema*

Camille E. F. Porter
Department of Biology, BYU
Master of Science

It is widely assumed that current phylogenetic methods are fairly accurate at recovering the evolutionary relationships among different species, but evaluating the relative success of this enterprise is a difficult task. This study addresses some fundamental questions associated with generating phylogenetic trees. The complete genomes of five species of *Steinernema* were sequenced and assembled. Genes were predicted in AUGUSTUS and orthologous genes were found from those data using OrthoMCL. I aligned 3890 genes in MAFFT and eliminated poorly aligned positions with GBlocks. I created individual trees for each gene as well as a supermatrix tree in PAUP*, using a closely related taxon from another genus, *Panagrellus redivivus*. In the resulting gene trees, I found only a small subset of all the possible topologies. I discovered that the supermatrix tree has the same topology as the topology with the most gene trees in the gene-topology distribution. There are only a small number of histories for all of the genes and many of the genes have the same lineage. I bootstrapped the gene-topology distribution and found that the best-supported topology was sampled 22.1% of the time. I show that many genes must be sampled in order to converge on the topology with the most support from the gene trees in this dataset.

Keywords: *Steinernema*, phylogeny, supermatrix, alignment editing, gene number

ACKNOWLEDGMENTS

I would like to thank the Sternberg lab for the sequencing, annotation, and gene prediction of the *Steinernema* species and for allowing me the first crack at their data. I would also like to thank Adler Dillman for his OrthoMCL analysis. Thanks goes to Kathleen Finlinson and Nathan Porter for help with the gene number statistical analysis, and Rachel Finlinson for help creating the figures. My gratitude goes to Dr. Byron Adams for continually supporting me and for helping me discover phylogenetics, an area for which I now have great passion. Appreciation goes to my committee members for their guidance and knowledge.

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES vi

INTRODUCTION 1

Analyzing Large Datasets 1

Taxon Sampling 2

Tree Searching 2

Steinernema 3

METHODS 4

Steinernema 4

Alignment 5

Phylogenetic Analysis 6

Number of Genes 6

RESULTS 7

Tree Topology Tests 9

Supermatrix 9

Number of Genes 9

DISCUSSION 10

Supermatrix 10

Number of Genes 11

Alignment Editing 11

CITATIONS 12

LIST OF FIGURES

Figure 1. Gene-topology distribution.. 15

Figure 2. The top 45 resolved topologies... 16

Figure 3. Supermatrix parsimony tree. 17

Figure 4. Probability of containing best supported topology.. 18

LIST OF TABLES

Table 1. A chart of all topologies... 19

INTRODUCTION

Phylogenetic trees are important because they allow us to visualize evolutionary history. It is widely assumed that current phylogenetic methods are fairly accurate at recovering the evolutionary relationships among different species, but evaluating the relative success of this enterprise is a difficult task. The study of phylogenetics involves making many assumptions about the methods we use, many of which are not fully tested. This study addresses some fundamental questions associated with generating phylogenetic trees, such as, how does a supermatrix compare to individual gene genealogies in terms of summarizing data? How does the number of genes sampled affect the probability of finding the optimal solution? How many genes must be sampled before the data converge on the best estimate? What is the effect of alignment editing on phylogenomic analyses? Given that each gene can have a different evolutionary history, what is the frequency of discordance (how often are gene trees discordant from the optimal solution)?

Analyzing Large Datasets

There are two methods that are commonly used to analyze large amounts of phylogenetic data: the supertree and supermatrix approaches. A supertree involves creating a tree for each gene and then combining the information from the trees into a single tree (Bininda-Emonds, 2004). One downside is that the final analysis does not use the character evidence directly; instead it uses the information from the individual tree topologies. This can lead to a loss in character information and evidence (De Queiroz et al., 1995). A supermatrix combines the alignment from each gene together and analyzes the concatenated alignment all at once. The phylogenetic signal of the supermatrix can be different than analyses of the individual genes. This is presumably because combining genes can reveal character support for relationships that are not found in the individual analyses (Gatesy et al., 1999; Lambkin, 2004).

Taxon Sampling

For the past 20 years there has been controversy over whether including more genes or more taxa will result in a more accurate phylogeny. Early studies showed that in cases of taxa with very different rates of evolution, it is beneficial to add more taxa to break up long branches (Hillis, 1996; Graybeal, 1998). Graybeal used highly divergent simulated data and found that adding taxa and adding characters increased accuracy, but adding taxa increased accuracy faster than adding characters. Poe and Swofford discovered that adding taxa can result in more or less accurate phylogenies, in different cases, and that adding characters can be the more favorable strategy (Poe and Swofford, 1999). There has been much controversy about the issue since the 1990s (Nabhan and Sarkar, 2010).

It is easier now than it was earlier to sample both more taxa and more characters. Baurain et al. state that “Our opinion is that it is no longer worthwhile to argue the relative benefits of gene versus taxon sampling but that progress in sequencing technology will lead to data sets rich in both genes and taxa” (Baurain et al., 2007).

Tree Searching

Phylogenetic searches of tree space optimize some criteria to measure how well a particular topology describes the data. Optimality methods work by finding the highest scoring tree for a specific sequence alignment, which is construed to be the best estimate of evolutionary relationships (Money and Whelan, 2012). As the number of species increases, it becomes harder to search all the trees and identify the optimal solution(s). In an analysis with 10 taxa, there are 282,137,824 possible rooted phylogenies (Felsenstein, 1978). For phylogenetic searches that contain more than 25 taxa, with contemporary computing power it is impossible to compare all the possible trees in a reasonable amount of time because there are far too many possible solutions. Tree searching is an NP-hard (non-deterministic polynomial-time hard) problem, which means for large amounts of taxa heuristic searches must be done, and only a subset of the possible trees can be searched. An exhaustive search can examine all possible trees, but can only be used for small numbers of taxa. An exhaustive search guarantees that the best trees for the given data will be found.

Steinernema

Steinernema is a genus of entomopathogenic nematodes frequently used for biological control of insect pests. The life cycle of *Steinernema* begins with the infective juvenile stage: a soil dwelling, non-feeding period (Goodrich-Blair and Clarke, 2007). *Steinernema* locates and then enters its host insect through natural openings. It then moves to the haemolymph and releases its mutualistic bacteria, *Xenorhabdus*, which produces toxins lethal to the insect. *Xenorhabdus* turns the insect into a nutrient soup that feeds both bacteria and nematode. After one to three generations of nematode reproduction, a new generation of infective juveniles is colonized by bacteria and leaves the insect host. *Steinernema* and *Xenorhabdus* have an obligate mutualistic relationship where each is required in order to hunt and kill insects (Poinar, 1993).

