Dynamic programming procedure for searching optimal models to estimate substitution rates based on the maximum-likelihood method

Chengjun Zhang[a], Jia Wang[a], Weibo Xie[a], Gang Zhou[a], Manyuan Long[b,1], and Qifa Zhang[a,1]

[a]National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China; and [b]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637

Contributed by Qifa Zhang, March 25, 2011 (sent for review May 12, 2010)

The substitution rate in a gene can provide valuable information for understanding its functionality and evolution. A widely used method to estimate substitution rates is the maximum-likelihood method implemented in the CODEML program in the PAML package. A limited number of branch models, chosen based on a priori information or an interest in a particular lineage(s), are tested, whereas a large number of potential models are neglected. A complementary approach is also needed to test all or a large number of possible models to search for the globally optional model(s) of maximum likelihood. However, the computational time for this search even in a small number of sequences becomes impractically long. Thus, it is desirable to explore the most probable spaces to search for the optimal models. Using dynamic programming techniques, we developed a simple computational method for searching the most probable optimal branch-specific models in a practically feasible computational time. We propose three search methods to find the optimal models, which explored O(n) (method 1) to O(n^2) (method 2 and method 3) models when the given phylogeny has *n* branches. In addition, we derived a formula to calculate the number of all possible models, revealing the complexity of finding the optimal branch-specific model. We show that in a reanalysis of over 50 previously published studies, the vast majority obtained better models with significantly higher likelihoods than the conventional hypothesis model methods.

likelihood-ratio test | natural selection | positive selection | synonymous substitution | nonsynonymous substitution

Estimating substitution rates is important in the investigation of functionality and evolution of genes. Natural selection can be also tested by comparing the substitution rates at synonymous and nonsynonymous sites, denoted usually as K_s and K_a, respectively (K_a = number of nonsynonymous substitutions per nonsynonymous site, Ks = number of synonymous substitutions per synonymous site). Such estimation is usually performed by analyzing the divergence of a protein-coding gene in a number of homologous sequences in different species.

The maximum-likelihood method is widely used for estimating the substitution rates of nucleotide sequences in protein-coding genes in molecular evolutionary analysis, although some of its techniques were recently debated (1, 2). The CODEML program in the PAML package (3) is among the most frequently used and utilizes a codon substitution model to infer evolutionary rates. Several approaches were incorporated into the program, including a site model, a clade model, a branch model, and a branch-site model. The widely used branch model allows estimation of the substitution rates with variable ratios of $\omega = K_a/K_s$ in different branches (lineages) in a phylogeny. Generally, $\omega > 1$ indicates positive selection, $\omega < 1$ indicates purifying selection with functional constraint, and $\omega \sim 1$ indicates neutral evolution (4).

The branch model was initially applied to the evolutionary analysis of the primate gene-encoding lysozyme (5). The analysis showed that the ω-parameter along the hominoid branch was significantly greater than 1, indicating that positive selection might have operated on it. This model has been widely used in molecular evolutionary studies and the functional analyses of genes, and it is particularly valuable to detect positive selection after gene duplications (3). For example, a branch model analysis of the *Drosophila* retroposed gene *Dntf-2r* detected positive selection (6). The use of this model revealed that three young chimeric genes, *jingwei, Adh-Twain*, and *Adh-Finnegan*, underwent both early rapid evolution and subsequent slow evolution of protein sequences resulting from increased functional constraints (7, 8). Branch model analysis on the NOD26-like intrinsic proteins also detected strong selective pressure on highly constrained functional proteins and many positive selective events that might change the gene's functions after the duplication and speciation events in the plants (9).

In the branch model analysis, a range of ω-values can be chosen. The one-ratio model (ORM) assumes that all branches have the same one ω-parameter, whereas the free-ratio model (FRM) assigns a different ω-parameter to each branch in the tree for estimation. Between ORM and FRM are a limited number of hypothesis models, assuming that some specific branches have specific ratios based on a priori available information or interest in a possible positive selection on a branch(s) implied by FRM analysis. These models were explored and compared by likelihood-ratio tests (LRTs) (5, 10). Obviously, in this approach, it is imperative to have some good a priori reasons to restrict the estimate of spaces to explore. As Pond and Frost pointed out (11), however, this approach has a disadvantage, because it is not always possible to derive suitable hypotheses when no useful information is available or when no branch can be focused on in the model search. As a model-searching approach to complement the current approach, there is thus a need to search all possible models for the best model that has a globally maximum likelihood. Because all models, except the ORM and FRM, need to be specified with ω-parameters for certain branches, however, the analysis often becomes impractical, especially because all possible models often require an intractably large number of repeated computations of likelihoods.

To solve these technical difficulties, we proposed to search the most probable spaces to determine the optimal branch-specific models that have likelihoods equal or close to the globally maximum likelihood over all possible models with the least degrees of freedom (12). We developed a two-step method to count all possible branch models to reveal the complexity of the computation using CODEML. Then, motivated by the dynamic programming that is widely used in computation (13), we developed three simple and rapid methods in search of the optimal branch models in the most probable spaces for the maximum likelihood. Finally, the proposed methods were assessed by the lysozyme sequences of primate species (5) and reanalysis of 50 previously published

Author contributions: C.Z., M.L., and Q.Z. designed research; C.Z., J.W., W.X., and G.Z. performed research; C.Z. and M.L. analyzed data; and C.Z., M.L., and Q.Z. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: mlong@uchicago.edu or qifazh@mail.hzau.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018621108/-/DCSupplemental.

studies. Through these analyses, we show that our simple methods can obtain globally better models with significantly higher likelihoods than the current approach that compares the models on the branches of particular interest. Because the current approach relies on the hypothesized branches of interest to test positive selection, we call it the "conventional hypothesis model."

Results

Large Number of Possible Branch Models. We calculated the number of all possible branch models using a two-step strategy, which is used in a program written using the Perl script (*SI Appendix*). In the first step, we defined a model that included the number of ω's and the branch number for each ω, recording this model in a configuration. For example, for a tree of four branches with three sequences, assuming two ω-values, ω_1 for one branch and ω_2 for the other three branches, we record this configuration as a vector (1 ω_1, 3 ω_2), or simply (1, 3). We developed a traversing algorithm to find all the configurations of a variety of ratios. In the second step, we calculated all possible branch models with each configuration following the two formulas that we derived, as shown below.

Imagine a phylogeny of six branches with four sequences (*SI Appendix*, Fig. 1). The models for this tree can be divided into six groups [ranging from ORMs, to two-ratio models, up to the six-ratio model (FRM)], and in each group, the models can be divided into several configurations. For example, it has three configurations in three-ratio models: the first configuration has one branch with ω_1, one branch with ω_2, and the other four branches with ω_3, expressed as (1, 1, 4); the second configuration has one branch with ω_1, two branches with ω_2, and the other three branches with ω_3, expressed as (1, 2, 3); and the third configuration has three two branches with ω_1, ω_2, and ω_3, respectively, expressed as (2, 2, 2).

The number of the models for the first configuration (1, 1, 4) can be calculated and expressed as K_{31}, the numbers of the models for the second and the third configurations [(1, 2, 3) and (2, 2, 2)] as K_{32} and K_{33}, respectively. In K_{32}, because the components in the configuration are not equal to each other, all possible combinations are

$$K_{32} = C_6^1 \times C_5^2 = 60$$

Because the first configuration has two different types (the numbers of branches) of components in K_{31} and the third configuration has three components each with the same number of branches in K_{33},

$$K_{31} = C_6^1 \times C_5^1 \div 2! = 15$$

Where the 2! is the denominator because we need only the combination, the order of arrangement does not matter. Similarly, we have

$$K_{33} = C_6^2 \times C_4^2 \div 3! = 15$$

In general, for a phylogeny with n branches, we use K_{mj} to denote the possible model numbers for the *jth* configuration with m ω-parameters; q_{ij} denotes the branch numbers of the *ith* ω-parameter of the *jth* configuration. By definition, we have

$$\sum_{i=1}^{m} q_{ij} = n, m \in (1 \text{ to } n).$$

When $q_{xj} \neq q_{yj}$ ($x \neq y$, $x, y \in$ (1 to m)), $q_{0j} = 0$), the formula to calculate K_{mj} can be expressed as

$$K_{mj} = \prod_{i=1}^{m-1} C_{n-\sum q_{(i-1)j}}^{q_{ij}} \quad [1]$$

When there exist x and y variables, let $q_{xj} = q_{yj}$ [$x \neq y$, $x, y \in$ (1 to m)], $q_{0j} = 0$ (A_g means having g groups and A_g components in the configuration, which have the same branch numbers), and thus we have

$$K_{mj} = \frac{\prod_{i=1}^{m-1} C_{n-\sum q_{(i-1)j}}^{q_{ij}}}{\prod_{l-1}^{g} A_l!} \quad [2]$$

By means of this approach, to illustrate the intractably large number of possible branch models visually, all configuration numbers and possible model numbers of phylogeny for 3, 4, 6, 8, 10, and 12 sequences are shown in Table 1 for all possible ω-values; an example of the details of the configuration and model is provided in *SI Appendix*.

Dynamic Programming Algorithms for Searching Optimal Branch Models. Despite present-day rapidly increasing computing powers, it is impractical to use the traversing algorithm to explore all models, as shown in Table 1. We developed three simplified methods for searching optimal models by using dynamic programming algorithms. We attempted to reduce computation to a practical workable level by exploring the most likely space that contains the maximum likelihood.

Method 1. Fig. 1*A* summarizes the procedure we propose. First, calculate all possible configurations for single-branch two-ratio models (SBTRMs), in which only one branch is labeled with ω_1 and all other branches are assumed to be background ratio ω_0. Obviously, the log likelihood (lnL) values for n SBTRMs need to be calculated when the analyzed phylogeny has n branches. Second, the lnL values of all n SBTRMs are compared and sorted from maximum to minimum; the model with the maximum lnL value is considered the optimal model within two-ratio models. The branch labeled with ω_1 in the maximum lnL value model is recorded as B_1, the branch labeled with ω_1 in the model that has the second greatest lnL value is recorded as B_2, and so on until B_n. Then, all the optimal models of the remaining variety of ratios are generated directly. For the optimal three-ratio model, branch B_1 is labeled as ω_1, branch B_2 is labeled as ω_2, and all other branches are assumed to have a background ratio ω_0 and optimal models for four ratios to an "$n - 1$" ratio as well. Finally, the $n - 2$ optimal models can be "predicted" in this way, and the likelihoods of these predicted models can be calculated and compared with each other to determine the final optimal model that has the maximum likelihood in the sense that the likelihood is significantly better than the likelihood of other optimal models and has the least degrees of freedom if there are more than one solutions that are not significantly different.

Method 2. This method can be described in $n - 2$ rounds with two steps in each round of iterations, as shown in Fig. 1*B*. The first step generates models and calculates InLs for all these models; the second step is to record the specific branch of the optimal model of this round, which is used for generating models in the next round. The models in the first round are all SBTRMs. The branch labeled with ω_1 in the maximum lnL value model is recorded as B_1. In the second round, $n - 1$ three-ratio models are generated by adding one more branch with one more ratio (ω_2) in addition to B_1, whereas all other $n - 2$ branches have the background ratio ω_0. The InLs for all $n - 1$ three-ratio models are calculated and compared with each other. The branch labeled ω_2 of the optimal

Table 1. Configurations and possible models

Sequence no.	Branch no.	Configuration no.	No. of possible models
3	4	3	15
4	6	9	203
6	10	40	115,975
8	14	133	190,899,322
10	18	383	6.821E + 11
12	22	1,000	4.507E + 15

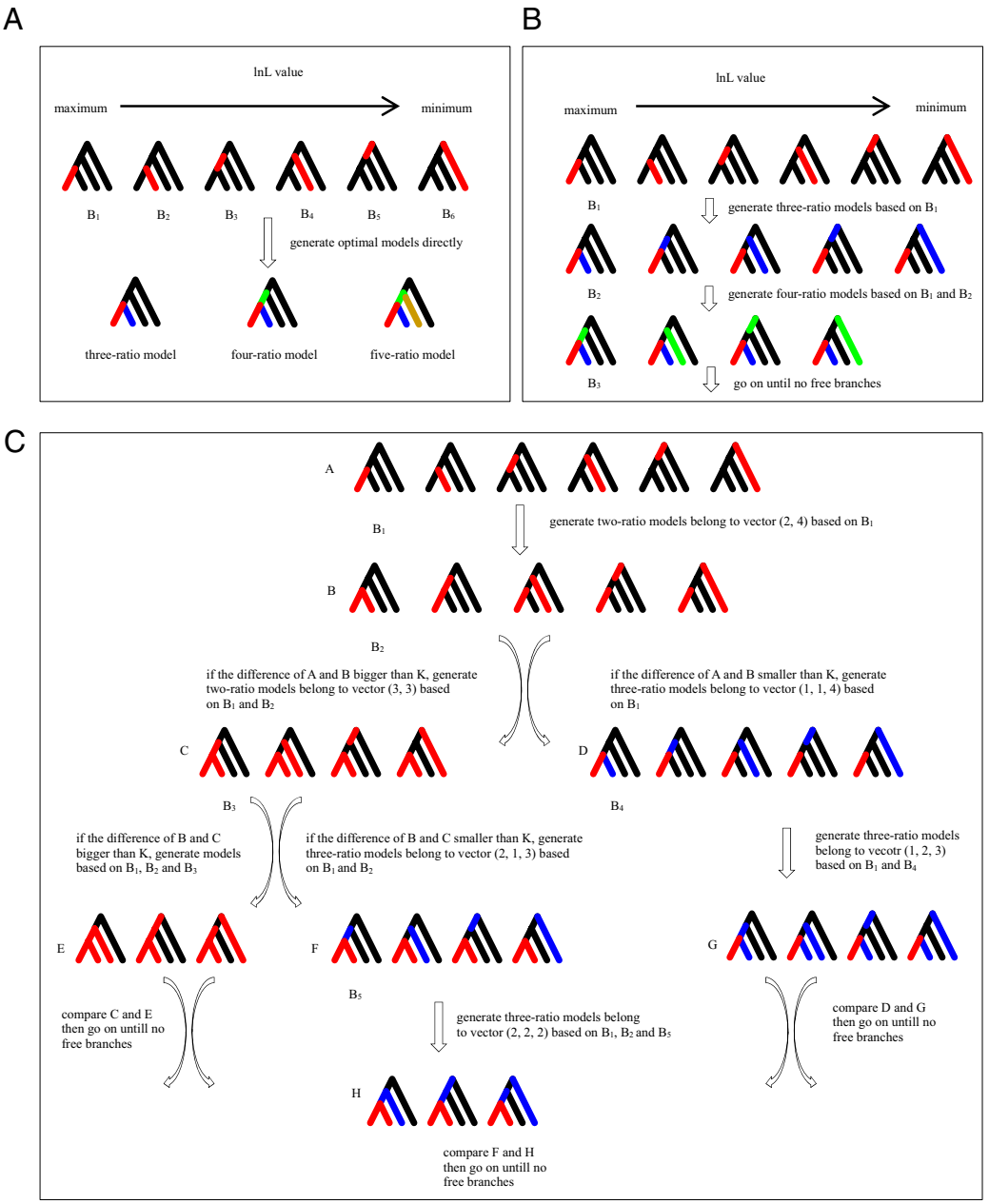


Fig. 1. Sketch of the proposed methods. (*A*) Method 1: Searching optimal models with more than two ω-parameters directly based on the sorted results of the SBTRM (2–5 ω-parameters exemplified). The different models with the same number of ω-parameters are arranged from high-likelihood to low-likelihood values. (*B*) Method 2: Searching optimal models with ω-parameters until there are no free branches, based on the maximum-likelihood value model from the last round. The different models with the same number of ω-parameters are also arranged from high-likelihood to low-likelihood values. (*C*) Method 3: Searching optimal models by iteration. In (*A*–*C*), one color stands for one ω-parameter.

model having the maximum lnL value in all $n - 1$ three-ratio models is recorded as B_2. This process is reiterated until all $n - 1$ ω's are calculated. In total, (n + 1)*n/2 models are generated and calculated; $n - 2$ optimal models of a variety of ratios are obtained and can be compared with each other, including the ORM and FRM, by LRT to determine the final optimal models.

Method 3. This is a modification of method 2 (Fig. 1*C*), to consider general cases of one ratio with more than one branch. First, similar to method 2, all the SBTRMs belonging to the configuration $(1, n - 1)$ are calculated in this step and the optimal model of SBTRMs (assumed to be A) is determined. This optimal model has only one branch B_1, which is labeled as ω_1. Then, in the second step, other $n - 1$ two-ratio models are generated, which have another branch labeled as ω_1 in addition to B_1; these models belong to the configuration $(2, n - 2)$. After calculation, the optimal model is found (assumed to be B). If the difference in the lnL values between A and B is greater than k (k is a threshold that can be defined by the user to decide if one model is better than another model with the same degree of freedom when they have different branches with same ratio, the default k = 0.5), the models belonging to configuration $(3, n - 3)$ are generated and calculated and the optimal model C is compared with B. Such iterations continue until the difference between the two optimal models is less than k. Clearly, the optimal model obtained from the penultimate iteration will become the final optimal two-ratio model. Note that the threshold value of k will determine the number of iterations; the more iterations calculated, the more the branch would be labeled with the same ω and the fewer would be

the number of large cycles that are needed. In the end, there will be no more than $n - 2$ optimal models of a variety of ratios obtained, and these can be compared with each other, including ORM and FRM, by LRT to find the final optimal models.

Evaluation of the Three Methods. To evaluate the three methods, we tested them using a dataset that has been tested in all possible ratio models of maximum-likelihood analysis. We first analyzed the datasets of the seven lysozyme sequences of primate species, which were used as an example for the maximum-likelihood analysis (5). We then randomly sampled the 50 previous studies (14–53) that used the branch model and reanalyzed their data using our methods. These studies covered a wide spectrum of phylogenetic breadth, ranging from 6 to 62 sequences, including both orthologous and paralogous groups (data in *SI Appendix*).

To describe the analysis of the lysozyme sequences in detail (*Methods*), we showed the results from the analysis of only one dataset of six sequences (the other six datasets of six sequences from each of the original seven sequences are summarized in the data in *SI Appendix*). The phylogeny of this dataset is shown in Fig. 2 (the remaining six are shown in *SI Appendix*, Fig. 2). The best models presented here means that the models have a maximum lnL value among a variety of ω-parameters, whereas the final best model is the model considered to be the best compared by the LRT among several best ones. The lnL values of the eight best models of this dataset are listed in Table 2 with several optimal models of the three methods and two hypothesis models as well (results from the other six datasets are shown in *SI Appendix*, Table 1). The eight best models with ORM and FRM were compared with each other by the LRT, and the best two-ratio model was considered to be the final best model, with an lnL value of −843.25. In the same way, all the optimal models of the three methods were compared with each other, and the lnL value of the final optimal models are −844.99 for methods 1 and 2 and −842.09 for method 3; the *P* values are shown in *SI Appendix*, Table 2. The final best model and final optimal models according to *SI Appendix*, Table 2 are evident and are marked in bold in Table 2. These results show that our simple methods obtained results very close to the results from a complete comparison.

The estimates of the substitution rate from the final best model, final optimal models, and hypothesis models are listed in Table 3. It is obvious that all these models, except the final optimal model of methods 1 and 2, suggest that positive selection operates on some lineages (K_a/K_s = 4.235–4.466), whereas the final optimal model of methods 1 and 2 indicates neutral evolution in most lineages (K_a/K_s = 1.075) and very strong purifying selection on Cja_marmoset branch (K_a/K_s = 0.0001). The results of final optimal models of methods 1 and 2 may well be wrong, but the final best model is not significantly better than the final optimal models of three methods (using df = 1) by the LRT. Conversely, the final optimal model of method 3 is significantly better than the two hypothesis models in the original computation (5) (P = 0.045 and P = 0.045, df = 1) and also significantly better than the final optimal model of methods 1 and 2 (P = 0.016, df = 1).

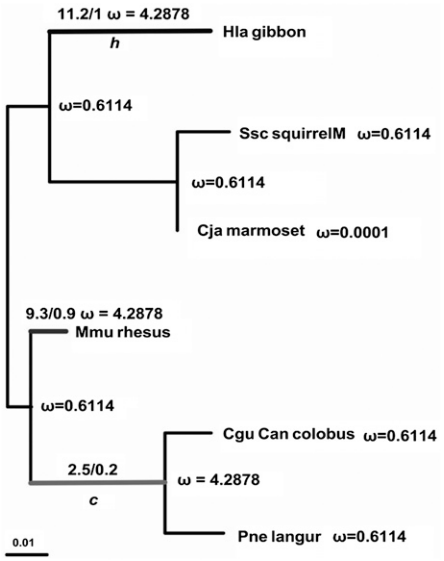


Fig. 2. Phylogeny of six lysozyme sequences, with the lineage *h* under positive selection and lineage *c* having a greater ω-value than the background in research (5). The branch length is estimated by the final optimal model of method 3; the number of nonsynonymous and synonymous sites and ω-parameters are labeled along the lineage.

In addition, the other six datasets all support the results presented above, indicating that the final optimal models are very close to the final best model in all seven datasets. Only once, in dataset 2, was the final best model significantly better than the final optimal models of all three methods (*SI Appendix*, Table 3). In these similar datasets, some of the final optimal models are significantly better than the hypothesis models, whereas none of the hypothesis models are significantly better than final optimal models; most of the final optimal models of the three methods in datasets 3 and 4 were significantly better than the hypothesis models (*SI Appendix*, Table 4). We calculated the seven sequences by the three methods and compared the final optimal models with the conventional hypothesis models. We reached the same conclusion that the six-sequence dataset showed.

Table 2. Maximum lnL values for various ratio models

		TRM	ThreeRM	FourRM	FiveRM	SixRM	SevenRM	EightRM	NineRM
Total model nos.		511	9,330	34,105	42,525	22,827	5,880	750	45
lnL of best models		**−843.25**	−841.74	−841.51	−841.36	−841.29	−841.29	−841.28	−841.28
lnL of hypothesis models‡		−844.10	−844.10						
Method 1	lnL values*	**−844.99**	−844.15	−842.66	−842.39	−841.78	−841.77	−841.69	−841.64
	Rank†	35	779	480	1,166	350	353	127	18
Method 2	lnL values	**−844.99**	−844.06	−842.66	−841.98	−841.78	−841.61	−841.35	−841.29
	Rank	35	643	480	412	350	150	18	4
Method 3	lnL values	−844.99	**−842.09**	−841.79	−841.74	−841.47	−841.42	−841.41	—
	Rank	35	4	20	94	40	37	30	

lnL value of the ORM is −847.33 and that of the FRM is −841.28. TRM, two-ratio model; ThreeRM, three-ratio model; FourRM, four-ratio model; FiveRM, five-ratio model; Six RM, six-ratio model; SevenRM, seven-ratio model; EightRM, eight-ratio model; NineRM, nine-ratio model.

*Number in bold is the lnL value for the final optimal (best) model of each method compared by the LRT (the *P* value is shown in *SI Appendix*, Table 2).

†Number in the Rank row indicates the relative position of the lnL value in all models. For example, the ThreeRM for method 3 has the lnL value −842.09, which is ranked in the fourth position from the highest one, −841.74.

‡The lnL values of the two hypothesis models are −844.097468 for TRM and −844.096995 for ThreeRM, both rounded to −844.10.

Our finding that most final optimal models detected by our methods are significantly better than the conventional hypothesis models was further confirmed by our subsequent studies of 50 gene families. We collected the sequences from these gene families from 40 original studies (14–53), and we then applied our methods to analyze these data and to compare them with the previous results of conventional hypothesis models using the maximum-likelihood method. These analyses are summarized in *SI Appendix*, Table 5. We found that in gene families (or cases) 40 and 45, the lnL value of the final optimal model our method detected and that of the conventional hypothesis model were congruent with each other; in case 38, there was no difference between the final optimal model and the current hypothesis model ($P > 0.05$). However, we were surprised to see that for the vast majority of the rest 47 cases, the lnL values for the final optimal models are significantly higher than the InL values for the conventional hypothesis models ($P < 0.001$). In these cases, 22 are significant at the level $P \leq 10^{-5}$ and 8 of them even at level $P \leq 10^{-10}$. More details of the conventional hypothesis models, our optimal models, and the 50 phylogenies are provided in the data in *SI Appendix*.

Discussion

In principle, the maximum-likelihood method was proposed to find the most probable estimates, given a phylogeny of homologous sequences. It is also clear that FRM cannot guarantee a parsimonious model. It is thus expected to find the globally most probable estimate by performing an exhaustive search of the most probable model from all possible models. Such a search is often impractically time-consuming, however, because of a huge number of possible models for a tree with even a small number of sequences. The problems in calculating all possible models were raised previously (54). Our method calculated the number of all possible models for a rooted tree in full agreement with the Bell number that was used to calculate the number in an unrooted tree (54). We proposed these simplified methods to find the most probable estimates of substitution rates with the least degrees of freedom in hypothesis testing compared with the FRM. The present study highlights the finding that the optimal models obtained from the three methods described in the following text via a dynamic programming approach are extremely close to the best model obtained from the traversing algorithm. The former simple methods use a reasonably short time, whereas the latter exhaustive search is often impractical in computing time for a large dataset, such as that used in this paper.

Compared with the previous analysis of the lysozyme dataset using the conventional hypothesis models (5), our simple method 3 obtained even significantly higher likelihoods than the previous two-ratio and three-ratio hypothesis models (−842.09 vs. −844.10, $P = 0.045$; −842.09 vs. −844.10, $P = 0.045$; Table 3). The advantage of our methods is further confirmed by our large-scale case analyses of 50 previously reported gene families using the conventional hypothesis method. In these 50 cases, we found that for 47 cases (94%), our final optional models had significantly higher likelihoods than the conventional hypothesis models and that there were only 3 cases not having significantly different likelihoods (*SI Appendix*, Table 5). The most significant differences were observed in the Chalcone Synthase Genes of Dendranthema (case 6: $2\Delta l = 198.91$, df = 11, $P <$ 1e-14), the Phytochrome Gene Family in Angiosperms (case 3: $2\Delta l = 206.25$, df = 8, $P <$ 1e-14), and the recently duplicated M_γ-type MADS-box genes in Petunia (case 13: $2\Delta l = 175.71$, df = 16, $P <$ 1e-14).

The compared models in the branch model should be nested, as suggested for the LRT (55). To make a more general comparison involving the models that do not meet such a condition, we also used the Akaike's information criterion (AIC) (56) method in analyses of these 50 cases, with the AIC values of the analyzed models in the data in *SI Appendix*. Again, except for 2 cases in which the final optimal model is congruent with the conventional hypothesis model, all other final optimal models have the lowest AIC value in 48 cases, even in the case (case 38) that failed in the LRT also getting a lower AIC than the conventional hypothesis model.

In additional, in the color vision gene (SWS2, case 17), in which $2\Delta l = 34.30$, df = 6, $P = 5.90$e-006, our optimal models suggest positive selection on the lineage *Sinocyclocheilus purpureus* (fix $\omega_{purpureus} = 1$ model vs. free $\omega_{purpureus}$ model: $2\Delta l = 5.74$, df =1, $P = 0.017$), which was not detected by the previous analysis using the conventional hypothesis method. These case analyses indicate that most previous reports missed the optional models and that the conventional hypothesis method can easily miss the globally most probable model. Our methods appear to be able to detect more significant models than the conventional hypothesis method.

Although the present methods provide simplified computational procedures for the maximum-likelihood analysis, caution should be urged in using these methods. The first caveat is that, like any other phylogeny-related study, if the phylogeny tree is inaccurate or incorrect (e.g., an incorrect inference of the orthologous-paralogous relationship), the estimates of the maximum-likelihood method, which is dependent on the tree, are meaningless. The second caution is that when many models explored by our methods detected a large ω-value in some lineages, this finding may not immediately suggest positive selection, because a statistical test for its significance is needed. The model comparison as implemented by the original branch model (5) is necessary using, for example, the nested model-based LRT or AIC discussed above. Third, we note here that method 3 seems to perform better than methods 1 and 2 in detecting final optimal models using the one gene-data analysis of lysozyme. We recommended using all three methods for more genes and comparing their performance. It would be a wise practice to start from method 1 when analyzing a large dataset to gain some useful insight because of its brief computation time.

Methods

Sequence. The sequences used in calculation of all possible models to evaluate our three methods are taken from previous work (5) and can be obtained in the PAML package in the example of lysozyme. For the reanalysis of the 50 previous studies, we utilized either available sequence alignments provided

Table 3. Substitution rate values of final best model, final optimal model, and hypothesis model

	Final best model*	Final optimal model (methods 1 and 2)†	Final optimal model (method 3)‡	Hypothesis TRM§	Hypothesis ThreeRM¶
lnL values	−843.25	−844.99	−842.09	−844.10	−844.10
ω_0	0.497	1.075	0.611	0.579	0.579
ω_1	4.466	0.0001	0.0001	4.224	4.333
ω_2	—	—	4.288	—	4.112
k	5.021	4.921	5.000	5.008	5.007

TRM, two-ratio model; ThreeRM, three-ratio model.

For the following phylogeny with markers for models (#1, ω_1; #2, ω_2):

*(((Ssc_squirrelM,Cja_marmoset),Hla_gibbon#1)#1,(Mmu_rhesus#1,(Cgu_Can_colobus,Pne_langur)#1))

†(((Ssc_squirrelM,Cja_marmoset#1),Hla_gibbon),(Mmu_rhesus,(Cgu_Can_colobus,Pne_langur)))

‡(((Ssc_squirrelM,Cja_marmoset#1),Hla_gibbon#2),(Mmu_rhesus#2,(Cgu_Can_colobus,Pne_langur)#2))

§(((Ssc_squirrelM,Cja_marmoset),Hla_gibbon#1),(Mmu_rhesus,(Cgu_Can_colobus,Pne_langur)#1))

¶(((Ssc_squirrelM,Cja_marmoset),Hla_gibbon#1),(Mmu_rhesus,(Cgu_Can_colobus,Pne_langur)#2))

in the literature or regenerated sequence realignments using MEGA 4.0 (57) when the original alignments were not available.

Calculating the Entire Range of Possible Models. We generated seven datasets of six sequences from these lysozyme sequences by deleting one sequence from seven. All possible models (115,975 possible models in one dataset) of these seven datasets were generated by the traversing algorithm (*SI Appendix*) and calculated. It took almost 4 d to finish all the calculations for one dataset, and according to this, it may take 160 d to calculate all possible 4,213,597 models of the seven sequences on the server (Dawning Information Industry), which has eight AMD Opteron 2376 processors with the operation system Linux AS 5. The phylogeny used in the calculations was built by MEGA 4.0 with the neighbor-joining method (57).

Exploring Optimal Models with Three Methods. Using the phylogeny and sequence, we performed analyses using the seven datasets with six sequences and seven sequence datasets of lysozymes (these databases are available upon request). The k value of method 3 is 0.5.

ACKNOWLEDGMENTS. We thank Shiping Wang for valuable discussion and support. We are grateful to Michael Sanderson and reviewers for valuable discussions. C.Z., J.W., W.X., G.Z., and Q.Z. were supported by grants from the National Special Key Project on Functional Genomics of Major Plants and Animals and National Natural Science Foundation of China. M.L. was supported by National Institutes of Health Grant R0IGM078070-01A1 and National Science Foundation Grant (MCB 1026200) as well as by the 111 Project of China.

1. Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA* 106:6700–6705.
2. Yang Z, Nielsen R, Goldman N (2009) In defense of statistical methods for detecting positive selection. *Proc Natl Acad Sci USA* 106:E95–E96, author reply E96.
3. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.
4. Li W (1997) *Molecular Evolution* (Sinauer, Sunderland, MA).
5. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573.
6. Betrán E, Long M (2003) *Dntf-2r*, a young Drosophila retroposed gene with specific male expression under positive Darwinian selection. *Genetics* 164:977–988.
7. Jones CD, Begun DJ (2005) Parallel evolution of chimeric fusion genes. *Proc Natl Acad Sci USA* 102:11373–11378.
8. Jones CD, Custer AW, Begun DJ (2005) Origin and evolution of a chimeric fusion gene in *Drosophila subobscura, D. madeirensis and D. guanche*. *Genetics* 170:207–219.
9. Liu Q, et al. (2009) Divergence in function and expression of the NOD26-like intrinsic proteins in plants. *BMC Genomics* 10:313.
10. Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 18:1585–1592.
11. Pond SL, Frost SD (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol* 22:478–485.
12. Goldman N (1993) Statistical tests of models of DNA substitution. *J Mol Evol* 36:182–198.
13. Bellman R (1957) *Dynamic Programming* (Princeton Univ Press, Princeton). (2003) Paperback edition (Dover, New York).
14. Jiggins FM, Hurst GD, Yang Z (2002) Host-symbiont conflicts: Positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Mol Biol Evol* 19:1341–1349.
15. Marcussen T, Oxelman B, Skog A, Jakobsen KS (2010) Evolution of plant RNA polymerase IV/V genes: Evidence of subneofunctionalization of duplicated NRPD2/ NRPE2-like paralogs in Viola (*Violaceae*). *BMC Evol Biol* 10:45.
16. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917.
17. Alba R, Kelmenson PM, Cordonnier-Pratt M-M, Pratt LH (2000) The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. *Mol Biol Evol* 17:362–373.
18. Huttley GA, et al. (2000) Adaptive evolution of the tumour suppressor *BRCA1* in humans and chimpanzees. Australian Breast Cancer Family Study. *Nat Genet* 25:410–413.
19. Yang J, Huang J, Gu H, Zhong Y, Yang Z (2002) Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (*Asteraceae*). *Mol Biol Evol* 19: 1752–1759.
20. Merritt TJ, Quattro JM (2001) Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. *Genetics* 159:689–697.
21. Yang Z (2002) Inference of selection from multiple species alignments. *Curr Opin Genet Dev* 12:688–694.
22. Yang J, Gu H, Yang Z (2004) Likelihood analysis of the chalcone synthase genes suggests the role of positive selection in morning glories (*Ipomoea*). *J Mol Evol* 58:54–63.
23. Aguileta G, Bielawski JP, Yang Z (2004) Gene conversion and functional divergence in the beta-globin gene family. *J Mol Evol* 59:177–189.
24. Schein M, Yang Z, Mitchell-Olds T, Schmid KJ (2004) Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species. *Mol Biol Evol* 21:659–669.
25. Narita Y, Oda S, Takenaka O, Kageyama T (2010) Lineage-specific duplication and loss of pepsinogen genes in hominoid evolution. *J Mol Evol* 70:313–324.
26. Larmuseau MH, Huyse T, Vancampenhout K, Van Houdt JK, Volckaert FA (2010) High molecular diversity in the rhodopsin gene in closely related goby fishes: A role for visual pigments in adaptive speciation? *Mol Phylogenet Evol* 55:689–698.
27. Arora R, et al. (2007) MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8:242.
28. Bemer M, Gordon J, Weterings K, Angenent GC (2010) Divergence of recently duplicated Mγ-type MADS-box genes in *Petunia*. *Mol Biol Evol* 27:481–495.
29. Schienman JE, Holt RA, Auerbach MR, Stewart CB (2006) Duplication and divergence of 2 distinct pancreatic ribonuclease genes in leaf-eating African and Asian colobine monkeys. *Mol Biol Evol* 23:1465–1479.
30. Yu L, et al. (2010) Adaptive evolution of digestive RNASE1 genes in leaf-eating monkeys revisited: New insights from ten additional colobines. *Mol Biol Evol* 27:121–131.
31. Zhang J (2006) Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet* 38:819–823.
32. Zhang J, Zhang YP, Rosenberg HF (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* 30:411–415.
33. Viaene T, et al. (2009) *Pistillata*—Duplications as a mode for floral diversification in (Basal) asterids. *Mol Biol Evol* 26:2627–2645.
34. Li Z, Gan X, He S (2009) Distinct evolutionary patterns between two duplicated color vision genes within cyprinid fishes. *J Mol Evol* 69:346–359.
35. Zhou D, et al. (2009) Duplication and adaptive evolution of the *COR15* genes within the highly cold-tolerant *Draba* lineage (Brassicaceae). *Gene* 441:36–44.
36. Miwa H, et al. (2009) Adaptive evolution of rbcL in *Conocephalum* (Hepaticae, bryophytes). *Gene* 441:169–175.
37. Zhao H, et al. (2009) The evolution of color vision in nocturnal mammals. *Proc Natl Acad Sci USA* 106:8980–8985.
38. Weadick CJ, Chang BS (2009) Molecular evolution of the betagamma lens crystallin superfamily: Evidence for a retained ancestral function in gamma N crystallins? *Mol Biol Evol* 26:1127–1142.
39. Wang Z, et al. (2009) Adaptive evolution of 5′HoxD genes in the origin and diversification of the cetacean flipper. *Mol Biol Evol* 26:613–622.
40. Dorus S, Freeman ZN, Parker ER, Heath BD, Karr TL (2008) Recent origins of sperm genes in *Drosophila*. *Mol Biol Evol* 25:2157–2166.
41. Schulenburg H, Boehnisch C (2008) Diversification and adaptive sequence evolution of *Caenorhabditis* lysozymes (Nematoda: Rhabditidae). *BMC Evol Biol* 8:114.
42. Zhang L (2008) Adaptive evolution and frequent gene conversion in the brain expressed X-linked gene family in mammals. *Biochem Genet* 46:293–311.
43. Zhang W, et al. (2008) Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). *Mol Phylogenet Evol* 47:175–195.
44. Storz JF, Hoffmann FG, Opazo JC, Moriyama H (2008) Adaptive functional divergence among triplicated alpha-globin genes in rodents. *Genetics* 178:1623–1638.
45. Muggia L, Schmitt I, Grube M (2008) Purifying selection is a prevailing motif in the evolution of ketoacyl synthase domains of polyketide synthases from lichenized fungi. *Mycol Res* 112:277–288.
46. Padhi A, Verghese B (2007) Evidence for positive Darwinian selection on the hepcidin gene of Perciform and Pleuronectiform fishes. *Mol Divers* 11:119–130.
47. Royer B, et al. (2007) Molecular evolution of the human *SRPX2* gene that causes brain disorders of the Rolandic and Sylvian speech areas. *BMC Genet* 8:72.
48. Ding K, McDonough SJ, Kullo IJ (2007) Evidence for positive selection in the C-terminal domain of the cholesterol metabolism gene *PCSK9* based on phylogenetic analysis in 14 primate species. *PLoS ONE* 2:e1098.
49. Hahn Y, Jeong S, Lee B (2007) Inactivation of *MOXD2* and *S100A15A* by exon deletion during human evolution. *Mol Biol Evol* 24:2203–2212.
50. Padhi A, Verghese B, Otta SK, Varghese B, Ramu K (2007) Adaptive evolution after duplication of penaeidin antimicrobial peptides. *Fish Shellfish Immunol* 23:553–566.
51. Zhang Q, et al. (2007) Rapid evolution, genetic variations, and functional association of the human spermatogenesis-related gene *NYD-SP12*. *J Mol Evol* 65:154–161.
52. Hou ZC, Xu GY, Su Z, Yang N (2007) Purifying selection and positive selection on the myxovirus resistance gene in mammals and chickens. *Gene* 396:188–195.
53. Wang Y, et al. (2007) Isolation and characterization of a putative class E gene from *Taihangia rupestris*. *J Integr Plant Biol* 49:343–350.
54. Sanderson MJ (1998) *Estimating rate and time in molecular phylogenies: beyond the molecular clock? Plant Molecular Systematics*, eds Soltis P, Soltis D, Doyle J (Chapman & Hall, New York), 2nd Ed, pp 242–264.
55. Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol* 15:1600–1611.
56. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
57. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.

Supporting Information

Supporting Information 1

Table S1. Log likelihood values of a variety of ratio models of the remaining six datasets

Supporting Information 2

Table S2. Likelihood ratio statistics for testing a variety of ratio models

Supporting Information 3

Table S3. Comparison between final best model and final optimal models

Supporting Information 4

Table S4. Comparison between final optimal models and hypothesis models

Supporting Information 5

Table S5 Comparing Final Optimal Model of OBSM with Hypothesis model suggested in 50 cases

Supporting Information 6

Figure S1. An example of a phylogeny with 4 species (or genes). S1, S2, S3, S4 denote the homologous sequences of 4 species (or genes); 1~6 are six branches. For three-ratio branch-models, there're three types of configurations (1, 1, 4), (1, 2, 3), (2, 2, 2). For the first configuration (1, 1, 4), we can choose branch 1 to set as ω_1, and branch 2 as ω_2, and the remaining 4 branches each as ω_3. Alternatively, we can choose branch 1 as ω_1, branch 3 as ω_2, and the remaining 4 branches each as ω_3. This procedure can go on until all the topologies are considered. For the second configuration (1, 2, 3), a similar procedure assigns ω_1, ω_2, ω_3 to one branch, two branches, the remaining three branches respectively.

Supporting Information 7

Figure S2. The phylogenies generated from 6 dataset (A-F)s each with 6 lysozyme sequences chosen from the 7 lysozyme sequences of the original lysozyme dataset (5) (Figure 2 in the main text is the phylogeny from the last one of the all 7 possible datasets). Red color: the branch with ω_h; Green color the branch with ω_c.

Supporting Information 8

Four Perl scripts were used in our work. The first part is used for searching the configuration and counting all the possible model numbers; the second part is used for

exploring all possible models; the third one is to do a likelihood ratio test based on the result of method 2; and the last one is to do a likelihood ratio test based on the result of method 3.

Supporting Information 9

The example of six sequences in detail with configurations and the number of possible models of each configuration.

