BIOINFORMATICS ORIGINAL PAPER

Vol. 21 no. 4 2005, pages 456–463
doi:10.1093/bioinformatics/bti191

RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees

A. Stamatakis[1,], T. Ludwig[2] and H. Meier[1]*

[1]Department of Computer Science, Technical University of Munich, Boltzmannstrasse 3, D-85748 München, Germany and [2]Department of Computer Science, University of Heidelberg, Im Neuenheimer Feld 348, D-69120 Heidelberg, Germany

Received on May 21, 2004; accepted on November 11, 2004
Advance Access publication December 17, 2004

ABSTRACT

Motivation: The computation of large phylogenetic trees with statistical models such as maximum likelihood or bayesian inference is computationally extremely intensive. It has repeatedly been demonstrated that these models are able to recover the true tree or a tree which is topologically closer to the true tree more frequently than less elaborate methods such as parsimony or neighbor joining. Due to the combinatorial and computational complexity the size of trees which can be computed on a Biologist's PC workstation within reasonable time is limited to trees containing approximately 100 taxa.

Results: In this paper we present the latest release of our program RAxML-III for rapid maximum likelihood-based inference of large evolutionary trees which allows for computation of 1.000-taxon trees in less than 24 hours on a single PC processor. We compare RAxML-III to the currently fastest implementations for maximum likelihood and bayesian inference: PHYML and MrBayes. Whereas RAxML-III performs worse than PHYML and MrBayes on synthetic data it clearly outperforms both programs on all real data alignments used in terms of speed and final likelihood values.

Availability Supplementary information: RAxML-III including all alignments and final trees mentioned in this paper is freely available as open source code at http://wwwbode.cs.tum/~stamatak

Contact: stamatak@cs.tum.edu

INTRODUCTION

In recent years there has been an astonishing accumulation of genetic information for many different organisms. This information can be used to infer evolutionary relationships (called a *phylogenetic tree* or *phylogeny*) among a collection of species. There are a variety of techniques that are used to compute these relationships, including the use of maximum likelihood (Felsenstein, 1981) which among bayesian methods is considered to represent one of the currently most accurate models. A useful review of traditional and bayesian approaches is available from Holder and Lewis (2003). Unfortunately, the number of possible tree topologies grows exponentially with the number of taxa and the computational cost of the likelihood function itself is high. Thus, the introduction of heuristics to reduce the search space in terms of potential tree topologies evaluated becomes inevitable for the computation of trees containing more than 15 to 20 organisms. However, heuristics for maximum likelihood-based phylogenetic tree calculations still remain computationally intensive, mainly due to the high cost of the likelihood function, which is invoked repeatedly for each analyzed tree topology.

Thus, to date only relatively small maximum likelihood-based trees could be computed on parallel computers: a 150-taxon tree with parallel fastDNAml (Stewart *et al*., 2001), and a 228-taxon tree using a parallel genetic algorithm (Brauer *et al*., 2002). However, large data alignments containing valuable phylogenetic information are available for example in the ARB (Ludwig *et al*., 2004) ssu rRNA (small subunit ribosomal RiboNucleic Acid) database which presently contains more than 30.000 sequences. Recently, a GRID-enabled version of fastDNAml has been used on a large alignment to compete in the High Performance Computing Challenge at Supercomputing 2003 conference (see http://www.sc-conference.org/sc2003/tech_hpc.php for details). We have however not been able to obtain the alignment or information about the size of the analysis.

In previous work (Stamatakis *et al*., 2002) we have introduced Subtree Equality Vectors (SEVs) to significantly accelerate the topology evaluation function of maximum likelihood-based phylogeny programs. We implemented SEVs in PAxML (Parallel A(x)ccelerated Maximum Likelihood), which was derived from parallel fastDNAml (Stewart *et al*., 2001). PAxML shows run time improvements of approximately 25% to 65% compared to parallel fastDNAml and yields exactly identical results at the same time.

*To whom correspondence should be addressed.

PAxML shows best accelerations for large alignments ($\geq$150 sequences) on inexpensive PC processor architectures.

One main goal of current work on RAxML-III (Randomized A(x)ccelerated Maximum Likelihood) is to obtain equally good or better likelihood values than PAxML and comparable state-of-the-art sequential programs in less time by deployment of improved search space heuristics. Another key objective is to enhance RAxML-III by a greater variety of evolutionary models and maximum likelihood-based estimation of model parameters. Finally, the RAxML-III algorithm is designed to allow for the implementation of relatively coarse-grained distributed and parallel (Stamatakis *et al*., 2004b) versions which do not rely on expensive hardware platforms. The parallel implementation of the preceding program version (RAxML-II) has been used to infer a 10.000-taxon phylogeny on a medium size PC cluster (Stamatakis *et al*., 2004b).