Steinernema are useful in this study because they have relatively small genomes (Grenier et al., 1997) making it possible to study all of the orthologous predicted genes in the genomes. The complete genomes of five species of *Steinernema*: *Steinernema. carpocapsae, S. scapterisci, S. monticolum, S. feltiae*, and *S. glaseri* are analyzed in this study. Based on its phylogenetic position relative to *Steinernema* (Adams et al., 2007), *Panagrellus redivivus* is used to polarize the homology statements and root the trees. With six species there are $(2*n-3)!! = (2n-3)!/((2^{n-2})*(n-2))! = (2*6-3)!/(2(^{6-2)}*(6-2))! =$ 954 possible tree topologies for rooted bifurcating trees (Felsenstein, 1978). It is useful to have a small number of species in this study because it makes it possible to do exhaustive searches of tree space.

METHODS

Steinernema

S. carpocapsae, S. scapterisci, S. monticolum, S. feltiae, and *S. glaseri* were sequenced and assembled in the Sternberg lab (Yook et al., 2012). After sequencing and assembly, AUGUSTUS (Stanke et al., 2006) was used for gene prediction. Then OrthoMCLv1.4 (Li et al., 2003) was used to predict orthologs with the default settings. There were 3890 genes in the OrthoMCL output that included only one sequence for each nematode; these were used in subsequent steps. Fasta files for each gene were prepared using a Java program that makes use of the OrthoMCL and AUGUSTUS results. The OrthoMCL results contained gene numbers and each species' gene name for that gene. The program read the OrthoMCL gene number and then searched the AUGUSTUS results for each species to find the correct gene.

Alignment

Each gene was aligned separately in MAFFTv6.821b (Katoh et al., 2002). The L-INS-i algorithm was run because it is the most accurate setting in MAFFT for data sets containing fewer than 200 species (Katoh et al., 2005).

Alignment accuracy greatly influences the resulting phylogeny (Ogden and Rosenberg, 2005; Simmons et al., 2011). In an earlier study on *Steinernema* phylogeny, it was shown that there can be greater topological variation due to different alignment construction parameters than due to the methods used to generate the phylogenies (Nguyen et al., 2001). Because there are too many genes to be able to go through the alignments individually and check them for accuracy, I performed an alignment quality control check (Talavera and Castresana, 2007) using GBlocks v0.91 (Castresana, 2000) to ensure that the alignments were objectively optimal. Strict settings were used—4 out of the 6 species' amino acids were required to make a conserved position for a column, 5 out of the 6 species' amino acids were required to create a flank position, 10 conserved amino acids were required to make a block, 8 consecutive non-conserved amino acids was the maximum allowed, and all gaps were removed. I used the batch feature of GBlocks.

In the optimized alignment, five genes were completely removed from the analysis by GBlocks because they were unable to be unambiguously aligned. Another supermatrix tree was made without using GBlocks to remove ambiguously aligned regions of the dataset. Before GBlocks was executed, all of the genes combined into a supermatrix contained 2,678,084 amino acids. 1,141,841 amino acids remained after

GBlocks was run. 43% of the amino acids were removed from the analysis. GBlocks concatenated the individual gene files into a supermatrix.

Phylogenetic Analysis

I constructed phylogenetic trees in PAUP* v4.0b10 (Swofford, 2002) under parsimony optimality criterion. Accordingly, I converted all 3,885 FASTA gene files and the supermatrix FASTA file to NEXUS using a modified python script (Sukumaran, 2008). A Perl script was used to append a PAUP* block to the end of each NEXUS file. The tree search parameters for each individual gene dataset, as well as the supermatrix, consisted of an exhaustive parsimony search enforcing a monophyletic root. The result was a separate tree file for each gene and another for the supermatrix. I inferred nodal support by bootstrap analysis (Felsenstein, 1985) of the supermatrix in PAUP* with 500 repetitions using a heuristic search with randomized additions.

To create the topological frequency distribution, I searched all of the tree files created by PAUP* for the specific line that included the tree. In many cases, PAUP* found more than one best tree for a gene. In these cases, all best trees were found. I wrote a Java program to search through all the genes for their trees and also kept track of the number of optimal trees for each gene. The Java program also counted the number of genes that had equivalent tree topologies.

Number of Genes

In order to determine the number of genes that must be sampled to get the tree supported by the majority of the genomic data, I bootstrapped the gene-topology distribution. The rounded Fraction Tree distribution was used as the sample distribution. The distribution was randomly sampled 5000 times with replacement. The assumption of random sampling will be violated at some point when enough genes have been sampled, but holds up well for smaller numbers of genes.

I tested to see what the probability was that one of the topologies for the genes was the correct one. Random sampling of genes from the gene distribution revealed that sampling a gene that supported the optimal solution occurred only 22% of the time. In order to understand the probability that the solution contains the topology with the most support (the topology that is most common in the distribution) among the set of solutions, I used the following algorithm:

Draw two genes from the distribution. Let A be the probability that the first gene is from the correct topology, and let B be the probability that the second gene is from the correct topology. The probability of A or B is p(A) + p(B) - p(A and B) if A and B are independent. If p(A) = p(B) = 0.221, then p(A or B) = 2(0.221) - 0.221^2 = 0.393. The probability of A or B or C is p(A) + p(B) + p(C) – p(A and B) – p(A and C) – p(B and C) + p(A)p(B)p(C). Continue the analysis to calculate for higher numbers of genes.

RESULTS

I made a topological frequency distribution depicting how many genes support each tree topology (Fig. 1). When a gene supported multiple best topologies I did not make consensus trees; instead, I created two metrics. The Sum Fraction (red) represents a fractionalized number of the genes that support a particular topology. If a gene had more than one best tree, each tree was counted as a fraction: 1/(the number of optimal trees for that gene). When the Sum Fraction genes are summed their total is 3885, the same as the number of genes included in the analysis. Polytomous Total (blue) on the y-axis represents the total number of genes that support a particular topology. Each topology for each gene was counted as one and then they were all added together. There were 7423 total topologies counted in the Polytomous Total, more than the number of genes included in the analysis. Some genes had more than one best topology and are counted as more than one. The Fraction Genes are a better estimate of gene number and are the relevant numbers used in most of the paper when discussing results. The Polytomous Total artificially inflates the weights of the genes that had the least amount of phylogenetic signal, and is therefore a less useful metric.