Supporting Information 10

supplementary data for 50 cases

Table S1 Maximum Log Likelihood Values in variety ratio models of remaining six data sets of lysozymes (5)

		TwoRM	ThreeRM	FourRM	FiveRM	SixRM	SevenRM	EightRM	NineRM
Total models		511	9330	34105	42525	22827	5880	750	45
Max_lnL		**-834.142582**	**-832.129502**	-831.864334	-831.675345	-831.626547	-831.62079	-831.618484	-831.618484
Method I	lnL	**-835.658129**	-834.595675	-833.51955	-832.210261	-831.938366	-831.883059	-831.798407	-831.726768
		30	670	1889	757	467	408	122	21
Method II	lnL	**-835.658129**	-834.142582	-833.508003	-832.817495	-831.938366	-831.734028	-831.692257	-831.624213
		30	365	1870	2454	467	99	31	3
Method III	lnL	**-834.142582**	-833.508003	-831.942625	-831.736076	-831.675345	-831.626547	-831.62079	/
		1	90	8	6	10	3	2	
Max_lnL		**-827.909168**	-826.390537	**-824.878946**	-824.654224	-824.63444	-824.629197	-824.629197	-824.629197
Method I	lnL	**-830.332163**	**-827.589533**	**-826.570755**	**-825.375631**	**-825.053261**	-824.887567	-824.75887	-824.632675
		47	142	763	508	561	461	99	8
Method II	lnL	**-830.332163**	**-827.589533**	**-826.535655**	**-825.375631**	**-824.881462**	**-824.657052**	-824.637086	-824.631779
		47	142	720	508	422	125	36	6
Method III	lnL	**-830.332163**	**-826.570755**	**-825.375631**	**-824.881462**	**-824.657052**	**-824.637086**	-824.631779	-824.629197
		47	23	56	126	52	24	10	4
Max_lnL		**-828.19381**	-826.908267	-825.613278	-825.538777	-825.484354	-825.464357	-825.464357	-825.464357
Method I	lnL	**-830.16028**	**-827.516759**	**-826.299258**	**-826.147571**	-825.962545	-825.902357	-825.624112	-825.486661
		37	43	48	497	824	717	160	9
Method II	lnL	**-830.16028**	**-827.516759**	**-826.299258**	**-825.813171**	-825.62463	-825.54421	-825.485788	-825.464357
		37	43	48	221	277	140	13	1
Method III	lnL	**-829.537486**	**-827.236751**	**-826.280292**	-825.813184	-825.624662	-825.518913	-825.464357	-825.464357
		19	22	45	222	278	53	2	1
Max_lnL		**-833.577481**	-832.725657	-831.770264	-831.661651	-831.570233	-831.559893	-831.556326	-831.555043

Method I	lnL	**-835.678272**	**-833.281509**	**-832.07163**	-832.054536	-831.871882	-831.702168	-831.636489	-831.621315
		57	79	19	413	695	286	77	14
Method II	lnL	**-835.678272**	**-833.281509**	**-832.07163**	-831.904479	-831.754968	-831.674178	-831.613467	-831.565594
		57	79	19	215	350	184	39	4
Method III	lnL	**-834.184208**	-832.866043	-832.705555	-832.562737	-832.469627	-832.3931	-832.415991	-832.329388
		6	13	337	2029	2990	1611	389	34
Max_lnL		**-826.875638**	-826.265078	-825.669814	-825.631417	-825.598444	-825.590278	-825.589132	-825.589132
Method I	lnL	**-829.704697**	-828.509613	-827.351366	-826.33224	-825.816763	-825.773872	-825.637962	-825.615311
		55	1180	2774	1693	429	359	68	11
Method II	lnL	**-829.704697**	-827.855504	**-826.682121**	-826.265078	-825.897248	-825.725926	-825.630145	-825.597695
		55	448	1018	1326	775	284	47	5
Method III	lnL	**-827.292062**	-826.871296	-826.51281	-826.334256	-826.239186	-826.208553	-826.200039	/
		3	104	662	1706	2134	1334	339	
Max_lnL		**-838.341428**	-837.491141	-836.91779	-836.89203	-836.872364	-836.870807	-836.86946	-836.86946
Method I	lnL	**-841.727813**	-839.998874	-838.931056	-838.00642	-837.243223	-836.900313	-836.896359	-836.895105
		90	1374	3497	3838	755	148	89	22
Method II	lnL	**-841.727813**	**-839.963719**	**-838.716878**	-837.764604	-837.330024	-837.107644	-836.89068	-836.870807
		90	1274	2814	2259	961	443	80	4
Method III	lnL	**-838.341428**	-837.915019	-837.684536	-837.466694	-837.448306	-837.447477	/	
		1	42	314	868	1648	1207		

Note: The pair lnL value of the hypothesis models ($\omega_h = \omega_c$ and $\omega_h \neq \omega_c$) for seven data sets are (-844.097468, -844.096995), (-835.868563, -835.867062), (-829.92476, -828.725264), (-829.470066, -828.554522), (-834.184208, -833.480516), (-828.354787, -827.855504), (-840.41675, -839.963719). The lnL value of ORM for seven data sets are -847.329662, -838.915094, -833.338934, -833.239572, -837.801279, -832.565192, -844.104086. The lnL values of final optimal models find out by three methods are marked with red color. The final best models are marked with blue color.

Table S2 Likelihood Ratio Statistics for comparing various ratio models

p-value	ThreeRM	FourRM	FiveRM	SixRM	SevenRM	EightRM	NineRM		
TwoRM	0.082(1)	0.175(2)	0.285(3)	0.416(4)	0.559(5)	0.685(6)	0.787(7)	All results	Data set I
ThreeRM		0.496(1)	0.68(2)	0.824(3)	0.923(4)	0.969(5)	0.989(6)		
FourRM			0.578(1)	0.802(2)	0.93(3)	0.978(4)	0.994(5)		
FiveRM				0.715(1)	0.932(2)	0.986(3)	0.997(4)		
SixRM					0.928(1)	0.994(2)	1(3)		
SevenRM						0.95(1)	0.998(2)		
EightRM							1(1)		
TwoRM	0.194(1)	0.096(2)	0.157(3)	0.17(4)	0.265(5)	0.359(6)	0.461(7)	Method I	
ThreeRM		0.084(1)	0.172(2)	0.192(3)	0.313(4)	0.426(5)	0.543(6)		
FourRM			0.464(1)	0.417(2)	0.621(3)	0.749(4)	0.847(5)		
FiveRM				0.271(1)	0.539(2)	0.708(3)	0.83(4)		
SixRM					0.875(1)	0.914(2)	0.966(3)		
SevenRM						0.693(1)	0.885(2)		
EightRM							0.767(1)		
TwoRM	0.171(1)	0.096(2)	0.11(3)	0.17(4)	0.238(5)	0.296(6)	0.387(7)	Method II	
ThreeRM		0.094(1)	0.125(2)	0.207(3)	0.297(4)	0.368(5)	0.476(6)		
FourRM			0.244(1)	0.417(2)	0.553(3)	0.627(4)	0.741(5)		
FiveRM				0.533(1)	0.692(2)	0.743(3)	0.848(4)		
SixRM					0.555(1)	0.652(2)	0.804(3)		
SevenRM						0.477(1)	0.726(2)		
EightRM							0.714(1)		
TwoRM	0.016(1*)	0.041(2*)	0.089(3)	0.133(4)	0.21(5)	0.306(6)		Method III	
ThreeRM		0.441(1)	0.704(2)	0.745(3)	0.857(4)	0.93(5)			
FourRM			0.742(1)	0.726(2)	0.865(3)	0.945(4)			
FiveRM				0.466(1)	0.731(2)	0.886(3)			
SixRM					0.759(1)	0.945(2)			
SevenRM						0.888(1)			
TwoRM	0.045(1*)	0.102(2)	0.177(3)	0.284(4)	0.411(5)	0.538(6)	0.654(7)	All results	Data set II
ThreeRM		0.466(1)	0.635(2)	0.8(3)	0.907(4)	0.961(5)	0.985(6)		
FourRM			0.539(1)	0.788(2)	0.922(3)	0.974(4)	0.992(5)		
FiveRM				0.755(1)	0.947(2)	0.99(3)	0.998(4)		
SixRM					0.915(1)	0.992(2)	0.999(3)		
SevenRM						0.946(1)	0.998(2)		
EightRM							1(1)		
TwoRM	0.145(1)	0.118(2)	0.075(3)	0.114(4)	0.183(5)	0.259(6)	0.345(7)	Method I	
ThreeRM		0.142(1)	0.092(2)	0.15(3)	0.246(4)	0.348(5)	0.453(6)		
FourRM			0.106(1)	0.206(2)	0.351(3)	0.487(4)	0.61(5)		
FiveRM				0.461(1)	0.721(2)	0.844(3)	0.915(4)		
SixRM					0.739(1)	0.869(2)	0.935(3)		
SevenRM						0.681(1)	0.855(2)		
EightRM							0.705(1)		

TwoRM	0.082(1)	0.116(2)	0.128(3)	0.114(4)	0.165(5)	0.243(6)	0.327(7)	Method II	
ThreeRM		0.26(1)	0.266(2)	0.221(3)	0.307(4)	0.428(5)	0.539(6)		
FourRM			0.24(1)	0.208(2)	0.315(3)	0.458(4)	0.583(5)		
FiveRM				0.185(1)	0.338(2)	0.522(3)	0.665(4)		
SixRM					0.523(1)	0.782(2)	0.89(3)		
SevenRM						0.773(1)	0.896(2)		
EightRM							0.712(1)		
TwoRM	0.26(1)	0.111(2)	0.186(3)	0.294(4)	0.412(5)	0.538(6)		Method III	
ThreeRM		0.077(1)	0.17(2)	0.3(3)	0.439(4)	0.582(5)			
FourRM			0.52(1)	0.765(2)	0.889(3)	0.958(4)			
FiveRM				0.727(1)	0.896(2)	0.973(3)			
SixRM					0.755(1)	0.947(2)			
SevenRM						0.915(1)			
TwoRM	0.081(1)	0.048(2*)	0.089(3)	0.162(4)	0.255(5)	0.363(6)	0.476(7)	All results	Data set III
ThreeRM		0.082(1)	0.176(2)	0.319(3)	0.474(4)	0.62(5)	0.741(6)		
FourRM			0.503(1)	0.783(2)	0.919(3)	0.974(4)	0.992(5)		
FiveRM				0.842(1)	0.975(2)	0.997(3)	1(4)		
SixRM					0.918(1)	0.995(2)	1(3)		
SevenRM						1(1)	1(2)		
EightRM							1(1)		
TwoRM	0.019(1*)	0.023(2*)	0.019(3*)	0.032(4*)	0.054(5)	0.084(6)	0.122(7)	Method I	
ThreeRM		0.153(1)	0.109(2)	0.167(3)	0.248(4)	0.341(5)	0.433(6)		
FourRM			0.122(1)	0.219(2)	0.339(3)	0.459(4)	0.567(5)		
FiveRM				0.422(1)	0.614(2)	0.745(3)	0.829(4)		
SixRM					0.565(1)	0.745(2)	0.84(3)		
SevenRM						0.612(1)	0.775(2)		
EightRM							0.615(1)		
TwoRM	0.019(1*)	0.022(2*)	0.019(3*)	0.028(4*)	0.045(5*)	0.077(6)	0.122(7)	Method II	
ThreeRM		0.147(1)	0.109(2)	0.144(3)	0.209(4)	0.316(5)	0.433(6)		
FourRM			0.128(1)	0.191(2)	0.289(3)	0.434(4)	0.577(5)		
FiveRM				0.32(1)	0.487(2)	0.688(3)	0.829(4)		
SixRM					0.503(1)	0.783(2)	0.919(3)		
SevenRM						0.842(1)	0.975(2)		
EightRM							0.918(1)		
TwoRM	0.006(1*)	0.007(2*)	0.012(3*)	0.023(4*)	0.044(5*)	0.077(6)	0.122(7)	Method III	
ThreeRM		0.122(1)	0.185(2)	0.281(3)	0.424(4)	0.567(5)	0.692(6)		
FourRM			0.32(1)	0.487(2)	0.688(3)	0.829(4)	0.914(5)		
FiveRM				0.503(1)	0.783(2)	0.919(3)	0.973(4)		
SixRM					0.842(1)	0.975(2)	0.997(3)		
SevenRM						0.918(1)	0.992(2)		
EightRM							0.943(1)		

TwoRM	0.109(1)	0.076(2)	0.15(3)	0.247(4)	0.362(5)	0.486(6)	0.604(7)	All results	Data set IV
ThreeRM		0.108(1)	0.254(2)	0.416(3)	0.577(4)	0.717(5)	0.823(6)		
FourRM			0.699(1)	0.879(2)	0.96(3)	0.99(4)	0.998(5)		
FiveRM				0.741(1)	0.928(2)	0.985(3)	0.997(4)		
SixRM					0.841(1)	0.98(2)	0.998(3)		
SevenRM						1(1)	1(2)		
EightRM							1(1)		
TwoRM	0.021(1*)	0.021(2*)	0.045(3*)	0.078(4)	0.13(5)	0.17(6)	0.229(7)	Method I	
ThreeRM		0.119(1)	0.254(2)	0.375(3)	0.52(4)	0.581(5)	0.669(6)		
FourRM			0.582(1)	0.714(2)	0.851(3)	0.853(4)	0.898(5)		
FiveRM				0.543(1)	0.783(2)	0.79(3)	0.858(4)		
SixRM					0.729(1)	0.713(2)	0.813(3)		
SevenRM						0.456(1)	0.66(2)		
EightRM							0.6(1)		
TwoRM	0.021(1*)	0.021(2*)	0.034(3*)	0.059(4)	0.1(5)	0.155(6)	0.226(7)	Method II	
ThreeRM		0.119(1)	0.182(2)	0.286(3)	0.413(4)	0.541(5)	0.662(6)		
FourRM			0.324(1)	0.509(2)	0.68(3)	0.804(4)	0.893(5)		
FiveRM				0.539(1)	0.764(2)	0.884(3)	0.952(4)		
SixRM					0.688(1)	0.87(2)	0.956(3)		
SevenRM						0.732(1)	0.923(2)		
EightRM							0.836(1)		
TwoRM	0.032(1*)	0.038(2*)	0.059(3)	0.098(4)	0.154(5)	0.228(6)	0.32(7)	Method III	
ThreeRM		0.167(1)	0.241(2)	0.358(3)	0.488(4)	0.617(5)	0.738(6)		
FourRM			0.334(1)	0.519(2)	0.677(3)	0.803(4)	0.897(5)		
FiveRM				0.539(1)	0.745(2)	0.874(3)	0.952(4)		
SixRM					0.646(1)	0.852(2)	0.956(3)		
SevenRM						0.741(1)	0.947(2)		
EightRM							1(1)		
TwoRM	0.192(1)	0.164(2)	0.28(3)	0.404(4)	0.544(5)	0.671(6)	0.775(7)	All results	Data set V
ThreeRM		0.167(1)	0.345(2)	0.51(3)	0.675(4)	0.801(5)	0.886(6)		
FourRM			0.641(1)	0.819(2)	0.936(3)	0.98(4)	0.994(5)		
FiveRM				0.669(1)	0.903(2)	0.976(3)	0.995(4)		
SixRM					0.886(1)	0.986(2)	0.999(3)		
SevenRM						0.933(1)	0.995(2)		
EightRM							0.96(1)		
TwoRM	0.029(1)	0.027(2*)	0.064(3)	0.107(4)	0.159(5)	0.232(6)	0.323(7)	Method I	
ThreeRM		0.12(1)	0.293(2)	0.42(3)	0.532(4)	0.655(5)	0.768(6)		
FourRM			0.853(1)	0.819(2)	0.864(3)	0.929(4)	0.97(5)		
FiveRM				0.546(1)	0.703(2)	0.841(3)	0.929(4)		
SixRM					0.56(1)	0.79(2)	0.919(3)		
SevenRM						0.717(1)	0.922(2)		
EightRM							0.862(1)		
TwoRM	0.029(1*)	0.027(2*)	0.056(3)	0.097(4)	0.156(5)	0.229(6)	0.313(7)		

ThreeRM		0.12(1)	0.252(2)	0.384(3)	0.523(4)	0.648(5)	0.753(6)		
FourRM			0.563(1)	0.729(2)	0.851(3)	0.922(4)	0.962(5)		
FiveRM				0.584(1)	0.794(2)	0.901(3)	0.954(4)		
SixRM					0.688(1)	0.868(2)	0.945(3)		
SevenRM						0.727(1)	0.897(2)		
EightRM							0.757(1)		
TwoRM	0.104(1)	0.228(2)	0.356(3)	0.489(4)	0.611(5)	0.739(6)	0.813(7)	Method III	
ThreeRM		0.571(1)	0.738(2)	0.851(3)	0.918(4)	0.97(5)	0.983(6)		
FourRM			0.593(1)	0.79(2)	0.891(3)	0.965(4)	0.98(5)		
FiveRM				0.666(1)	0.844(2)	0.961(3)	0.977(4)		
SixRM					0.696(1)	0.948(2)	0.964(3)		
SevenRM						2(1)	0.938(2)		
EightRM							0.677(1)		
TwoRM	0.269(1)	0.299(2)	0.477(3)	0.635(4)	0.766(5)	0.86(6)	0.921(7)	All results	Data set VI
ThreeRM		0.275(1)	0.531(2)	0.721(3)	0.853(4)	0.93(5)	0.969(6)		
FourRM			0.782(1)	0.931(2)	0.984(3)	0.997(4)	0.999(5)		
FiveRM				0.797(1)	0.96(2)	0.994(3)	0.999(4)		
SixRM					0.898(1)	0.991(2)	0.999(3)		
SevenRM						0.962(1)	0.999(2)		
EightRM							1(1)		
TwoRM	0.122(1)	0.095(2)	0.08(3)	0.1(4)	0.164(5)	0.228(6)	0.317(7)	Method I	
ThreeRM		0.128(1)	0.113(2)	0.146(3)	0.242(4)	0.332(5)	0.447(6)		
FourRM			0.153(1)	0.216(2)	0.368(3)	0.489(4)	0.628(5)		
FiveRM				0.31(1)	0.572(2)	0.708(3)	0.838(4)		
SixRM					0.77(1)	0.836(2)	0.94(3)		
SevenRM						0.602(1)	0.853(2)		
EightRM							0.831(1)		
TwoRM	0.054(1)	0.049(2*)	0.076(3)	0.107(4)	0.159(5)	0.227(6)	0.314(7)	Method II	
ThreeRM		0.126(1)	0.204(2)	0.271(3)	0.372(4)	0.487(5)	0.607(6)		
FourRM			0.361(1)	0.456(2)	0.591(3)	0.717(4)	0.825(5)		
FiveRM				0.391(1)	0.583(2)	0.736(3)	0.855(4)		
SixRM					0.558(1)	0.766(2)	0.897(3)		
SevenRM						0.662(1)	0.88(2)		
EightRM							0.799(1)		
TwoRM	0.359(1)	0.459(2)	0.59(3)	0.716(4)	0.826(5)	0.902(6)		Method III	
ThreeRM		0.397(1)	0.584(2)	0.738(3)	0.857(4)	0.93(5)			
FourRM			0.55(1)	0.761(2)	0.894(3)	0.96(4)			
FiveRM				0.663(1)	0.882(2)	0.966(3)			
SixRM					0.805(1)	0.962(2)			
SevenRM						0.896(1)			

TwoRM	0.192(1)	0.241(2)	0.407(3)	0.568(4)	0.709(5)	0.816(6)	0.89(7)	All results	Data set VII
ThreeRM		0.284(1)	0.549(2)	0.744(3)	0.871(4)	0.941(5)	0.975(6)		
FourRM			0.82(1)	0.956(2)	0.993(3)	0.999(4)	1(5)		
FiveRM				0.843(1)	0.979(2)	0.997(3)	1(4)		
SixRM					0.955(1)	0.997(2)	1(3)		
SevenRM						0.959(1)	0.999(2)		
EightRM							1(1)		
TwoRM	0.063(1)	0.061(2)	0.059(3)	0.062(4)	0.086(5)	0.14(6)	0.208(7)	Method I	
ThreeRM		0.144(1)	0.136(2)	0.138(3)	0.185(4)	0.287(5)	0.4(6)		
FourRM			0.174(1)	0.185(2)	0.255(3)	0.397(4)	0.539(5)		
FiveRM				0.217(1)	0.331(2)	0.528(3)	0.695(4)		
SixRM					0.408(1)	0.707(2)	0.874(3)		
SevenRM						0.929(1)	0.995(2)		
EightRM							0.96(1)		
TwoRM	0.06(1)	0.049(2*)	0.048(3*)	0.066(4)	0.1(5)	0.139(6)	0.205(7)	Method II	
ThreeRM		0.114(1)	0.111(2)	0.153(3)	0.222(4)	0.292(5)	0.403(6)		
FourRM			0.168(1)	0.25(2)	0.359(3)	0.455(4)	0.595(5)		
FiveRM				0.351(1)	0.518(2)	0.626(3)	0.775(4)		
SixRM					0.505(1)	0.644(2)	0.821(3)		
SevenRM						0.51(1)	0.789(2)		
EightRM							0.842(1)		
TwoRM	0.356(1)	0.518(2)	0.626(3)	0.775(4)	0.878(5)			Method III	
ThreeRM		0.497(1)	0.639(2)	0.817(3)	0.919(4)				
FourRM			0.509(1)	0.79(2)	0.925(3)				
FiveRM				0.848(1)	0.981(2)				
SixRM					0.968(1)				

Note: The bracketed number is the degree of freedom.

Table S3. The comparing between final best model and final optimal models

Data set II		**-835.658129(II)**	**-835.658129(II)**	**-834.142582(II)**		
	-832.129502(III)	0.0079(1*)	0.0079(1*)	0.0448(1*)		
Data set III		**-827.589533(III)**	**-826.570755(IV)**	**-825.375631(V)**	**-825.053261(VI)**	
	-824.878946(IV)	0.0199(1*)				
		-827.589533(III)	**-826.535655(IV)**	**-825.375631(V)**	**-824.881462(VI)**	**-824.657052(VII)**
	-824.878946(IV)	0.0199(1*)				
		-826.570755(III)	**-825.375631(IV)**	**-824.881462(V)**	**-824.657052(VI)**	**-824.637086(VII)**
	-824.878946(IV)					
Data set IV		**-827.516759(III)**	**-826.299258(IV)**	**-826.147571(V)**		
	-828.19381(II)					
		-827.516759(III)	**-826.299258(IV)**	**-825.813171(V)**		
	-828.19381(II)					
		-827.236751(III)	**-826.280292(IV)**			
	-828.19381(II)					
Data set V		**-833.281509(III)**	**-832.07163(IV)**	**-833.281509(III)**	**-832.07163(IV)**	**-834.184208(II)**
	-833.577481(II)					
Data set VI		**-829.704697(II)**	**-826.682121(IV)**	**-827.292062(II)**		
	-826.875638(II)	0.0174(1*)				
Data set VII		**-841.727813(II)**	**-839.963719(III)**	**-838.716878(IV)**	**-838.341428(II)**	
	-838.341428(II)	0.0093(1*)				

The green color means final best model of each data set, the red color means the final optimal model of method I, and the orange color means the final optimal model of method II while the blue color means the final optimal model of method III. When final best model is significant better than final optimal models, it's marked by asterisk

Table S4. The comparing between final optimal models and hypothesis models

Data sets	lnL value of Hypothesis models	lnL value of optimal models by three methods				
Data set II		**-835.658129(2-RM)**	**-835.658129(2-RM)**	**-834.142582(2-RM)**		
	-835.868563(2-RM)					
	-835.867062(3-RM)					
Data set III		**-827.589533(3-RM)**	**-826.570755(4-RM)**	**-825.375631(5-RM)**	**-825.053261(6-RM)**	
	-829.92476(2-RM)	0.0307(1*)	0.0349(2*)	0.028(3*)	0.045(4*)	
	-828.725264(3-RM)	/	0.0379(1*)	0.0351(2*)	0.0617(3)	
Data set IV		**-827.589533(3-RM)**	**-826.535655(4-RM)**	**-825.375631(5-RM)**	**-824.881462(6-RM)**	**-824.657052(7-RM)**
	-829.92476(2-RM)	0.0307(1*)	0.0349(2*)	0.028(3*)	0.039(4*)	0.0614(5)
	-828.725264(3-RM)	/	0.0364(1*)	0.0351(2*)	0.0529(3)	0.0867(4)
		-826.570755(3-RM)	**-825.375631(4-RM)**	**-824.881462(5-RM)**	**-824.657052(6-RM)**	**-824.637086(7-RM)**
	-829.92476(2-RM)	0.0096(1*)	0.0106(2*)	0.0178(3*)	0.0323(4*)	0.0605(5)
	-828.725264(3-RM)	/	0.0096(1*)	0.0351(2*)	0.0433(3*)	0.0853(4)
		-827.516759(3-RM)	**-826.299258(4-RM)**	**-826.147571(5-RM)**		
	-829.470066(2-RM)	0.0481(1*)	0.042(2*)	0.0841(3)		
	-828.554522(3-RM)	/	0.0337(1*)	0.0901(2)		
Data set V		**-827.516759(3-RM)**	**-826.299258(4-RM)**	**-825.813171(5-RM)**		
	-829.470066(2-RM)	0.0481(1*)	0.042(2*)	0.0625(3)		
	-828.554522(3-RM)	/	0.0337(1*)	0.0645(2)		
		-827.236751(3-RM)	**-826.280292(4-RM)**			
	-829.470066(2-RM)	0.0346(1*)	0.0412(2*)			
	-828.554522(3-RM)	/	0.0329(1*)			
		-833.281509(3-RM)	**-832.07163(4-RM)**	**-833.281509(3-RM)**	**-832.07163(4-RM)**	**-834.184208(2-RM)**
	-834.184208(2-RM)					
	-833.480516(3-RM)					
Data set VI		**-829.704697(2-RM)**	**-826.682121(4-RM)**	**-827.292062(2-RM)**		

	-828.354787(2-RM) -827.855504(3-RM)					
Data set VII		**-841.727813(2-RM)**	**-839.963719(3-RM)**	**-838.716878(4-RM)**	**-838.341428(2-RM)**	
	-840.41675 (2-RM) -839.963719(3-RM)					

The green color means hypothesis models, the red color means the models of method I, orange the models of method II while blue the models of method III. Only the significant better models' p-value is shown.

Table S5 Comparing Final Optimal Model of OBSM with Hypothesis model suggested in 50 cases

Cases	Hypothesis model	Final optimal model	2Δl	df	LRTs *P*-value
1 (12)	-4106.21	-4097.60	17.23	3	6.300e-004
2 (13)	-368.96	-348.67	40.58	19	2.700e-003
3 (14, 15)	-21407.46	-21304.34	206.25	8	0
4 (14, 15)	-29760.57	-29704.59	111.97	9	0
5 (15, 16) [A]	-9341.07	-9337.50	7.13	2	2.830e-002
6 (17)	-6426.88	-6327.43	198.91	11	0
7 (18, 19)	-4086.39	-4072.63	27.52	4	1.560e-005
8 (20)	-17761.37	-17647.19	228.34	27	1.98e-014
9 (21)	-4396.93	-4366.65	60.57	12	1.780e-008
10 (22)	-2460.12	-2446.34	27.56	7	2.600e-004
11 (23)	-2657.21	-2649.35	15.72	3	1.300e-003
12 (24)	-1563.67	-1559.60	8.13	3	4.340e-002
13 (25)	-21441.24	-21353.38	175.71	16	0
14 (26-29)	-1397.58	-1387.40	20.37	4	4.200e-004
15 (30)	-2557.24	-2542.53	29.43	8	2.670e-004
16 (31)	-3019.79	-2998.27	43.04	9	2.120e-006
17 (32)	-2604.31	-2587.16	34.30	6	5.900e-006
18 (33)	-2621.56	-2600.85	41.42	9	4.200e-006
19 (34)	-2800.78	-2774.63	52.31	9	3.96e-008
20 (35)	-2043.71	-2033.04	22.34	4	1.720e-004
21 (36)	-2518.06	-2503.41	29.30	4	6.807e-006*
22 (37)	-4664.00	-4649.42	29.15	9	6.108e-004
23 (38)	-1390.02	-1386.12	7.81	2	0.0201
24 (39)	-5374.34	-5307.31	134.06	22	0
25 (39)	-2277.11	-2264.78	24.66	2	4.411e-006
26 (40)	-2451.21	-2413.22	75.98	12	2.396e-011
27 (41)	-1187.19	-1184.11	6.18	1	1.294e-002
28 (42)	-1597.54	-1585.08	24.93	2	3.861e-006
29 (43)	-1439.86	-1431.59	16.54	3	8.774e-004
30 (43)	-1345.23	-1340.19	10.08	2	6.484e-003
31 (43)	-2961.58	-2943.89	35.37	11	2.149e-004
32 (43)[A]	-1570.58	-1567.76	5.65	1	0.0175
33 (43)	-2353.63	-2342.41	22.43	4	1.643e-004
34 (43)	-1768.02	-1738.75	58.53	8	9.039e-010
35 (43)	-6917.24	-6879.80	74.89	11	1.424e-011
36 (43)	-4245.82	-4222.10	47.44	14	1.629e-005
37 (43)	-3790.66	-3774.20	32.92	8	6.380e-005
38 (43)[A]	-5350.87	-5349.17	3.38	1	6.582e-002
39 (43)[A]	-2434.20	-2430.55	7.29	1	6.948e-003
40 (43)	-2332.73	-2332.73			/

41 (44)[A]	-1310.46	-1306.08	8.76	1	3.076e-003
42 (44)	-1692.47	-1675.31	34.32	6	5.835e-006
43 (45)[A]	-2170.44	-2167.84	5.22	1	2.234e-002
44 (46)	-1374.18	-1364.61	19.14	5	1.810e-003
45 (47)	-1432.48	-1432.48			/
46 (47)[A]	-315.41	-311.94	11.02	1	8.428e-003
47 (48)	-1493.08	-1473.71	38.72	7	2.210e-006
48 (49)[A]	-3214.20	-3209.92	8.56	2	1.387e-002
49 (50)	-11734.74	-11703.17	63.14	8	1.126e-010
50 (51)	-9963.96	-9920.25	87.43	13	4.321e-013

[A]The log likelihood of final optimal models of these cases is obtained from OBSM Method III.

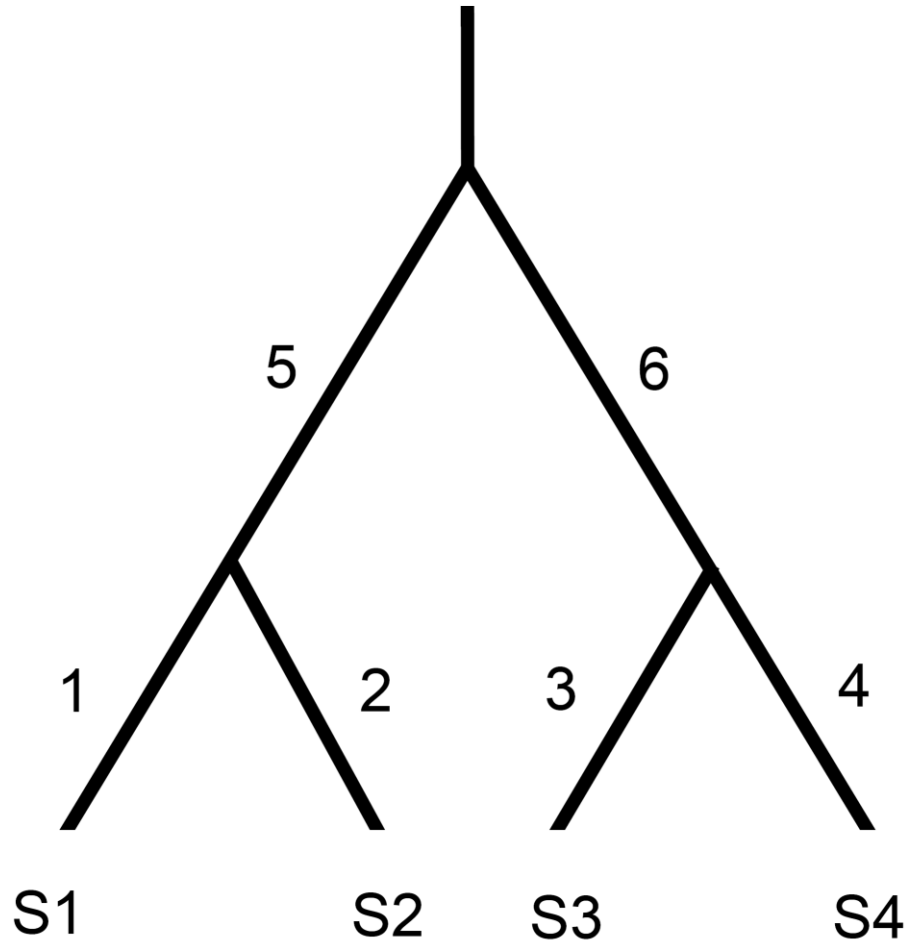


Figure S1. An example of a phylogeny with 4 species (or genes)

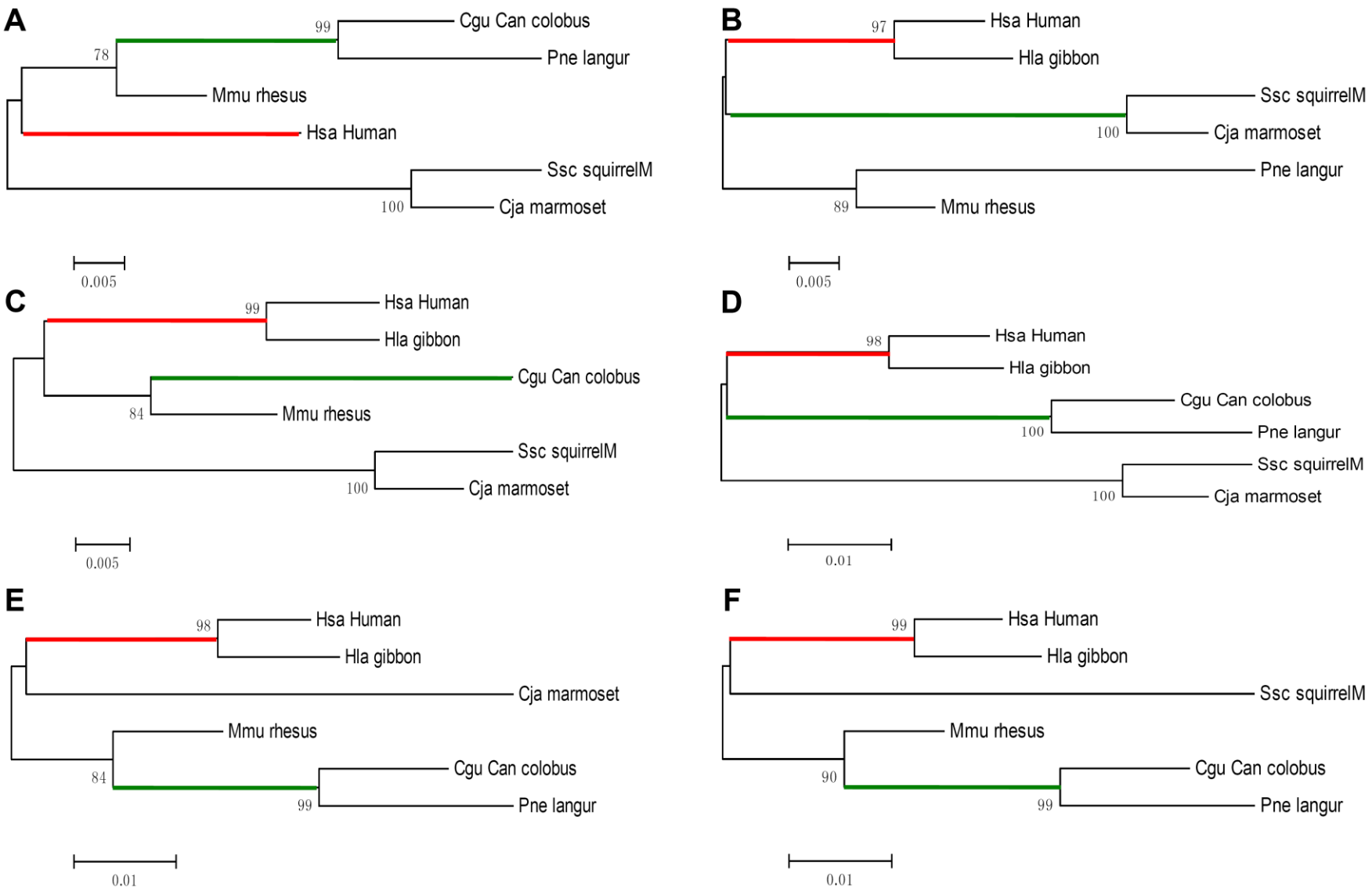


Figure S2. The phylogenies generated from 6 datasets.

