Phylogeny Estimation and Alignment via POY versus Clustal + PAUP*: A Response to Ogden and Rosenberg (2007)

SAMULI LEHTONEN

Department of Biology, Section of Biodiversity and Environmental Science, University of Turku, FI-20014 Turku, Finland; E-mail: samile@utu.fi

The performance of the traditional two-step approach (independent alignment and tree search) employed in molecular systematics was recently compared to direct optimization (DO; Wheeler, 1996) by Ogden and Rosenberg (2007a). In their analyses the two-step approach clearly outperformed DO both in alignment and topological accuracy, and the performance of DO was found to decrease with increasing search effort. The results suggested to Ogden and Rosenberg (2007a) that the two-step approach is superior to DO and therefore should be preferred when analyzing data sets similar to those generated in their study.

The settings they used in the DO and two-step analyses, however, were not comparable and biased their results in favor of the two-step approach. Specifically, more complete TBR branch swapping algorithms were used in the two-step analyses but not in the DO analyses. Moreover, dynamic homologies were approximated with static homologies during cladogram optimization. In addition, branch swapping during DO was limited to minimum cost trees, but the default setting of POY performs swapping on all trees found during the search. Finally, they carried out only four replicates in DO versus 100 in the PAUP* analyses. These features were also found in their more extensive analyses, which were based on four replicates, used static homologies during swapping, and were restricted to minimum trees only. These analytical shortcuts are required to perform large number of DO analyses in a short period of time, but in my opinion Ogden and Rosenberg’s (2007a) use of POY compromised the main objective of their study: the comparison of DO and the traditional approach.

MATERIAL AND METHODS

Inspired by the study of Ogden and Rosenberg (2007a), I generated 10 random topologies of 16 taxa (Fig. 1) under a Yule model (Steel and McKenzie, 2001) in Mesquite.
Figure 1. The 10 random Yule trees used to simulate the data of this study.
TABLE 1. Brief descriptions of commands used in POY analyses.

POY3 commands (De Laet and Wheeler, 2003)
- gap 1: Set indel cost to 1 (equal weighting).
- nooneasis: All random additions randomized.
- noleading: Leading and trailing gaps are counted and accounted for during builds and refinements (to prevent trivial alignments) but discounted when determining tree length.
- norandomizeoutgroup: The first terminal of the first data set is used as starting point for adding the other terminals.
- quick: Do not swap on non-minimal length trees during branch-swapping. This forces POY to swap only on minimum length trees despite the fact that other trees have been found in the swapping process.
- staticapprox: Use static approximation for sequence optimization while doing tree refinements. Static approximation consists of using a fixed implied alignment while doing branch-swapping and other kinds of tree refinement; every time a shorter tree is hit, a new implied alignment is calculated (based on that shorter tree) and this new alignment is used for further swapping.
- notbr: Turns off TBR cladogram search.
- replicates: Sets the number of random addition sequence searches or the number of jackknife pseudo-replicates.
- tbr: Perform TBR branch-swapping during replicate refinement.

POY4 commands (Varón et al., 2007)
transform((all, tcm:(1,1))): Set indel and substitution cost to 1 (equal weighting).
bild (100): Generate 100 random addition sequence Wagner trees.
select (unique): Selects only topologically unique trees (after collapsing zero-length branches) irrespective of their cost.
swap (threshold:10): Alternate SPR and TBR branch-swapping of all current trees as well as all new trees found that are within 10% of the minimum tree length.
select (): Selects only optimal trees.

(Maddison and Maddison, 2004). The maximum evolutionary distance between paired taxa was set to 2.0. For each topology 10 simulations were conducted using MySSP (Rosenberg, 2005a). Simulations were based on the Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al., 1985) with transition-transversion bias \( \kappa = 3.6 \), and with initial and expected base frequencies of A and T = 0.2 and G and C = 0.3. Initial sequence length was set to 2000 base pairs. Insertions and deletions were modeled using an expected rate of one insertion for every 100 substitutions and one deletion for every 40 substitutions, and the lengths of indels were determined from a Poisson distribution with a mean equal to four bases. Hence, the settings followed the simulations with maximum evolutionary distance of 2.0 made by Ogden and Rosenberg (2007a).