Table 1 has the same information as Figure 1 with a summary of all the topologies, the Fraction Genes and the Polytomous Total. I recovered 193 total topologies when I analyzed all possible genes. 97 of those topologies are at least partially unresolved, leaving only 96 fully resolved topologies. There are 11 topologies with at least 100 gene trees supporting them. 3245 of the 3885 genes supported the top 11 topologies. It is comforting to know that of the total number of topologies possible, only a small number of the topologies are supported by many genes. It means that only a relatively small number of histories exist for most genes, and most of the genes have similar histories.

Only 96 of the possible 954 topologies are found in the gene trees. 38 of those have less than one Fraction Gene supporting them, meaning they were only found when genes with low phylogenetic signal were analyzed, and would not be present in a meaningful consensus tree. It is evident that only a small number of the topologies are supported by a large number of genes (Fig. 2). Most of the genes depict *S. carpocapsae* and *S. scapterisci* as sister taxa. There are 15 possible topologies that pair the two. Of the top 18 best-supported genes, 15 of them pair *S. carpocapsae* with *S. scapterisci*. When you look at the Fraction Tree and Polotomous Total, there are no genes that are supported by at least one Fraction Tree that are not supported by at least two Polytomous Total trees. A topology would need to be supported by at least one Fraction Tree in order to be able to be found as the only solution for a gene. This means that there are not any genes that, if analyzed alone, will result in a phylogeny that is not found in any other gene.

Tree Topology Tests

Several tests were run on the supermatrix tree in PAUP*. The PTP test yielded a p-value of 0.001, suggesting rejection of the null hypothesis that the data are unfit to use in a phylogeny (Slowinski and Crother, 1998). CI and RI indices yielded values of 0.93 and 0.58, respectively. The branch lengths of the supermatrix tree in Figure 3 are roughly equal and symmetrical, suggesting that long-branch attraction isn't influencing topology (Bergsten, 2005).

Supermatrix

The parsimony analysis of the supermatrix resulted in only one best tree (Fig. 3). The bootstrap values are all 100 on each node, suggesting that the data highly supports the solution. The tree that is supported by the largest number of genes is the same tree that is the most parsimonious solution for the supermatrix. The topologies of the most parsimonious trees with or without GBlocks editing were congruent.

Number of Genes

The probability of arriving at the correct topology increases with gene sampling (Fig. 4). If two genes are sampled, the probability that the correct gene is among the set of solutions is .393, meaning that if two genes are chosen for a phylogeny, there is a

39.3% chance that the correct topology will be discovered. If three genes are sampled the probability is 57.7%, and for four it is 63.1%. At ten genes, the probability is 91.8% of sampling one of the genes that is concordant with the best-supported topology at least once.

The algorithm, continued indefinitely, will always produce a probability less than one. At some point, all the genes will have been sampled and of course the optimal topology will have been discovered. As more genes are sampled, they become less independent, violating the assumption of independence, so the analysis becomes less valid as the number of genes approaches the total number of genes in the genome. For this data set, more than 77.9% of the genes need to be included to ensure the presence of at least one gene that is concordant with the optimal topology. If fewer are selected, it is possible that none of the genes result in the most supported topology, though the probability of that is low with the use of many genes.

DISCUSSION

Most of the trees in the gene-topology distribution support only a small percentage of the many possible topologies. This should comfort taxonomists (Felsenstein, 1978) because almost all of the genes in the *Steinernema* genomes have strong phylogenetic signal. It would be interesting to see how this pattern holds up for both very deep and shallow phylogenetic analyses.

Supermatrix

The tree that was best supported by the supermatrix is the tree that is most commonly found in the gene-topology distribution. In this dataset the phylogenetic signal of the majority of the genes is the same as the supermatrix. The results show that both methodologies converged on the same optimal solution.

Number of Genes

Our results suggest that many genes should be used in order to accurately estimate a phylogeny. In this data set, it would be necessary to sample at least 25 genes to have a high probability of getting even one gene that is concordant with the most supported topology. It makes intuitive sense that adding more characters would increase phylogenetic accuracy. Adding more taxa makes the problem more complicated by increasing the number of possible trees while using less information per taxon to solve the problem. Adding more genes adds more informative data, which increases the probability of finding the correct solution. In this study where the taxa have roughly equal rates of evolution, it seems advantageous to add more characters to make an accurate phylogeny rather than adding more taxa.

Alignment Editing

Our results showed that the supertree topologies of the most parsimonious trees with or without GBlocks editing were congruent. This does not mean that alignment editing is unnecessary in all cases, however. It may be the case that with so much data, there was more phylogenetic signal than usual, and the best-supported tree was found with imperfect alignments in the supermatrix without alignment editing.

CITATIONS

Adams, B. J., Peat, S. M., Dillman, A. R. 2007. Phylogeny and evolution. Nematology Monographs and Perspectives 5, 693-733.

Baurain, D., Brinkmann, H., Philippe, H. 2007. Lack of Resolution in the Animal Phylogeny: Closely Spaced Cladogeneses or Undetected Systematic Errors? Mol. Biol. Evol. 24, 6-9.

Bergsten, J. 2005. A review of long-branch attraction. Cladistics 21, 163-193.

Bininda-Emonds, O. R. P. 2004. The evolution of supertrees. Trends Ecol. Evol. 19, 315-322.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-522.

De Queiroz, A., Donoghue, M. J., Kim, J. 1995. Separate Versus Combined Analysis OF Phylogenetic Evidence. Annu. Rev. Ecol. Syst. 26, 657-681.

Felsenstein, J. 1978. Number of Evolutionary Trees. Systematic Zoology 27, 27-33.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39, 783-791.

Gatesy, J., O'Grady, P., Baker, R. H. 1999. Corroboration among data sets in simultaneous analysis: Hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15, 271-313.

Goodrich-Blair, H., Clarke, D. J. 2007. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. Mol. Microbiol. 64, 260-268.

Graybeal, A. 1998. Is It Better to Add Taxa or Characters to a Difficult Phylogenetic Problem? Syst. Biol. 47, 9-17.