Perl scripts

```
## Perl Script 1 Start##

#!/usr/bin/perl
#
#after runing, two temp files will be there, and don't delete them, one temp file is needed
by supplementary file 2.pl.

print "Please input the sequences numbers:\n";
my $sequences_numbers=<STDIN>;
our $branches_numbers=2*($sequences_numbers-1);
my $temp="";
my $j;
my $i;
my $temp_branches_numbers;
my $branches=$branches_numbers;
my @temp;
my @temp2;
my $flag;
my $temp_arr;
my %hash;
my $temp_hash;
my $down_number;
my $blank_temp;
my $temp_key;
my $total_rel;
my @arr;

open (TEMP,">temp.pl");
&maina;
close TEMP;

##############################################################################################
###################
#first part
##############################################################################################
###################
while(&check_temp_pl){}
while(&check_temp_txt){}

&get_arr;
my $rel_file="factorization_of_".$branches_numbers."_branch.txt";
open (DECOMPOSITION,">$rel_file");
&main_2;
close DECOMPOSITION;

sub get_arr
{
    undef @arr;
```

```
    open (FILE,"<temp.txt");
    while(<FILE>){
        chomp;
        while($_=~s/\s//){}
        next if $_=~/\#/;
        $arr[$i++]=$_;
    }
    close FILE;
}
sub check_temp_pl
{
    if (-e "temp.pl"){
        my $tail=`grep end_maker temp.pl`;
        chomp $tail;
        if ($tail eq "#end_maker#"){
            system "perl temp.pl >temp.txt";
            return 0;
        }else{
            print "noend";
            sleep (2);
            return 1;
        }
    }else{
        print "nopl";
        sleep (2);
        return 1;
    }
}
sub check_temp_txt
{
    if (-e "temp.txt"){
        my $tail=`grep decompose_maker temp.txt`;
        chomp $tail;
        if ($tail eq "#decompose_maker#"){
            return 0;
        }else{
            sleep (2);
            return 1;
        }
    }else{
        sleep (2);
        return 1;
    }
}

sub main_2
{
    foreach $temp_key (@arr){
        chomp $temp_key;
```

```
next if $temp_key eq "";
($blank_temp,@temp)=split/\_/,$temp_key;
undef @temp2;
$flag="F";
foreach $temp_arr (sort {$a <=> $b} @temp){
    next if $temp_arr eq 1;
    @temp2=(@temp2,$temp_arr);
}
undef %hash;
my $k;
my $i;
for ($i=0;$i<=$#temp2;$i++){
    next if $temp2[$i] eq "";
    $k=$i+1 if $i<$#temp2;
    if ($temp2[$i] eq $temp2[$k]){
        $flag="T";
    }
    $temp_hash="hash"."_".$temp2[$i];
    $hash{$temp_hash}++;
}
$down_number=$branches;
my $all_rel=1;
my $rest_rel=0;
my ($diuqi,$value);
my $blank="";
if ($flag eq "T"){
    print DECOMPOSITION join("_",@temp),"\t";
    my $key;
    foreach $key (sort keys %hash){
        if ($hash{$key}>1){
            ($diuqi,$value)=split/\_/,$key;
            for ($valuesi=0;$valuesi<$hash{$key};$valuesi++){
                $rel=&combination($down_number,$value);
                $all_rel=$all_rel*$rel;
                $down_number=$down_number-$value;
            }
            $all_rel=$all_rel/(&factorial(1,$hash{$key}));
        }else{
            ($diuqi,$value)=split/\_/,$key;
            $rel=&combination($down_number,$value);
            $all_rel=$all_rel*$rel;
            $down_number=$down_number-$value;
        }
    }
    print DECOMPOSITION "*",$all_rel,"*\n";
    $total_rel=$total_rel+$all_rel;
}else{
    my $kk;
    for (my $i=0;$i<$#temp;$i++){
```

```
                next if $temp[$i] == 0;
                next if $temp[$i] eq "";
                next if $temp[$i] eq 1;
                $kk=$i+1 if $i<=$#temp;
                $rel=&combination($down_number,$temp[$i]);
                $all_rel=$all_rel*$rel;
                $rest_rel=$rest_rel+$temp[$i];
                $down_number=$branches-$rest_rel;
                last if $down_number eq $temp[$kk];
            }
            print DECOMPOSITION join("_",@temp),"\t$all_rel\n";
            $total_rel=$total_rel+$all_rel;
        }
    }
    $total_rel=$total_rel+2;
    print DECOMPOSITION $total_rel;
}

sub print_pre
{
    ($j)=@_;
    for ($i=0;$i<$j;$i++){
        $k=($i-1);
        if ($k eq -1){
            print   TEMP $blank."for (\$arr_$i=0;\$arr_$i<$branches;\$arr_$i++){\n";
        }else{
            print   TEMP $blank."for
(\$arr_$i=\$arr_$k;\$arr_$i<$branches;\$arr_$i++){\n";
        }
        $blank=$blank."\t";
    }
    for ($i=0;$i<$j;$i++){
        print TEMP "\$arr_$i";
    }
    print TEMP "\n";
    for ($i=0;$i<$j;$i++){
        $blank=~s/\t//;
        print   TEMP $blank."}\n";
    }
}

sub combination
{
    my ($down_number,$up_number)=@_;
    my $numerator=&factorial(1,$down_number);
    my $denominator_1=&factorial(1,$up_number);
    my $denominator_2=&factorial(1,($down_number-$up_number));
    my $combination=$numerator/($denominator_1*$denominator_2);
    return $combination;
```

```
}

sub factorial()
{
    my ($f_rel,$n)=@_;
    $f_rel=$f_rel*$n;
    return $f_rel=1 if $n eq 0;
    if ($n>1)
    {
        $n--;
        &factorial($f_rel,$n)    ;
    }else{
        return $f_rel;
    }
}

sub maina
{
for ($i=2;$i<$branches_numbers;$i++){
    print TEMP <<END;
foreach \$key (sort keys \%hash){
    print "\$key\\n";
}
END
    print TEMP "undef \%hash;\n";
    print TEMP "################ $i ratio models ###############\n";
    $temp_branches_numbers=$branches_numbers;
    for ($j=1;$j<($i-1);$j++){
        &circle_pre($j,$i,($j-1));
        $temp=$temp."\t";
    }
    &print_sub($j,$i,($j-1));
    for ($j=1;$j<=($i-1);$j++){
        &circle_end;
        $temp=~s/\t//;
    }
}
&print_paixu;
}

sub print_paixu
{
print TEMP<<END;
sub paixu
{
    \@array=\@_;
    my \$ttemp;
    foreach (sort { \$a <=> \$b } \@array){
        \$ttemp=\$ttemp."_\$_";
```

```
    }
    \$hash{\$ttemp}++;
}

foreach \$key (sort keys \%hash){
    print "\$key\\n";
}
print "#decompose_maker#";
#end_maker#
END
}

sub circle_pre
{
    my ($k,$dd,$k_1)=@_;
    if ($k_1 eq 0){
        print TEMP $temp,"for
(\$arr_$k=1;\$arr_$k<".(int($temp_branches_numbers/($dd-$k+1))+1).";\$arr_$k++){\n";
    }else{
        print TEMP $temp,"for
(\$arr_$k=\$arr_$k_1;\$arr_$k<".(int($temp_branches_numbers/($dd-$k+1))+1).";\$arr_$k+
+){\n";
    }
    $temp_branches_numbers=$temp_branches_numbers-1;
}

sub circle_end{
    print TEMP $temp,"}\n";
}

sub print_sub
{
    my ($k,$dd,$k_1)=@_;
    if ($k_1 eq 0){
        print TEMP $temp,"for
(\$arr_$k=1;\$arr_$k<".(int($temp_branches_numbers/($dd-$k+1))+1).";\$arr_$k++){\n";
    }else{
        print TEMP $temp,"for
(\$arr_$k=\$arr_$k_1;\$arr_$k<".(int($temp_branches_numbers/($dd-$k+1))+1).";\$arr_$k+
+){\n";
    }
    $temp_branches_numbers=$temp_branches_numbers-1;
    print TEMP $temp,"\t\$temp=$branches_numbers";
    for (my $kk=1;$kk<=$k;$kk++){
        print TEMP "-\$arr_$kk";
    }
    print TEMP ";\n";
    print TEMP $temp,"\t",qq|\&paixu(|;
    for ($kk=1;$kk<=$k;$kk++){
```

```
        print TEMP "\$arr_$kk,";
    }
    print TEMP qq|\$temp) if \$temp>0;\n|;
}

## Perl Script 1 End##

## Perl Script 2 Start##

#!/usr/bin/perl -w
#should run after supplementary file 1.pl
#
#two parameters are needed:
#
#first, the branch numbers (it should be the same with the input of supplementary file 1.pl)
#second, the phylogeny file name
#
#after runing, two files all_labled_trees.txt and all_models.txt will be there.
#
#Chengjun Zhang 2009-10-21
#

my $kk;
my @arr;
print "Please input the sequences numbers:\n";
my $sequences_numbers=<STDIN>;
my $branches_numbers=2*($sequences_numbers-1);
my $branches=$branches_numbers;
my $file="temp.txt";
chomp $file;
open (TEMP,">temp.pl");
print TEMP "&get_arr;\n";
open (FILE,"$file");
while(<FILE>)
{
    last if $_=~/\#/;
    chomp;
    my $tempp;
    ($tempp,@line)=split/\_/,$_;
    $line=$#line+1;
    print TEMP "###############$line Ratio Model
@line####################################\n";
    $last_round=$#line-1;
    &get_circles($line[0],$last_round,1,"");
}
close FILE;

sub get_circles
{
```

```
    my ($circles,$last_round,$index_number,$blank)=@_;
    my $i;
    my $k;
    $k=1;
    for ($i=0;$i<$circles;$i++){
        print TEMP $blank,"\$l=\$i_$index_number"."_$k+1;\n" if $i ne 0;
        if ($i eq 0){
            print TEMP $blank,"for
(\$i_$index_number"."_$i=1;\$i_$index_number"."_$i<=$branches;\$i_$index_number"."_$i+
+){\n";
        }else{
            print TEMP  $blank,"for
(\$i_$index_number"."_$i=\$l;\$i_$index_number"."_$i<=$branches;\$i_$index_number"."_$
i++){\n";
        }
        $k=$i;
        $blank=$blank."\t";
    }

    if ($index_number > 1){
        $temp_number=$index_number-1;
        print  TEMP $blank,"undef \@arr_new_$index_number;\n";
        print  TEMP $blank,"\@arr_new_$index_number=\@arr_new_$temp_number;\n";
    }



    if ($index_number eq 1){
        print TEMP $blank,"my \@arr_temp=\@arr;\n";
    }

    for ($i=0;$i<$circles;$i++){
        print  TEMP qq($blank).qq(next
if ).qq(\$arr_new_$index_number).qq([\$i_$index_number).qq(_$k]).qq( eq "";\n) if
$index_number>1;
        $k=$i;
    }
    print  TEMP qq($blank).qq(my \@arr_$index_number;\n);
    print  TEMP qq($blank).qq(undef \@arr_$index_number;\n);
    for ($i=0;$i<$circles;$i++){
        print  TEMP
qq($blank).qq{\@arr_$index_number=(\@arr_$index_number,\$arr_temp[\$i_$index_number}.q
q{_$k]);\n};
        $k=$i;
    }

    if ($last_round eq 0){
        print  TEMP $blank,"print \"";
        for ($i=1;$i<=$index_number;$i++){
            print  TEMP "\@arr_$i*";
```

```
        }
        print  TEMP "\\n\";\n";
    }
    for ($i=0;$i<$circles;$i++){
        if ($last_round>0){
            if ($index_number eq 1){
                print TEMP  $blank,"delete \$arr_temp[\$i_$index_number"."_$k];\n";
            }else{
                print  TEMP $blank,"delete
\$arr_new_$index_number"."[\$i"."_$index_number"."_$k];\n";
            }
        }
        $k=$i;
    }
    if ($index_number eq 1){
        print  TEMP $blank,"undef \@arr_new_1;\n";
        print  TEMP $blank,"\@arr_new_1=\@arr_temp;\n";
    }else{
        $temp_number=$index_number+1;
        print TEMP  $blank,"undef \@arr_new_$temp_number;\n";
        print  TEMP $blank,"\@arr_new_$temp_number=\@arr_new_$index_number;\n";
    }
    if ($last_round>0){
        $last_round=$last_round-1;
        $index_number=$index_number+1;
        &get_circles($line[$index_number-1],$last_round,$index_number,$blank);
    }
    for ($i=0;$i<$circles;$i++){
        $blank=~s/\t//;
        print  TEMP $blank,"}\n";
    }
}

print  TEMP <<END;
print "#decompose_maker#";
sub get_arr
{
    my \$i;
    for (\$i=1;\$i<=$branches;\$i++)  {
        \$arr[\$i]=\$i;
    }
}
END
print TEMP "#end_maker#\n";
close TEMP;
while(&check_temp_pl){}
system "perl temp.pl >temp1.txt";
while(&check_temp1_txt){}
my $all_branch=&all_branch;
```

```
my $temp_five;
open (CORRECT,">all_models.txt");
open (TREES,">all_labled_trees.txt");
open (FILE,"<temp1.txt");
print "Please input the tree file name(make sure in one line):\n";
my $tree_file=<STDIN>;
my $flag=0;
while(<FILE>){
    last if $_=~/\#/;
    chomp;
    @line=split/\*/,$_;
    undef %hash if $flag ne $#line;
    $flag=$#line;
    $k=0;
    undef @new_arr;
    my $original_tree=&check_original_tree if $tree_file;
    my ($edit_tree,$edit_tree_lines)=&get_edit_tree($original_tree);
    $temp_five=$all_branch;
    foreach $ratio (@line){
        @menbers=split/ /,$ratio;
        @menbers=sort {$a <=> $b} @menbers;
        $new_arr[$k]=join(" ",@menbers);
        $k++;
        $maker="#".$k;
        $edit_tree=&lable_tree($maker,$edit_tree,@menbers);
    }
    my ($lost,@remain)=split/\_/,$temp_five;
    $new_arr[$#new_arr+1]=join(" ",@remain) if $remain[0] ne "";
    @new_arr=sort {$a cmp $b} @new_arr;
    $hash_pre=join("*",@new_arr);
    $hash{$hash_pre}++;
    &one_line_tree($edit_tree) if $hash{$hash_pre} eq 1;
    next if substr($hash_pre,0,1) eq "*" or substr($hash_pre,0,1) eq " ";
    print CORRECT $hash_pre,"\n" if $hash{$hash_pre} eq 1;
}
close FILE;
close CORRECT;
close TREES;

sub one_line_tree
{
    my ($tree)=@_;
    my @tree=split/\n/,$tree;
    $tree=join("",@tree);
    print TREES $tree,"\n";
}

sub lable_tree
{
```

```
    my ($maker,$edit_tree,@menbers)=@_;
    my @tree=split/\n/,$edit_tree;
    foreach (@menbers){
        $del="_".$_;
        $temp_five=~s/$del//;
        $_=$_-1;
        my $last_letter=substr($tree[$_],-1,1);
        if ($last_letter eq ","){
            $tree[$_]=~s/,/$maker,/;
        }
        if ($last_letter eq ")"){
            $tree[$_]=~s/\)/$maker\)/;
        }
    }
        $tree=join("\n",@tree);
        return $tree;
}

sub check_original_tree
{
    open(TREE,"<$tree_file") or die "can't open the tree file";
    my $tree="";
    while(<TREE>)
    {
        $tree=$tree.$_;
    }
    close TREE;
    return $tree;
}

sub get_edit_tree
{
    my ($tree)=@_;
    my ($i,$j);
    while($tree=~s/,/+\n/){$i++;}
    while($tree=~s/\)/=\n/){$j++;}
    while($tree=~s/\+/,/){}
    while($tree=~s/\=/)/){}
    $i=$i+$j;
    return ($tree,$i);
}

sub check_temp_pl
{
    if (-e "temp.pl"){
        my $tail=`grep end_maker temp.pl`;
        chomp $tail;
        if ($tail eq "#end_maker#"){
            system "perl temp.pl >temp1.txt";
```

```
            return 0;
        }else{
            print "noend";
            sleep (2);
            return 1;
        }
    }else{
        print "nop1";
        sleep (2);
        return 1;
    }
}



sub all_branch
{
    my $string="";
    for (my $i=1;$i<=$branches;$i++){
        $string=$string."_".$i;
    }
    return $string;
}

sub check_temp1_txt
{
    if (-e "temp1.txt"){
        my $tail=`grep decompose_maker temp1.txt`;
        chomp $tail;
        if ($tail eq "#decompose_maker#"){
            return 0;
        }else{
            sleep (2);
            return 1;
        }
    }else{
        sleep (2);
        return 1;
    }
}

## Perl Script 2 End##

## Perl Script 3 Start##

#!/usr/bin/perl -w
#2009-8-30 edit from YangZiheng'S PAML chi2.c
#This script is to do a likelihood ratio test based on the result of Method II.
#
```

```
print "please input the result file name of method II\n";
my $file_name=<STDIN>;
chomp $file_name;

my %hash;

open (REL,"<$file_name");
while(<REL>){
    chomp;
    @line=split/\t/,$_;
    last if $#line <3;
    $hash{$line[1]}++;
    $temp_id=$hash{$line[1]};
    $id="l".$line[1];
    $hash{$id}{$temp_id}=$line[3];
}
for ($i=2;$i<=($line[1]>9?9:$line[1]);$i++){
    #the number is determined to 9 here just because nine ratio is enough for normal
analysis.
    $max=-10000;
    for ($j=1;$j<=$hash{$i};$j++){
        $k="l".$i;
        $max=$hash{$k}{$j} if $max<$hash{$k}{$j};
    }
    @p_value=(@p_value,$max);
}
print join("\t",@p_value);
print "\n";

@input=@p_value;
    for ($i=0;$i<$#input;$i++){
    $k=$i+1;
    for ($ll=0;$ll<$i;$ll++){
        print "\t";
    }
    for ($j=$k;$j<=$#input;$j++){
        $df=$j-$i;
        $chi2=2*($input[$j]-$input[$i]);
        $prob=1-&CDFChi2($chi2,$df);
        $prob=int($prob*1000+0.5)/1000;
        if ($prob<0.05){
            print "$prob($df *)\t";
        }else{
            print "$prob($df)\t";
        }
    }
    print "\n";
}
```

```
sub CDFChi2
{
    ($x,$v)=@_;
    return (&CDFGamma($x,($v)/2.0,0.5));
}

sub CDFGamma
{
    ($x,$alpha,$beta)=@_;
    return (&IncompleteGamma(($beta)*($x),$alpha,&LnGammaFunction($alpha)));
}
sub IncompleteGamma
{
   ($x, $alpha, $ln_gamma_alpha)=@_;
   my $i;
   $p=$alpha;
   $g=$ln_gamma_alpha;
   $accurate=1e-8;
   $overflow=1e30;
   my $factor;
   $gin=0;
   $rn=0;
   $a=0;
   $b=0;
   $an=0;
   $dif=0;
   $term=0;
   #$pn[6];

   return (0) if ($x==0) ;
   return (-1) if ($x<0 or $p<=0);

   $factor=exp($p*log($x)-$x-$g);
   goto l30 if ($x>1 and $x>=$p) ;
#   /* (1) series expansion */
   $gin=1;  $term=1;  $rn=$p;
 l20:
   $rn++;
   $term*=$x/$rn;
   $gin+=$term;
     goto l20 if ($term > $accurate) ;
   $gin*=$factor/$p;
   goto l50;
 l30:
#   /* (2) continued fraction */
   $a=1-$p;
   $b=$a+$x+1;
```

```
  $term=0;
  $pn[0]=1;
  $pn[1]=$x;
  $pn[2]=$x+1;
  $pn[3]=$x*$b;
  $gin=$pn[2]/$pn[3];
l32:
  $a++;
  $b+=2;
  $term++;
  $an=$a*$term;
  for ($i=0; $i<2; $i++) {
   $pn[$i+4]=$b*$pn[$i+2]-$an*$pn[$i];
  }
  goto l35 if ($pn[5] == 0) ;
  $rn=$pn[4]/$pn[5];
  $dif=abs($gin-$rn);
  goto l34 if ($dif>$accurate);
  goto l42 if ($dif<=$accurate*$rn) ;
l34:
  $gin=$rn;
l35:
  for ($i=0; $i<4; $i++){
       $pn[$i]=$pn[$i+2];
  }
  goto l32 if (abs($pn[4]) < $overflow);
  for ($i=0; $i<4; $i++){
   $pn[$i]/=$overflow;
 }
  goto l32;
l42:
  $gin=1-$factor*$gin;

l50:
  return ($gin);
}
sub LnGammaFunction
{
     ($alpha)=@_;
  $x=$alpha;
  $f=0;
  my $z;

  if ($x<7) {
      $f=1;
      $z=$x-1;
      while(++$z<7){
        $f*=$z;
      }
```

```
            $x=$z;
            $f=-log($f);
        }
        $z = 1/($x*$x);
        return  $f + ($x-0.5)*log($x) - $x + .918938533204673+
(((-.000595238095238*$z+.000793650793651)*$z-.002777777777778)*$z+.083333333333333)/$x
;
    }

## Perl Script 3 End##

## Perl Script 4 Start##

#!/usr/bin/perl -w
#2009-8-30 edit from YangZiheng'S PAML chi2.c
#This script is to do a likelihood ratio test based on the result of Method III.
#

print "please input the result file name of method III\n";
my $file_name=<STDIN>;
chomp $file_name;

my %hash;
$id_max=1;
$last_ratio=2;
$flag=1;
open (REL,"<$file_name");
while(<REL>){
    chomp;
    if ($_=~/www/){
        $flag++ ;
        next;
    }
    @line=split/\t/,$_ if $_!~/www/;
    last if $#line <3 and $_!~/www/;
    if ($last_ratio ne $line[1]){
        $id_max=$id_max-$flag;
        $ratio_id="l".$last_ratio;
        for ($i=0;$i<$id_max;$i++){
            $temp_id=$hash{$last_ratio}-$i;
            #print "$temp_id\t";
            #print "$ratio_id\twhy\t";
            #print "$hash{$ratio_id}{$temp_id}\n";
            delete $hash{$ratio_id}{$temp_id};
        }
        $hash{$last_ratio}-=$id_max;
        $flag=1;
        $last_ratio = $line[1];
    }
```

```
    $hash{$line[1]}++;
    $temp_id=$hash{$line[1]};
    $id="1".$line[1];
    $hash{$id}{$temp_id}=$line[3];
    $id_max=$line[0] if $line[0]>$id_max;
}
for ($i=2;$i<=($line[1]>9?9:$line[1]);$i++){
    $max=-10000;
    for ($j=1;$j<=$hash{$i};$j++){
        $k="1".$i;
        $max=$hash{$k}{$j} if $max<$hash{$k}{$j};
    }
    @p_value=(@p_value,$max);
}
print join("\t",@p_value);
print "\n";

@input=@p_value;
    for ($i=0;$i<$#input;$i++){
    $k=$i+1;
    for ($ll=0;$ll<$i;$ll++){
        print "\t";
    }
    for ($j=$k;$j<=$#input;$j++){
        $df=$j-$i;
        $chi2=2*($input[$j]-$input[$i]);
        $prob=1-&CDFChi2($chi2,$df);
        $prob=int($prob*1000+0.5)/1000;
        if ($prob<0.05){
            print "$prob($df *)\t";
        }else{
            print "$prob($df)\t";
        }
    }
    print "\n";
}


    sub CDFChi2
    {
        ($x,$v)=@_;
        return (&CDFGamma($x,($v)/2.0,0.5));
    }

    sub CDFGamma
    {
        ($x,$alpha,$beta)=@_;
        return (&IncompleteGamma(($beta)*($x),$alpha,&LnGammaFunction($alpha)));
    }
```

```
sub IncompleteGamma
{
   ($x, $alpha, $ln_gamma_alpha)=@_;
   my $i;
   $p=$alpha;
   $g=$ln_gamma_alpha;
   $accurate=1e-8;
   $overflow=1e30;
   my $factor;
   $gin=0;
   $rn=0;
   $a=0;
   $b=0;
   $an=0;
   $dif=0;
   $term=0;
   #$pn[6];

   return (0) if ($x==0) ;
   return (-1) if ($x<0 or $p<=0);

   $factor=exp($p*log($x)-$x-$g);
   goto l30 if ($x>1 and $x>=$p) ;
#   /* (1) series expansion */
   $gin=1;  $term=1;  $rn=$p;
 l20:
   $rn++;
   $term*=$x/$rn;
   $gin+=$term;
     goto l20 if ($term > $accurate) ;
   $gin*=$factor/$p;
   goto l50;
 l30:
#   /* (2) continued fraction */
   $a=1-$p;
   $b=$a+$x+1;
   $term=0;
   $pn[0]=1;
   $pn[1]=$x;
   $pn[2]=$x+1;
   $pn[3]=$x*$b;
   $gin=$pn[2]/$pn[3];
 l32:
   $a++;
   $b+=2;
   $term++;
   $an=$a*$term;
   for ($i=0; $i<2; $i++) {
    $pn[$i+4]=$b*$pn[$i+2]-$an*$pn[$i];
```

```
        }
        goto l35 if ($pn[5] == 0) ;
        $rn=$pn[4]/$pn[5];
        $dif=abs($gin-$rn);
        goto l34 if ($dif>$accurate);
        goto l42 if ($dif<=$accurate*$rn) ;
      l34:
        $gin=$rn;
      l35:
        for ($i=0; $i<4; $i++){
             $pn[$i]=$pn[$i+2];
        }
        goto l32 if (abs($pn[4]) < $overflow);
        for ($i=0; $i<4; $i++){
         $pn[$i]/=$overflow;
       }
        goto l32;
      l42:
        $gin=1-$factor*$gin;

      l50:
        return ($gin);
    }
    sub LnGammaFunction
    {
          ($alpha)=@_;
        $x=$alpha;
        $f=0;
        my $z;

        if ($x<7) {
            $f=1;
            $z=$x-1;
            while(++$z<7){
              $f*=$z;
            }
            $x=$z;
            $f=-log($f);
        }
        $z = 1/($x*$x);
        return  $f + ($x-0.5)*log($x) - $x + .918938533204673+
(((-.000595238095238*$z+.000793650793651)*$z-.002777777777778)*$z+.083333333333333)/$x
;
    }

## Perl Script 4 End##
```

The example configuration of six sequences

two ratio configuration

1_9 *10*

2_8 *45*

3_7 *120*

4_6 *210*

5_5 *126*

three ratio configuration

1_1_8 *45*

1_2_7 *360*

1_3_6 *840*

1_4_5 *1260*

2_2_6 *630*

2_3_5 *2520*

2_4_4 *1575*

3_3_4 *2100*

four ratio configuration

1_1_1_7 *120*

1_1_2_6 *1260*

1_1_3_5 *2520*

1_1_4_4 *1575*

1_2_2_5 *3780*

1_2_3_4 *12600*

1_3_3_3 *2800*

2_2_2_4 *3150*

2_2_3_3 *6300*

five ratio configuration

1_1_1_1_6 *210*

1_1_1_2_5 *2520*

1_1_1_3_4 *4200*

1_1_2_2_4 *9450*

1_1_2_3_3 *12600*

1_2_2_2_3 *12600*

2_2_2_2_2 *945*

six ratio configuration

1_1_1_1_1_5 *252*

1_1_1_1_2_4 *3150*

1_1_1_1_3_3 *2100*

1_1_1_2_2_3 *12600*

1_1_2_2_2_2 *4725*

seven ratio configuration

1_1_1_1_1_1_4 *210*

1_1_1_1_1_2_3 *2520*

1_1_1_1_2_2_2 *3150*

eight ratio configuration

1_1_1_1_1_1_1_3 *120*

1_1_1_1_1_1_2_2 *630*

nine ratio configuration

1_1_1_1_1_1_1_1_2 *45*

The total number of possible models is 115975 (including one-ratio model and free-ratio model).

Supplementary data for 50 cases

CASE 1: *wsp* genes from *Wolbachia*

Sequences: ftp://ftp.ebi.ac.uk/pub/databases/embl/align/ALIGN_000201
Alignment: Not change
[a]Phylogeny: Build by Clustalx1.83

[b]Previous study:
ORM:
lnL= -4122.011213 ω_0=0.2616 κ=5.84225 AIC=2p-2lnL=2p+8244.022426
TwoRM:
lnL= -4106.214548 ω_0=0.3303 ω_1= 0.1099 (clade in green color) κ=5.94271
AIC=2p+2+8212.429096=2p+8214.429096
FRM:
lnL= -4044.313860 AIC=2p+2*120+8088.62772=8328.62772

[c]OBSM Method I

TwoRM:
lnL= -4113.270031 ω_0=0.25081 ω_1= 999.0000 κ=5.84608
AIC=2p+2+8226.540062=2p+8228.540062
ThreeRM:
lnL= -4106.101581 ω_0=0.26773 ω_1=999.00000 ω_2=0.04649 κ=5.88205
AIC=2p+2*2+8212.203162=2p+8216.203162
FourRM:
lnL= -4102.095931 ω_0=0.28110 ω_1=999.00000 ω_2=0.05263 ω_3=0.05459 κ=5.92788
AIC=2p+2*3+8204.191862=2p+8210.191862
FiveRM:
lnL= -4097.601521 ω_0=0.29273 ω_1=0.05186 ω_2=999.00000 ω_3=0.05648 ω_4=0.05283 κ=5.93956
AIC=2p+2*4+8195.203042=2p+ 8203.203042

Compare:

LRT:
FourRM (OBSM) vs TwoRM (Hypothesis)
Df=2 2Δl=8.237234 p-value=0.01627
FiveRM (OBSM) vs TwoRM (Hypothesis)
Df=3 2Δl=17.226054 p-value=0.00063
FRM vs FiveRM
Df=116 2Δl=106.575322 p-value=0.7231
AIC: the FiveRM is best model

a: the phylogeny may be a little different with phylogeny used in previous study
b: the TwoRM is consider to be the good model in previous study
c: only SBTRMs (single branch two ratio models) and 3-rm, 4-rm, 5-rm are calculate along the method 1(too many branches to calculate and this three models are good enough to prove our method).

CASE 2: NRPD2/NRPE2-like Gene Family
Sequences: http://www.biomedcentral.com/content/supplementary/1471-2148-10-45-S4.TXT
Alignment: Not change
Phylogeny: is congruent with Fig .5 in original paper

Previous study:
ORM:
lnL= -375.434947 ω_0= 0.30837 κ= 2.52005
AIC=2p+750.869894
TwoRM (a): the branch with red arrow is ω_1 (one branch)
lnL=-372.101882 ω_0=0.2672 ω_1=999.0000 κ=2.52011
AIC=2p+2+744.203764=2p+746.203764
TwoRM (b): the branches with blue arrow are all ω_1 (three branches)
lnL=-368.960632 ω_0=0.2311 ω_1=999.0000 κ=2.53092
AIC=2p+2+737.921264=2p+739.921264
FRM:
lnL=-341.184883
AIC=2p+2*59+682.369766=2p+800.369766

[a]OBSM Method I:
TwoRM: same with TwoRM (a)
lnL=-372.101882 ω_0=0.2672 ω_1=999.0000 κ=2.52011
AIC=2p+2+744.203764=2p+746.203764
21-RM(final optimal model):
lnL=-348.670597 ω_0=0.1226 ω_1=999.0000 κ=2.66455
$\omega_{2\text{-}20}$ ~ (999.00000 999.00000 999.00000 0.00010 0.00010 0.00010 0.00010 0.00010 0.06952 999.00000 999.00000 999.00000 999.00000 999.00000 999.00000 239.96815 999.00000 999.00000 999.00000)
AIC=2p+2*20+697.341194=2p+737.341194
Compare:

LRT:
TwoRM (b) vs 21-RM
Df=19 2Δl=40.58007 p-value=0.0027
FRM vs 21-RM
Df=39 2Δl=14.971428 p-value=0.9998
AIC: the 21-RM is the best

a: more optimal models significant better than ORM is not shown.

CASE 3: Phytochrome Gene Family BDE
Sequences: according the Genbank id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 3B

[b]Previous study:
ORM:
lnL=-21407.461562 ω_0=0.0851 κ=1.96981
AIC=2p+42814.923124
TwoRM: the branch in bord is ω_1(one branch)
lnL=-21407.425858 ω_0=0.0853 ω_1=0.0789 κ=1.97135
AIC=2p+2+42814.851716=2p+42816.851716
FRM
LnL=-21302.383260
AIC=2p+2*18+42604.76652=2p+42640.76652

[c]OBSM Method I:
TwoRM:
lnL=-21376.483344 ω_0=0.0931 ω_1=0.0058 (branch in red)κ=2.01938
AIC=2p+2+42752.966688=2p+42754.966688
ThreeRM:
lnL=-21349.096363 ω_0=0.10388 ω_1=0.00574 ω_2=0.01537 κ=2.01938
AIC=2p+2*2+42698.192726=2p+42702.192726
NineRM(final optimal model):
lnL=-21304.336157 ω_0=0.08317 ω_1=0.00533 ω_2=0.01537 κ=2.05211
$\omega_{3\text{-}8}$~(0.04479 0.02142 0.19176 0.27542 0.11115 0.14108 0.24123)
AIC=2p+2*8+42608.672314=2p+42624.672314
Compare:
LRT:
Obviously, optimal models of OBSM method I are all significant better than ORM
Df=1 2Δl=61.956436 p-value<0.001
Df=2 2Δl=116.730398 p-value<0.001
Df=8 2Δl=206.25081 p-value<0.001
FRM vs NineRM
Df=10 2Δl=3.905794 p-value= 0.951
AIC: the NineRM is the best model

b: The two-ratios model gave about the same fit to the data as the one-ratio model (sentence in original paper).
c: more optimal models significant better than ORM is not shown.

CASE 4: Phytochrome Gene Family ACF
Sequences: according the Genbank id in paper Zea Mays (AY260865)
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 3A

Previous study:
ORM:
lnL=-29762.994812 ω_0=0.0883 κ=2.02011
AIC=2p+59525.989624
TwoRM: the branch in bord is ω_1(one branch)
lnL=-29760.573826 ω_0=0.0889 ω_1=0.0159 κ=2.04276
AIC=2p+2+59521.147652=2p+59523.147652
FRM：
lnL=-29695.459319
AIC=2p+2*26+59390.918638=2p+59442.918638
[b]OBSM Method I:
TwoRM:
lnL=-29741.542874 ω_0=0.08267 ω_1=0.28210 (branch in red) κ=2.02372
AIC=2p+2+59483.085748=2p+59485.085748
ThreeRM:
lnL=-29731.152707 ω_0=0.08641 ω_1=0.27317 ω_2=0.02539 κ=2.03679
AIC=2p+2*2+59462.305414=2p+59466.305414
11-RM(final optimal model):
lnL=-29704.587670 ω_0=0.09360 ω_1=0.26076 ω_2=0.01537 κ=2.06294 $\omega_{3\text{-}8}$~(0.06238 0.00782 0.05597 0.24491 0.06556 0.12880 0.16551 0.00385 0.05778)
AIC=2p+2*10+59409.17534=2p+59429.17534

Compare:
LRT:
Obviously, optimal models of OBSM method I are all significant better than TwoRM in original study
Df=1 2Δl=38.061904 p-value<0.001
Df=1 2Δl=58.842226 p-value<0.001
Df=9 2Δl=111.972312 p-value<0.001
FRM vs 11-RM
Df=16 2Δl=18.256702 p-value=0.309
AIC: the 11-RM is the best model

a: the alignment may have a little different
b: more optimal models significant better than ORM is not shown.

CASE 5: *BRCA1*
Sequences: according the Genbank id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: Tree is congruent with Fig 2

Previous study:
ORM:
lnL=-9343.871464 ω_0=0.6299 κ=4.42745
AIC=2p+18687.742928
TwoRM: assume the human and chimpanzee lineages are under positive selection
lnL=-9341.068346 ω_0=0.6125 ω_1=1.9228 κ=4.42646
AIC=2p+2+18682.136692=2p+18684.136692
FRM:
lnL=-9336.729565
AIC=2p+2*13+18673.45913=2p+18699.45913

OBSM Method I:
TwoRM: assume the human lineage is under positive selection
lnL=-9341.818380 ω_0=0.6170 ω_1=1.9122 κ=4.42674
AIC=2p+2+18683.63676=2p+18685.63676
OBSM Method II:
Same with Method I
OBSM Method III (k=0.5):
TwoRM: same with hypothesis model
lnL=-9341.068346 ω_0=0.6125 ω_1=1.9228 κ=4.42646
AIC=2p+2+18682.136692=2p+18684.136692
ThreeRM:
lnL=-9338.174830 ω_0=0.67931 ω_1=1.92326 (in bold) ω_2=0.36342(in red) κ=4.43057
AIC=2p+2*2+18676.34966=2p+18680.34966
FourRM (final optimal model):
lnL=-9337.502099 ω_0=0.70281 ω_1=1.92336 ω_2=0.36751 ω_3=0.50489 (in blue) κ=4.43149
AIC=2p+2*3+18675.004198=2p+18681.004198
Compare:
LRT:
ThreeRM of Method III vs Hypothesis TwoRM:
Df=1 2Δl=5.787032 p-value=0.01614
FourRM of Method III vs Hypothesis TwoRM:
Df=2 2Δl=7.132494 p-value=0.02826
FRM vs FourRM
Df=10 2Δl=1.545068 p-value=0.9987
AIC: the ThreeRM of Method III is the best model

a: the alignment may have a little difference

CASE 6: Chalcone Synthase Genes in Dendranthema
Sequences: according the Genbank id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: according the phylogeny with Fig 1

Previous study:
[b]Model A (ORM): $\omega_0=\omega_1=\omega_2=\omega_3$
lnL=-6445.168687 ω_0= 0.0533 κ=1.67328
AIC=2p+12890.337374
[b]Model C: $\omega_0=\omega_1=\omega_3, \omega_2$
lnL=-6427.273174 ω_0=0.0466 ω_2=0.4666 κ=1.67212
AIC=2p+2+12854.546348=2p+12856.546348
[b]Model F: $\omega_0, \omega_1, \omega_2, \omega_3$
lnL=-6426.880750 ω_0=0.0471 ω_1=177.6597 ω_2=0.5023 ω_3=0.0220 κ=1.67203
AIC=2p+2*3+12853.7615=2p+12859.7615
[b]Model H: $\omega_0=\omega_2, \omega_1, \omega_3$
lnL=-6444.974866 ω_0=0.0538 ω_1=0.57580 ω_3=0.03054 κ=1.67312
AIC=2p+2*2+12889.949732=2p+12893.949732
FRM：
lnL=-6312.561448
AIC=2p+2*40+12625.122896=2p+12705.122896
[c]OBSM Method I:
TwoRM:
lnL=-6424.527983 ω_0=0.0611 ω_{phchsa}=0.0045 (lineage marked with red arrow) κ=1.63192
AIC=2p+2+12849.055966=2p+12851.055966
ThreeRM:
lnL=-6409.685071 ω_0=0.0538 ω_{phchsa}=0.0045 ω_2=0.4195 (in red) κ=1.63300
AIC=2p+2*2+12819.370142=2p+12823.370142
15-RM(final optimal model):
lnL=-6327.428407 ω_0=0.08856 ω_{phchsa}=0.00381 ω_2=0.30785 (in red) κ=1.65631
$\omega_{3\text{-}14}$~(0.26757 0.00610 0.01269 0.00200 0.00863 0.03584 1.18773 0.00978 0.10479 0.00755 0.49616 0.09588)
AIC=2p+2*14+12654.856814=2p+12682.856814

Compare:
LRT:
Obviously, optimal models of OBSM method I are all significant better than mode F in original study
FRM vs 15-RM
Df=26 2Δl=29.733918 p-value=0.2787
AIC: the 15-RM is the best model
a: the alignment may have a little different
b: the model is congruent with Table 3 in original paper, model F is best among these models
c: more optimal models significant better than ORM is not shown.

CASE 7: Triosephosphate Isomerase
Sequences: according the Genbank id in paper, mouse NM_009415
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 1

Previous study:
ORM:
lnL=-4089.438559 ω_0=0.05236 κ=1.50940
AIC=2p+8178.877118
ThreeRM: (vs TwoRM df=2 p-value=0.0476)
lnL=-4086.394347 ω_0=0.04886 ω_B=0.51321 ω_A=0.25327 κ=1.49552
AIC=2p+2*2+8172.788694=2p+8176.788694
FRM: (vs ThreeRM df=14 2Δl=35.58811 p-value=0.0012)
lnL=-4068.600292 ω_0=0.08235 ω_B=0.09533 ω_A=383.10392 κ=1.50190
$\omega_{3\text{-}17}$~(0.00010 0.07190 0.11167 0.12394 0.05770 0.01807 999.00000 0.09457 0.03877 0.05280 0.01263 0.02210 0.04044 0.05025)
AIC=2p+2*16+8137.200584=2p+8169.200584
[b]OBSM Method I:
TwoRM:
lnL=-4083.894080 ω_0=0.0495 ω_1=999.0000 (in red) κ=1.50150
AIC=2p+2+8167.78816=2p+8169.78816
ThreeRM:
lnL=-4081.261109 ω_0=0.05455 ω_1=999.00000 ω_2=0.02006 κ=1.51481
AIC=2p+2*2+8162.522218=2p+8166.522218
7-RM(final optimal model):
lnL=-4072.634101 ω_0=0.05390 ω_1=999.00000 ω_2=0.02058 κ=1.50605 $\omega_{3\text{-}6}$=0.02192 1.83280 0.12169 0.00010
AIC=2p+2*6+8145.268202=2p+8157.268202

Compare:
LRT:
Obviously, optimal model TwoRM ThreeRM and 7-RM of OBSM method I are all significant better than ThreeRM in original study, and FRM is not significant better than final optimal model 7-RM of method I (df=10 , 2Δl=8.067618, p-value=0.6222)
AIC: the 7-RM is the best model
a: the alignment may have a little different
b: more optimal models significant better than ORM is not shown.

CASE 8: Chalcone Synthase Genes in Morning Glories (Ipomoea)
Sequences: according the Genbank id paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 1

Previous study:
ORM:
lnL=-17880.400682 ω_0=0.06646 κ=1.56773
AIC=2p+35760.801364
TwoRM:
lnL=-17878.564824 ω_0=0.06572 ω_a=0.15751 κ=1.56350
AIC=2p+2+35757.129648=35759.129648
7-RM:
lnL=-17761.365327 ω_0=0.03899 ω_a=0.19258 ω_b=0.07423 ω_c=0.09128 ω_A=0.22714
ω_B=0.03954 ω_C=0.12176 κ=1.58659
AIC=2p+2*6+35522.730654=2p+35534.730654
FRM:
lnL=-17574.394267
AIC=2p+2*87+35148.788534=2p+35322.788534
[b]OBSM Method I:

11-RM:
lnL=-17750.913944 ω_0=0.05864 ω_1=0.31287 ω_2=0.00063 κ=1.55469
$\omega_{3\text{-}10}$~(0.42278 7.87158 0.35731 0.00296 0.01615 0.01569 0.21626 0.16033)
AIC=2p+2*10+35501.827888=2p+35521.827888
17-RM:
lnL=-17713.043181 ω_0=0.06284 ω_1=0.31279 ω_2= 0.00079 κ=1.56033
$\omega_{3\text{-}16}$~(0.18931 0.78795 0.35979 0.00279 0.01517 0.01599 0.21449 0.16052 0.20426 0.01161 0.03188 0.01282 0.99054 1.12049)
AIC=2p+2*16+35426.086362=2p+35458.086362
23-RM:
lnL=-17682.501932 ω_0=0.06663 ω_1=0.31268 ω_2= 0.00011 κ=1.55667
$\omega_{3\text{-}22}$~(0.19474 0.00010 0.35606 0.00271 0.01530 0.02032 0.21387 0.16031 0.18937 0.01379 0.02993 0.01296 0.96491 1.11444 80.63478 0.02020 0.01631 44.84182 0.03889 0.31906)
AIC=2p+2*22+35365.003864=35409.003864
30-RM:
lnL=-17647.194774 ω_0=0.07302 ω_1=0.31061 ω_2=0.02090 κ=1.54646
$\omega_{3\text{-}29}$~(0.18670 0.00010 0.33513 0.00250 0.00893 0.03025 0.20192 0.16088 0.20693 0.01832 0.02034 0.01717 0.95975 1.12263 999.00000 0.02882 0.01367 999.00000 0.02850 0.31543 0.16603 0.82103 0.02126 0.26883 0.01184 0.23291 0.01319)
AIC=2p+2*29+35294.389548=2p+35352.389548
Compare:
LRT:
Obviously, The Optimal models of OBSM method I are all significant better than 7-RM in original study
Df=4 2Δl=20.902766 p-value=3.31E-04
Df=10 2Δl=96.644292 p-value=5.42E-07
Df=16 2Δl=157.72679 p-value=2.67E-10
Df=23 2Δl=228.341106 p-value=1.98E-14
FRM vs 30-RM (too many calculation, didn't finish searching)
Df=58 2Δl=145.601014 p-value=1.754e-009
AIC: the FRM is the best model
a: the alignment may have a little different
b: more optimal models significant better than ORM is not shown

CASE 9: Globin Gene Family
[a]Sequences: according the Genbank id in paper
[b]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 3

Previous study:
ORM:
lnL=-4401.552483 ω_0=0.23436 κ=2.15341
AIC=2p+8803.104966
[c]ThreeRM:
lnL=-4396.934657 ω_0=0.45619 ω_β=0.17242 ω_γ=0.21530 ω_ε=0.27402 κ=2.14474
AIC=2p+2*2+8793.869314=2p+8797.869314
FRM:
lnL=-4356.459281
AIC=2p+2*47+8712.918562=2p+8806.918562
[d]OBSM Method I:
TwoRM:
lnL= -4392.037887 ω_0=0.2241 ω_1=999.0000 κ=2.14232
AIC=2p+2+8784.075774=2p+8786.075774
ThreeRM:
lnL=-4385.218567 ω_0=0.21393 ω_1= 999.00000 ω_2=1.44661 κ=2.14455
AIC=2p+2*2+8770.437134=2p+8774.437134
15-RM(Final optimal model):
lnL=-4366.651880 ω_0=0.21374 ω_1=999.00000 ω_2=1.44624 κ=2.14788
$\omega_{3\text{-}14}$~(0.05223 0.06145 0.59826 999.00000 0.08814 0.78402 0.13002 0.40781 0.07391 3.57236 999.00000 0.00010)
AIC=2p+2*14+8733.30376=2p+8761.30376
Compare:
LRT:
Obviously, optimal models of OBSM method I are all significant better than FourRM in original study

FRM vs 15-RM
Df=33 2Δl=20.385198 p-value=0.9578
AIC: the 15-RM is the best model
a: few sequences (human beta and mouse beta) may not congruent with the original paper
b: the alignment may have a little different
c: we try 4 models, and chose the model of best lnL value among these four
d: more optimal models significant better than ORM is not shown.