Before the data were analyzed, all gaps were removed from the simulated sequences. Analyses using the traditional two-step method were run by aligning sequences with default values (gap opening = 15, gap extension = 6.66) in ClustalX (Thompson et al., 1997), and then analyzing multiple sequence alignments in PAUP* (Swofford, 1998) with 100 random additions and TBR branch-swapping; otherwise, default settings were used (e.g., gaps were treated as missing data). Three sets of DO analyses were performed, each one using equal costs for gap opening, gap extension, transversions, and transitions. First, the command line (-gap 1-nooneasis-noleading-norandomizeoutgroup-quick-staticapprox-notbr-replicates 4) of Ogden and Rosenberg (2007a) was run with POY v. 3.0.11. (Wheeler et al., 2003). These analyses are later referred to as “POY3-O&K” analyses. Second, I examined the impact of TBR branch-swapping by rerunning all the POY searches, but allowing TBR branch-swapping (command-notbr replaced with -tbr, otherwise the command line was identical). These analyses are referred to as “POY3-TBR” analyses. Third, in order to make a more equitable comparison between two-step method and DO, I ran a set of POY analyses using the default search strategy. Due to the extreme time demand of extensive DO analyses, I used a newer, much faster version of POY (Varón et al., 2007). The command line used was the basic search of the POY4 manual (Varón et al., 2007) with an additional command “transform” to set equal weighting [transform((all, tcm:(1,1))) build (100) select (unique) swap (threshold:10) select ()]. This search generates 100 random addition trees, discards duplicate trees, and performs SPR and TBR branch-swapping for all trees and for all new trees that are up to 10% longer than the tree under swapping. I consider this to be the default search strategy of POY and refer to this hereafter as POY4-basic. Commands employed in these POY analyses are briefly described in Table 1.

Performance of these approaches was quantified by determining the congruence of the inferred trees directly with the correct topologies by calculating the symmetric-difference distance (also known as Robinson-Foulds distance; Robinson and Foulds, 1981) between the inferred trees and true topologies with PAUP*. The alignments produced by Clustal and POY were compared to true alignments by calculating the proportion of aligned sites that are truly homologous (total alignment accuracy; Rosenberg, 2005b). The possible differences in the performance between Clustal+PAUP* and POY analyses were examined with SAS Statistical Package, version 9.1 (SAS Institute, Cary, NC, USA).

RESULTS AND DISCUSSION

Ogden and Rosenberg’s (2007a) statements that Clustal+PAUP* analyses outperform POY analyses and that more extensive POY analyses yield poorer results than less extensive POY analyses seem incorrect based on the following results of my study.

First, POY4-basic searches produced more correct topologies than Clustal+PAUP* 46% of the time (Table 2), whereas Clustal+PAUP* topologies were more correct than POY4-basic topologies 23% of the time. POY4-basic found the exactly correct topology 20 times
TABLE 2. The relative performance of Clustal + PAUP* and POY4-basic analyses (see text for details). No statistically significant difference was found in the performance of POY4-basic and Clustal + PAUP* when tree distances were compared, even though Clustal alignments are more accurate than POY-implied alignments. Note that default alignment parameters differ markedly between these two approaches.

<table>
<thead>
<tr>
<th></th>
<th>Clustal + PAUP*</th>
<th>POY4-basic</th>
<th>Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>outperforms</td>
<td>outperforms</td>
<td></td>
</tr>
<tr>
<td>Tree distance</td>
<td>23%</td>
<td>46%</td>
<td>31%</td>
</tr>
<tr>
<td>Alignment accuracy</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

and Clustal+PAUP* 15 times (out of 100). In those cases where POY4-basic outperformed Clustal+PAUP*, it did it on average by symmetric distance difference of 2.5, and in the cases where Clustal+PAUP* outperformed, the difference was on average 2.2. The worst symmetric difference distance for a POY4-basic topology was 6, but 12 for Clustal+PAUP* topology. In total, 12% of the Clustal+PAUP* topologies were more inaccurate than the most inaccurate POY4-basic topology.

Second, on average, the symmetric difference distance of Clustal+PAUP* topologies and true topologies was 3.13 (± 2.77 [standard deviation]), whereas the symmetric difference distance of POY4-basic topologies and true topologies was on average 2.49 (± 1.79 [standard deviation]; Table 3). However, the difference was not statistically significant (Wilcoxon two-sample test: Z = 0.91, P = 0.36, n = 100).

Third, POY3-TBR outperformed POY3-O&R 33% of the time (and was outperformed by POY3-O&R 24% of the time) and resulted in trees that were on average better (Table 3). POY3-TBR found the correct topology 13 times, and POY3-O&R 9 times. On average, the more extensive POY4-basic search found clearly better topologies and outperformed both POY3-O&R and POY3-TBR 41% of the time (and was outperformed by at least one of the two 19% of the time). The same trend was evident in tree lengths: 60% of the POY3-TBR trees were more parsimonious than POY3-O&R trees of a corresponding simulation, and 98% of the POY4-basic trees were more parsimonious than the corresponding POY3-TBR trees.

Hence, the more extensive POY searches clearly resulted in better topologies than less extensive POY searches and were able to find the exactly correct trees much more often. This means that Ogden and Rosenberg’s (2007a:186) justification to draw general conclusions from their less extensive POY searches was mistaken. It should also be noted that the performance of POY and Clustal+PAUP* varied more from topology to topology than between the different simulations on the same topology (Kruskal-Wallis test: χ² = 65.7, df = 9, P < 0.0001 for POY4-basic analyses, and χ² = 51.1, df = 9, P < 0.0001 for PAUP* analyses). This means that a study based on a relatively few topologies but many simulation replicates per topology may be biased. The results presented here show that the two-step approach is not better than DO in recovering topologies, as claimed by Ogden and Rosenberg (2007a).