RELATED WORK

A recent comparative survey (Williams and Moret, 2003) covers an important range of widely-used state-of-the-art statistical phylogeny programs such as fastDNAml (Olsen *et al*., 1994), MrBayes (Huelsenbeck and Ronquist, 2001), PAUP* (Swofford, 1999), and treepuzzle (Strimmer and Haeseler, 1996). The most important result of this paper is that MrBayes outperforms all other analyzed phylogeny programs in terms of speed and tree quality. MrBayes is a program for bayesian analysis of phylogenetic trees. However, this survey is entirely based on synthetic (simulated) data. As the results of this paper show additional experiments with real data can lead to distinct conclusions and a more differentiated image. Furthermore, the largest alignments of this survey contained only 60 sequences. Thus, the results do not necessarily apply to inference of large trees based on real data sets. In addition, this survey does not cover genetic algorithms (Lewis, 1998) which generally converge faster than MrBayes (Guindon and Gascuel, 2003).

More recently, Guindon and Gascuel (2003) published a paper about their new program PHYML, which is very fast and outperforms other recent approaches including MrBayes and genetic algorithms such as MetaPIGA (Lemmon and Milinkovitch, 2002) which -to the best of our knowledge- currently represents the most efficient genetic algorithm for phylogenetic analysis. Like RAxML-III, PHYML is a 'traditional' maximum likelihood program which seeks to find the optimal topology in respect to the likelihood value and is also capable of optimizing model parameters. The PHYML publication includes a comparative survey based on two large real world data sets comprising 218 and 500 taxa, as well as on 50 synthetic 100-taxon alignments.

Thus, -to the best of our knowledge- MrBayes and PHYML are currently the fastest and most accurate representatives of bayesian and 'traditional' approaches to phylogenetic tree inference using statistical models of nucleotide substitution. Therefore, the focus is on those two programs for assessing performance of RAxML-III within the context of this paper. Comparative surveys which assess performance of PHYML, MrBayes, and other common phylogeny programs can be found in the aforementioned survey (Williams and Moret, 2003) and paper about PHYML (Guindon and Gascuel, 2003).

ALGORITHM

The heuristics of RAxML-III belong to the class of algorithms, which optimize the likelihood of a starting tree already comprising all sequences. In contrast to other programs RAxML-III starts by building an initial parsimony tree with dnapars from Felsenstein's PHYLIP package (http://evolution.genetics.washington.edu) for two reasons:

Firstly, parsimony is related to maximum likelihood under simple evolutionary models (Tuffley and Steel, 1997), such that one can expect to obtain a starting tree with a relatively good likelihood value compared to random or neighbor joining starting trees. For example the 500_ZILLA parsimony starting tree showed a better likelihood than the final tree of PHYML (see Table 3).

Secondly, dnapars uses stepwise addition (Felsenstein, 1981) for tree building and is relatively fast. The stepwise addition algorithm enables the construction of distinct starting trees by using a randomized input sequence order. Thus, RAxML-III can be executed several times with different starting trees and thereby compute a set of distinct final trees. The set of final trees can be used to build a consensus tree and augment confidence into the final result since RAxML-III explores the search space from different starting points. To speed up computations, some optimization steps have been removed from dnapars.

The tree optimization process represents the second and most important part of the heuristics. RAxML-III performs standard subtree rearrangements by subsequently removing all possible subtrees from the currently best tree t_{best} and re-inserting them into neighboring branches up to a specified distance of nodes. RAxML-III inherited this optimization strategy from fastDNAml. One rearrangement step in fastDNAml consists of moving all subtrees within the currently best tree by the minimum up to the maximum distance of nodes specified (lower/upper rearrangement setting). This process is outlined for a single subtree (ST5) and a distance of 1 in Figure 1 and for a distance of 2 in Figure 2 (not all possible moves are shown). In fastDNAml the likelihood of each thereby generated topology is evaluated by exhaustive branch length optimizations. If one of those alternative topologies improves the likelihood t_{best} is updated accordingly and once again all possible subtrees are rearranged within t_{best}. This process of rearrangement steps is repeated until no better topology is found.

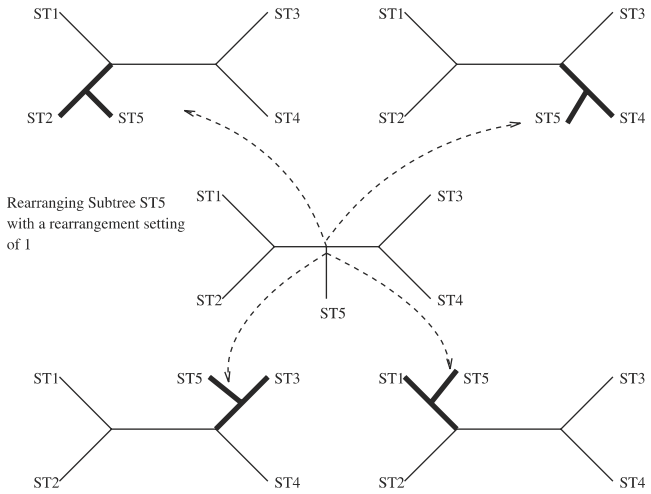

Fig. 1. Rearrangements traversing one node for subtree ST5, branches which are optimized by RAxML-III are indicated by bold lines.