Grenier, E., Catzeflis, F. M., Abad, P. 1997. Genome sizes of the entomopathogenic nematodes *Steinernema carpocapsae* and *Heterorhabditis bacteriophora* (Nematoda: Rhabditida). Parasitology 114, 497-501.

Hillis, D. M. 1996. Inferring complex phylogenies. Nature 383, 130-131.

Katoh, K., Kuma, K., Miyata, T., Toh, H. 2005. Improvement in the accuracy of multiple sequence alignment program MAFFT. Genome Informatics 16, 22-33.

Katoh, K., Misawa, K., Kuma, K., Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059-3066.

Lambkin, C. L. 2004. Partitioned Bremer support localises significant conflict in bee flies (Diptera : Bombyliidae : Anthracinae). Invertebr. Syst. 18, 351-360.

Li, L., Stoeckert, C. J., Roos, D. S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178-2189.

Money, D., Whelan, S. 2012. Characterizing the Phylogenetic Tree-Search Problem. Syst. Biol. 61, 228-239.

Nabhan, A. R., Sarkar, I. N. 2010. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Briefings in Bioinformatics 13, 122-134.

Nguyen, K. B., Maruniak, J., Adams, B. J. 2001. Diagnostic and phylogenetic utility of the rDNA internal transcribed spacer sequences of *Steinernema*. J. Nematol. 33, 73-82.

Ogden, T. H., Rosenberg, M. S. 2005. Multiple Sequence Alignment Accuracy and Phylgenetic Inference. Syst. Biol. 55, 314-328.

Poe, S., Swofford, D. L. 1999. Taxon Sampling Revisited. Nature 398, 299-300.

Poinar, G. O. 1993. Origins and Phylogenetic Relationships of the Entomophilic *Rhabditis*, *Heterorhabditis*, and *Steinernema*. Fundam. Appl. Nematol. 16, 333-338.

Simmons, M. P., Muller, K. F., Webb, C. T. 2011. The deterministic effects of alignment bias in phylogenetic inference. Cladistics 27, 402-416.

Slowinski, J. B., Crother, B. I. 1998. Is the PTP Test Useful? Cladistics 14, 297-302.

Stanke, M., Tzvetkova, A., Morgenstern, B. 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol. 7.

Sukumaran, J. 2008. Fasta to Nexus. Program and Scripts.

Swofford, D. L. 2002. Phylogenetic Analysis Using Parsimony (*and Other Methods). PAUP. Sinauer Associates, Sunderland, Massachusetts.

Talavera, G., Castresana, J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564-577.

Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., Davis, P., de la Cruz, N., Duong, A., Fang, R. H., Ganesan, U., Grove, C., Howe, K., Kadam, S., Kishore, R., Lee, R., Li, Y. L., Muller, H. M., Nakamura, C., Nash, B., Ozersky, P., Paulini, M., Raciti, D., Rangarajan, A., Schindelman, G., Shi, X. Q., Schwarz, E. M., Tuli, M. A., Van Auken, K., Wang, D., Wang, X. D., Williams, G., Hodgkin, J., Berriman, M., Durbin, R., Kersey, P., Spieth, J., Stein, L., Sternberg, P. W. 2012. WormBase 2012: more genomes, more data, new website. Nucleic Acids Research 40, D735-D741.

FIGURES AND TABLES

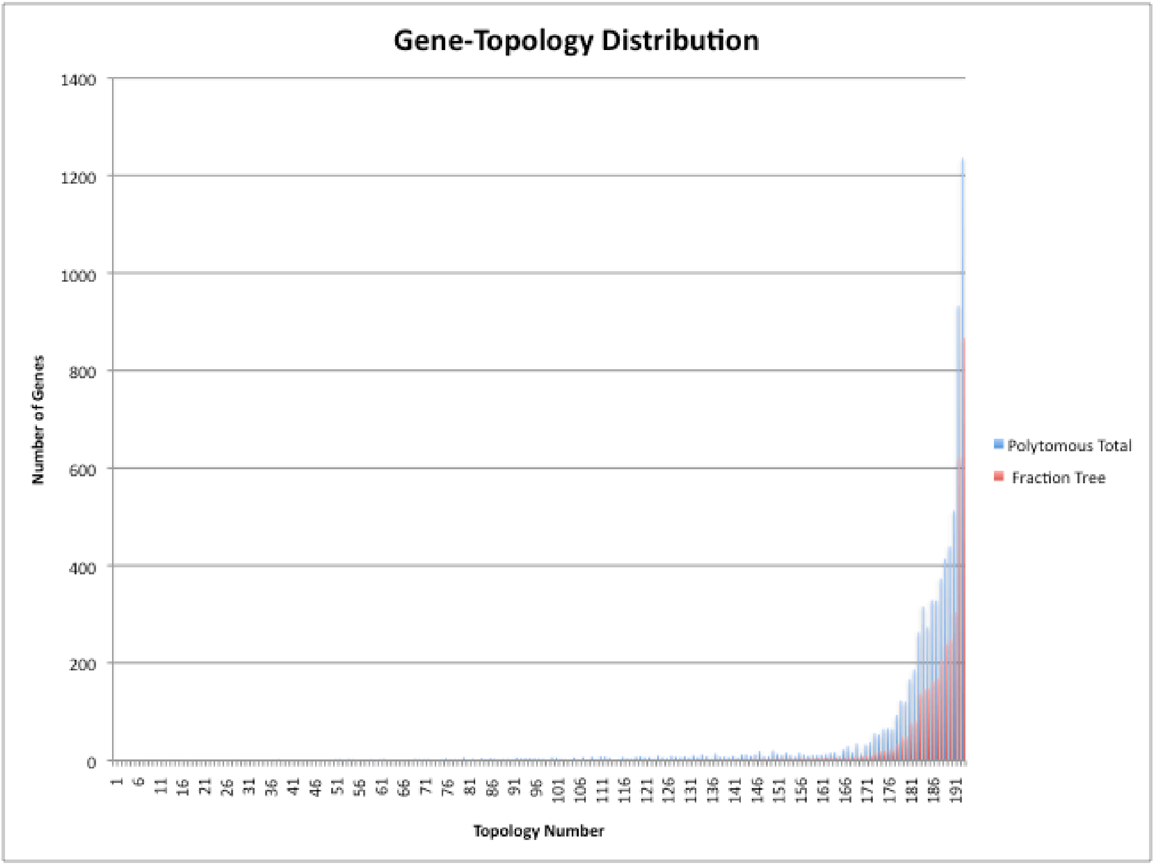

Figure 1. Gene-topology distribution shows the number of trees that support each gene. Fraction Gene was calculated by using the percent of the trees that supported a particular topology. Polytomous Total weights each topology the same and sums to more than the number of genes in the analysis. 97 of the topologies are at least partially unresolved.