CASE 10: Pollen-Specific Oleosin-Like Gene Family
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 4

[b]Previous study:
ORM:
lnL=-2460.115266 ω_0=0.67749 κ=1.64825
AIC=2p+4920.230532
FRM：
lnL=-2432.734251
AIC=2p+2*45+4865.468502=2p+4955.468502

[c]OBSM Method I:
TwoRM:
lnL=-2457.443865 ω_0=0=0.6643 ω_1=999.0000 κ=1.65263
AIC=2p+2+4914.88773=2p+4916.88773
ThreeRM:
lnL=-2454.763631 ω_0=0.6484 ω_1=999.00000 ω_2=999.00000 κ=1.65944
AIC=2p+2*2+4909.527262=2p+4913.527262
8-RM(Final optimal model):
lnL=-2446.335287 ω_0=0.72178 ω_1= 999.00000 ω_2=999.00000 κ=1.65607 $\omega_{3\text{-}7}$~(0.05638 0.00010 0.17969 0.00010 0.270210.06238)
AIC=2p+2*7+4892.670574=2p+4906.670574

Compare:
LRT:
Obviously, optimal models of OBSM method I are all significant better than ORM
Df=1 2Δl=5.342802 p-value=0.0208
Df=2 2Δl=10.70327 p-value=0.0047
Df=7 2Δl=27.559958 p-value=0.00026
FRM vs 8-RM
Df=38 2Δl=27.202072 p-value=0.903
AIC: the 8-RM is the best model
a: the alignment may have a little different
b: Comparisons of branch models revealed no signicant differences between different lineages in the gene phylogeny (sentence in original paper).
c: more optimal models significant better than ORM is not shown.

CASE 11: Pollen-Specific Oleosin-Like Gene Family
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 4

Previous study:
ORM:
lnL= -2662.201006 ω_0= 0.1945 κ= 2.24635
AIC=2p+5324.402012
Hypothesis_TwoRM:
lnL= -2657.211303 ω_0= 0.1706 ω_1= 2.5434 κ= 2.24962
AIC=2p+2+5314.422606=2p+5316.422606
FRM:
lnL=-2641.379441
AIC=2p+2*25+5282.758882=2p+5332.758882

OBSM Method I:
TwoRM: same with hypothesis TwoRM
lnL= -2657.211303 ω_0= 0.1706 ω_1= 2.5434 κ= 2.24962
AIC=2p+2+5314.422606=2p+5316.422606
ThreeRM:
lnL=-2652.975199 ω_0=0.18837 ω_1=2.40186 ω_2=0.00010 κ=2.24783
AIC=2p+2*2+5305.950398=2p+5309.950398
5-RM(Final optimal model, ωratio is labeled above the branch):
lnL= -2649.352240 ω_0= 0.21385 ω_1= 2.35867 ω_2=0.00010 κ=2.24539 $\omega_{3\text{-}4}$~(0.05109 0.00010)
AIC=2p+2*4+5298.70448=2p+5306.70448

Compare:
LRT:
Obviously, optimal models (ThreeRM and 5-RM) of OBSM method I are significant better than Hypothesis TwoRM
Df=1 2Δl=8.472208 p-value=0.0036
Df=3 2Δl=15.718126 p-value=0.0013
FRM vs 5-RM
Df=21 2Δl=15.945598 p-value= 0.7726
AIC: the 5-RM is the best model
a: the alignment may have a little different

CASE 12: Rhodopsin Gene
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2(not include outgroup)

[b]Previous study:
ORM:
lnL= -1563.668037 ω_0= 0.2655 κ= 2.81281
AIC=2p+3127.336074
FRM:
LnL=-1549.634709
AIC=2p+2*31+3099.269418=2p+3161.269418

OBSM Method I:
FourRM:
lnL=-1559.603182 ω_0=0.26495 ω_1=999.00000 $\omega_{2,3}$=0.00010 κ=2.80964
AIC=2p+2*3+3119.206364=2p+3125.206364

Compare:
LRT:
FRM vs FourRM
Df=3 2Δl=8.12971 p-value=0.0434
AIC: the FourRM is the best model .
a: the alignment may have a little different
b: no likelihood ratio test was significant in the comparison between the ω ratio of model M0 and the two-ratio model that estimates two different ω ratios (sentence in original paper).

CASE 13: M_gamma_type_MADS_Box_Genes
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2 (original paper)

Previous study:
ORM:
lnL= -21447.371961 ω_0= 0.51565 κ= 1.65870
AIC=2p+42894.743922
TwoRM:
lnL= -21441.236742 ω_0=0.47932 ω_1=0.76610 κ=1.66651
AIC=2p+2+42882.473484=2p+42884.473484
FRM:
lnL= -21330.684136
AIC=2p+2*67+42661.368272=2p+42795.368272

[b]OBSM Method I:
TwoRM:
lnL= -21412.775657 ω_0=0.54361 ω_1=0.00351 κ=1.67500
AIC=2p+2+42825.551314=2p+42827.551314
18-RM:
lnL= -21353.379969 ω_0=0.56269 ω_1=0.00010 ω_2= 0.00404 κ=1.71444
$\omega_{3\text{-}18}$~(0.04736 0.08443 0.08913 214.44949 0.00668 1.12431 0.27295 0.16102 1.40938 0.93848 0.10296 1.73498 0.15243 999.00000 999.00000 1.93258)
AIC=2p+2*17+42706.759938=2p+42740.759938

Compare:
LRT:
Obviously, optimal models of OBSM method I are all significant better than hypothesis model.
FRM vs 18-RM
Df=50 2Δl=45.391666 p-value=0.6585
AIC: 18-RM is the best model
a: the alignment result show that it may not suitable for such research (too distant to deduce its ancestral sequences).
b: more optimal models significant better than ORM is not shown.

CASE 14: Digestive_RNASE1_Genes
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 6

[b]Previous study:
ORM:
lnL= -1409.406777 ω_0= 0.66410 κ= 2.95678
AIC=2p+2818.813554
SixRM:
lnL= -1397.582193 ω_0=0.37396 ω_1=1.07877 ω_5=1.08460 $\omega_{2,3,4}$=999.0000 κ=2.96104
AIC=2p+2*5+2795.164386=2p+2805.164386
FRM:
lnL=-1379.257362
AIC=2p+2*57+2758.514724=2p+2872.514724

[c]OBSM Method I:
SixRM: (marked with red arrow and number)
lnL= -1394.828338 ω_0=0.62326 $\omega_{2,3,4}$=0.00010 $\omega_{1,5}$=999.0000 κ= 2.96773
AIC=2p+2*5+2789.656676=2p+2799.656676
TenRM:
lnL= -1387.396695 ω_0=0.75823 ω_1= 0.12173 ω_2=0.08177 κ= 2.96773
$\omega_{3\text{-}6}$=0.00010 $\omega_{7\text{-}9}$=999.0000
AIC=2p+2*9+2774.79339=2p+2792.79339

Compare:
LRT:
Obviously, optimal SixRM of OBSM method I is significant better than Hypothesis SixRM, and final optimal model TenRM is also significant better than Hypothesis SixRM:
Df=4 2Δl= 20.370996 p-value= 0.00042
FRM vs TenRM
Df=48 2Δl= 16.278666 p-value= 0.9999
AIC: the TenRM is the best model

a: the alignment may have a little different
b: Only Hypothesis SixRM (Model D in original paper) is calculate while other two models (Model B and C) is confused described.
c: more optimal models significant better than ORM is not shown.

CASE 15: Pistillata PI genes
Sequences: according the id in paper, Matrix A
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is manual edit according Fig 1 (20)

Previous study:
ORM:
lnL= -2562.634189 ω_0= 0.29816 κ= 1.35424
AIC=2p+5125.268378
[b]Hypothesis_TwoRM:
lnL= -2557.238920 ω_0= 0.3360 ω_1= 0.0750 (red branches) κ= 1.35376
AIC=2p+2+5114.47784
FRM:
lnL=-2536.246228
AIC=2p+2*29+5072.492456=2p+5130.492456

[c]OBSM Method I:
TwoRM:
lnL= -2558.902183 ω_0= 0.32180 ω_1= 0.01833 κ= 1.35176
AIC=2p+2+5117.804366=2p=5119.804366
FourRM:
lnL=-2555.807051 ω_0=0.30531 ω_1=999.00000 ω_2=0.01985 ω_3=2.56347 κ= 1.36062
AIC=2p+2*3+5111.614102=2p+5117.614102
TenRM:
lnL= -2542.526246 ω_0= 0.36921 ω_1= 0.04096 ω_2=273.72251 κ= 1.37044
$\omega_{3\text{-}9}$~(2.33347 0.80354 0.11693 0.04545 999.00000 0.14308 0.13157)
AIC=2p+2*9+5085.052492=2p+5103.052492

Compare:
LRT:
Final optimal model is significant better than Hypothesis model.
Df=2 2Δl=2.863738 p-value= /
Df=8 2Δl=29.425348 p-value=2.670e-004
FRM vs TenRM
Df=20 2Δl=12.560036 p-value=0.89545
AIC: the TenRM is the best model
a: the alignment may have a little different
b: Three hypothesis models are tried and the best one is show here.
c: more optimal models significant better than ORM is not shown.

CASE 16: SWS1_HeShunping
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2a (not include outgroup Danio_rerio)

Previous study:
ORM:
lnL= -3022.465018 ω_0= 0.43018 κ= 2.31164
AIC=2p+6044.930036
TwoRM:
lnL= -3019.786644 ω_0= 0.5369 ω_1= 0.3157 κ= 2.31152
AIC=2p+2+6039.573288=2p+6041.573288
FRM：
lnL=-2989.152966
AIC=2p+2*63+5978.305932=2p+6104.305932

[b]OBSM Method I:
TwoRM:
lnL= -3017.843433 ω_0= 0.36791 ω_1=0.93827(red arrow) κ= 2.31570
AIC=2p+2+ 6035.686866=2p+6037.686866
ThreeRM:
lnL= -3014.933243 ω_0=0.40355 ω_1= 0.93886 ω_2=0.13203 κ= 2.32292
AIC=2p+2*2+6029.866486=2p+6033.866486
11-RM:
lnL= -2998.267273 ω_0= 0.42293 ω_1= 0.93878 ω_2=0.13461 κ= 2.32208
$\omega_{3\text{-}11}$~(0.00010 0.88031 0.17243 2.74113 0.06632 0.00010 0.00010 0.00010)
AIC=2p+2*10+5996.534546=2p+6016.534546

Compare:
LRT:
Obviously, optimal models of OBSM method I are significant better than Hypothesis TwoRM
Df=1 2Δl=9.706802 p-value=0.00184
Df=9 2Δl=43.038742 p-value=2.121e-006
FRM vs 11-RM
Df=53 2Δl=18.228614 p-value= 0.9999
AIC: the 11-RM is the best model
a: the alignment may have a little different and some CDS region marked in GenBank is confused and re-predict by Fgenesh-M
b: more optimal models significant better than ORM is not shown.

CASE 17: SWS2_HeShunping
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2b (not include outgroup Danio_rerio)

Previous study:
ORM:
lnL= -2606.878710 ω_0= 0.35308 κ= 2.65541
AIC=2p+5213.75742
TwoRM:
lnL= -2604.305376 ω_A= 0.4883 ω_B= 0.2517 κ= 2.65793
AIC=2p+2+5208.610752=2p+5210.610752
FRM：
LnL=-2576.232247
AIC=2p+2*39+5152.464494=2p+5230.464494

[b]OBSM Method I:
[c]TwoRM:
lnL= -2600.116139 ω_0= 0.31892 ω_1= 999.00000 (red) κ= 2.65897
AIC=2p+2+5200.232278=2p+5202.232278
ThreeRM:
lnL= -2595.542652 ω_0=0.28778 ω_1=999.00000 ω_2=3.43585 κ=2.66214
AIC=2p+2*2+5191.085304=2p+5195.085304
EightRM:
lnL= -2587.157466 ω_0= 0.38075 ω_1= 999.00000 ω_2= 3.37710 κ= 2.66500
$\omega_{3\text{-}7}$~(0.00010 999.00000 0.14775 0.00010 0.17214)
AIC=2p+2*7+5174.314932=2p+5188.314932

Compare:
LRT:
Obviously, TwoRM ThreeRM and EightRM of OBSM method I are all significant better than TwoRM in original study

FRM vs 8-RM
Df=32 2Δl=21.850438 p-value= 0.9113
AIC: the EightRM is the best model
a: the alignment may have a little different
b: more optimal models significant better than ORM is not shown.
c: when we fix the ω_1 =1, the likelihood ratio test show the branch is significant larger than 1(p-value=0.0166)

CASE 18: COR15 gene family
[a]Sequences: according the id in paper
[b]Alignment: MEGA4 (only CDS)
Phylogeny: is built by ClustalX v1.83

[c]Previous study:
ORM:
lnL= -2625.334652 ω_0= 0.50507 κ= 1.68299
AIC=2p+5250.669304
TwoRM A (clade model):
lnL= -2625.315516 ω_0= 0.51190 ω_1= 0.48956 (clade a and b) κ= 1.68277
AIC=2p+2+5250.631032=2p+5252.631032
TwoRM B:
lnL= -2621.561591 ω_0= 0.48746 ω_1= 999.00000 κ= 1.68444
AIC=2p+2+5243.123182=2p+5245.123182
ThreeRM (clade model):
lnL= - 2622.053852 ω_0= 0.48832 ω_1= 0.95778(clade a) ω_2= 0.28658(clade b) κ= 1.68368
AIC=2p+2*2+5244.107704=2p+5248.107704
ThreeRM:
lnL= - 2623.293414 ω_0= 0.52395 ω_a= 0.69154 ω_b= 0.20211 κ= 1.68367
AIC=2p+2*2+5246.586828=2p+5250.586828
FRM:
LnL= -2588.725430
AIC=2p+2*50+5177.45086=2p+5277.45086

[d]OBSM Method I:
TwoRM:
lnL= -2621.186909 ω_0= 0.54334 ω_1= 0.13860 κ= 1.68683
AIC=2p+2+5242.373818=2p+5244.373818
ThreeRM:
lnL= -2617.640992 ω_0=0.52420 ω_1= 0.13933 ω_2= 999.00000 κ=1.68850
AIC=2p+2*2+5235.281984=2p+5239.281984
11-RM:
lnL= -2600.853632 ω_0= 0.63718 ω_1= 0.15782 ω_2= 999.00000 κ= 1.67854
$\omega_{3\text{-}10}$~(0.04557 999.00000 999.00000 2.64062 0.20135 0.22141 0.15076 0.00010)
AIC=2p+2*10+5201.707264=2p+5221.707264

Compare:
LRT:
Obviously, ThreeRM and 11-RM of OBSM method I are all significant better than TwoRM B in original study.
Df=1 2Δl=7.841198 p-value=0.0051
Df=9 2Δl=41.415918 p-value=4.201e-006
FRM vs 11-RM
Df=40 2Δl=24.256404 p-value=0.9765
AIC: the 11-RM is the best model
a: the Genbank id showed in the original paper is confused with fig 2 in the original paper.
b: the alignment may have a little different
c: since the variety of sequences and phylogeny, the values of Hypothesis models suggested in original research can't reappear closely. We calculate several hypothesis models according the original paper implied.
d: more optimal models significant better than ORM is not shown.

CASE 19: rbcL genes in Conocephalum
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2

[b]Previous study:
ORM:
lnL= -2809.960994 ω_0= 0.04252 κ= 2.26661
AIC=2p+5619.921988
TwoRM A:
lnL= -2800.778895 ω_0= 0.0381 ω_1= 999.0000 (red branch) κ= 2.26984
AIC=2p+2+5601.55779=2p+5603.55779
TwoRM B (clade model):
lnL= -2803.416120 ω_0= 0.03830 ω_1= 0.85378 (clade F) κ= 2.26967
AIC=2p+2+5606.83224=2p+5608.83224
FRM:
lnL=-2769.503035
AIC=2p+2*27+5539.00607=2p+5593.00607

[c]OBSM Method I:
TwoRM:
lnL= -2798.002656 ω_0= 0.08036 ω_1= 0.01317 (blue branch) κ= 2.27228
AIC=2p+2+5596.005312=2p+5598.005312
11-RM:
lnL= -2774.625477 ω_0=0.04157 ω_1=0.00010 ω_2= 0.02645 κ= 2.28439
$\omega_{3\text{-}10}$~(999.00000 999.00000 999.00000 0.35504 999.00000 999.00000 999.00000 999.00000)
AIC=2p+2*10+5549.250954=2p+5569.250954

Compare:
LRT:
Obviously, TwoRM and 11-RM of OBSM method I are all better than TwoRM A in original study.
Df=1 2Δl=7.841198 p-value=0.0051
Df=9 2Δl=52.306836p-value=3.955e-008
FRM vs 11-RM
Df=17 2Δl=10.244884p-value=0.8930
AIC: the 11-RM is the best model
a: the alignment may have a little different
b: We calculate several hypothesis models according the original paper implied and only two optimal model is shown.
d: more optimal models significant better than ORM is not shown.

CASE 20: M_LWS_gene
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2

[b]Previous study:
ORM:
lnL= -2044.437533 ω_0= 0.07246 κ= 5.62795
AIC=2p+4088.875066
TwoRM A: (fruit bats)
lnL= -2044.376785 ω_0= 0.07171 ω_1= 0.10955 κ= 5.63077
AIC=2p+2+4088.75357=2p+4090.75357
TwoRM B: (Yinpterochiroptera):
lnL= -2043.712372 ω_0= 0.07429 ω_1= 0.00010 κ= 5.62842
AIC=2p+2+4087.424744=2p+4089.424744
TwoRM C: (Yangochiroptera):
lnL= -2044.209877 ω_0= 0.07124 ω_1= 0.14036 κ= 5.62884
AIC=2p+2+4088.419754=2p+4090.419754
FRM:
lnL=-2022.899266
AIC=2p+2*31+4045.798532=2p+4107.798532

OBSM Method I:
TwoRM:
lnL= -2040.314104 ω_0= 0.06467 ω_1=0.35723 κ= 5.63068
AIC=2p+2+4080.628208=2p+4082.628208
ThreeRM:
lnL= -2037.013979 ω_0=0.06209 ω_1=0.35761 ω_2= 999.00000 κ= 5.63350
AIC=2p+2*2+4074.027958=2p+4078.027958
SixRM:
lnL= -2033.039125 ω_0=0.07333 ω_1=0.34682 ω_2= 999.00000 κ= 5.64782
$\omega_{3\text{-}5}$~(0.02076 0.00010 0.00010)
AIC=2p+2*5+4066.07825=2p+4076.07825

Compare:
LRT:
Obviously, TwoRM and SixRM of OBSM method I are all significant better than ORM and TwoRM B in original study.
Df=1 2Δl=6.796536 p-value=0.0091
Df=1 2Δl=13.396786 p-value=2.521e-004
Df=4 2Δl=22.341504 p-value=1.714e-004
FRM vs SixRM
Df=26 2Δl=20.279718 p-value= 0.77811
AIC: SixRM is the best model
a: the alignment may have a little different
b: We calculate several hypothesis models according the original paper implied

CASE 21: Gama N Crystallin Superfamily
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 1b

[b]Previous study:
ORM:
lnL= -2520.154970 ω_0= 0.05602 κ= 2.16197
AIC=2p+5040.30994
ThreeRM:
lnL= -2518.060965 ω_0= 0.05518 ω_1= 0.03996 ω_2= 0.08011 κ= 2.17468
AIC=2p+2*2+5036.12193=2p+5040.12193
FRM:
lnL=-2490.212643
AIC=2p+2*32+4980.425286=2p+5044.425286

[c]OBSM Method I:
TwoRM:
lnL= -2514.560860 ω_0= 0.06322 ω_1= 0.00368 κ= 2.13310
AIC=2p+2+5029.12172=2p+5031.12172
ThreeRM:
lnL= -2512.520642 ω_0= 0.06098 ω_1= 0.00389 ω_2=0.48310 κ= 2.12556
AIC=2p+2*2+5025.041284=2p+5029.041284
SevenRM:
lnL= -2503.413056 ω_0=0.06976 ω_1=0.00386 ω_2= 0.60896 κ= 2.09140
$\omega_{3\text{-}6}$~(0.00010 0.01776 0.00010 999.00000)
AIC=2p+2*6+5006.826112=2p+5018.826112

Compare:
LRT:
Obviously, TwoRM, ThreeRM and 11-RM of OBSM method I are all significant better than ThreeRM (clade model) in original study.
Df=4 2Δl=29.295818 p-value=6.807e-006
FRM vs SevenRM
Df=26 2Δl=26.400826 p-value=0.4412
AIC: the SevenRM is the best model
a: the alignment may have a little different. This is a typical case that the sequences used in previous study may have some change (the CDS region) in the later version in GenBank.
b: We calculate several hypothesis models according the original paper implied and only ThreeRM (clade model) is shown.
c: more optimal models significant better than ORM is not shown.

CASE 22: HoxD Genes
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2

[b]Previous study:
ORM:
lnL= -4670.009057 ω_0= 0.10712 κ= 4.57654
AIC=2p+9340.018114
Hypothesis model D:
lnL= -4664.528087 ω_0= 0.10328 ω_c=ω_{tm}= 1.10715 κ= 4.58076
AIC=2p+2+9329.056174=2p+9331.056174
Hypothesis model E:
lnL= -4663.998915 ω_0= 0.10332 ω_c= 0.77848 ω_{tm}= 347.95623 κ= 4.58077
AIC=2p+2*2+9327.99783=2p+9331.99783
FRM:
lnL=-4633.711096
AIC=2p+2*67+9267.422192=2p+9401.422192

[c]OBSM Method I:
TwoRM:
lnL= -4666.625301 ω_0= 0.10437 ω_c= 0.77759 κ= 4.57886
AIC=2p+2+9333.250602=2p+9335.250602
ThreeRM:
lnL= -4663.476101 ω_0=0.10078 ω_c= 0.78253 ω_2=999.00000 (red arrow) κ= 4.56168
AIC=2p+2*2+9326.952202=9330.952202
12-RM:
lnL= -4649.422772 ω_0=0.09676 ω_1=0.78520 ω_2= 999.00000 κ= 4.56555
$\omega_{3\text{-}11}$~(0.00010 1.11937 999.00000 0.04225 0.22779 999.00000 999.00000 0.03805 0.56985)
AIC=2p+2*11+9298.845544=2p+9320.845544

Compare:
LRT:
12-RM of OBSM method I is significant better than two hypothesis models suggested in original study.
Df=10 2Δl=30.21063 p-value=7.912e-004
Df=9 2Δl=29.152286 p-value=6.108e-004

FRM VS 12-RM
Df=56 2Δl=31.423352 p-value=0.9967
AIC: the 12-RM is the best model
a: the alignment may have a little different.
b: We calculate several hypothesis models according the original paper implied and only optimal two models are shown.
c: more optimal models significant better than ORM is not shown.

CASE 23: Sperm Genes
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2a

Previous study:
ORM:
lnL= -1393.366296 ω_0= 0.24026 κ= 1.42403
AIC=2p+2786.732592
TwoRM:
lnL= -1390.023095 ω_0= 0.21576 ω_1= 2.01228 κ= 1.42746
AIC=2p+2+2780.04619=2p+2782.04619
FRM:
lnL=-1385.132916
AIC=2p+2*9+2770.265832=2p+2788.265832

OBSM Method I:
TwoRM:
It's same with hypothesis TwoRM.
FourRM: (branches are labeled by arrows)
lnL= -1386.115852 ω_0=0.35322 ω_1=2.00656 ω_2= 0.13203 ω_3=0.09261 κ= 1.45027
AIC=2p+2*3+2772.231704=2p+2778.231704

Compare:
LRT:
FourRM of OBSM method I is significant better than hypothesis TwoRM in original study.
Df=2 2Δl=7.814486 p-value=0.0201
FRM vs FourRM
Df=6 2Δl=1.965872 p-value=0.9228
AIC: the FourRM is the best model
a: the alignment may have a little different.

CASE 24: Celegent_Lysozymes_lys
Sequences: according the id in paper
Alignment: MEGA4 and edit by manual to adjust with original paper
Phylogeny: is congruent with Fig 6B

[a]Previous study:
ORM:
lnL= -5385.853900 ω_0= 0.09985 κ= 1.62418
AIC=2p+10771.7078
TwoRM A:
lnL= -5374.337116 ω_0= 0.09365 ω_J= 999.00000 κ= 1.63526
AIC=2p+2+10748.674232=2p+10750.674232
TwoRM B:
lnL= -5376.391808 ω_0= 0.10530 ω_{AJ}= 0.00010 κ= 1.62222
AIC=2p+2+10752.783616=2p+10754.783616
FRM:
LnL=-5297.096110
AIC=2p+2*43+10594.19222=2p+10680.19222

[b]OBSM Method I:
ThreeRM:
lnL= -5365.417863 ω_0= 0.09867 ω_J=999.00000 ω_{AJ}=0.00010 κ= 1.63309
AIC=2p+2*2+10730.835726=2p+10734.835726
24-RM:
lnL= -5307.308567 ω_0=0.08185 ω_J=999.00000 ω_{AJ}=0.00010 κ= 1.60670
$\omega_{3\text{-}23}$~(0.01039 0.46570 999.00000 0.24787 1.84484 0.10786 999.00000 0.26308 0.22000 0.60653 0.02760 0.03657 0.02596 0.04064 0.39669 0.03612 0.52866 0.22819 0.42078 1.11020 0.36133)
AIC=2p+2*23+10614.617134=2p+10660.617134

Compare:
LRT:
Obviously, ThreeRM and 24-RM of OBSM method I are all significant better than hypothesis models in original study.
Df=1 2Δl=17.838506 p-value= 2.405e-005
Df=22 2Δl=134.057098 p-value=0
FRM VS 24-RM
Df=20 2Δl=20.424914 p-value=0.4316
AIC: the 24-RM is the best model
a: We calculate several hypothesis models according the original paper implied and only optimal two are shown.
b: more optimal models significant better than ORM is not shown.

CASE 25: Celegent_Lysozymes_ilys
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 4B

[b]Previous study:
ORM:
lnL= -2285.487402 ω_0= 0.08775 κ= 1.46754
AIC=2p+4570.974804
TwoRM A: (Branch J)
lnL= -2277.110133 ω_0= 0.07059 ω_1= 999.00000 κ= 1.47874
AIC=2p+2+4554.220266=2p+4556.220266
TwoRM B: (Branch N)
lnL= -2277.411278 ω_0= 0.07774 ω_1= 2.74348 κ= 1.49366
AIC=2p+2+4554.822556=2p+4556.822556
FRM:
lnL= -2255.780988
AIC=2p+2*17+4511.561976=2p+4545.561976

OBSM Method I:
ThreeRM:
lnL= -2267.532752 ω_0=0.06000 ω_1= 2.84147 ω_2= 999.00000 κ= 1.50711
AIC=2p+2*2+4535.065504=2p+4539.065504
FourRM:
lnL= -2264.778762 ω_0=0.07076 ω_1=3.30886 ω_2= 999.00000 ω_3= 0.01957 κ= 1.52597
AIC=2p+2*3+4529.557524=2p+4535.557524

Compare:
LRT:
ThreeRM and FourRM of OBSM method I are all significant better than TwoRM A and TwoRM B in original study.
ThreeRM vs A Df=1 2Δl=19.154762 p-value=1.205e-005
FourRM vs A Df=2 2Δl=24.662742 p-value=4.411e-006
FRM VS FourRM
Df=142Δl=17.995548 p-value=0.20698
AIC: the FourRM is the best model
a: the alignment may have a little different
b: We calculate several hypothesis models according the original paper implied and only two of them are shown

CASE 26: X_Linked_Gene_Family
[a]Sequences: according the id in paper
[b]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 2

[c]Previous study:
ORM:
lnL= -2456.267806 ω_0= 0.28262 κ= 2.70747
AIC=2p+4912.535612
TwoRM A:
lnL= -2454.333403 ω_0= 0.27450 ω_1= 999.00000 κ= 2.71278
AIC=2p+2+4908.666806=2p+4910.666806
ThreeRM:
lnL= -2453.178179 ω_0=0.26442 ω_1= 999.00000 ω_2= 0.57204 κ= 2.65807
AIC=2p+2*2+4906.356358=2p+4910.356358
FourRM:
lnL= -2451.212063 ω_0=0.25053 ω_1=0.56358 ω_2= 999.00000 ω_3= 0.84446 κ= 2.70908
AIC=2p+2*3+4902.424126=2p+4908.424126
FRM:
lnL=-2396.264656
AIC=2p+2*66+4792.529312=2p+4924.529312

[d]OBSM Method I:
TwoRM:
lnL= -2448.331000 ω_0= 0.30898 ω_1= 0.01533 κ= 2.66214
AIC=2p+2+4896.662=2p+4898.662
ThreeRM:
lnL= -2442.218704 ω_0=0.33152 ω_1= 0.01505 ω_2= 0.03336 κ= 2.65807
AIC=2p+2*2+4884.437408=2p+4888.437408
16-RM:
lnL= -2413.221621 ω_0=0.25372 ω_1=0.01310 ω_2=0.03481 κ= 2.63900
$\omega_{3\text{-}15}$~(1.06662 0.03730 0.98333 1.91290 999.00000 0.00010 0.00010 154.56875 0.88120 0.65897 99.08026 0.00010 0.59155)
AIC=2p+2*15+4826.443242=2p+4856.443242

Compare:
LRT:
ThreeRM and FourRM of OBSM method I are all significant better than TwoRM A and TwoRM B in original study.
ThreeRM vs A Df=1 2Δl=19.154762 p-value=1.205e-005
FourRM vs 16-RM Df=12 2Δl=75.980884 p-value=2.396e-011

FRM vs 16-RM
Df=512Δl=33.91393 p-value= 0.96863
AIC: the 16-RM is the best model
a: some id given by original paper have improved and we select the longest CDS.
b: the alignment may have a little different
c: We calculate several hypothesis models according the original paper implied and only three of them are shown
d: more optimal models significant better than ORM is not shown.

CASE 27: PISTILLATA_like_genes
Sequences: according the id in paper
[a]Alignment: MEGA4 (only CDS)
Phylogeny: is congruent with Fig 4B

[b]Previous study:
ORM:
lnL= -1187.751149 ω_0= 0.13369 κ= 1.53112
AIC=2p+2375.502298
TwoRM A:
lnL= -1187.259645 ω_0= 0.11798 ω_1= 0.16143 κ= 1.54170
AIC=2p+2+2374.51929=2p+2376.51929
ThreeRM B:
lnL= -1187.194399 ω_0= 0.11385 ω_1= 0.16165 ω_2=0.13723 κ= 1.49366
AIC=2p+2*2+2374.388798=2p+2378.388798
FRM:
lnL= -1153.603318
AIC=2p+2*111+2307.206636=2p+2529.206636

OBSM Method I:
TwoRM A:
lnL= -1184.105906 ω_0= 0.12624 ω_1= 999.00000 κ= 1.53211
AIC=2p+2+2368.211812=2p+2370.211812

Compare:
LRT:
TwoRM of OBSM method I is significant better than hypothesis models in original study.
Df=1 $2\Delta l$=6.307478 p-value= 1.202e-002
Df=1 $2\Delta l$=6.176986 p-value= 1.294e-002
FRM vs TwoRM
Df=110 $2\Delta l$=61.005176 p-value= 0.99995
AIC: the TwoRM is the best model
a: the alignment may have a little different
b: We calculate several hypothesis models according the original paper implied and best two of them are shown

CASE 28: triplicated_alpha_globin_genes
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 4

[b]Previous study:
ORM:
lnL= -1612.072779 ω_0= 0.31077 κ= 2.31811
AIC=2p+3224.145558
TwoRM:
lnL= -1598.147350 ω_0= 0.1913 ω_1= 1.1151 κ= 2.36482
AIC=2p+2+3196.2947=2p+3198.2947
ThreeRM:
lnL= -1597.543353 ω_0=0.1902 ω_1= 0.8270 ω_2= 1.6875 κ= 2.37466
AIC=2p+2*2+3195.086706=2p+3199.086706
FRM:
lnL=-1580.053550
AIC=2p+2*15+3160.1071=2p+3190.1071

OBSM Method I:
TwoRM:
lnL= -1603.272372 ω_0= 0.27191 ω_1= 999.00000 κ= 2.34649
AIC=2p+2+3206.544744=2p+3208.544744
ThreeRM:
lnL= -1593.631244 ω_0=0.20754 ω_1= 1.45908 ω_2= 999.00000 κ= 2.40139
AIC=2p+2*2+3187.262488=2p+3191.262488
FourRM:
lnL= -1591.690155 ω_0=0.26197 ω_1=1.45196 ω_2= 999.00000 ω_3= 0.13052 κ= 2.45056
AIC=2p+2*3+3183.38031=2p+3189.38031
SixRM:
lnL= -1585.078668 ω_0=0.16188 ω_1=1.47085 ω_2= 999.00000 κ= 2.43738
ω_3= 0.13318 ω_4=0.82982
AIC=2p+2*5+3170.157336=2p+3180.157336

Compare:
LRT:
ThreeRM, FourRM and FiveRM of OBSM method I are all significant better than TwoRM and ThreeRM in original study.
Df=1 2Δl=11.706396 p-value=6.229e-004
Df=2 2Δl=24.92937 p-value=3.861e-006
FRM vs SixRM
Df=102Δl=10.050236 p-value=0.43609
AIC: the SixRM is the best model
a: the sequences may have a little change
b: it seems there're some difference between our results with original paper, this difference may be cause by a parameter change in control file.

CASE 29: Ketoacyl synthase domains Clade I
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -1442.533873 ω_0= 0.00516 κ= 3.51222
AIC=2p+2885.067746
TwoRM:
lnL= -1439.863723 ω_0=0.00010 ω_1= 0.00718 κ=3.47665
AIC=2p+2+2879.727446=2p+2881.727446
FRM:
lnL=-1427.795754
AIC=2p+2*29+2855.591508=2p+2913.591508

[c]OBSM Method I:
TwoRM:
lnL= -1437.170953 ω_0= 0.00392 ω_1= 0.23541 κ=3.58453
AIC=2p+2+2874.341906=2p+2876.341906
ThreeRM:
lnL= -1434.465677 ω_0=0.00333 ω_1= 0.23891 ω_2= 0.26447 κ= 3.61716
AIC=2p+2*2+2868.931354=2p+2872.931354
FiveRM:
lnL= -1431.592282 ω_0=0.00243 ω_1=0.23772 ω_2=0.27729 κ= 3.49458
ω_3= 0.00010 ω_4=0.01551
AIC=2p+2*4+2863.184564=2p+2871.184564

Compare:
LRT:
ThreeRM and FiveRM of OBSM method I are all significant better than TwoRM in original study.
Df=1 2Δl=10.796092 p-value=1.017e-003
Df=3 2Δl=16.542882 p-value=8.774e-004
FRM vs FiveRM
Df=25 2Δl=7.593056 p-value=0.99968
AIC: FiveRM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try several models according the original paper, but it seems there're some difference
c: the FourRM of OBSM method I have smaller log likelihood -1440.119958.

CASE 30: Ketoacyl synthase domains Clade II
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

Previous study:
ORM:
lnL= -1347.126347 ω_0= 0.01994 κ= 3.44062
AIC=2p+2694.252694
TwoRM:
lnL= -1345.232158 ω_0= 0.01450 ω_1= 0.04716 κ=3.42820
AIC=2p+2+2690.464316
FRM:
lnL=-1336.208517
AIC=2p+2*14+2672.417034=2p+2700.417034

[b]OBSM Method I:
TwoRM:
lnL=-1344.601473 ω_0= 0.01561 ω_1= 0.07851 κ=3.45083
AIC=2p+2+2689.202946=2p+2691.202946
ThreeRM:
lnL= -1342.629160 ω_0=0.01886 ω_1= 0.07786 ω_2= 0.00010 κ= 3.45195
AIC=2p+2*2+2685.25832=2p+2689.25832
FourRM:
lnL= -1340.193687 ω_0=0.01102 ω_1=0.07873 ω_2=0.00010 ω_3= 0.07279 κ= 3.43848
AIC=2p+2*3+2680.387374=2p+2686.387374

Compare:
LRT:
ThreeRM and FiveRM of OBSM method I are all significant better than TwoRM in original study.
Df=1 2Δl=5.205996 p-value=0.0225
Df=2 2Δl=10.076942 p-value=6.484e-003
FRM vs FourRM
Df=1 1 2Δl=7.97034 p-value= 0.7159
AIC: the FourRM is the best model
a: the sequences may have a little change (may be shorter than original paper)

CASE 31: Ketoacyl synthase domains Clade III
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -2969.405876 ω_0= 0.01015 κ= 1.87659
AIC=2p+5938.811752
TwoRM: (branch b)
lnL= -2966.588079 ω_0= 0.01106 ω_1= 0.00032 κ=1.94088
AIC=2p+2+5933.176158=2p+5935.176158
TwoRM: (clade a)
lnL= -2961.581034 ω_0= 0.00338 ω_1= 0.02002 κ=2.00419
AIC=2p+2+5923.162068=2p+5925.162068
FRM:
lnL=-2940.835126
AIC=2p+2*25+5881.670252=2p+5931.670252

[c]OBSM Method I:
TwoRM:
lnL=-2963.922594 ω_0= 0.01210 ω_1= 0.00010 κ=1.85404
AIC=2p+2+5927.845188=2p+5929.845188
ThreeRM:
lnL= -2957.525489 ω_0=0.01474 ω_1= 0.00010 ω_2= 0.00076 κ= 1.91676
AIC=2p+2*2+5915.050978=2p+5919.050978
13-RM:
lnL= -2943.893767 ω_0=0.00623 ω_1=0.00010 ω_2=0.00075 κ=1.95339
$\omega_{3\text{-}12}$~(0.02852 999.00000 999.00000 0.06094 0.00601 0.02369 999.00000 0.00156 999.00000 0.02487)
AIC=2p+2*12+5887.787534=2p+5911.787534

Compare:
LRT:
ThreeRM and 13-RM of OBSM method I are all significant better than TwoRM (calde meodel) in original study.
Df=1 2Δl=8.11109 p-value=4.400e-003
Df=11 2Δl=35.374534 p-value=2.149e-004
FRM vs 13-RM
Df=13 2Δl=6.117282 p-value=0.9417
AIC: the 13-RM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try six models according the original paper, and only branch model b and clade model a is shown.
c: more optimal models significant better than ORM is not shown.

CASE 32: Ketoacyl synthase domains Clade IV
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

Previous study:
ORM:
lnL=-1572.043286 ω_0= 0.02194 κ= 2.52282
AIC=2p+3144.086572
TwoRM:
lnL= -1570.583052 ω_0=0.01887 ω_1= 0.07576 κ=2.47106
AIC=2p+2+3141.166104
FRM:
lnL=-1566.875766
AIC=2p+2*10+3133.751532=2p+3153.751532

OBSM Method I:
TwoRM:
lnL=-1570.351702 ω_0= 0.02856 ω_1= 0.00010 κ=2.36845
AIC=2p+2+3140.703404=2p+3142.703404
OBSM Method II:
same with method I
OBSM Method III:
TwoRM (k=0.5):
lnL= -1567.760247 ω_0= 0.0347 ω_1= 0.0011 κ=2.44028
AIC=2p+2+3135.520494=2p+3137.520494

Compare:
LRT:
TwoRM of OBSM method III is significant better than TwoRM in original study.
Df=1 2Δl=5.64561 p-value=0.0175
FRM vs TwoRM
Df=9 2Δl=1.768962 p-value=0.9946
AIC: the TwoRM of Method III is the best model
a: the sequences may have a little change (may be shorter than original paper)

CASE 33: Ketoacyl synthase domains Clade V
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -2356.171114 ω_0= 0.01622 κ= 4.75052
AIC=2p+4712.342228
TwoRM: (branch a)
lnL= -2355.599899 ω_0= 0.01526 ω_1= 0.14417 κ=4.81586
AIC=2p+2+4711.199798=2p+4713.199798
TwoRM: (clade a)
lnL= -2353.627003 ω_0= 0.00831 ω_1= 0.02135 κ=4.81184
AIC=2p+2+4707.254006=2p+4709.254006
FRM:
lnL=-2336.130641
AIC=2p+2*26+4672.261282=2p+4724.261282

OBSM Method I:
TwoRM:
lnL=-2348.967332 ω_0=0.02106 ω_1= 0.00010 κ=4.70463
AIC=2p+2+4697.934664=2p+4699.934664
FourRM:
lnL= -2345.674448 ω_0=0.01878 ω_1= 0.00010 ω_2= 74.91367 ω_3=174.69408 κ= 4.82045
AIC=2p+2*3+4691.348896=2p+4697.348896
SixRM:
lnL= -2342.410437 ω_0=0.01991 ω_1=0.00010 ω_2=67.12320 κ=4.93669
$\omega_{3\text{-}5}$~(212.64653 0.05116 0.00587)
AIC=2p+2*5+4684.820874=2p+4694.820874

Compare:
LRT:
TwoRM, FourRM and SixRM of OBSM method I are all significant better than TwoRM (calde meodel) in original study.
Df=2 2Δl=15.90511 p-value=3.518e-004
Df=4 2Δl=22.433132 p-value=1.643e-004
FRM vs SixRM
Df=21 2Δl=12.559592 p-value=0.9233
AIC: the SixRM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try four models according the original paper, and only branch model a and clade model a is shown.