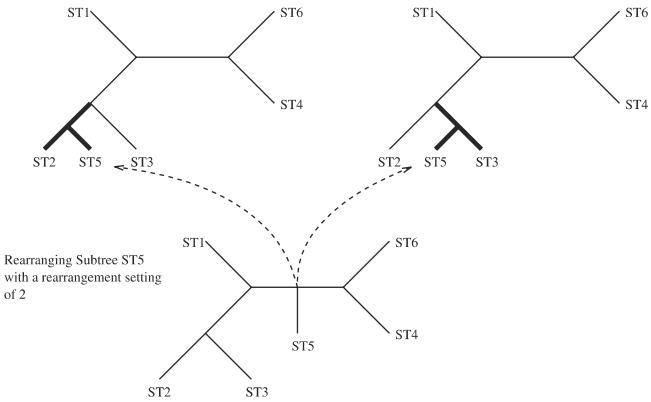

Fig. 2. Example rearrangements traversing two nodes for subtree ST5, branches which are optimized by RAxML-III are indicated by bold lines.

The rearrangement process of RAxML-III differs in two major points: In fastDNAml after each insertion of a subtree into an alternative branch the branch lengths of the entire tree are optimized. As depicted in Figure 1 with bold lines RAxML-III only optimizes the three local branches adjacent to the insertion point either analytically or by the Newton-Raphson method before computing its likelihood value. Since the likelihood of the tree strongly depends on the topology per se this fast pre-scoring can be used to establish a small list of potential alternative trees which are very likely to improve the score of t_{best}. RAxML-III uses a list of size 20 to store the best 20 trees obtained during one rearrangement step. This list size proves to be a practical value in terms of speed and thoroughness of the search. After completion of one rearrangement step the algorithm performs global branch length optimizations on those 20 best topologies only. Due to the capability to analyze

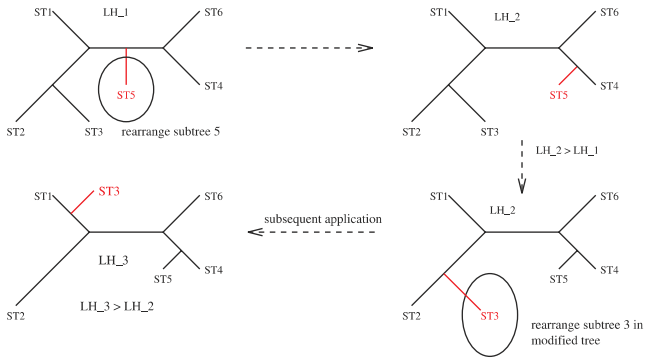

Fig. 3. Example for subsequent application of topological improvements during one rearrangement step.

significantly more alternative and diverse topologies in less time a higher upper rearrangements setting can be used e.g. 5 or 10 which results in significantly improved final trees.

Another important change especially for the initial optimization phase, i.e. the first 3-4 rearrangement steps, consists in the subsequent application of topological improvements during one rearrangement step. If during the insertion of one specific subtree into an alternative branch a topology with a better likelihood is encountered this tree is kept immediately and all subsequent subtree rearrangements of the current step are performed on the improved topology. The mechanism is outlined in Figure 3 for a subsequent application of topological improvements via subtree rearrangements of ST5 and ST3 on the same initial tree. This enables rapid initial optimization of random starting trees as depicted e.g. for two alignments containing 150 taxa in Figures 6 and 7. The exact implementation of the RAxML-III algorithm is indicated in the C-like pseudocode below. The algorithm is passed the user/parsimony starting tree `t`, the initial rearrangement setting `rStart` (default: 5) and the maximum rearrangement setting `rMax` (default: 21). Initially the rearrangement stepwidth ranges from `rL = 1` to `rU = rStart`. Fast analytical local branch length optimization `a` is turned off when functions `rearr(...)`, which actually performs the rearrangements, and `ptimizeList20()` fail to yield an improved tree for the first time. As long as the tree does not improve the lower and upper rearrangement parameters `rL, rU` are incremented by `rStart`. The program terminates when the upper rearrangement setting is greater or equal to the maximum rearrangement setting, i.e. `rU >= rMax`.

```
RAxML-III(tree t, int rStart, int rMax)
{
 int rL, rU;
 boolean a = TRUE;
 boolean impr = TRUE;
 while(TRUE)
  {
```

```
   if(impr)
    {
     rL = 1;
     rU = rStart;
     rearr(t, rL, rU, a);
    }
   else
    {
     if(!a)
      {
       a = FALSE;
       rL = 1;
       rU = rStart;
      }
     else
      {
       rL += rStart;
       rU += rStart;
      }
     if(rU < rMax)
      rearr(t, rL, rU, a);
     else
      goto end;
    }
   impr = optimizeList20();
  }
 end:
}
```