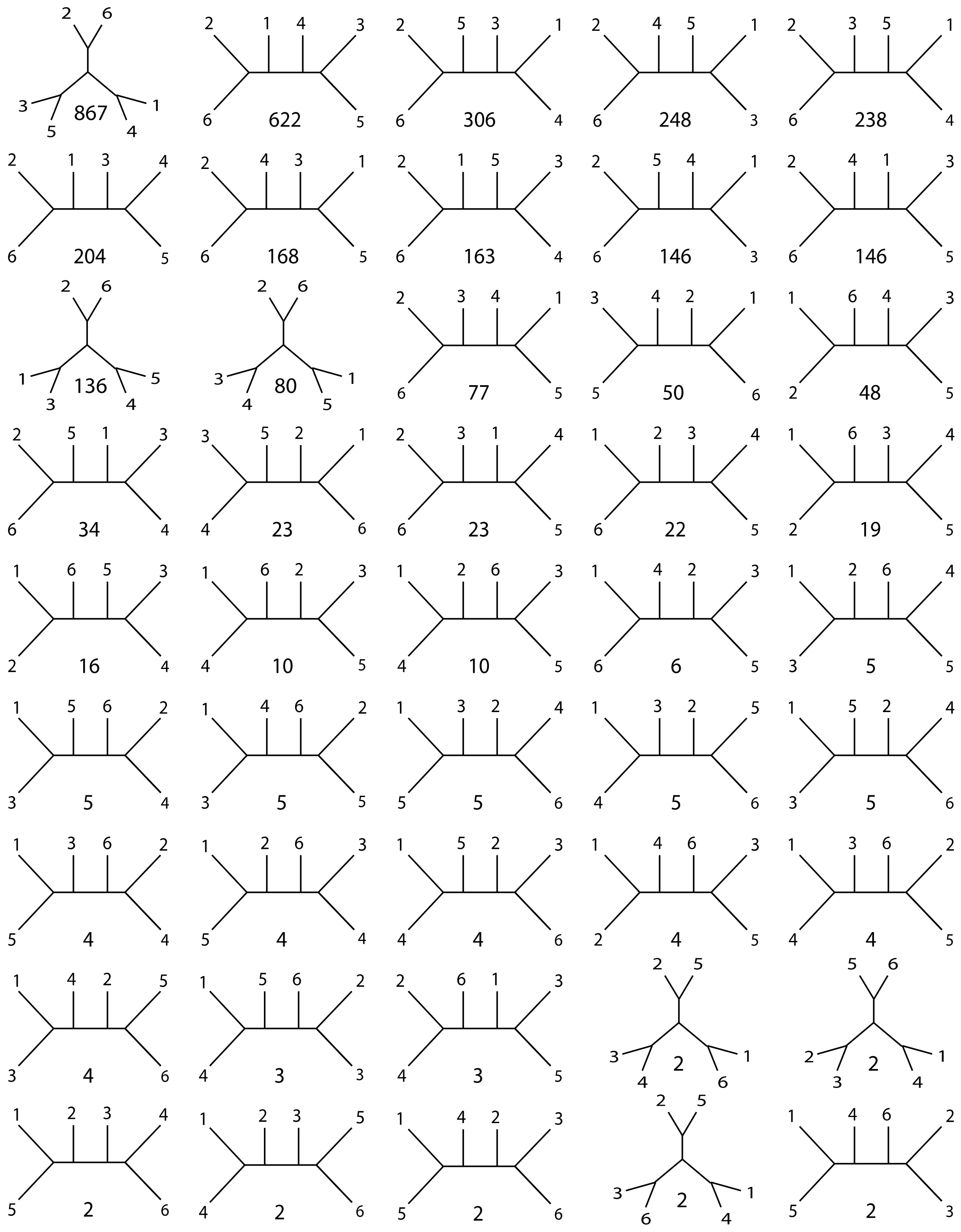

Figure 2. The top 45 resolved topologies and the number of rounded Fraction Genes that support each (shown unrooted). 1 = *Panagrellus redivivus*, 2 = *Scapterisci carpocapsae*, 3 = *S. feltiae*, 4 = *S. glaseri*, 5 = *S. monticolum*, 6 = *S. scapterisci*.