CASE 34: Ketoacyl synthase domains Clade VI
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -1776.430586 ω_0= 0.04141 κ= 4.22245
AIC=2p+3552.861172
TwoRM: (clade b)
lnL= -1768.018146 ω_0= 0.12261 ω_1= 0.02464 κ=4.70537
AIC=2p+2+3536.036292=2p+3538.036292
FRM:
lnL=-1738.570051
AIC=2p+2*15+3477.140102=2p+3507.140102

[c]OBSM Method I:
TwoRM:
lnL= -1763.087057 ω_0=0.08059 ω_1=0.00408 κ=5.01164
AIC=2p+2+ 3526.174114=2p+3528.174114
FourRM:
lnL= -1755.050278 ω_0=0.06845 ω_1= 0.00010 ω_2= 0.02008ω_3=999.00000 κ= 5.08283
AIC=2p+2*3+3510.100556=2p+3516.100556
10-RM:
lnL= -1738.752202 ω_0=0.04541 ω_1=0.00010 ω_2=0.01973 κ=5.27260
$\omega_{3\text{-}9}$~(999.00000 1.32052 0.54685 999.00000 0.17221 999.00000 0.00010)
AIC=2p+2*9+3477.504404=2p+3495.504404

Compare:
LRT:
TwoRM, FourRM and 10-RM of OBSM method I are all significant better than TwoRM (calde meodel) in original study.
Df=2 2Δl=25.935736 p-value=2.334e-006
Df=8 2Δl=58.531888 p-value=9.039e-010
FRM vs 10-RM
Df=6 2Δl=0.364302 p-value=0.9991
AIC: the 10-RM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try four models according the original paper, and only clade model b is shown.
c: more optimal models significant better than ORM is not shown.

CASE 35: Ketoacyl synthase domains Clade VII
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -6919.600817 ω_0= 0.04925 κ= 2.76949
AIC=2p+13839.201634
TwoRM: (clade a)
lnL= -6917.240527 ω_0= 0.05346 ω_1= 0.03344 κ= 2.75765
AIC=2p+2+13834.481054=2p+13836.481054
TwoRM: (branch a)
lnL= -6917.612101 ω_0= 0.04847 ω_1= 999.00000 κ= 2.78387
AIC=2p+2+13835.224202=2p+13837.224202
FRM:
lnL=-6862.802488
AIC=2p+2*39+13725.604976=2p+13803.604976

[c]OBSM Method I:
TwoRM:
lnL= -6906.202348 ω_0= 0.04641 ω_1=0.80833 κ= 2.80510
AIC=2p+2+13812.404696=2p+13814.404696
ThreeRM:
lnL= -6902.812662 ω_0= 0.04821 ω_1= 0.80835 ω_2= 0.01364 κ= 2.80989
AIC=2p+2*2+13805.625324=2p+13809.625324
13-RM:
lnL= -6879.796276 ω_0= 0.04593 ω_1=0.81655 ω_2=0.01367 κ= 2.98419
$\omega_{3\text{-}12}$~(0.28053 999.00000 0.00117 0.01360 0.00010 0.01618 0.00010 999.00000 0.18339 45.39811)
AIC=2p+2*12+13759.592552=2p+13783.592552

Compare:
LRT:
TwoRM, ThreeRM and 13-RM of OBSM method I are all significant better than TwoRM (calde meodel) in original study.
Df=1 2Δl=28.85573 p-value=7.797e-008
Df=11 2Δl=74.888502 p-value=1.424e-011
FRM vs 13-RM
Df=27 2Δl=33.987576 p-value=0.1664
AIC: the 13-RM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try ten models according the original paper, and two of them are shown.
c: more optimal models significant better than ORM is not shown.

CASE 36: Ketoacyl synthase domains Clade VIII
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

Previous study:
ORM:
lnL= -4250.179073 ω_0= 0.00844 κ= 0.80641
AIC=2p+8500.358146
TwoRM:
lnL= -4245.816088 ω_0= 0.00805 ω_1= 999.00000 κ= 0.80984
AIC=2p+2+8491.632176=2p+8493.632176
FRM:
lnL=-4215.743015
AIC=2p+2*43+8431.48603=2p+8517.48603

[b]OBSM Method I:
TwoRM:
lnL= -4245.816088 ω_0= 0.00805 ω_1= 999.00000 κ= 0.80984
AIC=2p+2+8491.632176=2p+8493.632176
ThreeRM:
lnL= -4242.093240 ω_0= 0.00744 ω_1= 936.71839ω_2= 0.14035 κ= 0.80719
AIC=2p+2*2+8484.18648=2p+8488.18648
16-RM:
lnL= -4222.098324 ω_0= 0.00626 ω_1=0.03564 ω_2=0.17681 κ= 0.86957
$\omega_{3\text{-}15}$~(0.20547 48.59960 2.33064 30.88344 0.00062 0.00084 0.00012 0.00010 0.13668 46.83536 0.00010 0.00010 0.00393)
AIC=2p+2*15+8444.196648=2p+8474.196648

Compare:
LRT:
ThreeRM and 16-RM of OBSM method I are all significant better than TwoRM in original study.
Df=1 2Δl=7.445696 p-value=6.359e-003
Df=14 2Δl=47.435528 p-value=1.629e-005
FRM vs 16-RM
Df=28 2Δl=12.710618 p-value=0.994
AIC: the 16-RM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: more optimal models significant better than ORM is not shown.

CASE 37: Ketoacyl synthase domains Clade IX
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -3801.247660 ω_0= 0.03241 κ= 2.98811
AIC=2p+7602.49532
TwoRM: (branch b)
lnL= -3799.831950 ω_0= 0.03418 ω_1= 0.00980 κ= 2.99947
AIC=2p+2+7599.6639=2p+7601.6639
TwoRM: (clade c)
lnL= -3790.660485 ω_0= 0.01990 ω_1= 0.06138 κ= 3.04410
AIC=2p+2+7581.32097=2p+7583.32097
FRM:
lnL=-3770.506192
AIC=2p+2*26+7541.012384=2p+7593.012384

[c]OBSM Method I:
TwoRM:
lnL= -3793.903278 ω_0= 0.03598 ω_1= 0.00343 κ= 2.98488
AIC=2p+2+7587.806556=2p+7589.806556
ThreeRM:
lnL= -3787.908586 ω_0= 0.03335 ω_1= 0.00344 ω_2=999.00000 κ= 2.99748
AIC=2p+2*2+7575.817172=2p+7579.817172
10-RM:
lnL= -3774.202946 ω_0= 0.04314 ω_1=0.00371 ω_2= 999.00000 κ= 3.08358
$\omega_{3\text{-}9}$~(0.08615 0.11755 0.00437 0.01750 0.00911 0.01522 0.00977)
AIC=2p+2*9+7548.405892=2p+7566.405892

Compare:
LRT:
ThreeRM and 10-RM of OBSM method I are all significant better than TwoRM (clade model) in original study.
Df=1 2Δl=5.503798 p-value=1.898e-002
Df=8 2Δl=32.915078 p-value=6.380e-005
FRM vs 10-RM
Df=172Δl=7.393508 p-value=0.9778
AIC: the 10-RM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try six models according the original paper, and two of them are shown.
c: more optimal models significant better than ORM is not shown.

CASE 38: Ketoacyl synthase domains Clade X
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1
[b]Previous study:
ORM: lnL= -5368.317970 ω_0= 0.02285 κ= 1.88448
AIC=2p+10736.63594
TwoRM: (clade b)
lnL= -5350.866731 ω_0= 0.00635 ω_1= 0.09546 κ= 1.90825
AIC=2p+2+10701.733462=2p+10703.733462
FRM:
lnL=-5341.104326
AIC=2p+2*21+10682.208652=2p+10724.208652
[c]OBSM Method I:
TwoRM:
lnL= -5361.540123 ω_0= 0.01938 ω_1= 0.81478 κ= 1.89504
AIC=2p+2+ 10723.080246=2p+10725.080246
ThreeRM:
lnL= -5357.617433 ω_0=0.03022 ω_1= 0.78823 ω_2=0.00366 κ= 1.87018
AIC=2p+2*2+10715.234866=2p+10719.234866
SevenRM:
lnL= -5348.280085 ω_0= 0.02861 ω_1=0.64904 ω_2= 0.00589 κ= 1.90195
$\omega_{3\text{-}6}$~(0.01076 999.00000 0.19920 0.00213)
AIC=2p+2*6+10696.56017=2p+10708.56017
[c]OBSM Method II:
TwoRM:
It was same with Method I.
ThreeRM:
lnL= -5356.377877 ω_0=0.01532 ω_1= 0.81557 ω_2=999.00000 κ= 1.90255
AIC=2p+2*2+ 10712.755754=2p+10716.755754
SixRM:
lnL= -5346.670985 ω_0= 0.00621 ω_1=0.64300 ω_2= 999.00000 κ= 1.90844
$\omega_{3\text{-}5}$~(0.20178 0.06858 0.05229)
AIC=2p+2*5+10693.34197=2p+10703.34197
[d]OBSM Method III: (k=0.5)
TwoRM:
lnL= -5349.174539 ω_0= 0.00422 ω_1= 0.10201 κ= 1.91186
AIC=2p+2+ 10698.349078=2p+10700.349078

Compare:
LRT:
In this case, hypothesis TwoRM (clade model) is obviously significant better than some optimal models explored by OBSM Method I or Method II, but is not better than final models, the final optimal model of Method III is better than this hypothesis model, but not significant if set the df=1.
Df=1 2Δl=3.384384 p-value=6.582e-002
FRM vs TwoRM
Df=202Δl=16.140426p-value=0.7078
AIC: the TwoRM of Method III is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we tried four models according the original paper, and only the best one is shown.
c: more optimal models significant better than ORM is not shown.
d: we also tried set k=0.2 and k=1, the results also didn't get better, but interestingly, we observed a sudden log likelihood increase in NineRM (k=1) and in EightRM (k=0.2) when the previous optimal models have very slim log likelihood difference.

CASE 39: Ketoacyl synthase domains Clade XI
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -2436.063975 ω_0= 0.00849 κ= 1.81073
AIC=2p+4872.12795
TwoRM: (clade model)
lnL= -2434.195583 ω_0=0.01402 ω_1=0.00530 κ= 1.83238
AIC=2p+2+4868.391166=2p+4870.391166
FRM:
lnL=-2427.362661
AIC=2p+2*17+4854.725322=2p+4888.725322

OBSM Method I:
No model is significant better than ORM.
[c]OBSM Method II:
FourRM:
lnL= -2430.452590 ω_0=0.00406 ω_1=681.43273 ω_2= 926.92276 ω_3= 0.06654 κ= 1.83103
AIC=2p+2*3+4860.90518=2p+4866.90518
OBSM Method III: (k=0.5)
TwoRM:
lnL= -2430.552363 ω_0=0.00391 ω_1=33.20042 κ= 1.83445
AIC=2p+2+4861.104726=2p+4863.104726

Compare:
LRT:
The hypothesis clade model is not significant better than ORM but is very close, the p value is 5.323e-002.
FourRM of Method II, TwoRM of Method III of OBSM are all significant better than ORM.
FourRM of Method II vs ORM Df=3 2Δl=11.22277 p-value=1.058e-002
TwoRM of Method III vs ORM Df=1 2Δl=11.023224 p-value=8.998e-004
And FourRM of Method II, TwoRM of Method III of OBSM are all significant better than TwoRM in original study.
FourRM of Method II vs TwoRM Df=2 2Δl=7.485986 p-value=2.368e-002
TwoRM of Method III vs TwoRM Df=1 2Δl=7.28644 p-value=6.948e-003
FRM vs TwoRM
Df=162Δl=6.379404 p-value=0.9834
AIC: the TwoRM of the Method III is the best model

a: the sequences may have a little change (may be shorter than original paper)
b: we try three models according the original paper, and the best model is shown.
c: more optimal models significant better than ORM is not shown.

CASE 40: Ketoacyl synthase domains Clade XII
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -2334.288011 ω_0= 0.00388 κ= 3.16771
AIC=2p+4668.576022
TwoRM: (branch a)
lnL= -2332.734490 ω_0=0.00780 ω_1=0.00027 κ= 3.19185
AIC=2p+2+4665.46898=2p+4667.46898
FRM:
lnL=-2331.698694
AIC=2p+2*11+4663.397388=2p+4685.397388

[c]OBSM Method:
TwoRM:
lnL= -2332.734490 ω_0=0.00780 ω_1=0.00027 κ= 3.19185
AIC=2p+2+4665.46898=2p+4667.46898

Compare:
LRT:
In this case, we try all three OBSM methods and no model is found out that significant better than ORM. The maximum log likelihood in TwoRM suggest by three methods are all congruent with hypothesis model.
Df=1 2Δl=3.107042 p-value=7.795e-002
FRM vs FRM
Df=10 2Δl=2.071592 p-value=0.9957
AIC: the TwoRM is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try four models according the original paper, and the best model is shown.
c: we try all three method, and no model is significant better than ORM.

CASE 41: Hepcidin Gene in mammal
Sequences: according the id in paper
[a]Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -1310.463212 ω_0= 0.34291 κ= 2.32131
AIC=2p+2620.926424
FRM:
lnL=-1306.357211
AIC=2p+2*24+2612.714422=2p+2660.714422

[c]OBSM Method I:
TwoRM:
lnL= -1309.645825 ω_0=0.36984 ω_1=0.18439 κ= 2.32462
AIC=2p+2+2619.29165=2p+2621.29165
[c]OBSM Method II:
TwoRM:
lnL= -1308.499025 ω_0= 0.3883 ω_1= 0.1342 κ= 2.44803
AIC=2p+2+2616.99805=2p+2618.99805
OBSM Method III: (k=0.5)
TwoRM:
lnL= -1306.082389 ω_0= 0.49775 ω_1=0.14775 κ= 2.35330
AIC=2p+2+2612.164778=2p+2614.164778

Compare:
LRT:
TwoRM of Method II vs ORM Df=1 2Δl=3.928374 p-value=4.748e-002
TwoRM of Method III vs TwoRM Df=1 2Δl=8.761646 p-value=3.076e-003
FRM vs TwoRM
Df=23 2Δl=-0.549644 p-value=1
AIC: the TwoRM of Method III is the best model

a: the alignment may have a little different
b: no hypothesis model, and free ratio model is not significant better than ORM.
c: in this case, there're two pair of sequences are same with each other respectively, and the log likelihood of some models are so instable that the optimal two model of Method I and Method II is not the same.

CASE 42: Hepcidin Gene in Pleuronectiformes and Perciformes
Sequences: according the id in paper
[a]Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -1692.468704 ω_0= 0.56045 κ= 2.55100
AIC=2p+3384.937408
FRM:
lnL=-1659.752558
AIC=2p+2*60+3319.505116=2p+3439.505116

[c]OBSM Method I:
TwoRM:
lnL= -1688.457755 ω_0=0.59137 ω_1= 0.00010 κ= 2.55587
AIC=2p+2+3376.91551=2p+3378.91551
ThreeRM:
lnL= -1685.432393 ω_0=0.56484 ω_1= 0.00010 ω_2= 999.00000 κ= 2.55776
AIC=2p+2*2+3370.864786=2p+3374.864786
SevenRM(final optimal model):
lnL= -1675.308796 ω_0=0.65674 ω_1=0.00010 ω_2= 999.00000 κ=2.56295
$\omega_{2\text{-}6}$ ~ (0.13097 0.13689 0.00010 2.60183)
AIC=2p+2*6+3350.617592=2p+3362.617592

Compare:
LRT:
TwoRM, ThreeRM and SevenRM of OBSM Method I are all significant better than ORM.
Df =1 2Δl=8.021898 p-value=4.622e-003
Df =2 2Δl=14.072622 p-value= 8.794e-004
Df =6 2Δl=34.319816 p-value=5.835e-006
FRM vs SevenRM
Df =54 2Δl=31.112476 p-value=0.9947
AIC: the SevenRM is the best model
a: the alignment may have a little different
b: no hypothesis model, and free ratio model is not significant better than ORM.
c: more optimal models significant better than ORM is not shown.

CASE 43: SRPX2_gene
Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 4

Previous study:
ORM:
lnL= -2172.869055 ω_0= 0.05323 κ= 27.93800
AIC=2p+4345.73811
TwoRM: (Human lineage)
lnL= -2170.444813 ω_0=0.04253 ω_1=999.00000 κ= 27.96774
AIC=2p+2+4340.889626=2p+4342.889626
FRM:
lnL= -2166.238379
AIC=2p+2*10+4332.476758=2p+4352.476758

OBSM Method I:
Same with hypothesis model.
OBSM Method II:
FourRM:
lnL= -2167.312563 ω_0=0.02178 ω_1=999.00000 ω_2= 0.90182 ω_3= 0.49328 κ= 28.13717
AIC=2p+2*3+4334.625126=2p+4340.625126
OBSM Method III: (k=0.5)
TwoRM:
lnL= -2167.835109 ω_0=0.02177 ω_1=0.96753 (red arrow) κ= 28.14329
AIC=2p+2+4335.670218=2p+4337.670218

Compare:
LRT:
FourRM of Method II, TwoRM of Method III of OBSM are all significant better than hypothesis model.
FourRM of Method II vs TwoRM Df=2 2Δl=6.2645 p-value=4.362e-002
TwoRM of Method III vs TwoRM Df=1 2Δl=5.219408 p-value=2.234e-002
FRM vs TwoRM
Df=9 2Δl=3.19346 p-value= 0.9561
AIC: the TwoRM of Method III is the best model

CASE 44: Cholesterol Metabolism Gene
Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 3

Previous study:
ORM:
lnL= -1377.506105 ω_0= 0.38583 κ= 8.01562
AIC=2p+2755.01221
[a]TwoRM:
lnL= -1374.183004 ω_0= 0.35223 ω_g= 999.00000 κ= 7.99858
AIC=2p+2+2748.366008=2p+2750.366008
FRM:
lnL=-1360.065636
AIC=2p+2*24+2720.131272=2p+2768.131272

[b]OBSM Method I:
TwoRM:
Same with hypothesis model
ThreeRM:
lnL= -1371.448491 ω_0=0.32659 ω_1=999.00000 ω_2= 999.00000 κ= 7.99851
AIC=2p+2*2+2742.896982=2p+2746.896982
SevenRM:
lnL= -1364.613059 ω_0=0.49116 ω_1=999.00000 ω_2=999.00000 κ= 7.99626
$\omega_{3\text{-}6}$~(0.09249 0.06185 0.00010 0.15082)
AIC=2p+2*6+2729.226118=2p+2741.226118

Compare:
LRT:
ThreeRM and SevenRM of OBSM Method I are all significant better than hypothesis.
Df=1 2Δl=5.469026 p-value=1.936e-002
Df=5 2Δl=19.13989 p-value=1.810e-003
FRM vs SevenRM
Df=18 2Δl=9.094846 p-value=0.9575
AIC: the SevenRM is the best model
a: we try several models according the original paper, and the best model is shown.
b: more optimal models significant better than hypothesis models are not shown.

CASE 45: MOXD2
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 5A

[b]Previous study:
ORM:
lnL= -1437.821252 ω_0= 0.22956 κ= 7.46279
AIC=2p+2875.642504
TwoRM:
lnL= -1432.478425 ω_0=0.17723 ω_1=2.84197 κ= 7.49018
AIC=2p+2+2864.95685=2p+2866.95685
FRM:
lnL= -1430.473242
AIC=2p+2*9+2860.946484=2p+2878.946484

OBSM Method I, II and III:
TwoRM:
Same with hypothesis model

Compare:
LRT:
All the three methods and the hypothesis model suggest the same model.
FRM vs TwoRM
Df=8 2Δl=4.010366 p-value=0.8562
AIC: the best model is the TwoRM

a: the id given by original paper is too confused to get identical sequences and the data we got have some changes (have some region missing and shorter than original paper)
b: we try several models according the original paper, and the best model is shown.

CASE 46: S100A15A
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -316.981093 ω_0= 0.10216 κ= 4.55632
AIC=2p+633.962186
TwoRM:
lnL= -315.411448 ω_0=0.07520 ω_G=0.53909 κ=4.60778
AIC=2p+2+630.822896=2p+632.822896
TwoRM:
lnL= -315.562527 ω_0=0.07124 ω_H=0.34965 κ=4.66719
AIC=2p+2+631.125054=2p+633.125054
FRM:
lnL=-310.909248
AIC=2p+2*9+621.818496=2p+639.818496

OBSM Method I:
TwoRM:
lnL= -314.968454 ω_0=0.13024 ω_1=0.00010 κ= 4.45714
AIC=2p+2+629.936908=2p+631.936908
OBSM Method II:
Same with Method I
OBSM Method III: (k=0.5)
TwoRM:
lnL= -311.941308 ω_0= 0.2752 ω_1= 0.0001 κ= 4.58316
AIC=2p+2+623.882616=2p+625.882616

Compare:
LRT:
TwoRM of OBSM Method III is significant better than hypothesis model.
Df=1 2Δl=11.023224 p-value=8.428e-003
FRM vs TwoRM
Df=8 2Δl=2.06412 p-value=0.97898
AIC: the TwoRM of Method III is the best model
a: the sequences may have a little change (may be shorter than original paper)
b: we try three models according the original paper, and the best model is shown.
c: more optimal models significant better than ORM is not shown.

CASE 47: Penaeidin antimicrobial peptides
Sequences: according the id in paper
[a]Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -1493.075007 ω_0= 0.79029 κ= 1.64528
AIC=2p+2986.150014
FRM:
lnL=-1461.569406
AIC=2p+2*69+2923.138812=2p+3061.138812

[c]OBSM Method I:
TwoRM:
lnL= -1485.990663 ω_0=0.92749 ω_1=0.00010 κ= 1.66723
AIC=2p+2+2971.981326=2p+2973.981326
FourRM:
lnL= -1482.637071 ω_0=0.86416 ω_1=0.00010 ω_2= 999.00000 κ= 1.66892
AIC=2p+2*3+2965.274142=2p+2971.274142
EightRM:
lnL= -1473.714700 ω_0=1.05542 ω_1=0.00010 ω_2=999.00000 κ= 1.65460
$\omega_{3\text{-}7}$~(0.25444 0.00010 999.00000 0.00010 0.21093)
AIC=2p+2*7+2947.4294=2p+2961.4294

Compare:
LRT:
TwoRM FourRM and Eight of OBSM Method I are all significant better than ORM.
Df=1 2Δl=14.168688 p-value=1.671e-004
Df=3 2Δl=20.875872 p-value=1.117e-004
Df=7 2Δl=38.720614 p-value=2.210e-006
FRM vs EightRM
Df=62 2Δl=24.290588 p-value=0.999995
AIC: the EightRM is the best model
a: the alignment may have a little different
b: no model is significant better than ORM in original study
c: more optimal models significant better than ORM is not shown.

CASE 48: NYD SP12
Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[a]Previous study:
ORM:
lnL= -3222.614087 ω_0= 0.62004 κ= 4.50420
AIC=2p+6445.228174
TwoRM:
lnL= -3214.198480 ω_0=0.4589 ω_1=4.8091 κ= 4.51495
AIC=2p+2+6428.39696=2p+6430.39696
FRM:
lnL=-3209.257860
AIC=2p+2*11+6418.51572=2p+6440.51572
OBSM Method I:
TwoRM:
lnL= -3217.680711 ω_0=0.54921 ω_1=999.00000 κ= 4.50692
AIC=2p+2+6435.361422=2p+6437.361422
ThreeRM:
lnL= -3215.065138 ω_0=0.49872 ω_1=999.00000 ω_2= 3.56377 κ= 4.51073
AIC=2p+2*2+6430.130276=2p+6434.130276
OBSM Method II:
Same with Method I
OBSM Method III: (k=0.5)
TwoRM:
lnL= -3213.688700 ω_0=0.4524 ω_1=4.9861 κ= 4.51422
AIC=2p+2+6427.3774=2p+6429.3774
FourRM:
lnL= -3209.920138 ω_0=0.2135 ω_1=4.8632 ω_2= 0.8290 ω_3=0.0001 κ= 4.51602
AIC=2p+2*3+6419.840276=2p+6425.840276

Compare:
LRT:
FourRM of OBSM Method III are all significant better than hypothesis model
Df=2 $2\Delta l$=8.556684 p-value=1.387e-002
FRM vs FourRM
Df=8 $2\Delta l$=1.324556 p-value=0.9952
AIC: the FourRM of Method III is the best model
a: we try three models according the original paper, and the best model is shown.

CASE 49: Myxovirus resistance gene
Sequences: according the id in paper
[a]Alignment: MEGA4
Phylogeny: is congruent with Fig 1

[b]Previous study:
ORM:
lnL= -11761.353476 ω_0= 0.30812 κ= 2.40593
AIC=2p+23522.706952
TwoRM: (Branch duck)
lnL= -11743.731336 ω_0=0.27584 ω_1=30.81443 κ= 2.42547
AIC=2p+2+23487.462672=2p+23489.462672
TwoRM: (Branch chicken)
lnL= -11734.736366 ω_0=0.26402 ω_1=1.24721 κ= 2.40555
AIC=2p+2+23469.472732=2p+23471.472732
FRM:
lnL=-11697.659375
AIC=2p+2*21+23395.31875=2p+23437.31875

[c]OBSM Method I:
TwoRM:
Same with hypothesis model (chicken)
ThreeRM:
lnL= -11730.837896 ω_0=0.25735 ω_1=0.74941 ω_2=0.70825 κ= 2.41932
AIC=2p+2*2+23461.675792=2p+23465.675792
TenRM:
lnL= -11703.168029 ω_0= 0.28169 ω_1=0.74325 ω_2=0.73485 κ= 2.45458
$\omega_{3\text{-}7}$~(0.00010 0.27599 0.14481 0.15634 0.50811 0.53044 0.06674)
AIC=2p+2*9+23406.336058=2p+23424.336058

Compare:
LRT:
ThreeRM and EightRM of OBSM Method I are all significant better than hypothesis model (chicken).
Df=1 2Δl=7.79694 p-value=5.233e-003
Df=8 2Δl=63.136674 p-value=1.126e-010
FRM vs TenRM
Df=12 2Δl=11.017308 p-value= 0.5274
AIC: the TenRM is the best model
a: the alignment may have a little different
b: we try several models according the original paper, and the best model is shown.
c: more optimal models significant better than hypothesis models are not shown.

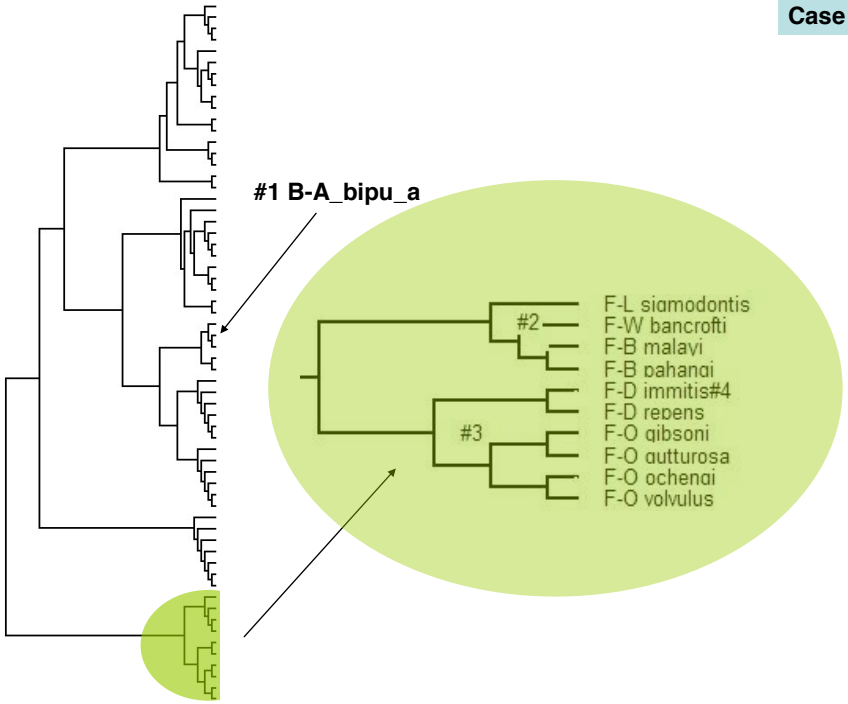
CASE 50: Last Case
[a]Sequences: according the id in paper
Alignment: MEGA4
Phylogeny: is congruent with Fig 2

[b]Previous study:
ORM:
lnL= -9972.141967 ω_0= 0.11693 κ= 1.44964
AIC=2p+19944.283934
TwoRM: (Branch GRCD1)
lnL= -9964.958754 ω_0=0.11208 ω_1=0.35186 κ= 1.44837
AIC=2p+2+19929.917508=2p+19931.917508
TwoRM: (Branch VvMADS4)
lnL= -9963.962501 ω_0=0.12148 ω_1=0.02850 κ= 1.44404
AIC=2p+2+19927.925002=2p+19929.925002
FRM:
lnL=-9905.929279
AIC=2p+2*53+19811.858558=2p+19917.858558

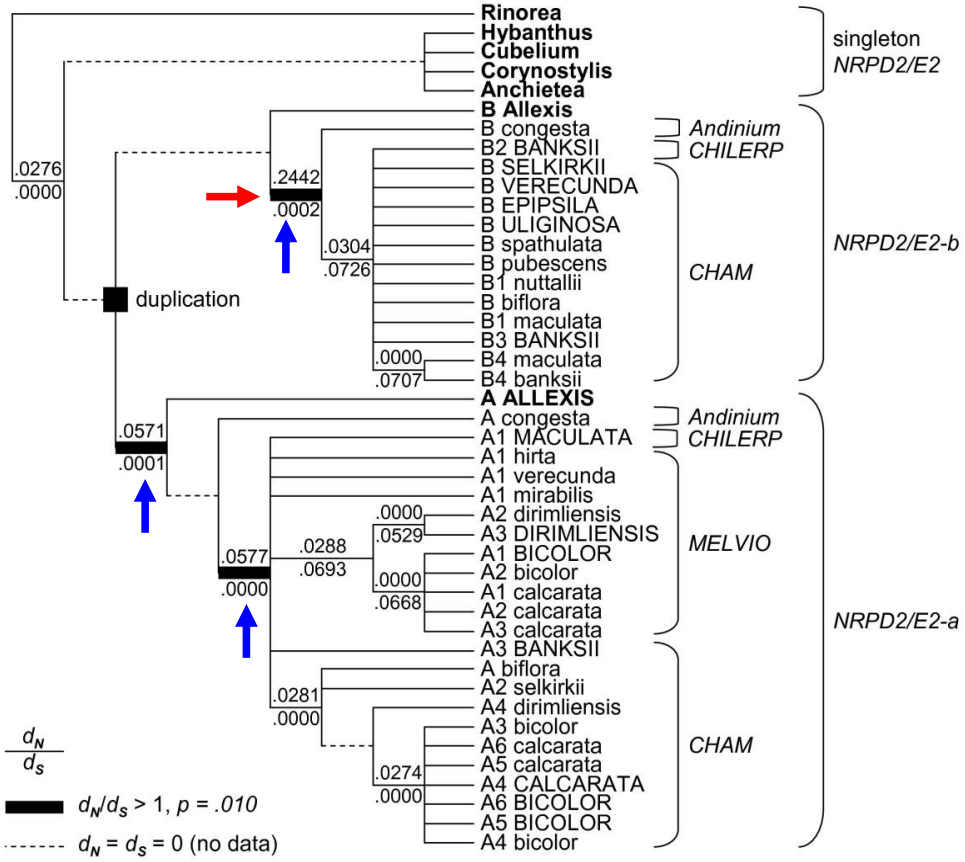
[c]OBSM Method I:
TwoRM:
Same with hypothesis model (Branch VvMADS4)
ThreeRM:
lnL=-9957.697019 ω_0=0.11654 ω_1=0.02935 (VvMADS4) ω_2=0.33541 (GRCD1) κ=1.44278
AIC=2p+2*2+19915.394038=2p+19919.394038
16-RM:
lnL= -9920.246557 ω_0= 0.11331 ω_1=0.03090 ω_2=0.34508 κ= 1.42649
$\omega_{3\text{-}15}$~(0.01912 2.74301 0.28361 0.26519 0.05442 0.58769 0.02559 0.01744 0.03899 0.06861 0.17050 0.36419 0.19309)
AIC=2p+2*15+19840.493114=2p+19870.493114

Compare:
LRT:
ThreeRM of 15-RM of OBSM Method I are all significant better than hypothesis model
Df=3 2Δl= 12.530964 p-value=5.769e-003
Df=14 2Δl= 87.431888 p-value=4.321e-013
FRM vs 15-RM
Df=38 2Δl= 28.634556 p-value=0.8642
AIC: the 16-RM is the best model
a: the sequences may have a little change (some id given by original paper have changed)
b: we try several models according the original paper, and the best two model are shown.
c: more optimal models significant better than ORM is not shown.

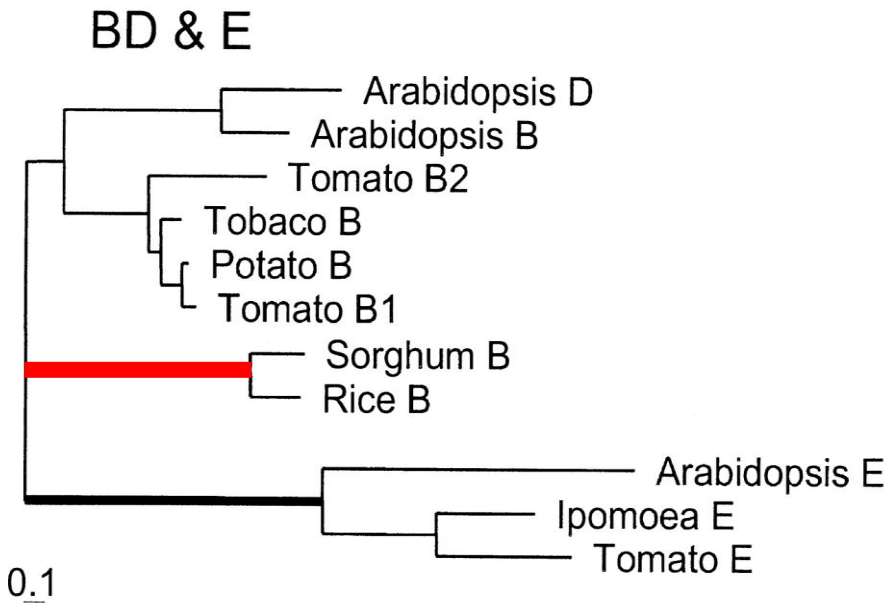
Case 1


#1 B-A_bipu_a
#2
#3
F-L siamodontis
F-W bancrofti
F-B malayi
F-B pahangi
F-D immitis#4
F-D repens
F-O gibsoni
F-O gutturosa
F-O ochengi
F-O volvulus

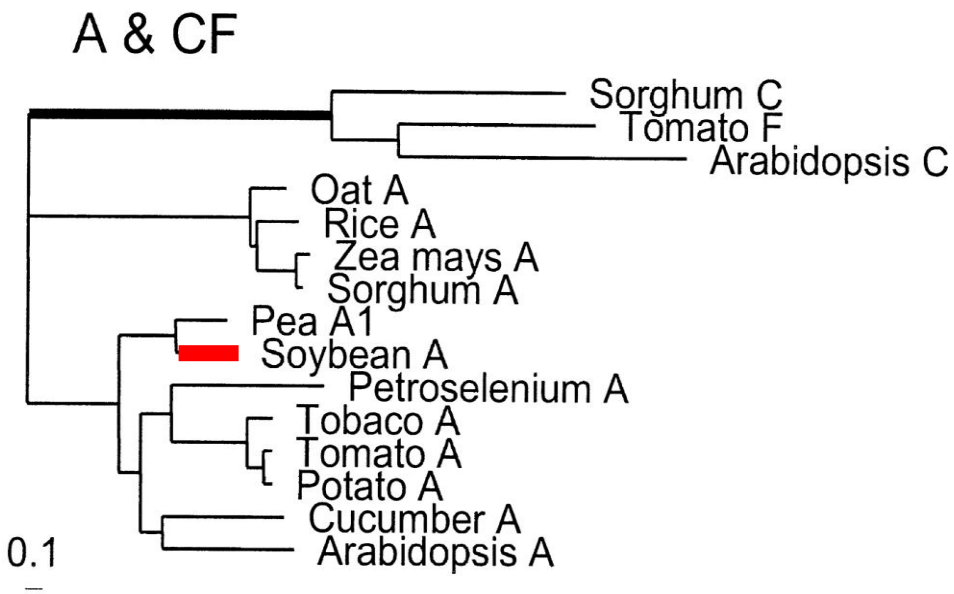
Case 2


Rinorea
Hybanthus
Cubelium
Corynostylis
Anchietea
B Allexis
B congesta
B2 BANKSII
B SELKIRKII
B VERECUNDA
B EPIPSILA
B ULIGINOSA
B spathulata
B pubescens
B1 nuttallii
B biflora
B1 maculata
B3 BANKSII
B4 maculata
B4 banksii
A ALLEXIS
A congesta
A1 MACULATA
A1 hirta
A1 verecunda
A1 mirabilis
A2 dirimliensis
A3 DIRIMLIENSIS
A1 BICOLOR
A2 bicolor
A1 calcarata
A2 calcarata
A3 calcarata
A3 BANKSII
A biflora
A2 selkirkii
A4 dirimliensis
A3 bicolor
A6 calcarata
A5 calcarata
A4 CALCARATA
A6 BICOLOR
A5 BICOLOR
A4 bicolor
.0276
.0000
.2442
.0002
.0304
.0726
.0000
.0707
duplication
.0571
.0001
.0577
.0000
.0288
.0693
.0000
.0529
.0000
.0668
.0281
.0000
.0274
.0000
singleton
NRPD2/E2
Andinium
CHILERP
CHAM
NRPD2/E2-b
Andinium
CHILERP
MELVIO
CHAM
NRPD2/E2-a
dN/dS
dN/dS > 1, p = .010
dN = dS = 0 (no data)

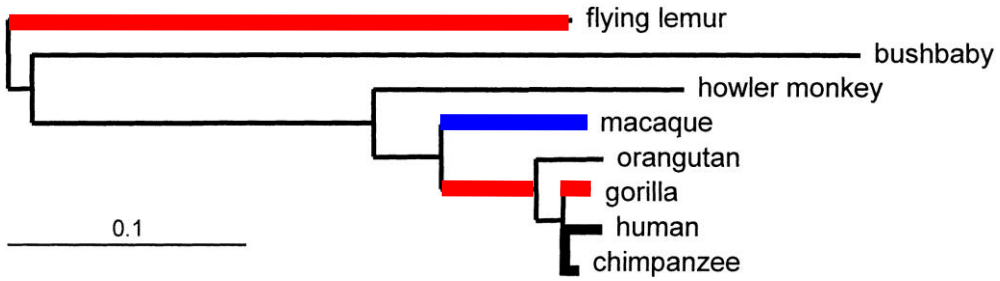
Case 3


BD & E
Arabidopsis D
Arabidopsis B
Tomato B2
Tobaco B
Potato B
Tomato B1
Sorghum B
Rice B
Arabidopsis E
Ipomoea E
Tomato E
0.1

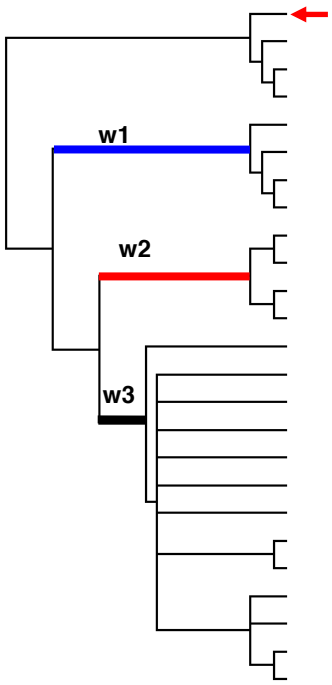
Case 4


A & CF
Sorghum C
Tomato F
Arabidopsis C
Oat A
Rice A
Zea mays A
Sorghum A
Pea A1
Soybean A
Petroselenium A
Tobaco A
Tomato A
Potato A
Cucumber A
Arabidopsis A
0.1

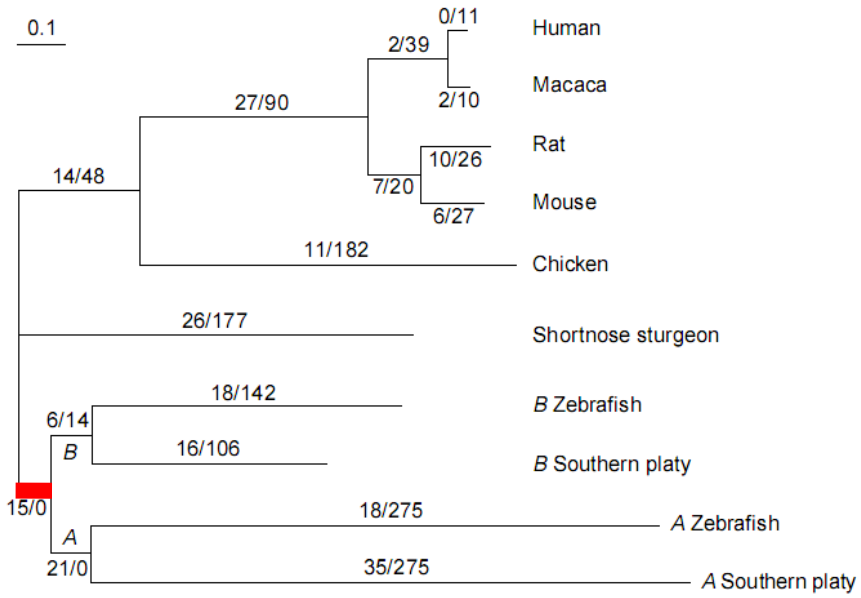
Case 5


flying lemur
bushbaby
howler monkey
macaque
orangutan
gorilla
human
chimpanzee
0.1