RESULTS

Test data and platforms

For conducting experiments alignments comprising 150, 200, 250, 500, 1.000, and 2.025 taxa (150_ARB, . . . , 2025_ARB) have been extracted from the ARB small subunit ribosomal ribonucleic acid (ssu rRNA) database. Those alignments contain organisms from the domains Eukarya, Bacteria and Archaea. In addition, the 101 and 150 sequence data sets (101_SC, 150_SC) which can be downloaded at http://www.indiana.edu/~rac/hpc/fastDNAml were used. Those data sets have been used by C. Stewart *et al*. to conduct performance analysis of parallel fastDNAml. The larger 101_SC and 150_SC alignments have proved to be very hard to optimize, in terms of convergence to best-known likelihood values, especially for MrBayes with random starting trees (see Figure 4). According to a personal communication with C. Stewart this is due to the fact that these two data sets contain several hard-to-classify fungi which randomly scatter throughout the final trees. Furthermore, two well-known real data sets comprising 218 and 500 sequences (218_RDPII, 500_ZILLA) were included into the test set. Those two alignments are considered to be 'classic' real data benchmarks. In particular the 500_ZILLA alignment has been studied extensively under the parsimony criterion (Chase *et al*., 1993).

We also used 50 synthetic (simulated) 100-taxon alignments (100_SIM_1, . . . , 100_SIM_50) with a length of 500 base pairs each. The respective true reference trees and alignments are available at http://www.lirmm.fr/w3ifa/MAAS and were originally used to assess accuracy of PHYML (Guindon and Gascuel, 2003). Details on the generation of those data sets which contain e.g. varying sequence divergence rates are also available in the respective paper. Finally, we generated 10 synthetic 4000-taxon (4000_SIM_1, . . . , 4000_SIM_10) alignments using the r8s program (Sanderson, 2003) to generate a random tree with the following command:

```
begin rates;
  simulate diversemodel=bdback
    ntaxa=4000 seed=3049;
  simulate charevol=yes infinite=yes
    startrate=1 minrate=0.1 maxrate=2;
  changerate=0.5 model=NORMAL;
  describe plot=phylo_description;
end;
```

Furthermore, we invoked Seq-Gen (Rambaut and Grassly, 1997)

```
seq-gen -m HKY -l 2000 -s x -t 2.0
```

with scaling factor `x` ranging from 0.1 to 1.0 to obtain the respective synthetic alignments. For sake of completeness the number of base pairs in each alignment is provided in Table 1.

We compiled MrBayes, PHYML, and RAxML-III with the native Intel compiler `icc -O3` and executed the programs on a cluster of unloaded Intel Xeon 2.4GHz processors equipped with 4GB of main memory at our laboratory. Since PHYML and RAxML-III are directly comparable and both significantly faster than MrBayes we mainly focus on those two programs for performance analysis of compute-intensive large data sets and complex models of nucleotide substitution. We include data from sequential executions of MrBayes to show that the MC^3 (Metropolis-Coupled Markov Chain Monte Carlo simulation) chain does generally not attain stationarity within acceptable time limits, i.e. less than 24 hours, for real data sets containing more than 250 taxa. However, simple comparison of intermediate or final likelihood values does certainly not represent the only criterion for conducting a fair performance assessment of maximum likelihood and bayesian inference. Our intention is to emphasize that coupling those methods induces substantial benefits.

Small simulated data

For synthetic data we executed MrBayes for 100.000 generations using 4 MC^3 chains and recommended random starting trees. We specified a sample and print frequency of 500 and used the last 50 trees to build a majority-rule consensus tree. Those relatively fast settings for MrBayes prove to be sufficient to obtain good accuracy values since analyses for synthetic data converge much faster to a

Table 1. Alignment lengths

data	#bp
101_SC	1858
150_SC	1269
218_RDPII	4182
500_ZILLA	759
150_ARB	3188
200_ARB	3270
250_ARB	3638
500_ARB	4030
1000_ARB	5547
2025_ARB	1517
100_SIM	500
4000_SIM	2000