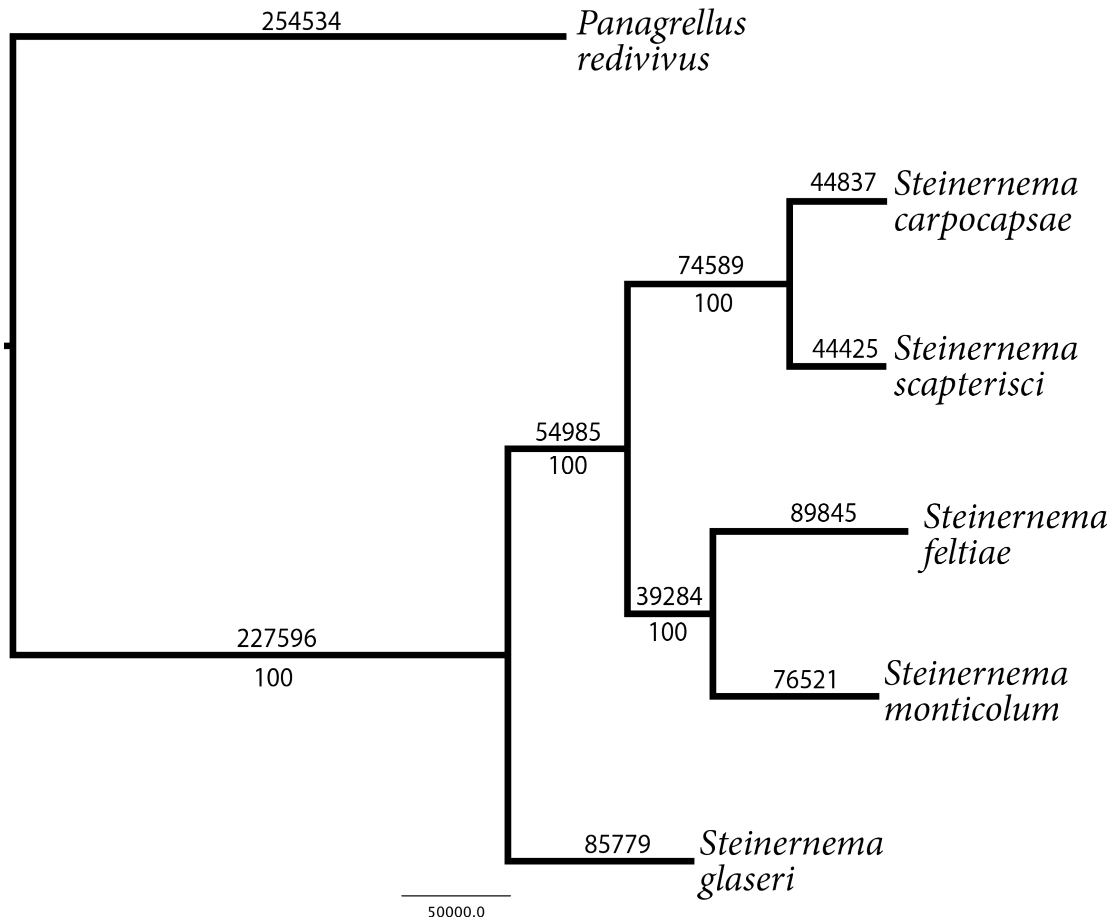

Figure 3. Supermatrix parsimony tree. Node labels are branch lengths and bootstrap values.

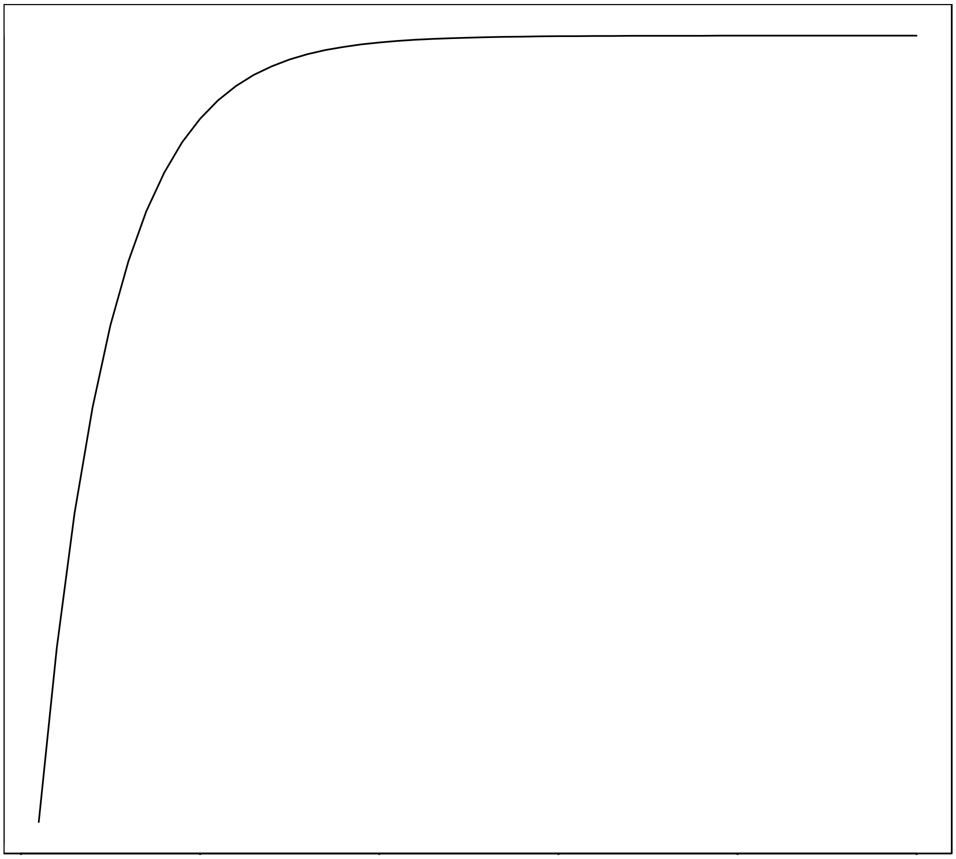

Figure 4. Probability of containing best supported topology. The probability that selecting a certain number of genes will yield at least one gene with the best-supported topology for up to 50 genes.

Table 1. A chart of all topologies. Fraction Gene was calculated by using the percent of the trees that supported a particular topology. Polytomous Total weights each topology the same and sums to more than the number of genes in the analysis. All topologies are rooted.

Tree Layouts	Polytomous Total	Fraction Tree	Unresolved or Resolved
(1,(((2,6),(3,5)),4))	1236	866.5749371	resolved
(1,((2,6),((3,5),4)))	933	622.4125994	resolved
(1,((((2,6),5),3),4))	512	305.1721071	resolved
(1,((((2,6),4),5),3))	439	248.3988206	resolved
(1,((((2,6),3),5),4))	414	238.2342102	resolved
(1,((2,6),(3,(4,5))))	373	203.6025109	resolved
(1,((((2,6),4),3),5))	328	167.9403411	resolved
(1,((2,6),((3,4),5)))	329	163.2882052	resolved
(1,((((2,6),5),4),3))	274	146.3986689	resolved
(1,(((2,6),4),(3,5)))	316	146.3028772	resolved
(1,(((2,6),(4,5)),3))	263	136.4934852	resolved
(1,(((2,6),(3,4)),5))	187	80.39462659	resolved
(1,((((2,6),3),4),5))	167	77.26429681	resolved
(1,((2,((3,5),4)),6))	121	49.74258468	resolved
(1,(2,(((3,5),4),6)))	123	47.94811526	resolved
(1,(((2,6),5),(3,4)))	93	33.94648636	resolved
(1,((2,((3,4),5)),6))	65	23.26746903	resolved
(1,(((2,6),3),(4,5)))	66	22.84984876	resolved
(1,((2,(3,(4,5))),6))	65	21.59014409	resolved
(1,(2,((3,(4,5)),6)))	53	19.37774538	resolved
(1,(2,(((3,4),5),6)))	56	15.61978893	resolved
(1,(((2,(3,5)),6),4))	38	10.08481051	resolved
(1,((2,((3,5),6)),4))	33	9.676623377	resolved
(1,(2,3,4,5,6))	15	8.323232323	unresolved
(1,(((2,6),3,5),4))	35	8.108910534	unresolved
(1,((2,3,5,6),4))	17	6.470851371	unresolved
(1,((2,6),(3,4,5)))	30	6.469155844	unresolved
(1,((2,(3,5),6),4))	23	5.77202381	unresolved
(1,(((2,(3,5)),4),6))	10	5.514969241	resolved
(1,((2,6),3,4,5))	17	5.374116162	unresolved
(1,((2,((4,5),6)),3))	16	5.278329514	resolved
(1,((((2,4),6),5),3))	14	5.23452381	resolved
(1,((((2,5),6),4),3))	12	5.202030812	resolved