It has been shown that topological accuracy tends to increase by decreasing alignment error, although in some reported cases the quality of the tree and alignment seem to be independent (Ogden and Rosenberg, 2006). Interestingly, in these simulations POY-implied alignments were always poorer than the corresponding Clustal alignments as measured by total alignment accuracy (Table 2), despite the fact that no significant difference in topological accuracy was found. This suggests that POY-implied alignments behave somehow differently than Clustal alignments. Total alignment accuracy score is based on aligned sites only, and gaps that are aligned with gaps or nucleotides are ignored (Ogden and Rosenberg, 2006), but it is difficult to say whether this could have a different impact on POY and Clustal alignment accuracy scores. Alignments are highly sensitive to the parameters employed (Morrison and Ellis, 1997), and the parameters used here in Clustal and POY analyses were drastically different. Clustal alignment accuracy varied from 81.4% to 16.9% with an average of 50%, and the accuracy of POY4-basic alignments varied from 63.7% to 14.2% with an average of 36.2%. Thus, both methods produced relatively incorrect alignments on average but were still able to reconstruct quite accurate trees. It is my opinion (but not only mine; see, e.g., Morrison and Ellis, 1997) that the ability to correctly resolve evolutionary relationships of the sequences (i.e., topological accuracy) allows more reasonable comparison of different phylogenetic reconstruction methods, optimality criteria, and parameter settings than sequence alignments as such. Furthermore, the main usage of POY is to recover phylogenies without the need to align sequences.

The most extensive DO analyses of the present study were performed with POY4, which has faster algorithms and better heuristics than POY3 used by Ogden and Rosenberg (2007a) and Varón et al. (2007). Although the programs differ, they are expected to do the same thing. The great speed-up provided by POY4 made it possible to fairly compare DO and two-step approaches over relatively many simulations. Yet, my test runs in HP CP4000 BL ProLiant supercluster (Finnish IT Center

TABLE 3. Means and standard deviations of symmetric difference distances (tree distance) and total alignment accuracies (n = 100 in each case). Lower values in tree distance indicate better congruence. In alignment accuracy higher values are better with a maximum value of 1 (= 100% similarity between true and hypothesized alignments). On average, POY4-basic topologies are the best, although no statistically significant difference between POY4-basic and Clustal+PAUP* topological accuracy was found.

<table>
<thead>
<tr>
<th></th>
<th>Clustal+PAUP* (default)</th>
<th>POY3 (O&amp;R)</th>
<th>POY3 (TBR)</th>
<th>POY4 (basic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tree distance</td>
<td>3.13 ± 2.77</td>
<td>4.2 ± 2.71</td>
<td>3.57 ± 2.34</td>
</tr>
<tr>
<td></td>
<td>Alignment accuracy</td>
<td>0.50 ± 0.16</td>
<td>0.32 ± 0.14</td>
<td>0.32 ± 0.13</td>
</tr>
</tbody>
</table>
for Science [CSC]) suggest that reanalyzing Ogden and Rosenberg’s (2007a) 15,400 simulations would require years of nonstop computing in a supercomputer cluster with a basic POY4 search and probably decades with a comparable search in POY3. Thus, analyzing thousands of simulated data sets, even with default POY settings, is not very practical. But in any real study where months or even years are spent in collecting data, a few more hours used for the data analysis should not matter much.

It should be remembered that by treating gaps as missing data all the phylogenetic information related to indel events is lost. For this reason it is generally advised to code gaps as fifth character state (Giribet and Wheeler, 1999; Simmons and Ochoterena, 2000; Ogden and Rosenberg, 2007b), but in this study the default setting of PAUP∗ was used. PAUP∗ might have performed better if the indel information would have been used. POY4-basic analyses were not especially extensive either, and more sophisticated refining techniques could have been used (see, for example, Goloboff, 1999). Therefore, neither PAUP∗ nor POY analyses in this study were optimal, and probably both approaches could perform better.

Both in this study and Ogden and Rosenberg’s (2007a) study, the less extensive POY searches were clearly outperformed by traditional Clustal+PAUP∗ analyses, but this does not mean that the traditional two-step approach would be a better choice than DO, as suggested by Ogden and Rosenberg (2007a). It merely shows that effective tree search strategy is highly important, especially in the DO, but even a default POY search recovers topologies at least equally well as default Clustal+PAUP∗ approach. More research by means of more realistic simulations, e.g., several data partitions of varying evolutionary rates analyzed with unbiased search strategies, might further illuminate the strengths and pitfalls of the two-step approach and DO.

ACKNOWLEDGMENTS

Michael S. Rosenberg kindly provided code for calculating total alignment accuracy, and Toni Lehtonen translated it to the programming language Python for use in this study. I thank Palvi Salo for her help with statistical tests and reviewers of previous drafts of this paper their helpful critiques. This study was funded by an Academy of Finland grant to Hanna Tuomisto.

REFERENCES


First submitted 5 February 2008; reviews returned 25 March 2008; final acceptance 7 May 2008

Associate Editor: Tim Collins