Table 2. Topological accuracy and execution times for PHYML & RAxML-III on simulated data

data	PHYML (RF)	secs	RAxML (RF)	secs
4000_SIM_1	0.065	18944	0.065	9152
4000_SIM_2	0.039	22273	0.037	50609
4000_SIM_3	0.033	24907	0.027	97962
4000_SIM_4	0.030	30870	0.031	85080
4000_SIM_5	0.028	24182	0.035	91178
4000_SIM_6	0.027	32614	0.031	176686
4000_SIM_7	n/a	n/a	0.028	144519
4000_SIM_8	0.027	34750	0.032	185454
4000_SIM_9	0.026	18828	0.036	78061
4000_SIM_10	n/a	n/a	0.034	64690

peak likelihood value or stationary chain than respective real data experiments. The average RF-rate (Robinson and Foulds, 1979) on the 50 simulated 100-taxon trees (100_SIM_1-100) for PHYML is 0.0796, 0.0808 for RAxML-III, 0.0818 for RAxML-III with a less exhaustive search setting and 0.0741 for MrBayes. The average execution time of RAxML-III was 131.05 seconds and 29.27 seconds for the faster search. PHYML required an average of 35.21 seconds and MrBayes 945.32 seconds. The experiments illustrate that there seems to be no apparent difference between PHYML and RAxML-III for small synthetic data.

Large simulated data

In Table 2 we list the normalized Robinson-Foulds distance and execution time in seconds of PHYML and RAxML-III for 10 synthetic 4000-taxon alignments. For this test series we used the most recent linux binary version of PHYML (v2.1b1) since our source code version constantly exited with a segmentation fault. Performance results of PHYML for 4000_SIM_7 and 4000_SIM_10 are not available (n/a) because we encountered a tree parsing problem with the respective output trees. It is evident, that PHYML clearly outperforms RAxML-III on large synthetic data for branch length scaling factor $x \geq 0.5$, i.e. on the 4000_SIM_5-9 alignments. However, trees scaled by $x \leq 0.5$ appear to be more realistic in a biological context (Bininda-Emonds and Sanderson, 2001).

Real data and fixed model

To facilitate testing we used the HKY85 (Hasegawa *et al*., 1985) model of sequence evolution and a fixed transition/transversion (tr/tv) ratio for these experiments. All alignments including the best topologies are available at http://wwwbode.cs.tum.edu/~stamatak. Since the tr/tv ratio is defined differently in PHYML we scaled it accordingly for the test runs. The manual for PAML (Yang, 1997) which is available at http://bcr.musc.edu/manuals/pamlDOC.pdf contains a nice description of differences in the tr/tv ratio definitions among various maximum likelihood programs on page 20. For real data sets MrBayes was executed over 2.000.000 generations using 4 MC^3 chains and random starting trees. Furthermore, we used a sample and print frequency of 5000. To enable a fair comparison we evaluated all 400 of MrBayes output trees as well as the final PHYML results with fastDNAml. For MrBayes we report the value of the topology with the best likelihood and the execution time at that point. The trees of this test series are evaluated with fastDNAml and fixed tr/tv ratios, due to the availability of reference trees obtained by large scale parallel analyses with PAxML.

In Table 3 we summarize the final likelihood values and execution times in seconds for PHYML, MrBayes, and RAxML-III. Since overall execution times of RAxML-III might appear long compared to to those of PHYML we indicate the likelihood and the time at which RAxML-III passed the final likelihood obtained by PHYML in column R>PHY. Finally, in the last two columns we list the final likelihood values and execution times in hours (!) obtained with PAxML which is essentially equivalent to parallel fastDNAml. The results were obtained from parallel runs on the HeLiCs (Heidelberg Linux Cluster System: http://helics.uni-hd.de) compute cluster and the highest feasible rearrangement setting, in terms of acceptable computation times. The enormous improvement of execution times illustrates the algorithmic progress in the field over the last two years. The long overall execution times of RAxML-III in comparison to PHYML are due to the asymptotic convergence of likelihood over time which is typical for the tree optimization process. A particularly extreme case of slow asymptotic convergence has been observed for 500_ZILLA (Stamatakis *et al*., 2004a). Therefore, the comparatively small differences in final likelihood values which are usually below 1% should not be underestimated, in terms of the computational effort required to obtain those values. The application of the Kishino–Hasegawa likelihood ratio test shows that all final RAxML-III trees are significantly better than respective PHYML trees.