(1,(((2,(4,6)),3),5))	12	5.149456976	resolved
(1,(2,(3,4,5),6))	12	5.048809524	unresolved
(1,((2,4,5,6),3))	10	4.883838384	unresolved
(1,(((2,(5,6)),3),4))	13	4.614285714	resolved
(1,(((2,(4,6)),5),3))	16	4.557142857	resolved
(1,((((2,4),6),3),5))	8	4.308333333	resolved
(1,((2,((3,4),6)),5))	12	4.303221289	resolved
(1,(((2,(3,6)),5),4))	16	4.277633478	resolved
(1,(2,(((3,5),6),4)))	12	4.185353535	resolved
(1,((((2,5),6),3),4))	14	4.023845599	resolved
(1,(((2,6),4,5),3))	21	3.883813409	unresolved
(1,(((2,(5,6)),4),3))	9	3.862745098	resolved
(1,(((2,5,6),3),4))	9	3.757575758	unresolved
(1,(2,((3,5),4),6))	20	3.522059885	unresolved
(1,((((2,3),6),5),4))	13	3.367766955	resolved
(1,(((2,3,6),5),4))	10	3.209830447	unresolved
(1,(((2,4),6),(3,5)))	13	2.920238095	resolved
(1,(((2,6),3,4),5))	13	2.876587302	unresolved
(1,((2,3,4,6),5))	6	2.558080808	unresolved
(1,(((2,4,6),5),3))	10	2.466901154	unresolved
(1,(((2,5),(3,4)),6))	7	2.448459384	resolved
(1,(((2,4,6),3),5))	9	2.443091631	unresolved
(1,(((2,3),(5,6)),4))	8	2.397619048	resolved
(1,((2,(3,4,5)),6))	15	2.369191919	unresolved
(1,((2,(3,(4,6))),5))	4	2.369047619	resolved
(1,((2,(3,(5,6))),4))	9	2.366666667	resolved
(1,(((2,6),4),3,5))	13	2.266123642	unresolved
(1,((2,6),(3,5),4))	7	2.170833333	unresolved
(1,(((2,(3,6)),4),5))	11	2.145941558	resolved
(1,(((2,5,6),4),3))	6	1.983333333	unresolved
(1,(((2,5),(3,6)),4))	9	1.924747475	resolved
(1,((2,(4,5),6),3))	7	1.917857143	unresolved
(1,((((2,3),6),4),5))	8	1.914880952	resolved
(1,((2,(4,6)),(3,5)))	10	1.822076023	resolved
(1,(((2,(3,4)),6),5))	5	1.769047619	resolved
(1,((2,5),((3,4),6)))	7	1.734173669	resolved
(1,(((2,(4,5)),6),3))	11	1.649162847	resolved
(1,(2,((3,(5,6)),4)))	4	1.646464646	resolved
(1,(((2,3,6),4),5))	7	1.563636364	unresolved

(1,((2,(3,6)),(4,5)))	7	1.510335498	resolved
(1,(2,((3,4),5),6))	9	1.491269841	unresolved
(1,(2,(3,5),4,6))	8	1.465277778	unresolved
(1,((2,4,6),3,5))	4	1.455357143	unresolved
(1,(((2,3),6),(4,5)))	5	1.370941558	resolved
(1,((2,((3,6),5)),4))	8	1.296969697	resolved
(1,((2,3),4,5,6))	3	1.272727273	unresolved
(1,((2,5,6),(3,4)))	2	1.25	unresolved
(1,(((2,4),(3,5)),6))	6	1.242857143	resolved
(1,((2,4,6),(3,5)))	9	1.199044012	unresolved
(1,(2,((3,4,5),6)))	9	1.166973304	unresolved
(1,(((2,3,5),6),4))	3	1.1625	unresolved
(1,(2,((3,5),4,6)))	8	1.104599567	unresolved
(1,((((2,3),5),4),6))	2	1.1	resolved
(1,((2,(3,4),6),5))	7	1.09710657	unresolved
(1,(2,3,(4,5),6))	3	1.085227273	unresolved
(1,((((2,5),3),6),4))	6	1.051190476	resolved
(1,((((2,3),4),6),5))	2	1.035714286	resolved
(1,((((2,4),5),6),3))	3	1	resolved
(1,(2,((3,5),6),4))	4	0.952020202	unresolved
(1,(2,(((3,4),6),5)))	6	0.900840336	resolved
(1,(2,((3,4),(5,6))))	7	0.883751869	resolved
(1,((2,3,6),4,5))	2	0.833333333	unresolved
(1,(((2,6),5),3,4))	4	0.827020202	unresolved
(1,(((2,6),3),4,5))	4	0.722727273	unresolved
(1,(2,((3,5),(4,6))))	5	0.68358396	resolved
(1,((((2,5),4),6),3))	5	0.654411765	resolved
(1,((2,(5,6)),(3,4)))	5	0.653221289	resolved
(1,(((2,3),5,6),4))	5	0.62034632	unresolved
(1,(((2,5),6),(3,4)))	6	0.61512605	resolved
(1,((2,3,6),(4,5)))	3	0.606060606	unresolved
(1,(((2,5),(4,6)),3))	4	0.571078431	resolved
(1,((2,4),((3,5),6)))	4	0.567857143	resolved
(1,((2,(3,5),4),6))	3	0.5625	unresolved
(1,((2,(3,5)),(4,6)))	4	0.557393484	resolved
(1,((2,6),(3,4),5))	5	0.552173703	unresolved
(1,(2,(3,4,5,6)))	3	0.545454545	unresolved
(1,(((2,(4,5)),3),6))	5	0.53219697	resolved
(1,(((2,4,5),6),3))	2	0.522727273	unresolved