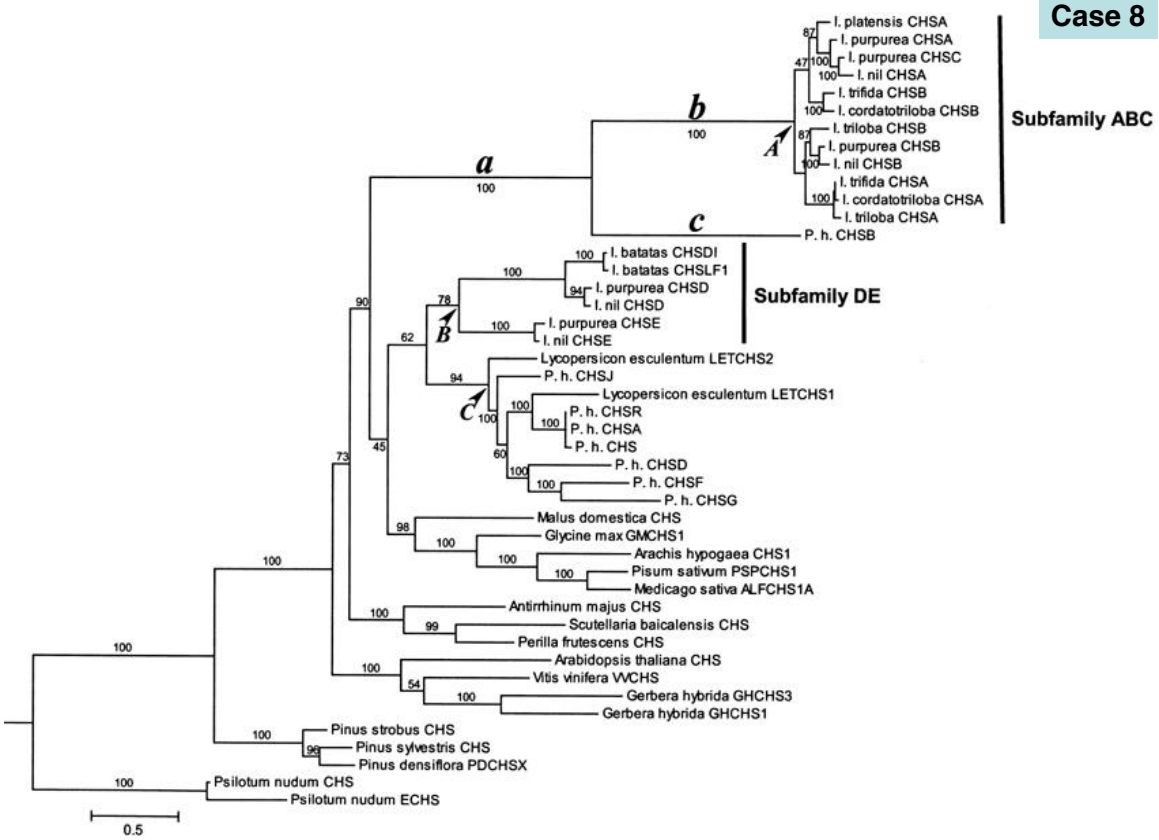
Case 6



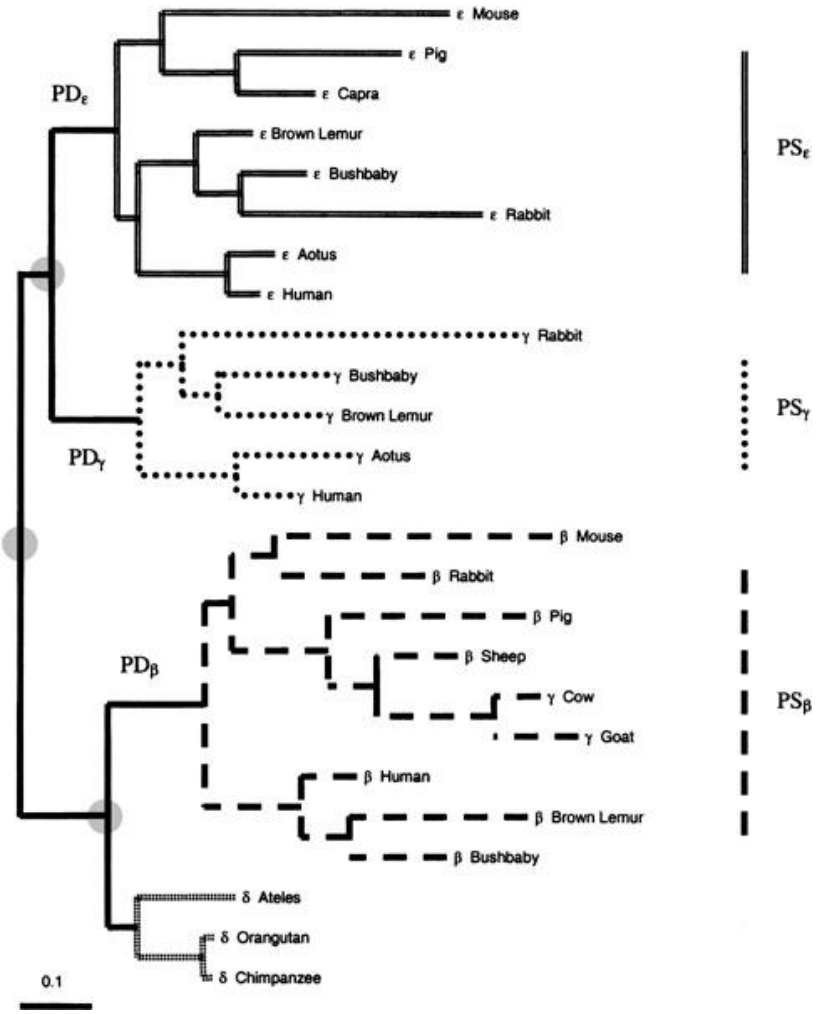
Case 7


0.1
0/11
2/39
Human
Macaca
2/10
27/90
Rat
10/26
14/48
7/20
Mouse
6/27
11/182
Chicken
26/177
Shortnose sturgeon
18/142
B Zebrafish
6/14
B
16/106
B Southern platy
15/0
18/275
A Zebrafish
A
21/0
35/275
A Southern platy

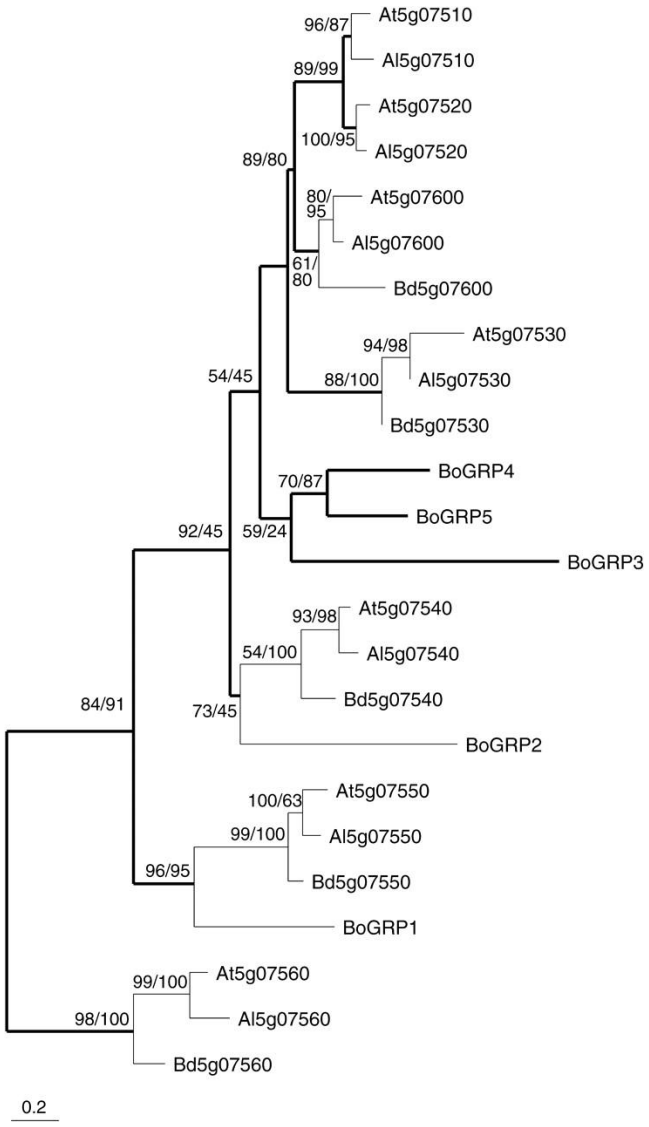
Case 8


I. platensis CHSA
I. purpurea CHSA
I. purpurea CHSC
I. nil CHSA
I. trifida CHSB
I. cordatotriloba CHSB
I. triloba CHSB
I. purpurea CHSB
I. nil CHSB
I. trifida CHSA
I. cordatotriloba CHSA
I. triloba CHSA
P. h. CHSB
Subfamily ABC
I. batatas CHSDI
I. batatas CHSLF1
I. purpurea CHSD
I. nil CHSD
I. purpurea CHSE
I. nil CHSE
Subfamily DE
Lycopersicon esculentum LETCHS2
P. h. CHSJ
Lycopersicon esculentum LETCHS1
P. h. CHSR
P. h. CHSA
P. h. CHS
P. h. CHSD
P. h. CHSF
P. h. CHSG
Malus domestica CHS
Glycine max GMCHS1
Arachis hypogaea CHS1
Pisum sativum PSPCHS1
Medicago sativa ALFCHS1A
Antirrhinum majus CHS
Scutellaria baicalensis CHS
Perilla frutescens CHS
Arabidopsis thaliana CHS
Vitis vinifera VVCHS
Gerbera hybrida GHCHS3
Gerbera hybrida GHCHS1
Pinus strobus CHS
Pinus sylvestris CHS
Pinus densiflora PDCHSX
Psilotum nudum CHS
Psilotum nudum ECHS
0.5
a
b
c
A
B
C

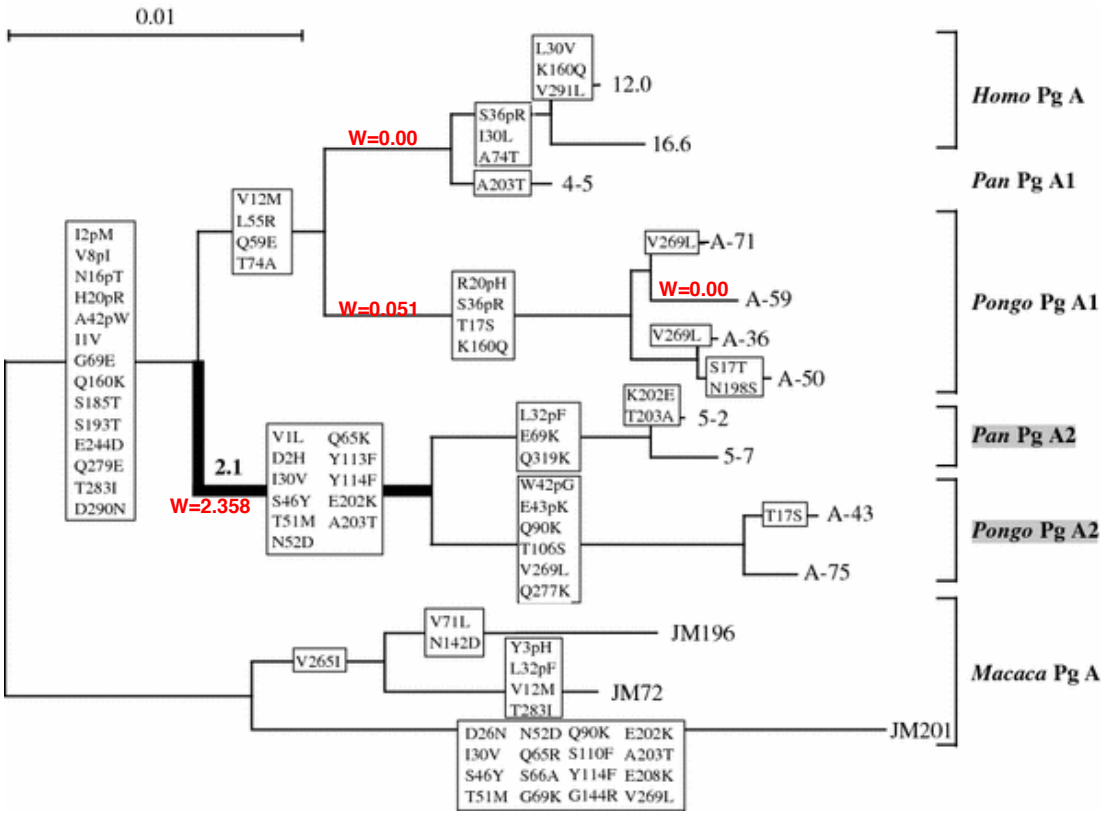
Case 9


ε Mouse
ε Pig
ε Capra
ε Brown Lemur
ε Bushbaby
ε Rabbit
ε Aotus
ε Human
PDε
PSε
γ Rabbit
γ Bushbaby
γ Brown Lemur
γ Aotus
γ Human
PDγ
PSγ
β Mouse
β Rabbit
β Pig
β Sheep
γ Cow
γ Goat
β Human
β Brown Lemur
β Bushbaby
PDβ
PSβ
δ Ateles
δ Orangutan
δ Chimpanzee
0.1

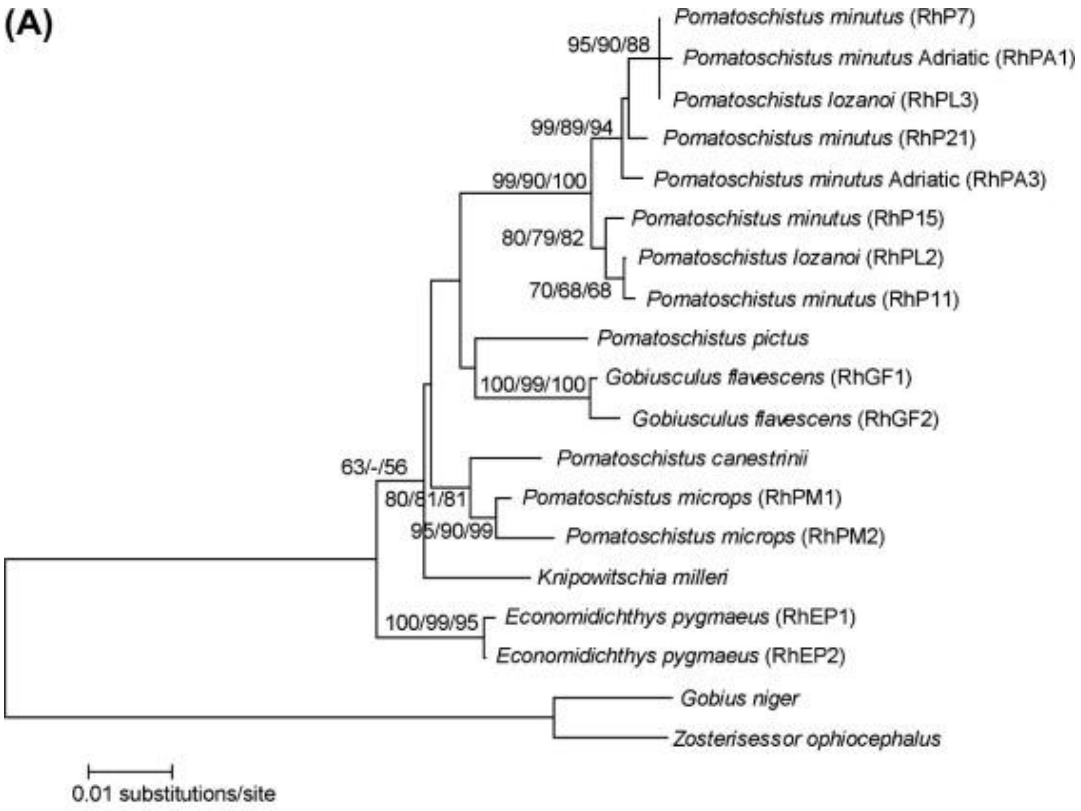
Case 10


96/87
At5g07510
Al5g07510
89/99
At5g07520
100/95
Al5g07520
89/80
80/
95
At5g07600
Al5g07600
61/
80
Bd5g07600
94/98
At5g07530
54/45
88/100
Al5g07530
Bd5g07530
70/87
BoGRP4
BoGRP5
92/45
59/24
BoGRP3
93/98
At5g07540
54/100
Al5g07540
84/91
Bd5g07540
73/45
BoGRP2
100/63
At5g07550
99/100
Al5g07550
96/95
Bd5g07550
BoGRP1
99/100
At5g07560
98/100
Al5g07560
Bd5g07560
0.2

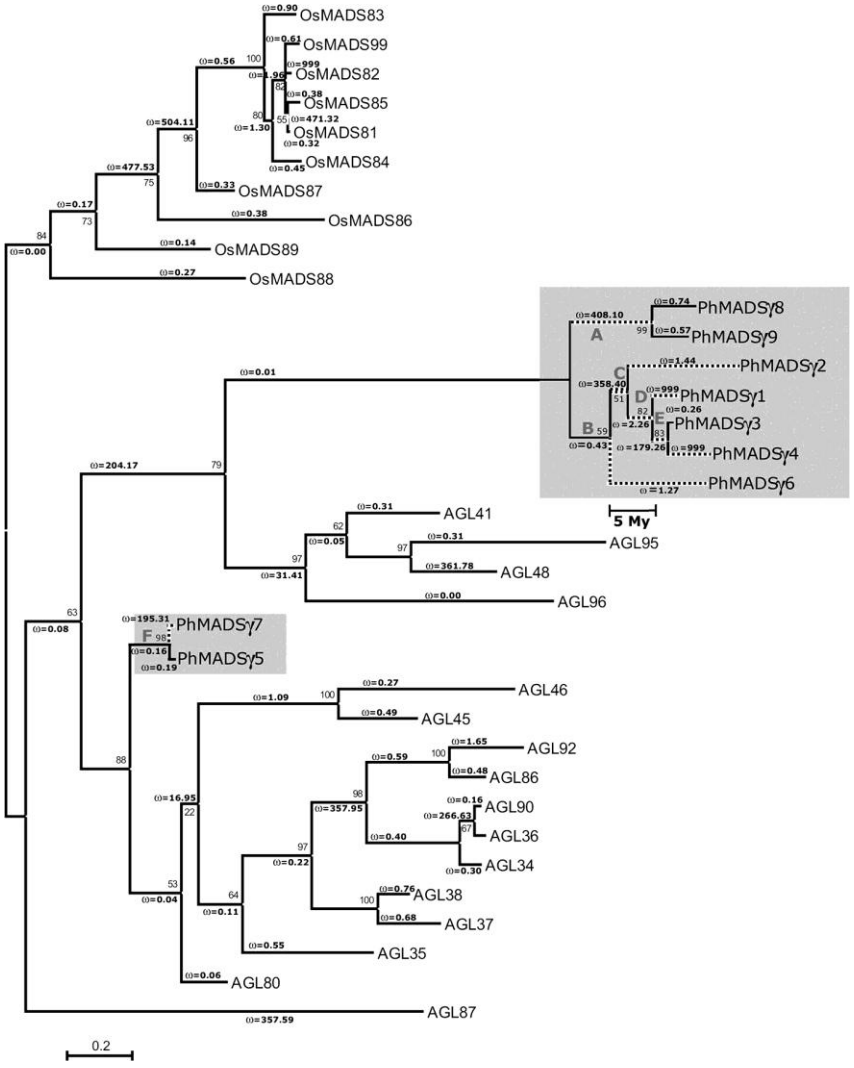
Case 11


0.01
L30V
K160Q
V291L
12.0
S36pR
I30L
A74T
16.6
W=0.00
A203T
4-5
Homo Pg A
Pan Pg A1
V12M
L55R
Q59E
T74A
I2pM
V8pI
N16pT
H20pR
A42pW
I1V
G69E
Q160K
S185T
S193T
E244D
Q279E
T283I
D290N
V269L
A-71
W=0.00
A-59
R20pH
S36pR
T17S
K160Q
W=0.051
V269L
A-36
S17T
N198S
A-50
Pongo Pg A1
K202E
T203A
5-2
L32pF
E69K
Q319K
5-7
Pan Pg A2
2.1
W=2.358
V1L
D2H
I30V
S46Y
T51M
N52D
Q65K
Y113F
Y114F
E202K
A203T
W42pG
E43pK
Q90K
T106S
V269L
Q277K
T17S
A-43
A-75
Pongo Pg A2
V71L
N142D
JM196
V265I
Y3pH
L32pF
V12M
T283I
JM72
JM201
Macaca Pg A
D26N N52D Q90K E202K
I30V Q65R S110F A203T
S46Y S66A Y114F E208K
T51M G69K G144R V269L

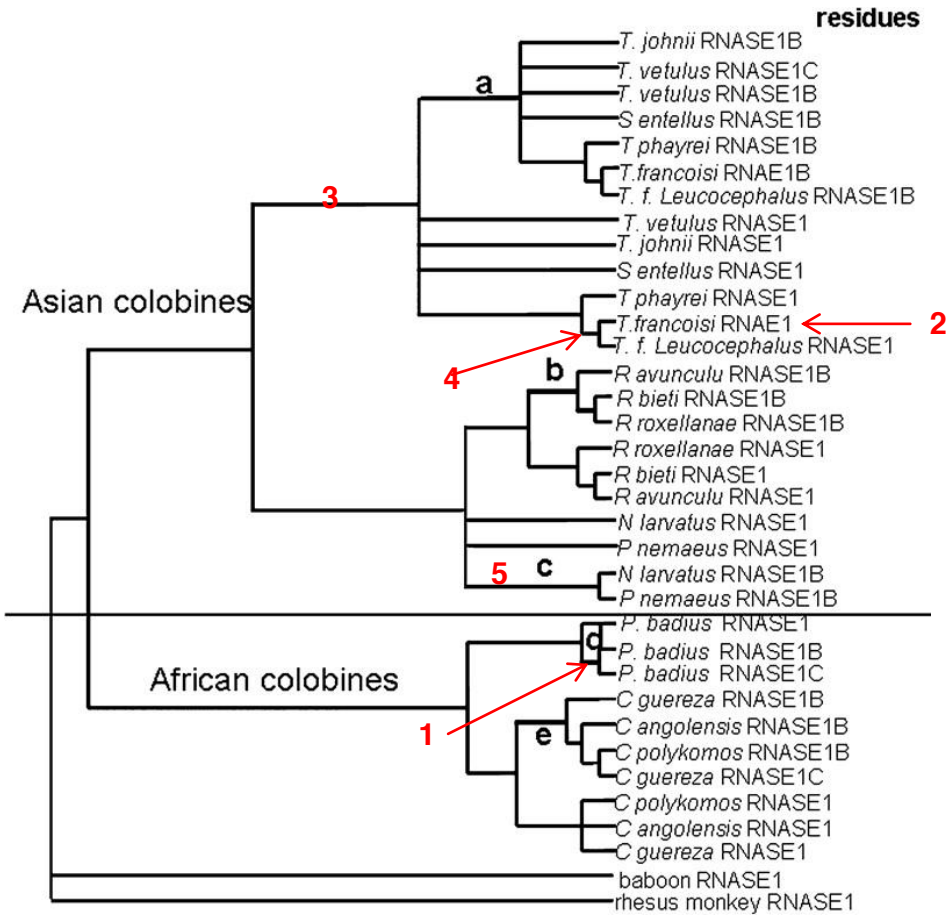
Case 12


(A)
Pomatoschistus minutus (RhP7)
95/90/88
Pomatoschistus minutus Adriatic (RhPA1)
Pomatoschistus lozanoi (RhPL3)
99/89/94
Pomatoschistus minutus (RhP21)
99/90/100
Pomatoschistus minutus Adriatic (RhPA3)
Pomatoschistus minutus (RhP15)
80/79/82
Pomatoschistus lozanoi (RhPL2)
70/68/68
Pomatoschistus minutus (RhP11)
Pomatoschistus pictus
100/99/100
Gobiusculus flavescens (RhGF1)
Gobiusculus flavescens (RhGF2)
63/-/56
Pomatoschistus canestrinii
80/81/81
Pomatoschistus microps (RhPM1)
95/90/99
Pomatoschistus microps (RhPM2)
Knipowitschia milleri
100/99/95
Economidichthys pygmaeus (RhEP1)
Economidichthys pygmaeus (RhEP2)
Gobius niger
Zosterisessor ophiocephalus
0.01 substitutions/site

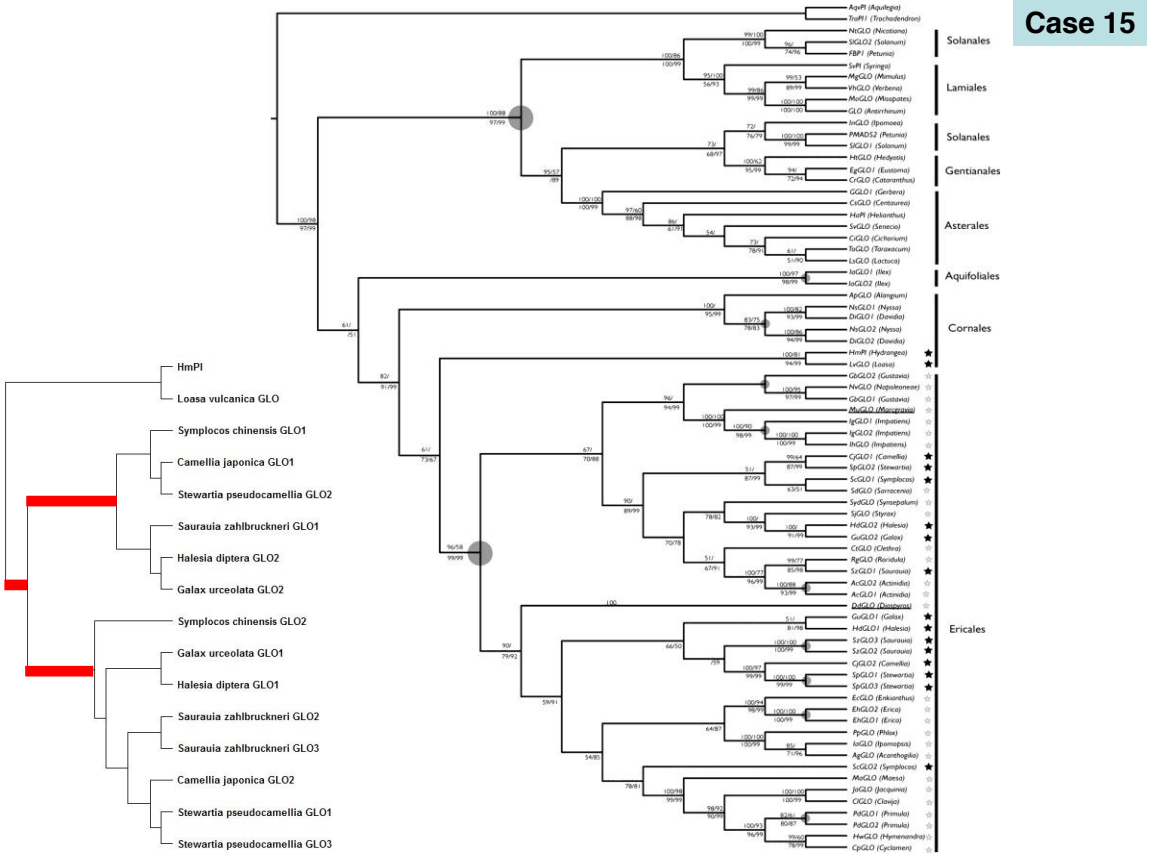
Case 13


OsMADS83
OsMADS99
OsMADS82
OsMADS85
OsMADS81
OsMADS84
OsMADS87
OsMADS86
OsMADS89
OsMADS88
PhMADSγ8
PhMADSγ9
PhMADSγ2
PhMADSγ1
PhMADSγ3
PhMADSγ4
PhMADSγ6
A
B
C
D
E
F
5 My
AGL41
AGL95
AGL48
AGL96
PhMADSγ7
PhMADSγ5
AGL46
AGL45
AGL92
AGL86
AGL90
AGL36
AGL34
AGL38
AGL37
AGL35
AGL80
AGL87
0.2

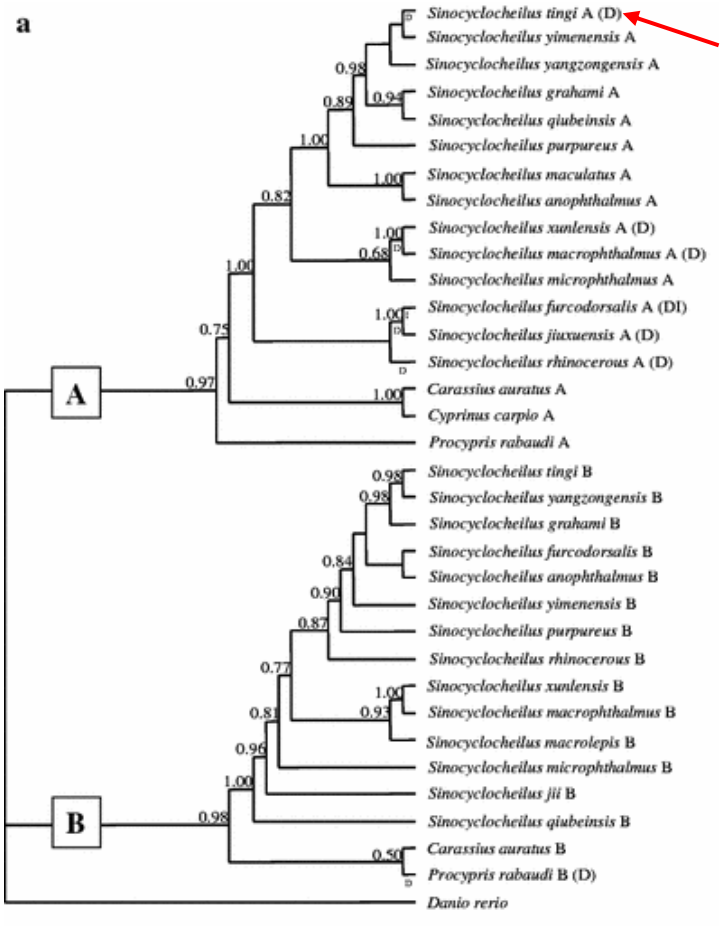
Case 14


residues
T. johnii RNASE1B
T. vetulus RNASE1C
T. vetulus RNASE1B
S entellus RNASE1B
T phayrei RNASE1B
T.francoisi RNAE1B
T. f. Leucocephalus RNASE1B
T. vetulus RNASE1
T. johnii RNASE1
S entellus RNASE1
T phayrei RNASE1
T.francoisi RNAE1
T. f. Leucocephalus RNASE1
R avunculu RNASE1B
R bieti RNASE1B
R roxellanae RNASE1B
R roxellanae RNASE1
R bieti RNASE1
R avunculu RNASE1
N larvatus RNASE1
P nemaeus RNASE1
N larvatus RNASE1B
P nemaeus RNASE1B
P. badius RNASE1
P. badius RNASE1B
P. badius RNASE1C
C guereza RNASE1B
C angolensis RNASE1B
C polykomos RNASE1B
C guereza RNASE1C
C polykomos RNASE1
C angolensis RNASE1
C guereza RNASE1
baboon RNASE1
rhesus monkey RNASE1
a
b
c
d
e
1
2
3
4
5
Asian colobines
African colobines

Case 15


HmPI
Loasa vulcanica GLO
Symplocos chinensis GLO1
Camellia japonica GLO1
Stewartia pseudocamellia GLO2
Saurauia zahlbruckneri GLO1
Halesia diptera GLO2
Galax urceolata GLO2
Symplocos chinensis GLO2
Galax urceolata GLO1
Halesia diptera GLO1
Saurauia zahlbruckneri GLO2
Saurauia zahlbruckneri GLO3
Camellia japonica GLO2
Stewartia pseudocamellia GLO1
Stewartia pseudocamellia GLO3
Solanales
Lamiales
Solanales
Gentianales
Asterales
Aquifoliales
Cornales
Ericales

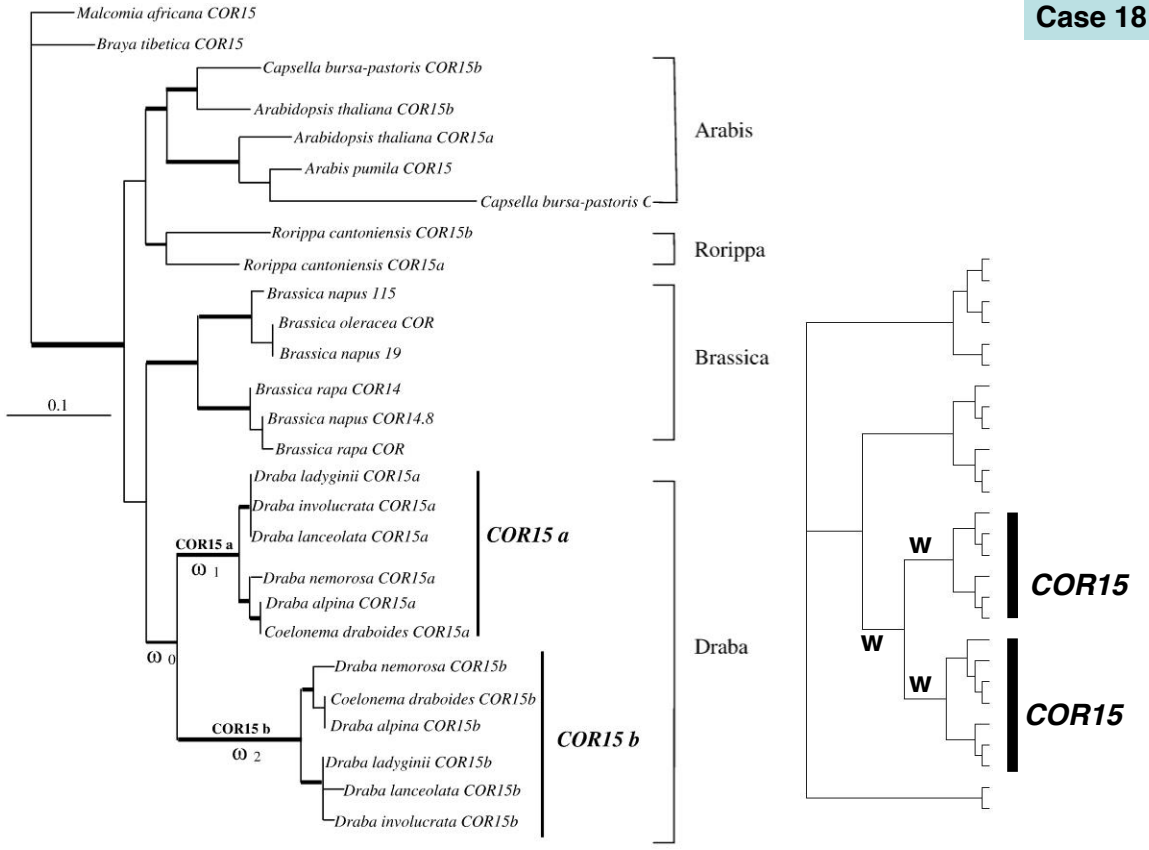
Case 16


a
Sinocyclocheilus tingi A (D)
Sinocyclocheilus yimenensis A
Sinocyclocheilus yangzongensis A
Sinocyclocheilus grahami A
Sinocyclocheilus qiubeinsis A
Sinocyclocheilus purpureus A
Sinocyclocheilus maculatus A
Sinocyclocheilus anophthalmus A
Sinocyclocheilus xunlensis A (D)
Sinocyclocheilus macrophthalmus A (D)
Sinocyclocheilus microphthalmus A
Sinocyclocheilus furcodorsalis A (DI)
Sinocyclocheilus jiuxuensis A (D)
Sinocyclocheilus rhinocerous A (D)
Carassius auratus A
Cyprinus carpio A
Procypris rabaudi A
Sinocyclocheilus tingi B
Sinocyclocheilus yangzongensis B
Sinocyclocheilus grahami B
Sinocyclocheilus furcodorsalis B
Sinocyclocheilus anophthalmus B
Sinocyclocheilus yimenensis B
Sinocyclocheilus purpureus B
Sinocyclocheilus rhinocerous B
Sinocyclocheilus xunlensis B
Sinocyclocheilus macrophthalmus B
Sinocyclocheilus macrolepis B
Sinocyclocheilus microphthalmus B
Sinocyclocheilus jii B
Sinocyclocheilus qiubeinsis B
Carassius auratus B
Procypris rabaudi B (D)
Danio rerio
A
B

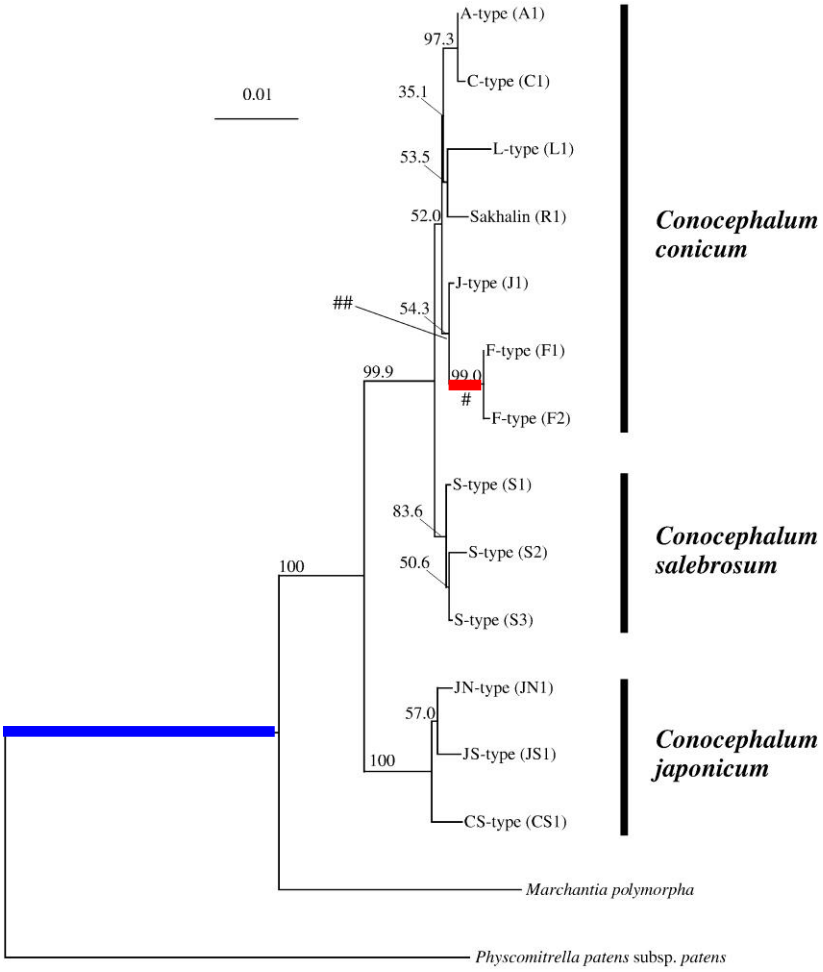
Case 17


b
Sinocyclocheilus grahami A
Sinocyclocheilus yimenensis A
Sinocyclocheilus tingi A
Sinocyclocheilus qiubeinsis A
Sinocyclocheilus yangzongensis A
Sinocyclocheilus anophthalmus A
Sinocyclocheilus purpureus A
Sinocyclocheilus maculatus A
Sinocyclocheilus macrolepis A
Sinocyclocheilus jii A
Sinocyclocheilus macrophthalmus A
Sinocyclocheilus xunlensis A
Sinocyclocheilus microphthalmus A
Sinocyclocheilus furcodorsalis A
Sinocyclocheilus rhinocerous A
Cyprinus carpio A
Sinocyclocheilus xunlensis B
Sinocyclocheilus furcodorsalis B
Carassius auratus B
Sinocyclocheilus jii B
Procypris rabaudi B
Danio rerio
0.98
1.00
1.00
S94L
1.00
0.97
0.98
1.00
A
0.75
94S
0.99
1.00
0.96
0.96
B
0.77
94S

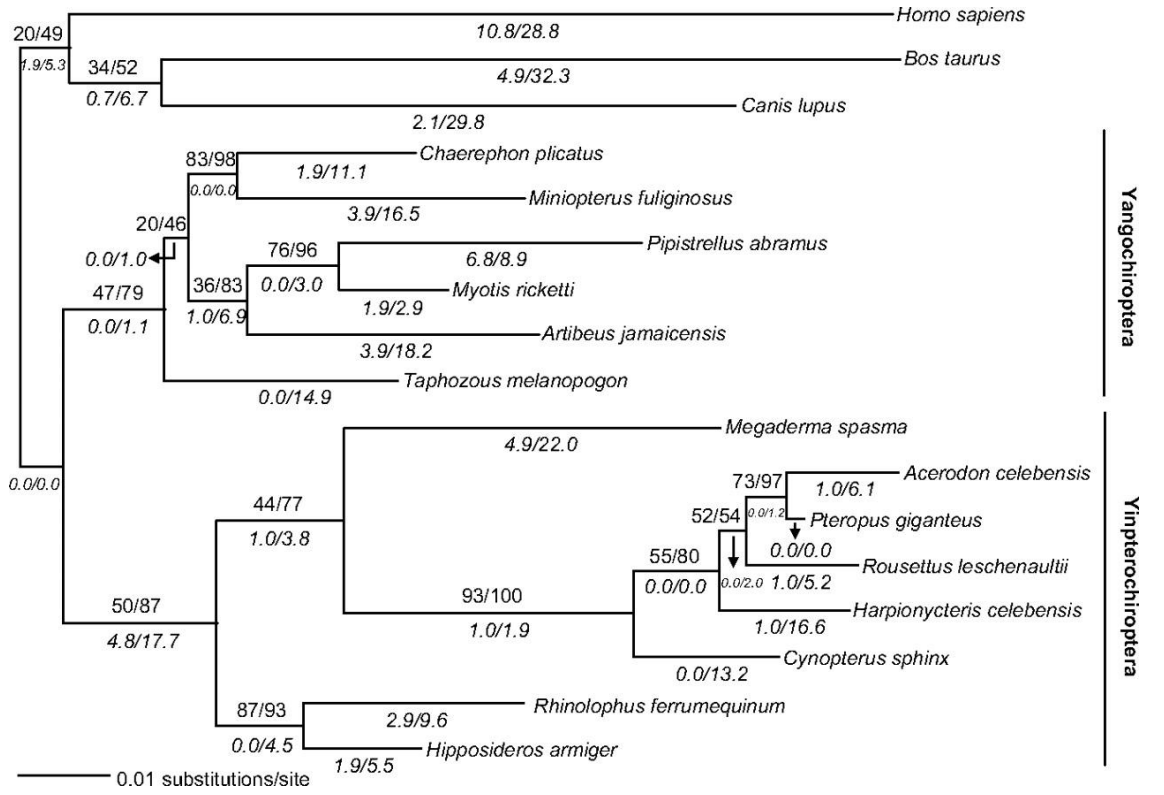
Case 18


Malcomia africana COR15
Braya tibetica COR15
Capsella bursa-pastoris COR15b
Arabidopsis thaliana COR15b
Arabidopsis thaliana COR15a
Arabis pumila COR15
Capsella bursa-pastoris C
Arabis
Rorippa cantoniensis COR15b
Rorippa cantoniensis COR15a
Rorippa
Brassica napus 115
Brassica oleracea COR
Brassica napus 19
Brassica rapa COR14
Brassica napus COR14.8
Brassica rapa COR
Brassica
0.1
Draba ladyginii COR15a
Draba involucrata COR15a
Draba lanceolata COR15a
COR15 a
COR15 a
ω 1
Draba nemorosa COR15a
Draba alpina COR15a
Coelonema draboides COR15a
ω 0
Draba nemorosa COR15b
Coelonema draboides COR15b
Draba alpina COR15b
COR15 b
ω 2
Draba ladyginii COR15b
Draba lanceolata COR15b
Draba involucrata COR15b
COR15 b
Draba
W
W
W
COR15
COR15