Two examples which underline how bayesian analysis can benefit from 'traditional' methods are outlined in Figures 4

Table 3. PHYML, MrBayes, RAxML-III execution times and likelihood values for real data sets

data	PHYML	secs	MrBayes	secs	RAxML	secs	R > PHY	secs	PAxML	hrs
101_SC	−74097.6	153	−77191.5	40527	−73919.3	617	−74046.9	31	−73975.9	47
150_SC	−44298.1	158	−52028.4	49427	−44142.6	390	−44262.9	33	−44146.9	164
150_ARB	−77219.7	313	−77196.7	29383	−77189.7	178	−77197.6	67	−77189.8	300
200_ARB	−104826.5	477	−104856.4	156419	−104742.6	272	−104809.0	99	−104743.3	775
250_ARB	−131560.3	787	−133238.3	158418	−131468.0	1067	−131549.4	249	−131469.0	1947
500_ARB	−253354.2	2235	−263217.8	366496	−252499.4	26124	−252986.4	493	−252588.1	7372
1000_ARB	−402215.0	16594	−459392.4	509148	−400925.3	50729	−401571.9	1893	−402282.1	9898
218_RDPII	−157923.1	403	−158911.6	138453	−157526.0	6774	−157807.9	244	n/a	n/a
500_ZILLA	−22186.8	2400	−22259.0	96557	−21033.9	29916	−22036.9	67	n/a	n/a

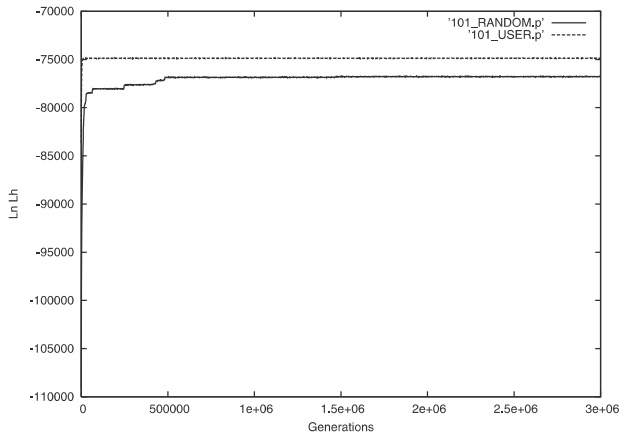

Fig. 4. Convergence behavior of MrBayes for 101_SC with user and random starting trees over 3.000.000 generations.

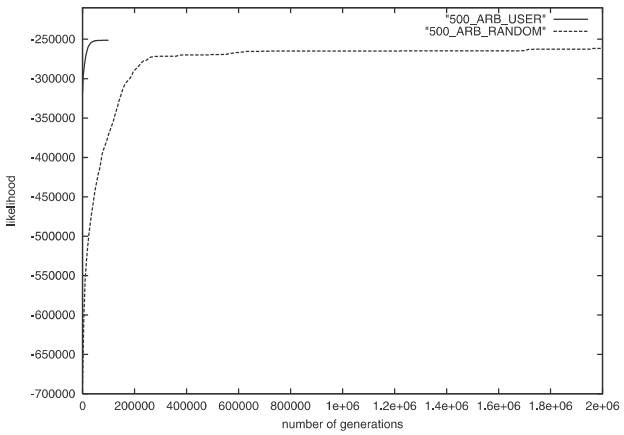

Fig. 5. Convergence behavior of MrBayes for 500_ARB with user and random starting trees.

and 5. In those figures we plot MrBayes likelihood values over generation numbers with RAxML- and random starting trees for 101_SC and 500_ARB respectively. Furthermore, Figure 4 reveals one of the main problems of MC^3 analysis

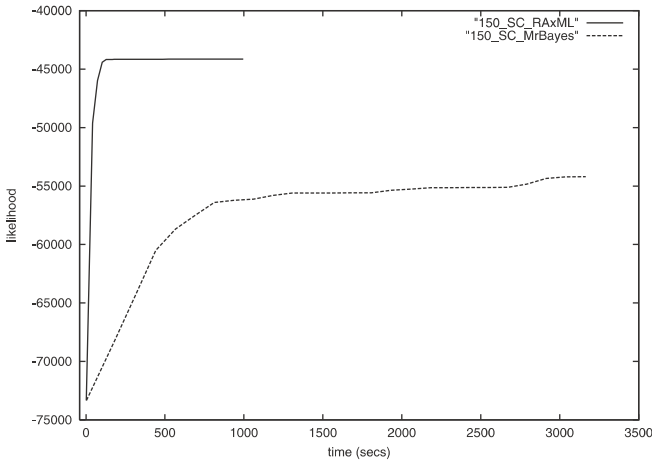

Fig. 6. 150_SC likelihood improvement over time of RAxML-III and MrBayes for the same random starting tree.

(Huelsenbeck *et al.*, 2002): When to stop the chain? In the example the run with the random starting tree seems to have reached apparent stationarity, although the tree is far from optimal. Therefore, 'good' starting tree obtained by 'traditional' methods can be useful to significantly accelerate computations and serve as reference point. This justifies the work on fast 'traditional' maximum likelihood methods despite the emergence and great impact of bayesian methods (Huelsenbeck *et al.*, 2001). Thus, we do not see RAxML-III as concurrence to MrBayes, but rather as useful tool to improve bayesian inference and vice versa. Finally, in order to demonstrate the rapid tree optimization capabilities of RAxML-III in Figures 6 and 7 we plot the likelihood improvement over time of RAxML-III and MrBayes for the same random starting trees.