(1,((2,6),3,(4,5)))	4	0.5125	unresolved
(1,((2,4),(3,(5,6))))	1	0.5	resolved
(1,(2,((3,6),(4,5))))	7	0.436511609	resolved
(1,((2,4),(3,5),6))	3	0.425	unresolved
(1,(((2,3),(4,5)),6))	3	0.418560606	resolved
(1,((2,3,4,5),6))	3	0.378787879	unresolved
(1,(((2,5),3,6),4))	5	0.378679654	unresolved
(1,((((2,3),5),6),4))	3	0.3625	resolved
(1,((((2,4),5),3),6))	2	0.356060606	resolved
(1,(((2,4,5),3),6))	2	0.356060606	unresolved
(1,((2,(3,5,6)),4))	3	0.328282828	unresolved
(1,(2,(((3,6),5),4)))	3	0.328282828	resolved
(1,((((2,5),3),4),6))	3	0.326190476	resolved
(1,(2,(((3,6),4),5)))	4	0.302139037	resolved
(1,((2,4),(3,5,6)))	2	0.272727273	unresolved
(1,((2,3),(4,5,6)))	2	0.272727273	unresolved
(1,(2,3,(4,5,6)))	1	0.25	unresolved
(1,(2,(3,5,6),4))	1	0.25	unresolved
(1,(2,(3,4),5,6))	1	0.25	unresolved
(1,((((2,4),3),5),6))	1	0.25	resolved
(1,(2,(3,(4,5),6)))	4	0.240782828	unresolved
(1,((2,(3,6),5),4))	3	0.233838384	unresolved
(1,(((2,4),3,5),6))	2	0.222727273	unresolved
(1,((2,5),((3,6),4)))	3	0.218805704	resolved
(1,((2,5),(3,4),6))	2	0.214285714	unresolved
(1,(((2,3,5),4),6))	2	0.2	unresolved
(1,((2,5),(3,(4,6))))	2	0.196078431	resolved
(1,(2,((3,4),5,6)))	3	0.194083694	unresolved
(1,(2,(3,((4,5),6))))	4	0.191562114	resolved
(1,(2,(3,(4,(5,6)))))	2	0.176923077	resolved
(1,((2,((3,6),4)),5))	4	0.171186656	resolved
(1,(2,(3,4),(5,6)))	3	0.167939903	unresolved
(1,(2,(3,4,(5,6))))	2	0.167832168	unresolved
(1,(2,((3,(4,6)),5)))	1	0.166666667	resolved
(1,(((2,5),6),3,4))	1	0.166666667	unresolved
(1,(2,((3,4),6),5))	1	0.166666667	unresolved
(1,((2,5,6),3,4))	1	0.166666667	unresolved
(1,((2,(3,6),4),5))	2	0.160714286	unresolved
(1,(2,(3,(4,5)),6))	2	0.158730159	unresolved

(1,(((2,4),3,6),5))	2	0.147727273	unresolved
(1,(((2,4),(3,6)),5))	2	0.147727273	resolved
(1,(2,(3,(4,6),5)))	1	0.142857143	unresolved
(1,(((2,5),3),(4,6)))	1	0.142857143	resolved
(1,((2,(4,6)),3,5))	2	0.12406015	unresolved
(1,((2,(4,5)),(3,6)))	3	0.107954545	resolved
(1,((((2,5),4),3),6))	2	0.106060606	resolved
(1,(((2,5),4),(3,6)))	2	0.106060606	resolved
(1,((2,3),(4,(5,6))))	1	0.1	resolved
(1,(((2,4),(5,6)),3))	1	0.1	resolved
(1,((2,(4,(5,6))),3))	1	0.1	resolved
(1,((2,3,5),4,6))	1	0.1	unresolved
(1,(((2,5),3),4,6))	1	0.1	unresolved
(1,((2,3,(4,5)),6))	2	0.085227273	unresolved
(1,((2,3),((4,5),6)))	2	0.085227273	resolved
(1,(((2,5),4,6),3))	1	0.083333333	unresolved
(1,(2,((3,5,6),4)))	2	0.078282828	unresolved
(1,((2,(3,4),5),6))	2	0.077030812	unresolved
(1,((2,(3,5)),4,6))	1	0.0625	unresolved
(1,((2,((4,6),5)),3))	1	0.0625	resolved
(1,((2,(4,6),5),3))	1	0.0625	unresolved
(1,((2,(3,4,6)),5))	2	0.052139037	unresolved
(1,(2,((3,4,6),5)))	2	0.052139037	unresolved
(1,((2,3,(5,6)),4))	1	0.047619048	unresolved
(1,(((2,3),(4,6)),5))	1	0.035714286	resolved
(1,(((2,3,4),6),5))	1	0.035714286	unresolved
(1,(((2,5),3,4),6))	1	0.022727273	unresolved
(1,((2,5),3,4,6))	1	0.022727273	unresolved
(1,((2,4),3,5,6))	1	0.022727273	unresolved
(1,(((2,4),5),3,6))	1	0.022727273	unresolved
(1,(((2,4),5),(3,6)))	1	0.022727273	resolved
(1,((2,(4,5)),3,6))	1	0.022727273	unresolved
(1,(((2,5),4),3,6))	1	0.022727273	unresolved
(1,((2,4),((3,6),5)))	1	0.022727273	resolved
(1,((2,5),(3,4,6)))	1	0.022727273	unresolved
(1,(((2,3),4,5),6))	1	0.022727273	unresolved
(1,(((2,3),6),4,5))	1	0.022727273	unresolved
(1,((2,(3,6)),4,5))	1	0.022727273	unresolved
(1,(2,(3,6),4,5))	1	0.022727273	unresolved

(1,((2,4,5),(3,6)))	1	0.022727273	unresolved
(1,((2,(4,5,6)),3))	1	0.022727273	unresolved
(1,(2,((3,6),4,5)))	1	0.022727273	unresolved
(1,(2,(3,(4,5,6))))	1	0.022727273	unresolved