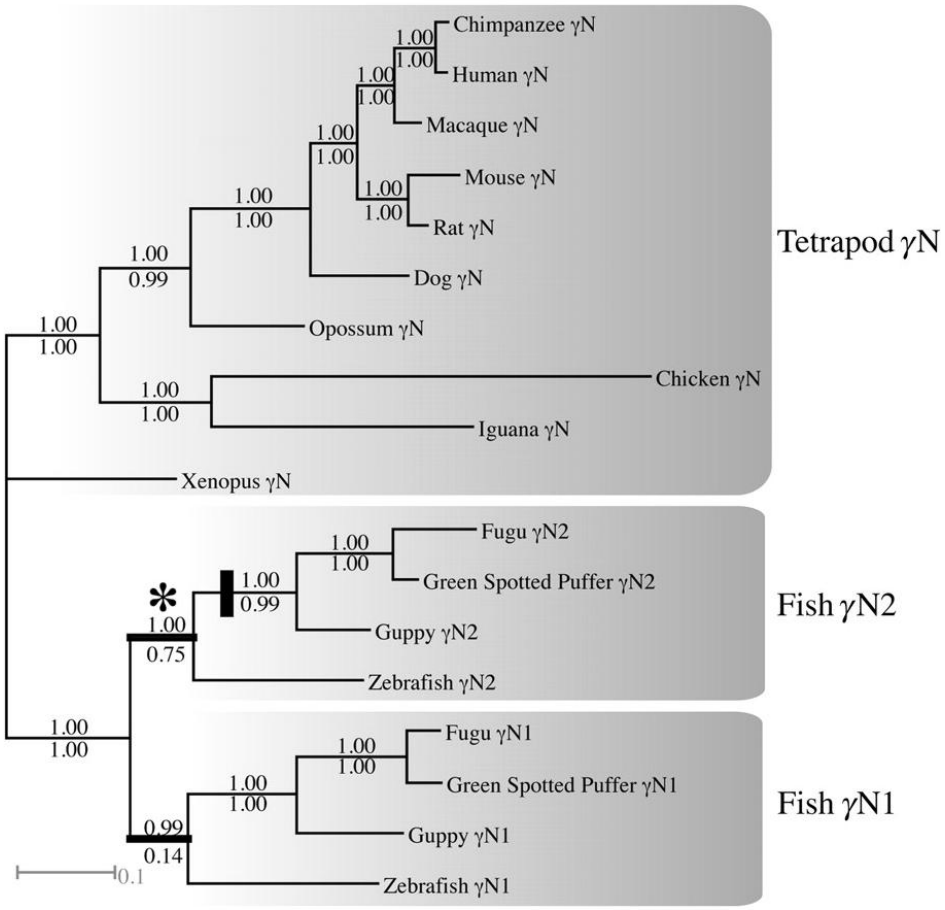
Case 19


A-type (A1)
97.3
C-type (C1)
0.01
35.1
L-type (L1)
53.5
Sakhalin (R1)
52.0
J-type (J1)
##
54.3
F-type (F1)
99.9
99.0
#
F-type (F2)
Conocephalum conicum
S-type (S1)
83.6
S-type (S2)
50.6
S-type (S3)
100
Conocephalum salebrosum
JN-type (JN1)
57.0
JS-type (JS1)
100
CS-type (CS1)
Conocephalum japonicum
Marchantia polymorpha
Physcomitrella patens subsp. patens

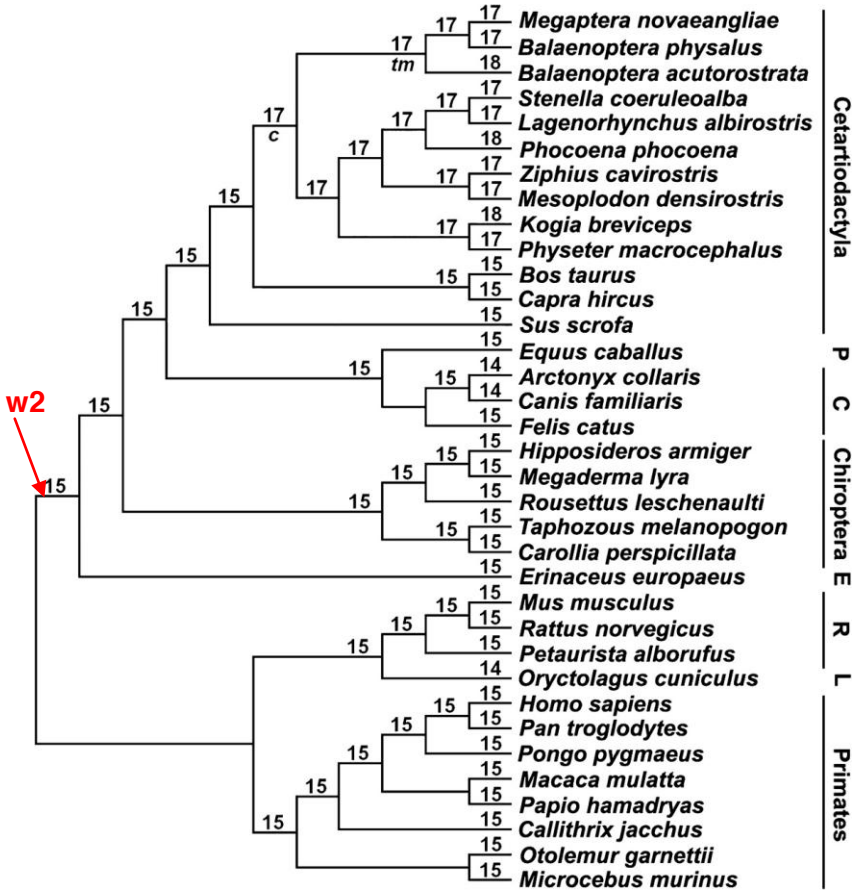
Case 20


Homo sapiens
Bos taurus
Canis lupus
Chaerephon plicatus
Miniopterus fuliginosus
Pipistrellus abramus
Myotis ricketti
Artibeus jamaicensis
Taphozous melanopogon
Megaderma spasma
Acerodon celebensis
Pteropus giganteus
Rousettus leschenaultii
Harpionycteris celebensis
Cynopterus sphinx
Rhinolophus ferrumequinum
Hipposideros armiger
Yangochiroptera
Yinpterochiroptera
0.01 substitutions/site

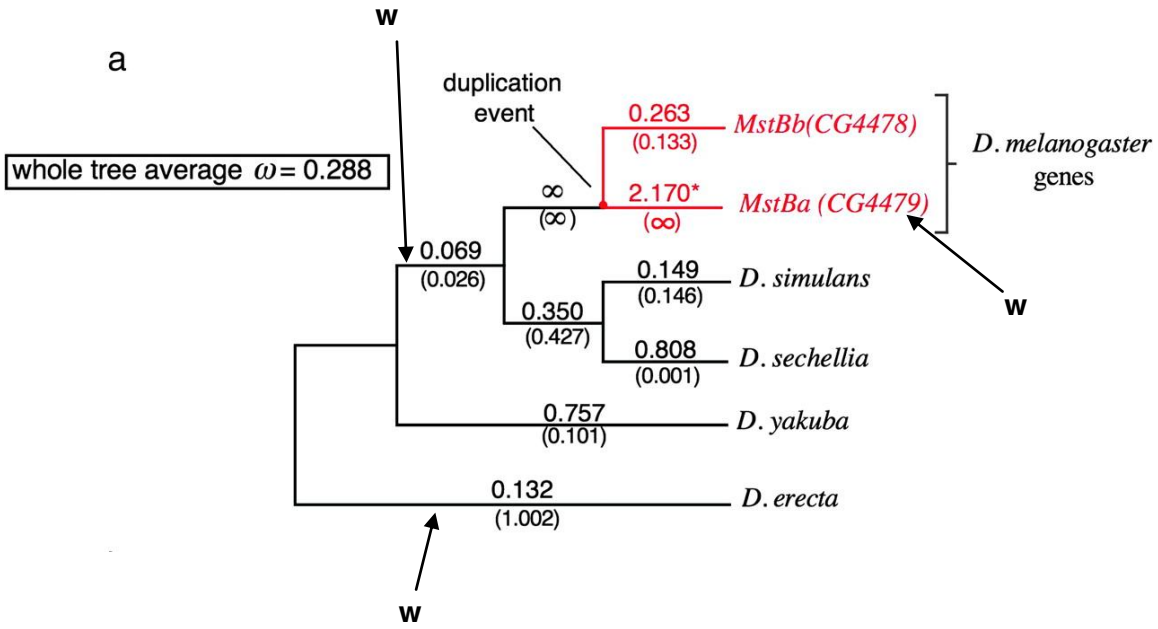
Case 21


Chimpanzee γN
Human γN
Macaque γN
Mouse γN
Rat γN
Dog γN
Opossum γN
Chicken γN
Iguana γN
Xenopus γN
Tetrapod γN
Fugu γN2
Green Spotted Puffer γN2
Guppy γN2
Zebrafish γN2
Fish γN2
Fugu γN1
Green Spotted Puffer γN1
Guppy γN1
Zebrafish γN1
Fish γN1
0.1

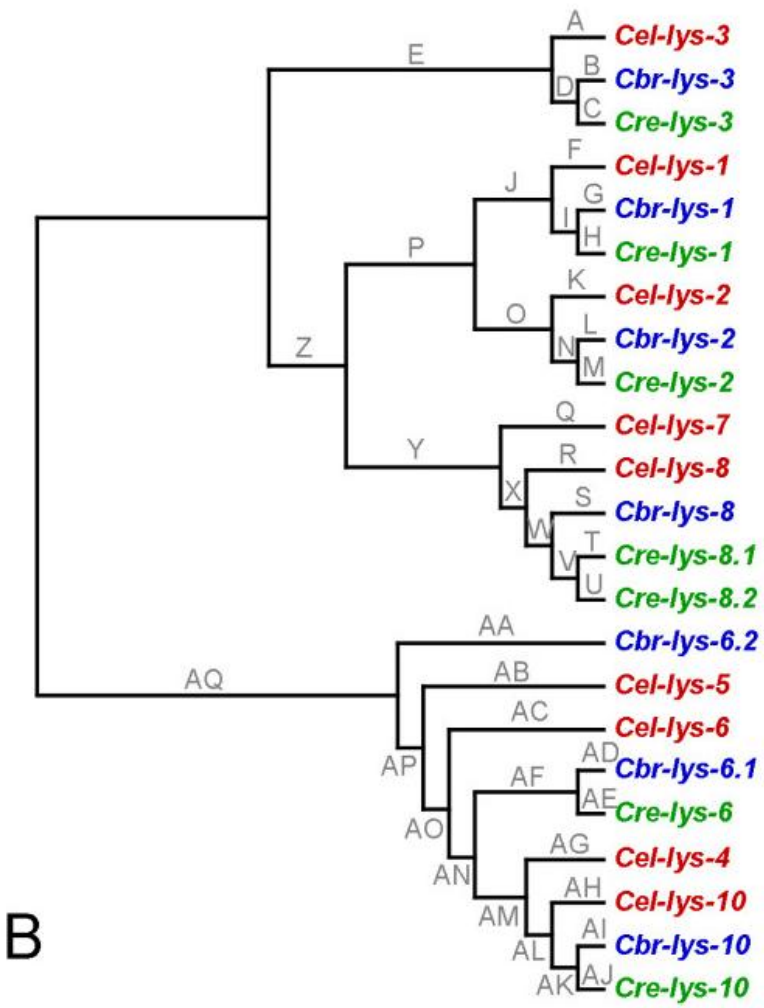
Case 22


w2
Megaptera novaeangliae
Balaenoptera physalus
Balaenoptera acutorostrata
Stenella coeruleoalba
Lagenorhynchus albirostris
Phocoena phocoena
Ziphius cavirostris
Mesoplodon densirostris
Kogia breviceps
Physeter macrocephalus
Bos taurus
Capra hircus
Sus scrofa
Equus caballus
Arctonyx collaris
Canis familiaris
Felis catus
Hipposideros armiger
Megaderma lyra
Rousettus leschenaulti
Taphozous melanopogon
Carollia perspicillata
Erinaceus europaeus
Mus musculus
Rattus norvegicus
Petaurista alborufus
Oryctolagus cuniculus
Homo sapiens
Pan troglodytes
Pongo pygmaeus
Macaca mulatta
Papio hamadryas
Callithrix jacchus
Otolemur garnettii
Microcebus murinus
tm
c
Cetartiodactyla
P
C
Chiroptera
E
R
L
Primates

Case 23

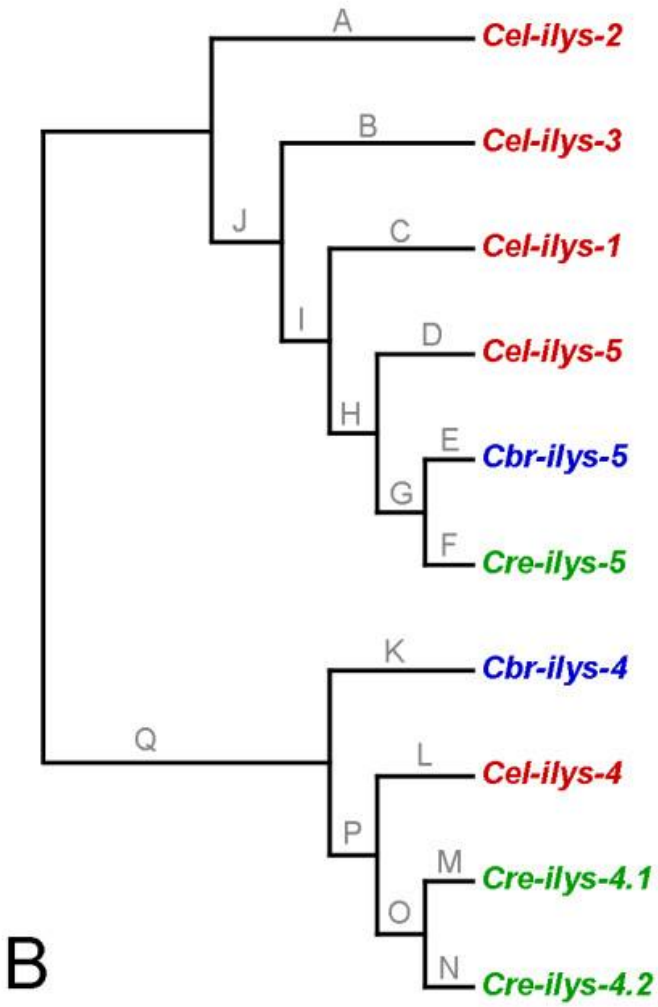

a
w
duplication
event
whole tree average ω = 0.288
0.263
(0.133)
MstBb(CG4478)
D. melanogaster
genes
∞
(∞)
2.170*
(∞)
MstBa (CG4479)
0.069
(0.026)
0.149
(0.146)
D. simulans
w
0.350
(0.427)
0.808
(0.001)
D. sechellia
0.757
(0.101)
D. yakuba
0.132
(1.002)
D. erecta
w

Case 24


A
Cel-lys-3
E
B
D
Cbr-lys-3
C
Cre-lys-3
F
Cel-lys-1
J
G
Cbr-lys-1
I
H
Cre-lys-1
P
K
Cel-lys-2
O
L
Cbr-lys-2
Z
N
M
Cre-lys-2
Q
Cel-lys-7
Y
R
Cel-lys-8
X
S
Cbr-lys-8
W
T
Cre-lys-8.1
V
U
Cre-lys-8.2
AA
Cbr-lys-6.2
AQ
AB
Cel-lys-5
AC
Cel-lys-6
AP
AD
Cbr-lys-6.1
AF
AE
Cre-lys-6
AO
AG
Cel-lys-4
AN
AH
Cel-lys-10
AM
AI
Cbr-lys-10
AL
AJ
AK
Cre-lys-10

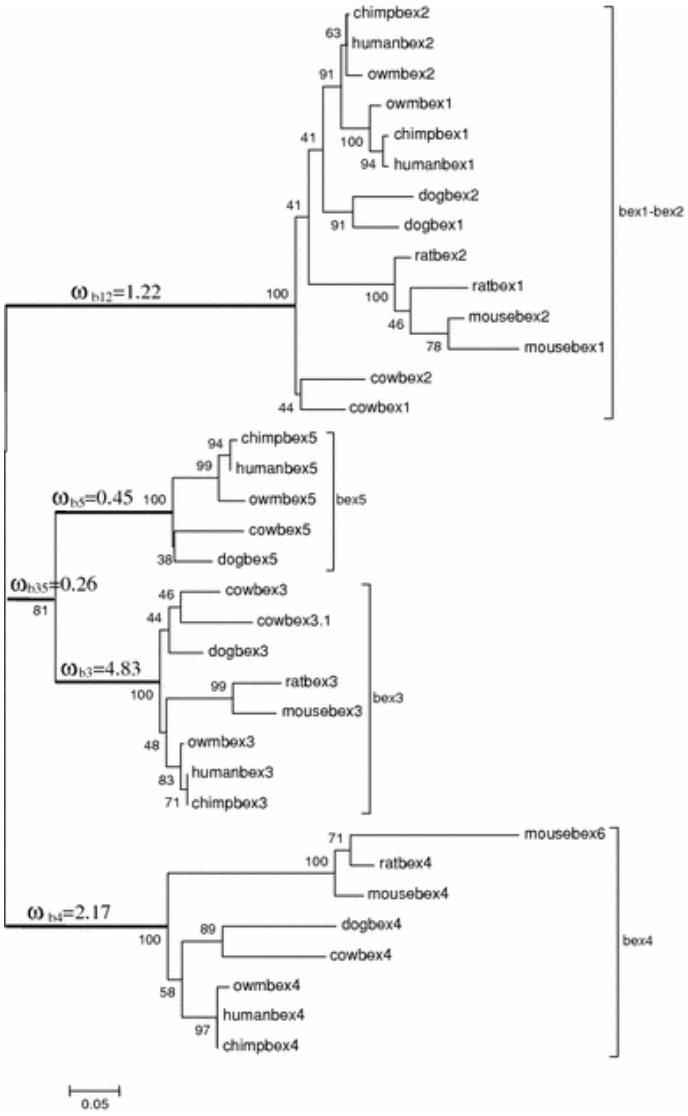
B

Case 25

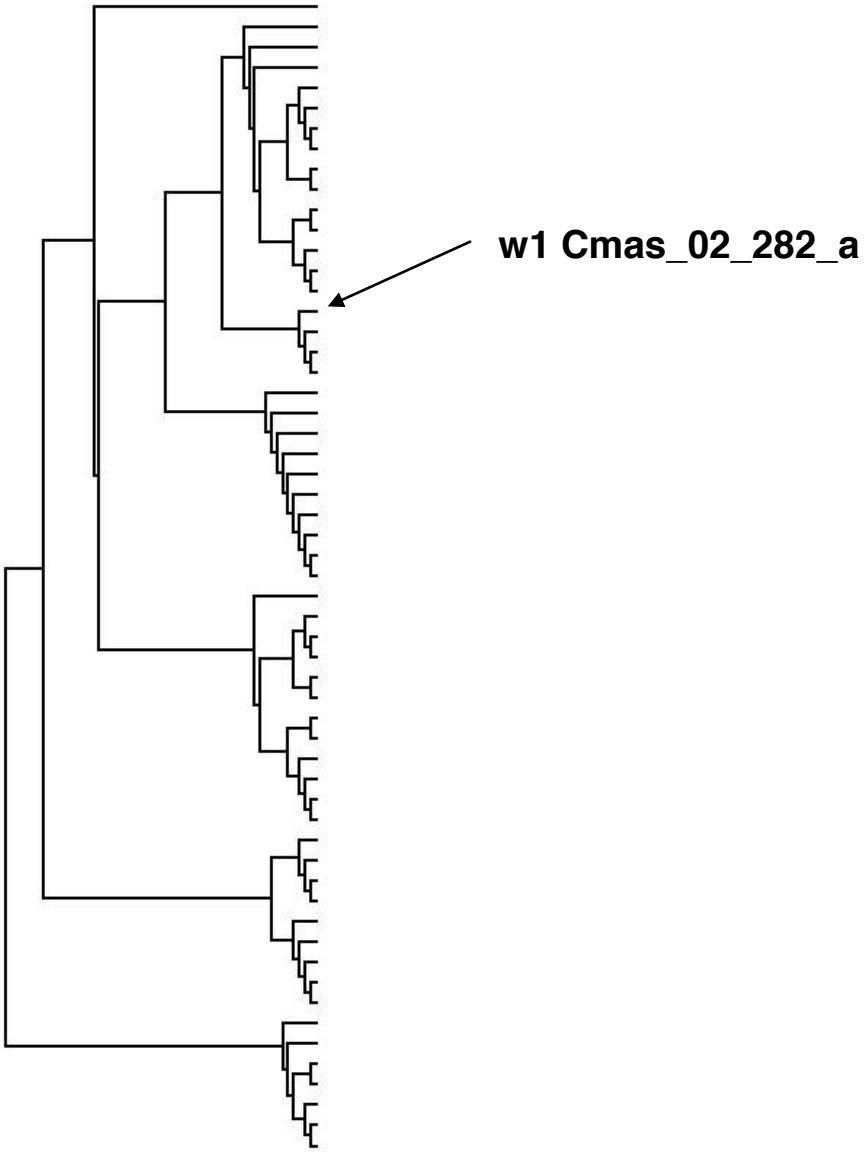

A
Cel-ilys-2
B
Cel-ilys-3
J
C
Cel-ilys-1
I
D
Cel-ilys-5
H
E
Cbr-ilys-5
G
F
Cre-ilys-5
K
Cbr-ilys-4
Q
L
Cel-ilys-4
P
M
Cre-ilys-4.1
O
N
Cre-ilys-4.2

B

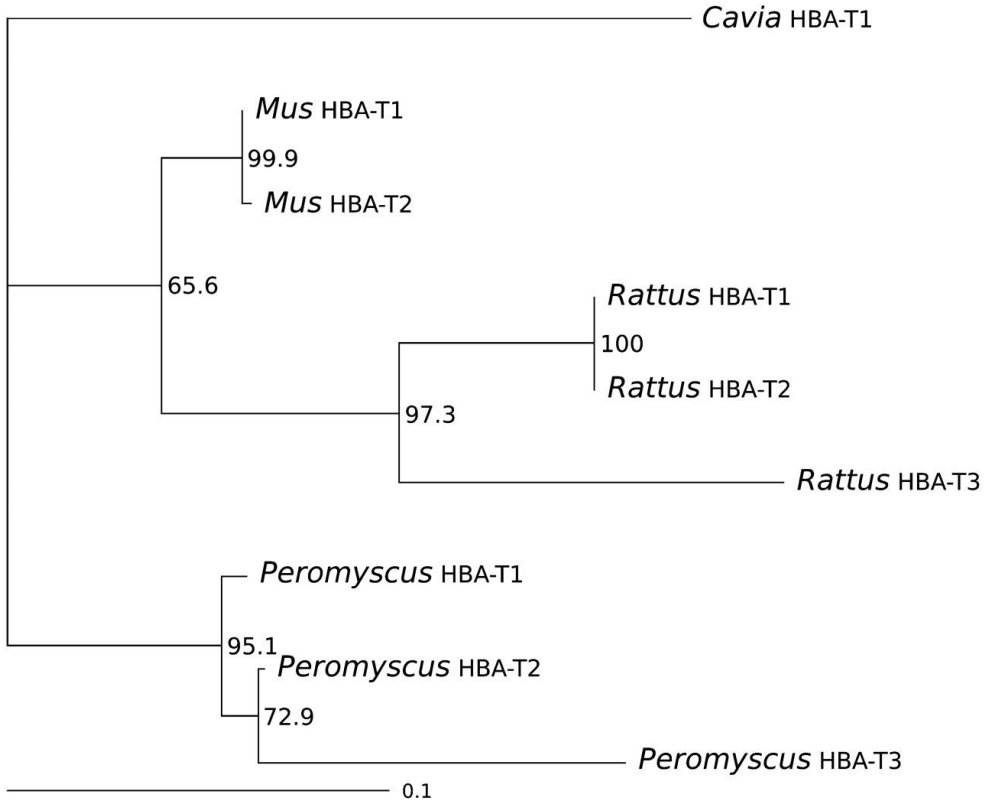
Case 26


chimpbex2
humanbex2
owmbex2
owmbex1
chimpbex1
humanbex1
dogbex2
dogbex1
ratbex2
ratbex1
mousebex2
mousebex1
cowbex2
cowbex1
bex1-bex2
ω b12=1.22
chimpbex5
humanbex5
owmbex5
cowbex5
dogbex5
bex5
ω b5=0.45
ω b35=0.26
cowbex3
cowbex3.1
dogbex3
ratbex3
mousebex3
owmbex3
humanbex3
chimpbex3
bex3
ω b3=4.83
mousebex6
ratbex4
mousebex4
dogbex4
cowbex4
owmbex4
humanbex4
chimpbex4
bex4
ω b4=2.17
0.05

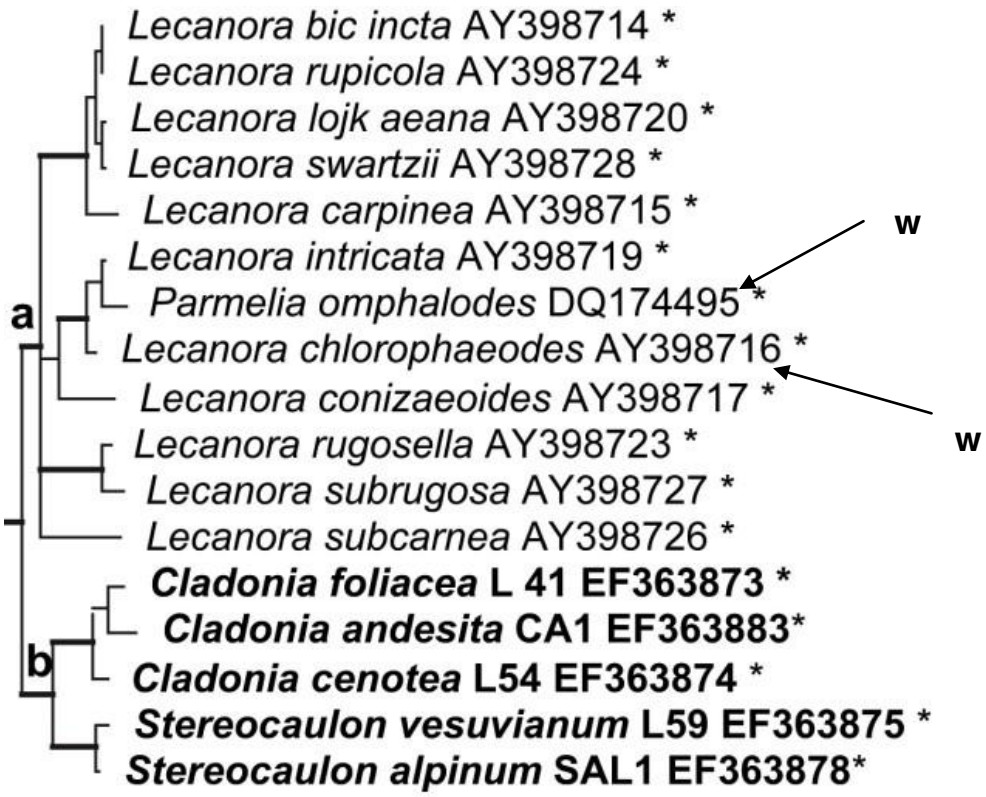
Case 27



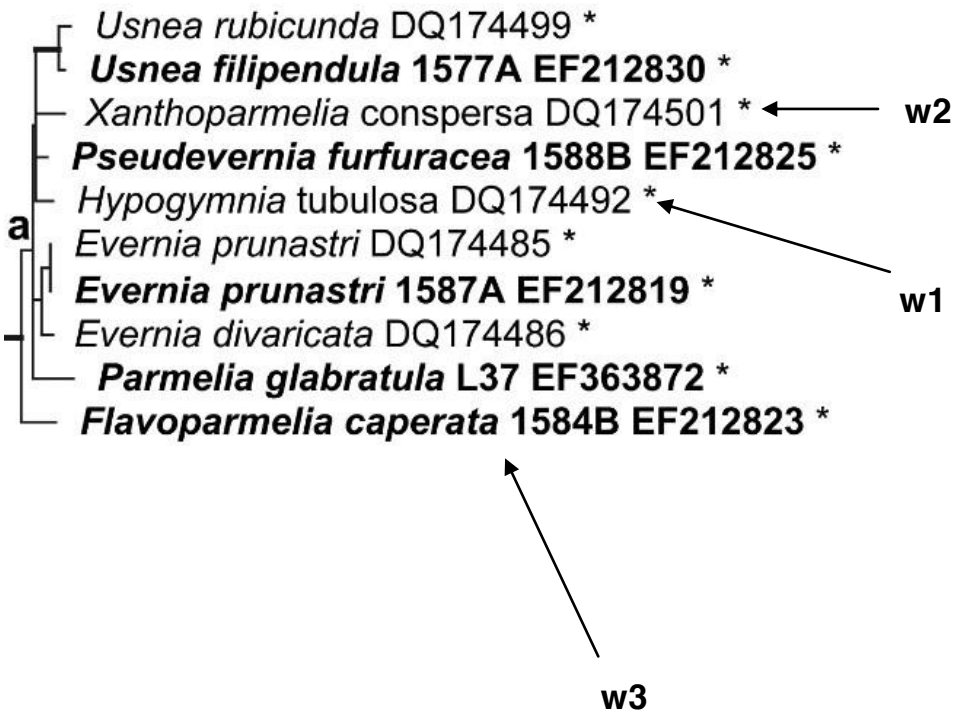
Case 28


Cavia HBA-T1
Mus HBA-T1
99.9
Mus HBA-T2
65.6
Rattus HBA-T1
100
Rattus HBA-T2
97.3
Rattus HBA-T3
Peromyscus HBA-T1
95.1
Peromyscus HBA-T2
72.9
Peromyscus HBA-T3
0.1

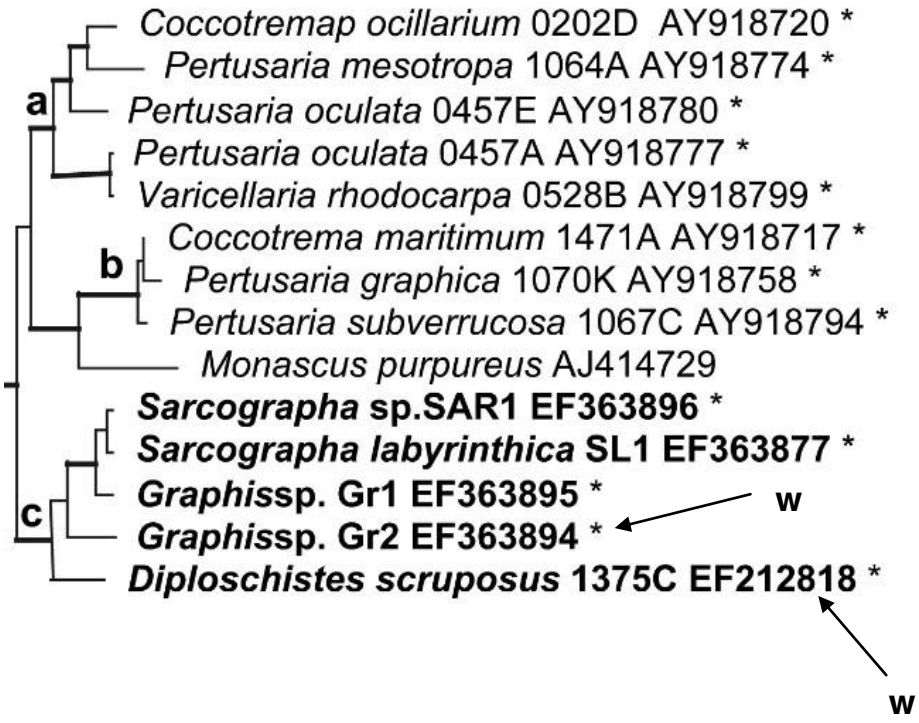
Case29


Lecanora bic incta AY398714 *
Lecanora rupicola AY398724 *
Lecanora lojk aeana AY398720 *
Lecanora swartzii AY398728 *
Lecanora carpinea AY398715 *
Lecanora intricata AY398719 *
Parmelia omphalodes DQ174495 *
Lecanora chlorophaeodes AY398716 *
Lecanora conizaeoides AY398717 *
Lecanora rugosella AY398723 *
Lecanora subrugosa AY398727 *
Lecanora subcarnea AY398726 *
Cladonia foliacea L 41 EF363873 *
Cladonia andesita CA1 EF363883*
Cladonia cenotea L54 EF363874 *
Stereocaulon vesuvianum L59 EF363875 *
Stereocaulon alpinum SAL1 EF363878*
a
b
w
w

Case30


Usnea rubicunda DQ174499 *
***Usnea filipendula* 1577A EF212830** *
Xanthoparmelia conspersa DQ174501 *
***Pseudevernia furfuracea* 1588B EF212825** *
Hypogymnia tubulosa DQ174492 *
Evernia prunastri DQ174485 *
***Evernia prunastri* 1587A EF212819** *
Evernia divaricata DQ174486 *
***Parmelia glabratula* L37 EF363872** *
***Flavoparmelia caperata* 1584B EF212823** *
a
w2
w1
w3

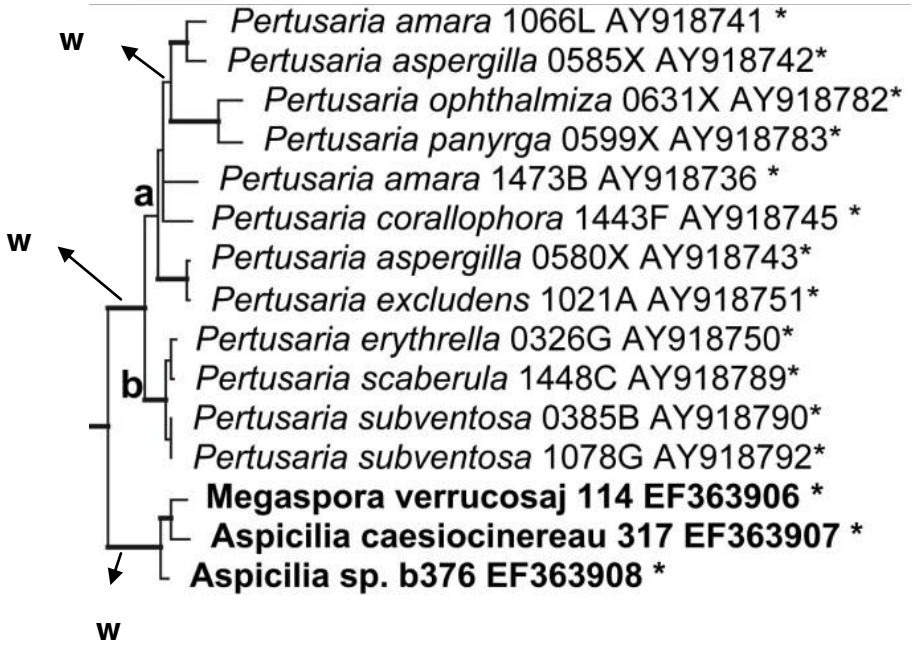
Case31


Coccotremap ocillarium 0202D AY918720 *
Pertusaria mesotropa 1064A AY918774 *
a
Pertusaria oculata 0457E AY918780 *
Pertusaria oculata 0457A AY918777 *
Varicellaria rhodocarpa 0528B AY918799 *
Coccotrema maritimum 1471A AY918717 *
b
Pertusaria graphica 1070K AY918758 *
Pertusaria subverrucosa 1067C AY918794 *
Monascus purpureus AJ414729
Sarcographa sp.SAR1 EF363896 *
Sarcographa labyrinthica SL1 EF363877 *
Graphissp. Gr1 EF363895 *
w
c
Graphissp. Gr2 EF363894 *
Diploschistes scruposus 1375C EF212818 *
w

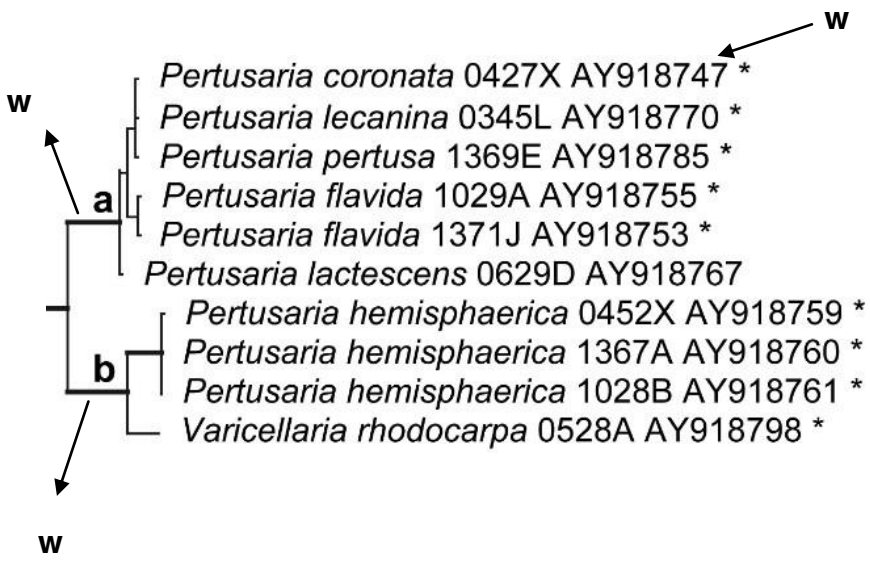
Case32


Lecanora flotowiana AY398718 *
Lecanora nashii AY398722 *
a
Lecanora macrocyclos AY398721 *
Lecanora straminea AY398725 *
Lecidella stigmatea LS1 EF363898 *
Pertusaria amara 1066A AY918339 *
Teloschistes flavicans EF363880*
w1

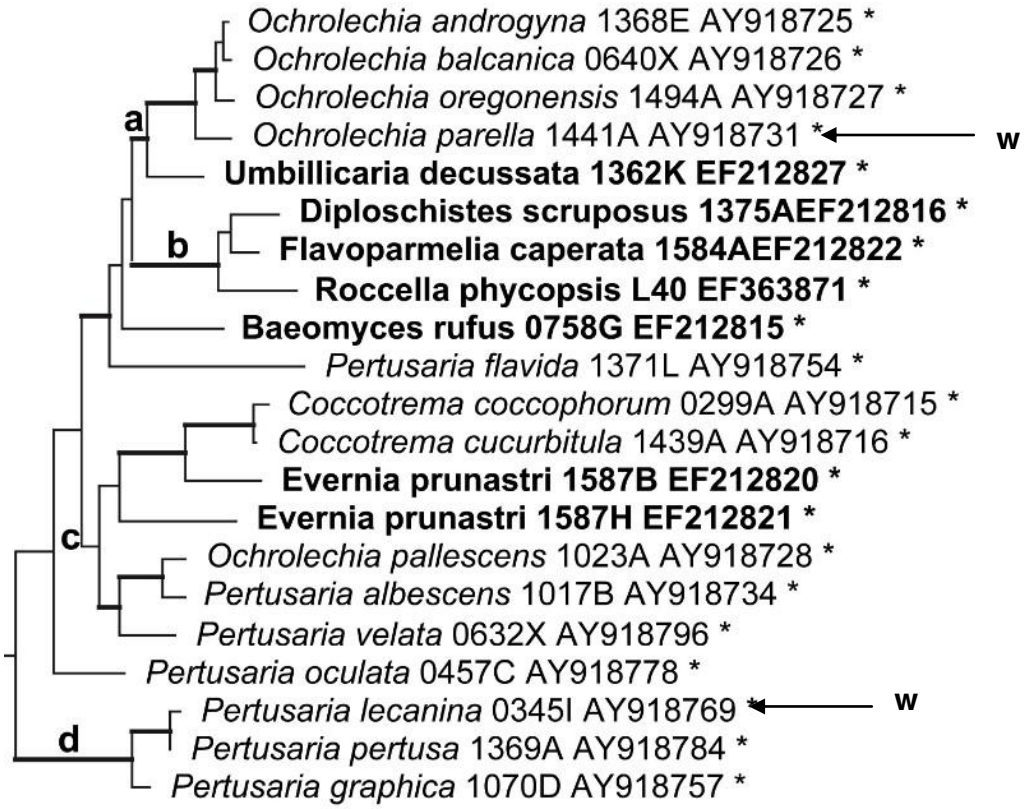
Case 33


w
w
w
a
b
Pertusaria amara 1066L AY918741 *
Pertusaria aspergilla 0585X AY918742*
Pertusaria ophthalmiza 0631X AY918782*
Pertusaria panyrga 0599X AY918783*
Pertusaria amara 1473B AY918736 *
Pertusaria corallophora 1443F AY918745 *
Pertusaria aspergilla 0580X AY918743*
Pertusaria excludens 1021A AY918751*
Pertusaria erythrella 0326G AY918750*
Pertusaria scaberula 1448C AY918789*
Pertusaria subventosa 0385B AY918790*
Pertusaria subventosa 1078G AY918792*
Megaspora verrucosaj 114 EF363906 *
Aspicilia caesiocinereau 317 EF363907 *
Aspicilia sp. b376 EF363908 *