Real data and estimated model

In this series of real data tests we compare PHYML and RAxML-III performance on the HKY85 and General Time Reversible (GTR (Lanave *et al.*, 1984)) model of nucleotide substitution. We let both programs estimate the tr/tv

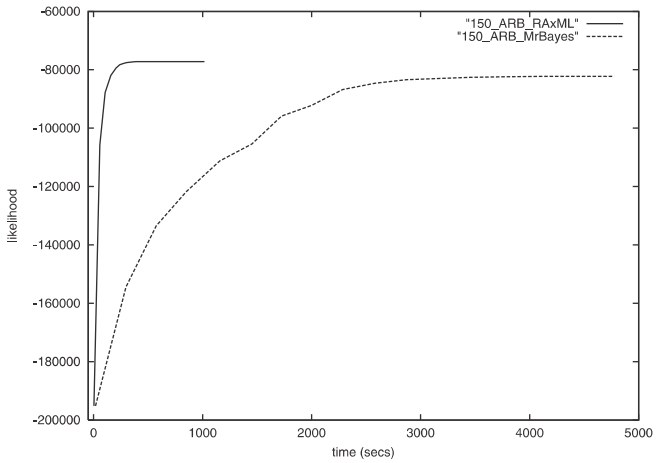

Fig. 7. 150_ARB likelihood improvement over time of RAxML-III and MrBayes for the same random starting tree.

Table 4. Performance of PHYML and RAxML-III for HKY85 and GTR models of evolution with model parameter optimization

data	PHYML	secs	RAxML	secs	R > PHY
HKY85					
101_SC	−74035	104	−73908	71	21
150_SC	−44315	85	−44219	64	26
150_ARB	−76881	190	−76863	94	49
200_ARB	−104316	282	−104270	185	120
250_ARB	−131013	405	−130926	342	116
500_ARB	−252224	1453	−251781	1049	420
1000_ARB	−400881	3908	−399732	3633	1666
2025_ARB	−372746	9749	−371472	8426	4779
218_RDPII	−156895	230	−156663	331	126
GTR					
101_SC	−73814	131	−73638	119	49
150_SC	−44139	132	−44043	157	60
150_ARB	−76500	235	−76490	203	144
200_ARB	−103789	714	−103758	352	262
250_ARB	−130518	526	−130353	416	218
500_ARB	−250858	1170	−250238	1516	688
1000_ARB	−398731	4727	−397612	5731	2469
2025_ARB	−370539	5299	−369197	10771	5558
218_RDPII	−155881	316	−155748	406	268

ratio (HKY85) and the substitution rates (GTR) along with the tree topology. To save some CPU hours we used a version of RAxML-III which terminates immediately when the tree fails to improve for the first time. To ensure a fair comparison we evaluated the likelihood of all final trees with PHYML. Typically final likelihood values obtained by different programs for the same tree and model of nucleotide substitution differ due to numerical differences in implementations. Results are summarized in the same style as in the previous Section in Table 4 for the HKY85 and GTR models of sequence evolution respectively.

DISCUSSION

We have presented the most recent version of our program RAxML-III for maximum likelihood-based inference of phylogenetic trees. The code incorporates the HKY85 and GTR models of DNA sequence evolution and is able to optimize all free model parameters. Furthermore, RAxML-III is able to perform a maximum likelihood estimate of per-site evolutionary rates for HKY85. In order to accelerate computations the number of distinct evolutionary rates can be categorized into a user-specified amount of rate categories. The program performs worse than PHYML and MrBayes on synthetic data. However, on real data it outperforms PHYML, MrBayes, and PAxML in terms of required execution time and final likelihood values. In addition, we provide failure scenarios for MrBayes on real data sets and argue that traditional and bayesian inference should be combined to circumvent intrinsic problems of either approach. Along with the RAxML-III source code we provide a large real-data benchmark set which includes best-known reference trees and execution times on a specific architecture/compiler combination. This data collection is intended to serve other researchers as reference data set to assess performance of maximum likelihood programs. The advantage of RAxML-III over PHYML consist in a more exhaustive analysis of search space which results in improved final likelihood values and in the ability to generate distinct random starting trees. In addition, the parallelization of the algorithm is straight-forward and RAxML-III has significantly lower memory requirements than MrBayes and PHYML (Stamatakis *et al*., 2004b). On the other hand RAxML-III provides significantly less modeling flexibility than MrBayes and PHYML, i.e. is not able to handle protein sequence data and does not provide estimation of the proportion of invariable sites or the Γ model of rate heterogeneity. An implementation of the missing model features is planned in a future version of the program.

Our results show that both RAxML-III (on real data) and PHYML (on simulated data) represent very fast and accurate conventional maximum likelihood programs, which allow for sequential inference of large trees within reasonable times on standard PC architectures.

ACKNOWLEDGEMENT

We would especially like to thank Stephane Guindon for his help on PHYML.

REFERENCES

Bininda-Emonds,O.R.P. and Sanderson,M.J. (2001) An assessment of the accuracy of MRP supertree construction. *Syst. Biol.*, **50**, 565–579.

Brauer,M.J., Holder,M.T., Dries,L.A., Zwickl,D.J., Lewis,P.O. and Hillis,D.M. (2002) Genetic algorithms and parallel processing in maximum-likelihood phylogeny inference. *Mol. Biol. Evol.*, **19**, 1717–1726.

Chase,M.W., Soltis,D.E., Olmstead,R.G., Morgan,D., Les,D.H., Mishler,B.D., Duvall,M.R., Price,R.A., Hills,H.G., Qiu,Y.L. *et al*. (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Garden*, **80**, 528–580.

Felsenstein,J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–376.

Guindon,S. and Gascuel,O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704.

Hasegawa,M., Kishino,H. and Yano,T. (1985) Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **22**, 160–174.

Holder,M.T. and Lewis,P.O. (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.*, **4**, 275–284.

Huelsenbeck,J.P. and Ronquist,F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Huelsenbeck,J.P., Ronquist,F., Nielsen,R. and Bollback,J.P. (2001) Bayesian inference and its impact on evolutionary biology. *Science*, **294**, 2310–2314.

Huelsenbeck,J.P., Larget,B., Miller,R.E. and Ronquist,F. (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.*, **51**, 673–688.

Lanave,C., Preparata,G., Saccone,C. and Serio,G. (1984) A new method for calculating evolutionary substitution rates. *J. Mol. Evol.*, **20**, 86–93.

Lemmon,A. and Milinkovitch,M. (2002) The metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. *Proc. Natl Acad. Sci. USA*, **99**, 10516–10521.

Lewis,P. (1998) A genetic algorithm for maximum likelihood phylogeny inference using nucleotide sequence data. *Mol. Biol. Evol.*, **15**, 277–283.

Ludwig,W., Stunk,O., Westram,R., Richter,L., Meier,H., Yadhukumar, Buchner,A., Lai,T., Steppi,S., Jobb,G. *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.*, **32**, 1363–1371.

Olsen,G., Matsuda,H., Hagstrom,R. and Overbeek,R. (1994) fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.*, **10**, 41–48.

Rambaut,A. and Grassly,N.C. (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.*, **13**, 235–238.

Robinson,D. and Foulds,L. (1979) Comparison of weighted labeled trees. In Horadam,A.F. and Wallis,W.D. (eds), *Isomorphic Factorisations VI: Automorphisms, Combinatorial Mathematics VI, Lecture Notes in Mathematics*, **748**, Springer, Berlin, pp. 119–126.

Sanderson,M.J. (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, **19**, 301–302.

Stamatakis,A., Ludwig,T., Meier,H. and Wolf,M.J. (2002) Accelerating parallel maximum likelihood-based phylogenetic tree computations using subtree equality vectors. In *Proceedings of 15th IEEE/ACM Supercomputing Conference (SC2002)*, Baltimore, MD, November.

Stamatakis,A., Ludwig,T. and Meier,H. (2004a) New fast and accurate heuristics for inference of large phylogenetic trees. In *Proceedings of 18th IEEE/ACM International Parallel and Distributed Processing Symposium* (*IPDPS'04*), Santa Fe, NM, April 26–30.

Stamatakis,A., Ludwig,T. and Meier,H. (2004b) Parallel inference of a 10.000-taxon phylogeny with maximum likelihood. In *Proceedings of Euro-Par 2004*, Pisa, Italy, August 31–September 3, (to be published).

Stewart,C., Hart,D., Berry,D., Olsen,G., Wernert,E. and Fischer,W. (2001) Parallel implementation and performance of fastDNAmL—a program for maximum likelihood phylogenetic inference. In *Proceedings of 14th IEEE/ACM Supercomputing Conference* (*SC2001*), Denver, CO, May 18.

Strimmer,K. and Haeseler,A.V. (1996) Quartet puzzling: a maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, **13**, 964–969.

Swofford,D. (1999) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Tuffley,C. and Steel,M. (1997) Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.*, **59**, 581–607.

Williams,T.L. and Moret,B.M.E. (2003) An investigation of phylogenetic likelihood methods. In *Proceedings of 3rd IEEE International Symposium on Bioinformatics and Bioengineering (BIBE'03)*, Bethesda, MD, March 10–12.

Yang,Y. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.*, **13**, 555–556.