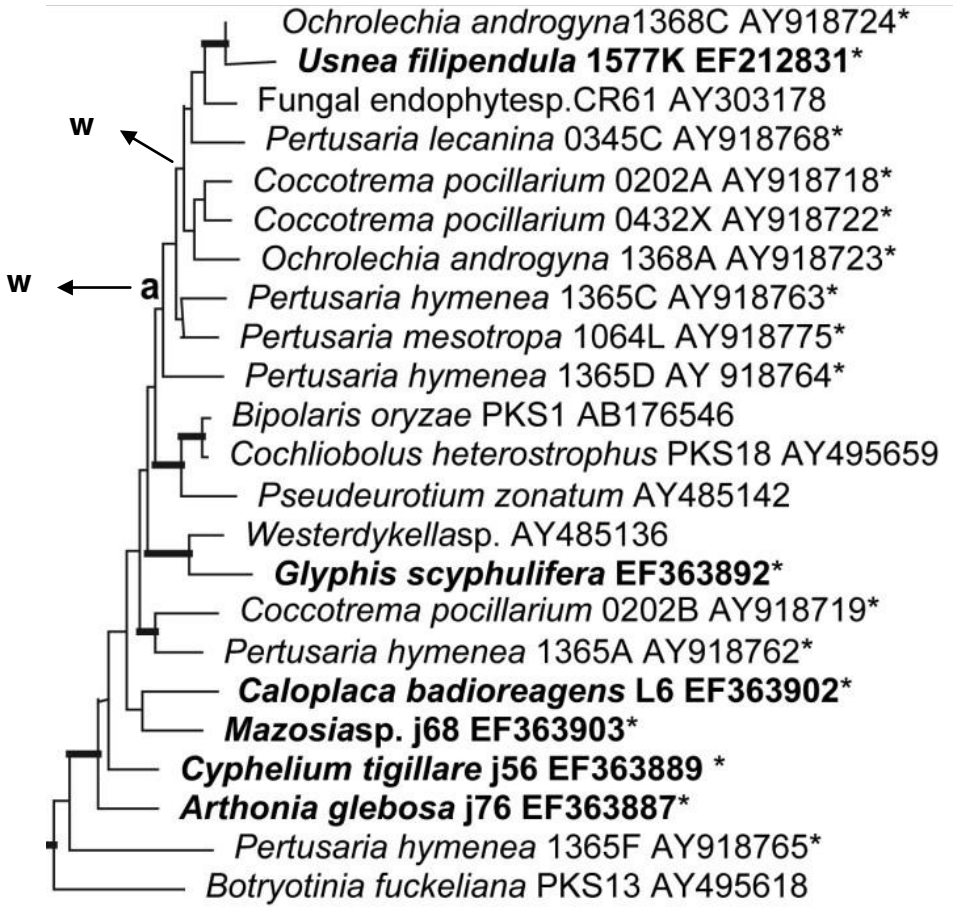
Case34


w
w
Pertusaria coronata 0427X AY918747 *
Pertusaria lecanina 0345L AY918770 *
Pertusaria pertusa 1369E AY918785 *
a
Pertusaria flavida 1029A AY918755 *
Pertusaria flavida 1371J AY918753 *
Pertusaria lactescens 0629D AY918767
Pertusaria hemisphaerica 0452X AY918759 *
Pertusaria hemisphaerica 1367A AY918760 *
b
Pertusaria hemisphaerica 1028B AY918761 *
Varicellaria rhodocarpa 0528A AY918798 *
w

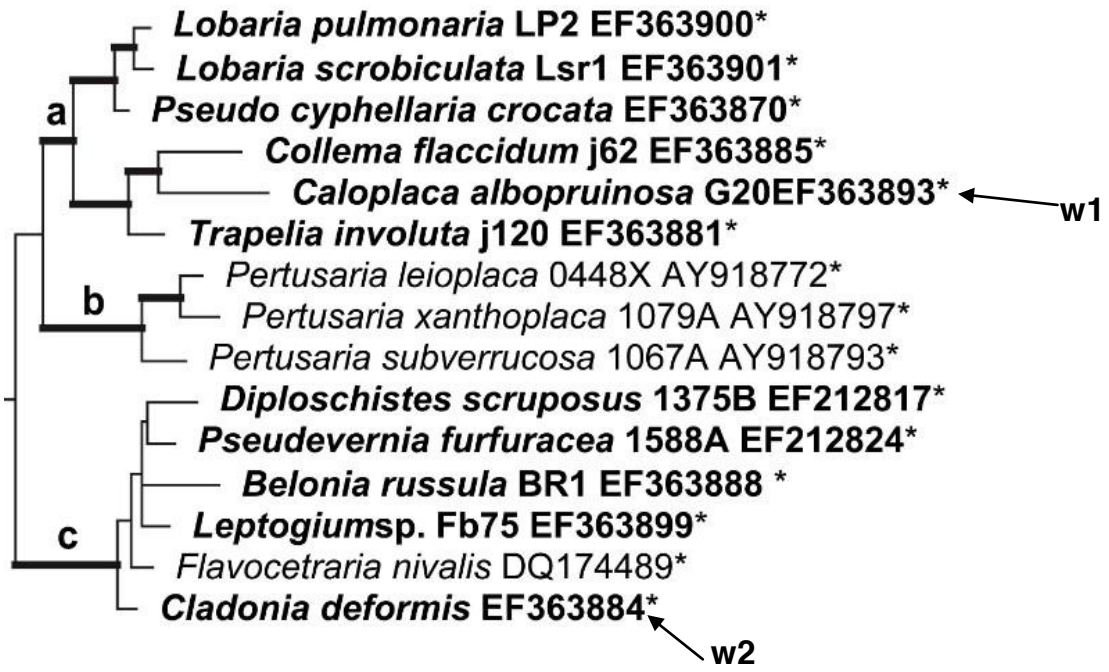
Case35


Ochrolechia androgyna 1368E AY918725 *
Ochrolechia balcanica 0640X AY918726 *
Ochrolechia oregonensis 1494A AY918727 *
Ochrolechia parella 1441A AY918731 *
w
Umbillicaria decussata 1362K EF212827 *
Diploschistes scruposus 1375AEF212816 *
Flavoparmelia caperata 1584AEF212822 *
Roccella phycopsis L40 EF363871 *
Baeomyces rufus 0758G EF212815 *
Pertusaria flavida 1371L AY918754 *
Coccotrema coccophorum 0299A AY918715 *
Coccotrema cucurbitula 1439A AY918716 *
Evernia prunastri 1587B EF212820 *
Evernia prunastri 1587H EF212821 *
Ochrolechia pallescens 1023A AY918728 *
Pertusaria albescens 1017B AY918734 *
Pertusaria velata 0632X AY918796 *
Pertusaria oculata 0457C AY918778 *
Pertusaria lecanina 0345I AY918769 *
w
Pertusaria pertusa 1369A AY918784 *
Pertusaria graphica 1070D AY918757 *
a
b
c
d

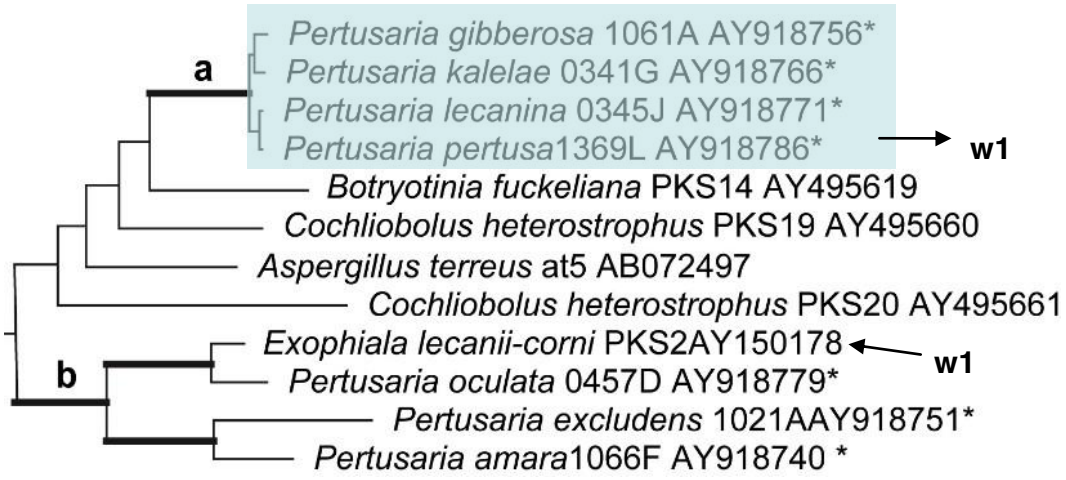
Case36


Ochrolechia androgyna1368C AY918724*
Usnea filipendula 1577K EF212831*
Fungal endophytesp.CR61 AY303178
Pertusaria lecanina 0345C AY918768*
Coccotrema pocillarium 0202A AY918718*
Coccotrema pocillarium 0432X AY918722*
Ochrolechia androgyna 1368A AY918723*
Pertusaria hymenea 1365C AY918763*
Pertusaria mesotropa 1064L AY918775*
Pertusaria hymenea 1365D AY 918764*
Bipolaris oryzae PKS1 AB176546
Cochliobolus heterostrophus PKS18 AY495659
Pseudeurotium zonatum AY485142
Westerdykellasp. AY485136
Glyphis scyphulifera EF363892*
Coccotrema pocillarium 0202B AY918719*
Pertusaria hymenea 1365A AY918762*
Caloplaca badioreagens L6 EF363902*
Mazosiasp. j68 EF363903*
Cyphelium tigillare j56 EF363889 *
Arthonia glebosa j76 EF363887*
Pertusaria hymenea 1365F AY918765*
Botryotinia fuckeliana PKS13 AY495618
w
w
a

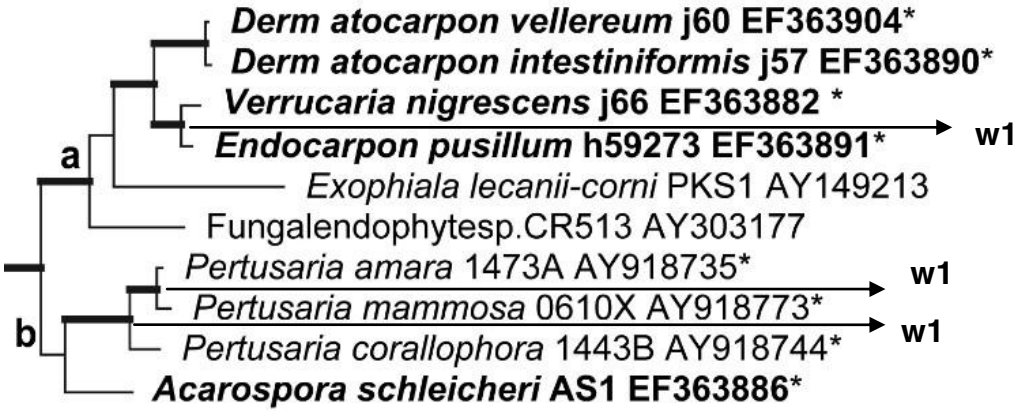
Case 37


Lobaria pulmonaria LP2 EF363900*
Lobaria scrobiculata Lsr1 EF363901*
a
Pseudo cyphellaria crocata EF363870*
Collema flaccidum j62 EF363885*
Caloplaca albopruinosa G20EF363893*
w1
Trapelia involuta j120 EF363881*
Pertusaria leioplaca 0448X AY918772*
b
Pertusaria xanthoplaca 1079A AY918797*
Pertusaria subverrucosa 1067A AY918793*
Diploschistes scruposus 1375B EF212817*
Pseudevernia furfuracea 1588A EF212824*
Belonia russula BR1 EF363888 *
Leptogiumsp. Fb75 EF363899*
c
Flavocetraria nivalis DQ174489*
Cladonia deformis EF363884*
w2

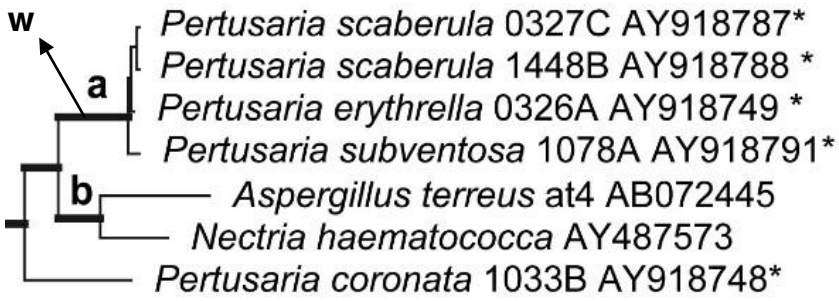
Case 38


a
Pertusaria gibberosa 1061A AY918756*
Pertusaria kalelae 0341G AY918766*
Pertusaria lecanina 0345J AY918771*
Pertusaria pertusa1369L AY918786*
w1
Botryotinia fuckeliana PKS14 AY495619
Cochliobolus heterostrophus PKS19 AY495660
Aspergillus terreus at5 AB072497
Cochliobolus heterostrophus PKS20 AY495661
Exophiala lecanii-corni PKS2AY150178
w1
b
Pertusaria oculata 0457D AY918779*
Pertusaria excludens 1021AAY918751*
Pertusaria amara1066F AY918740 *

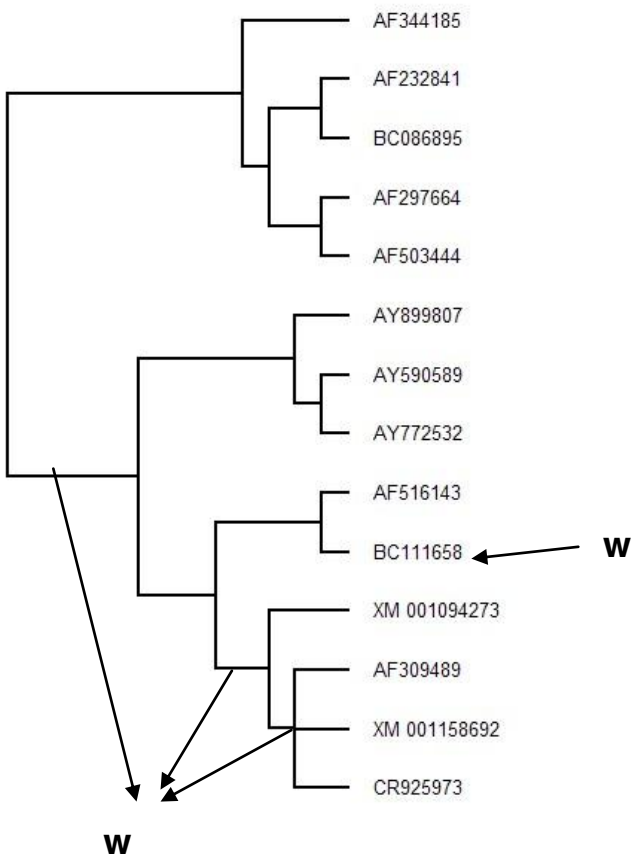
Case 39


Derm atocarpon vellereum j60 EF363904*
Derm atocarpon intestiniformis j57 EF363890*
Verrucaria nigrescens j66 EF363882 *
Endocarpon pusillum h59273 EF363891*
w1
a
Exophiala lecanii-corni PKS1 AY149213
Fungalendophytesp.CR513 AY303177
Pertusaria amara 1473A AY918735*
w1
Pertusaria mammosa 0610X AY918773*
w1
b
Pertusaria corallophora 1443B AY918744*
Acarospora schleicheri AS1 EF363886*

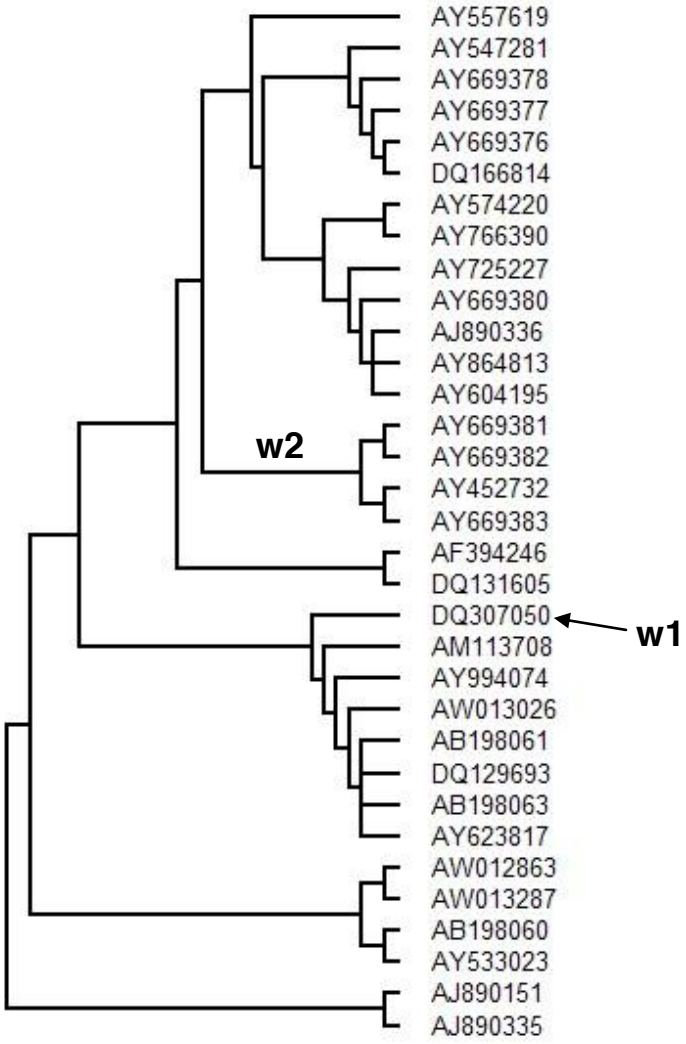
Case 40


w
a
b
Pertusaria scaberula 0327C AY918787*
Pertusaria scaberula 1448B AY918788 *
Pertusaria erythrella 0326A AY918749 *
Pertusaria subventosa 1078A AY918791*
Aspergillus terreus at4 AB072445
Nectria haematococca AY487573
Pertusaria coronata 1033B AY918748*

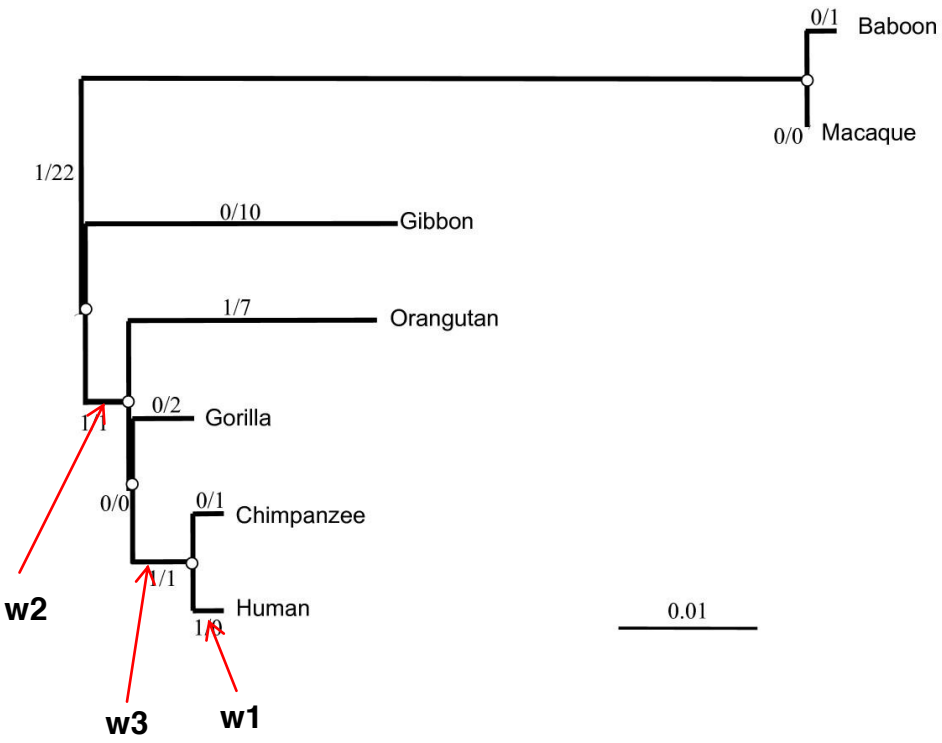
Case 41


AF344185
AF232841
BC086895
AF297664
AF503444
AY899807
AY590589
AY772532
AF516143
BC111658
XM 001094273
AF309489
XM 001158692
CR925973
w
w

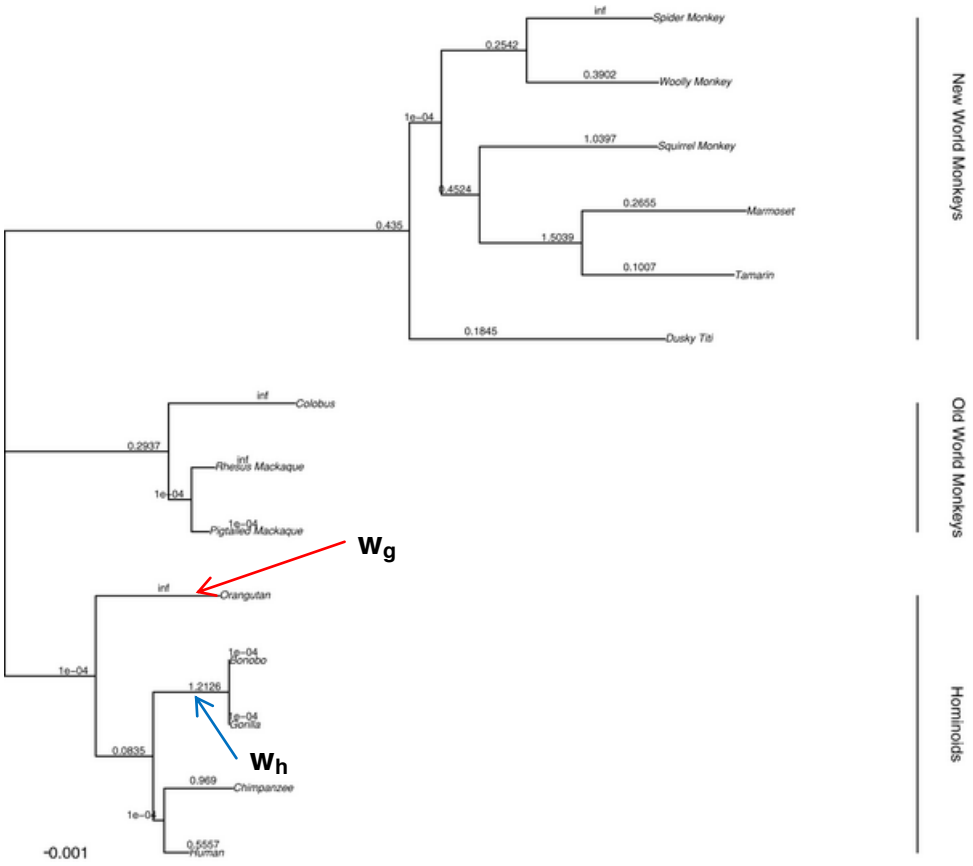
Case 42


AY557619
AY547281
AY669378
AY669377
AY669376
DQ166814
AY574220
AY766390
AY725227
AY669380
AJ890336
AY864813
AY604195
AY669381
AY669382
AY452732
AY669383
AF394246
DQ131605
DQ307050
AM113708
AY994074
AW013026
AB198061
DQ129693
AB198063
AY623817
AW012863
AW013287
AB198060
AY533023
AJ890151
AJ890335
w2
w1

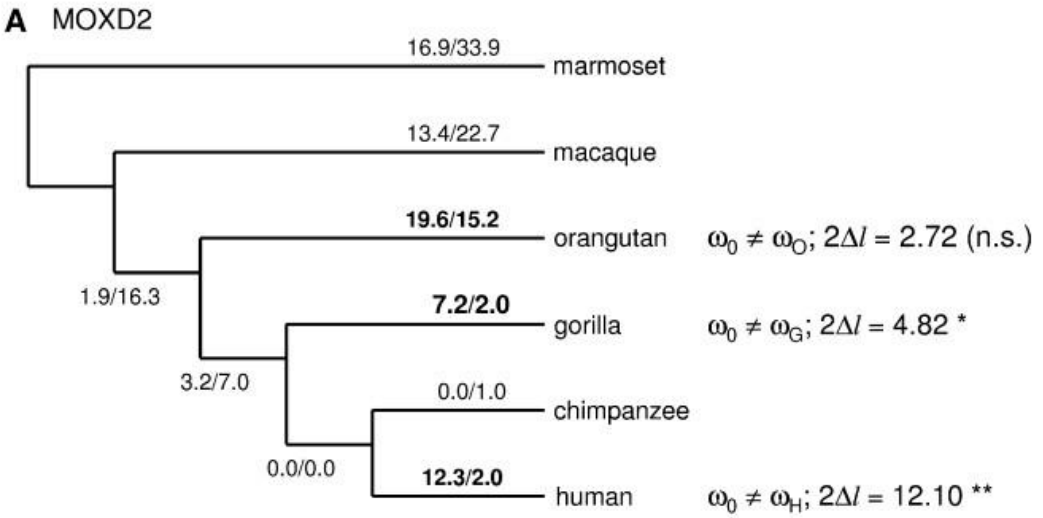
Case 43


0/1
Baboon
0/0
Macaque
1/22
0/10
Gibbon
1/7
Orangutan
1/1
0/2
Gorilla
0/0
0/1
Chimpanzee
1/1
Human
1/0
0.01
w2
w3
w1

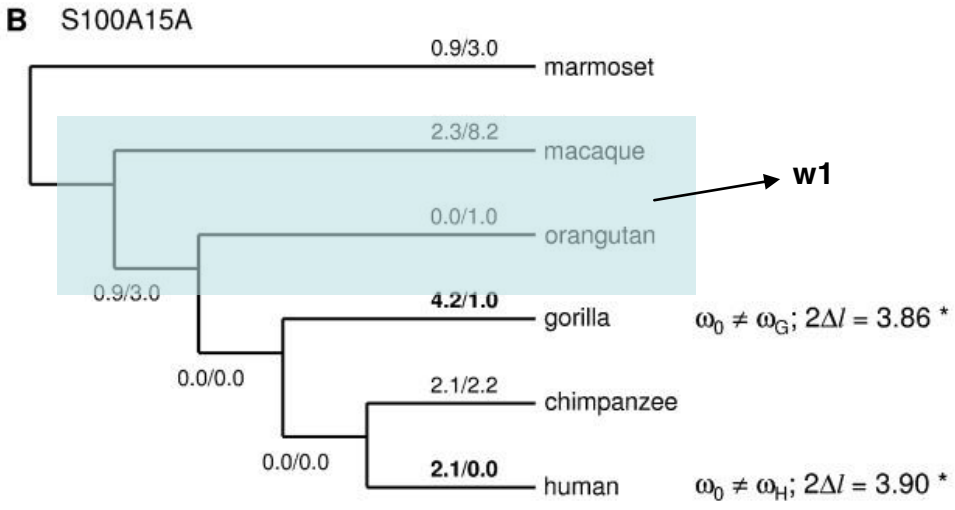
Case 44


inf
Spider Monkey
0.2542
0.3902
Woolly Monkey
1e−04
1.0397
Squirrel Monkey
0.4524
0.2655
Marmoset
0.435
1.5039
0.1007
Tamarin
0.1845
Dusky Titi
New World Monkeys
inf
Colobus
0.2937
inf
Rhesus Mackaque
1e−04
1e−04
Pigtailed Mackaque
Old World Monkeys
w_g
inf
Orangutan
1e−04
1e−04
Bonobo
1.2126
1e−04
Gorilla
0.0835
w_h
0.969
Chimpanzee
1e−04
0.5557
Human
Hominoids
-0.001

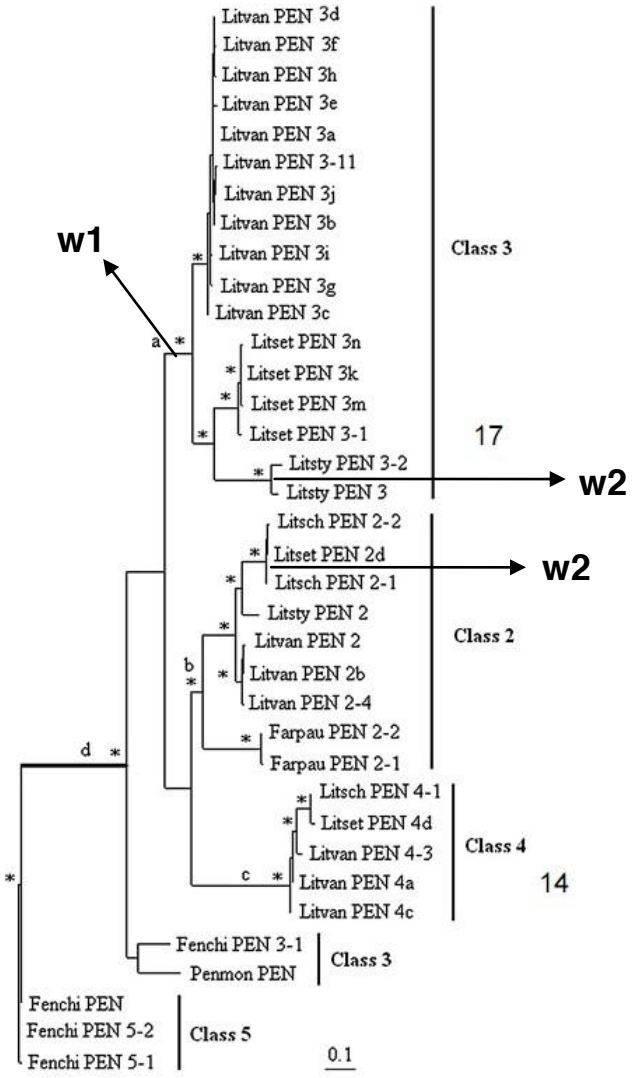
Case 45


A MOXD2
16.9/33.9
marmoset
13.4/22.7
macaque
19.6/15.2
orangutan
ω0 ≠ ωO; 2Δl = 2.72 (n.s.)
1.9/16.3
7.2/2.0
gorilla
ω0 ≠ ωG; 2Δl = 4.82 *
3.2/7.0
0.0/1.0
chimpanzee
0.0/0.0
12.3/2.0
human
ω0 ≠ ωH; 2Δl = 12.10 **

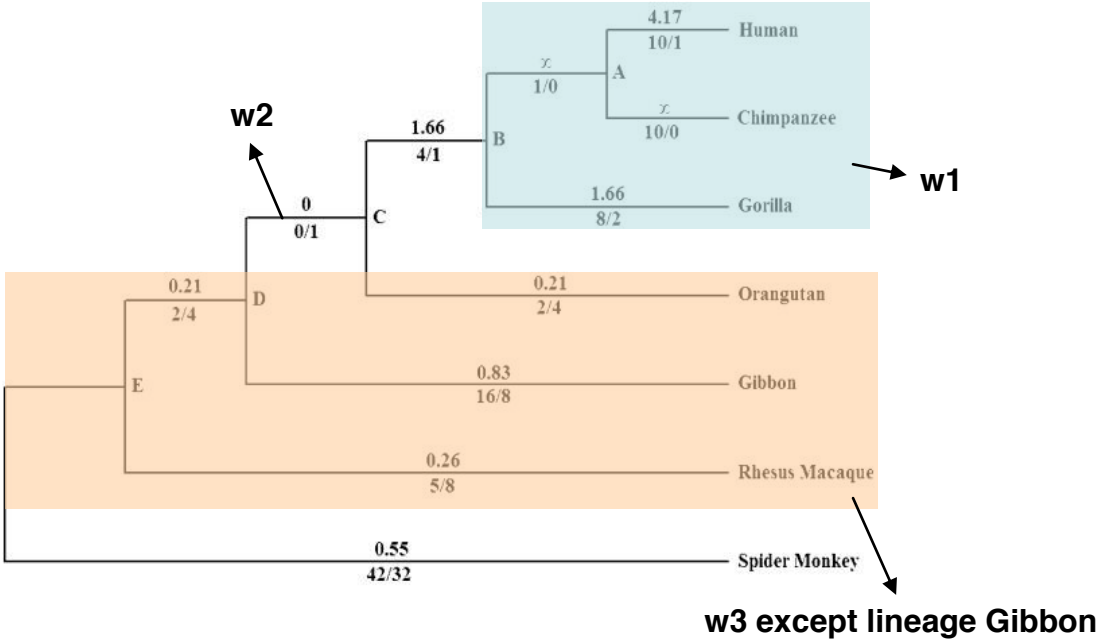
Case 46


B S100A15A
0.9/3.0
marmoset
2.3/8.2
macaque
w1
0.0/1.0
orangutan
0.9/3.0
4.2/1.0
gorilla
$\omega_0 \neq \omega_G$; $2\Delta l = 3.86$ *
0.0/0.0
2.1/2.2
chimpanzee
0.0/0.0
2.1/0.0
human
$\omega_0 \neq \omega_H$; $2\Delta l = 3.90$ *

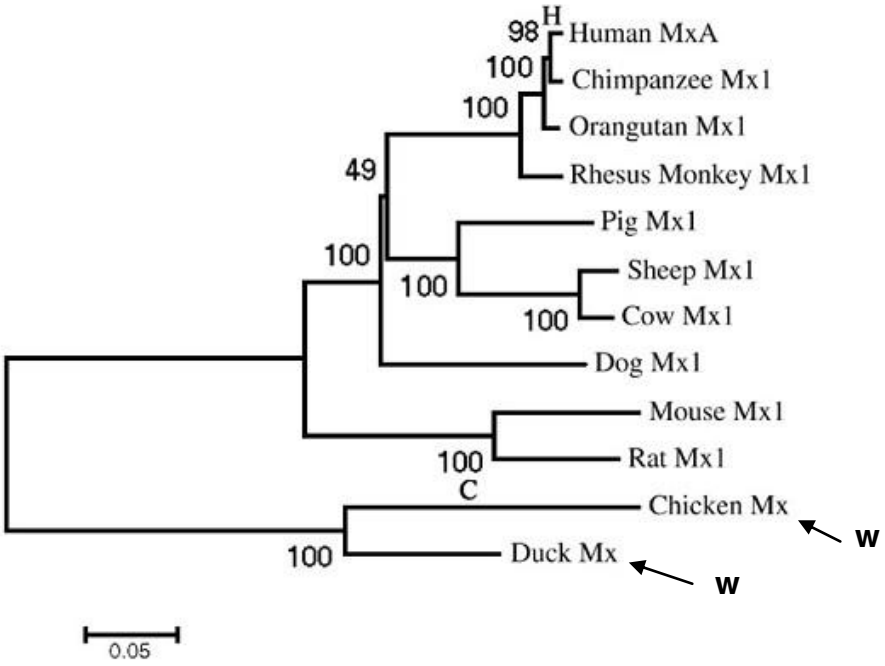
Case 47


Litvan PEN 3d
Litvan PEN 3f
Litvan PEN 3h
Litvan PEN 3e
Litvan PEN 3a
Litvan PEN 3-11
Litvan PEN 3j
Litvan PEN 3b
Litvan PEN 3i
Litvan PEN 3g
Litvan PEN 3c
Class 3
w1
a
Litset PEN 3n
Litset PEN 3k
Litset PEN 3m
Litset PEN 3-1
17
Litsty PEN 3-2
Litsty PEN 3
w2
Litsch PEN 2-2
Litset PEN 2d
w2
Litsch PEN 2-1
Litsty PEN 2
Litvan PEN 2
Litvan PEN 2b
Litvan PEN 2-4
Class 2
b
Farpau PEN 2-2
Farpau PEN 2-1
d
Litsch PEN 4-1
Litset PEN 4d
Litvan PEN 4-3
Litvan PEN 4a
Litvan PEN 4c
Class 4
14
c
Fenchi PEN 3-1
Penmon PEN
Class 3
Fenchi PEN
Fenchi PEN 5-2
Fenchi PEN 5-1
Class 5
0.1

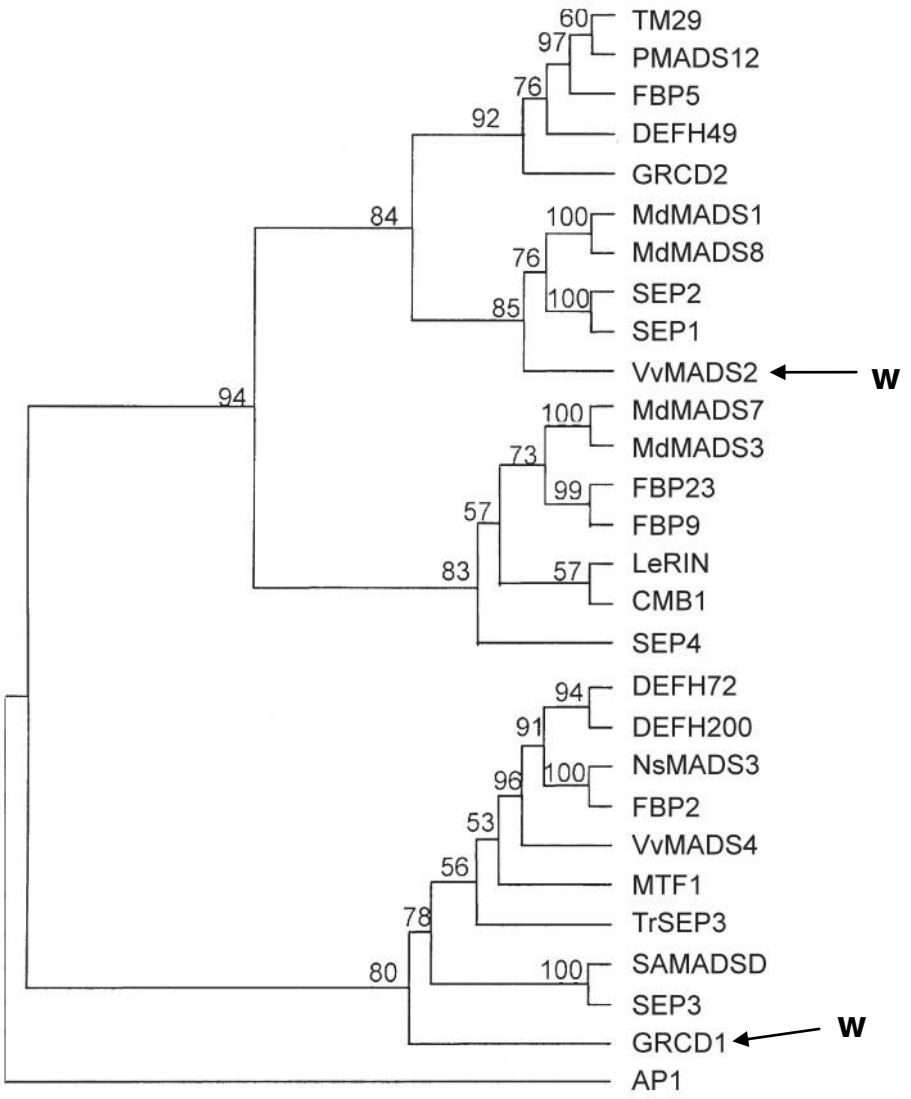
Case 48


4.17
10/1
Human
χ
1/0
A
χ
10/0
Chimpanzee
w2
1.66
4/1
B
w1
0
0/1
C
1.66
8/2
Gorilla
0.21
2/4
D
0.21
2/4
Orangutan
E
0.83
16/8
Gibbon
0.26
5/8
Rhesus Macaque
0.55
42/32
Spider Monkey
w3 except lineage Gibbon

Case 49


H
98
Human MxA
100
Chimpanzee Mx1
100
Orangutan Mx1
Rhesus Monkey Mx1
49
Pig Mx1
100
Sheep Mx1
100
Cow Mx1
100
Dog Mx1
Mouse Mx1
100
Rat Mx1
C
Chicken Mx
w
100
Duck Mx
w
0.05

Case 50


60
97
76
92
84
100
76
100
85
94
100
73
99
57
83
57
94
91
100
96
53
56
78
100
80
TM29
PMADS12
FBP5
DEFH49
GRCD2
MdMADS1
MdMADS8
SEP2
SEP1
VvMADS2
w
MdMADS7
MdMADS3
FBP23
FBP9
LeRIN
CMB1
SEP4
DEFH72
DEFH200
NsMADS3
FBP2
VvMADS4
MTF1
TrSEP3
SAMADSD
SEP3
GRCD1
w
AP1