In response to comment on ‘A congruence index for testing topological similarity between trees’

Damien M. de Vienne1,*, Tatiana Giraud1 and Olivier C. Martin2,3

1Univ Paris-Sud, Laboratoire de Recherche en Informatique, UMR8623, Orsay F-91405; CNRS, Orsay F-91405, 2Univ Paris-Sud, UMR8626, LPTMS, Orsay F-91405; CNRS, Orsay F-91405 and 3Univ Paris-Sud, UMR8120, Laboratoire de Génétique Végétale du Moulon, Gif-sur-Yvette F-91190, France

Received on September 11, 2008; revised and accepted on October 09, 2008

Advance Access publication October 17, 2008

Associate Editor: Martin Bishop

Contact: damien.de-vienne@lri.fr

In a paper entitled ‘A congruence index for testing topological similarity between trees’, we proposed a method for testing the congruence between two trees, the null hypothesis being that the trees were not more congruent (topologically similar) than expected by chance (de Vienne et al., 2007). The test is based on the size of the Maximum Agreement Subtree (MAST). As highlighted by Kupczok and von Haeseler (2008), our approach treats the discrete variable of the size of the MAST as a continuous one. Another, even more important, simplification of our method is that we replaced the histograms of the MAST sizes at different tree sizes by a ‘master’ distribution obtained by a global fit; this is really what allowed for a fast (though not exact) estimate of \( P \)-values for any tree size. In their comment, Kupczok and von Haeseler (2008) show that, when performing tests at the significance level of 5%, our procedure results in some cases in too liberal a test, the true critical MAST size there being in fact 1 unit larger than the value estimated by our method. The point is well taken, and the data of these authors provides a tabulation of the critical MAST sizes using the 5% significance level for trees with at most 100 leaves.

However, other significance levels are also of interest and more generally one wants to know how accurate the \( P \)-value estimate is when using our congruence index. To determine the level of accuracy, we have computed the exact \( P \)-values for tree sizes between 7 and 100 leaves (by generating 10 000 pairs of random trees for each number of leaves and calculating the size of the MAST) and compared them to the ones given by our index. Here, ‘exact’ is up to sampling uncertainties, which we render as small as possible. Figure 1 represents the estimated \( P \)-values against the exact \( P \)-values in the small \( P \)-value region, for all our data, i.e. for tree sizes between 7 and 100. By comparing with the straight line, we see that the estimated value tends to be smaller than the exact one; as a consequence, when testing the null hypothesis, if the estimated \( P \)-value is close to the imposed significance level \( \alpha \), the exact \( P \)-value may be above \( \alpha \) and the estimated one below, leading to too liberal a test as pointed out by Kupczok and Haeseler (2008). For illustration, we indicate in the figure the cases where this happens when \( \alpha = 5\% \) by adding crosses inside the symbols.

Nevertheless, we also see from this figure that the exact \( P \)-value is rarely larger than twice the estimated one, so if our index provides a small \( P \)-value then the exact \( P \)-value is also small.

At small \( P \)-values, we see a large spread in Figure 1 which means that tests in that regime will sometimes be too liberal and sometimes too conservative. To look into this matter, it is useful to consider the data at fixed tree size. In Figure 2, we plot the exact and estimated \( P \)-values for trees having 30 leaves as a function of MAST size (data collected from 1 million pairs of random trees).

The \( x \)-axis finishes at the smallest MAST size having an estimated \( P \)-value below 5%; the test at this level is too liberal, and not surprisingly the estimated \( P \)-value is smaller than the exact one. However, for larger MAST sizes, the estimated \( P \)-value becomes higher than the true one, rendering our test slightly too conservative. This pattern holds for the trees with more than 30 leaves, while smaller trees lead to too liberal a test for all MAST sizes. But again the important point is that the estimated \( P \)-values are relatively close to exact \( P \)-values. Our index thus gives a good qualitative measure of the similarity between trees, even if the \( P \)-values obtained are only semi-quantitative. Roughly speaking, if the estimated \( P \)-value is close to the significance level required, then the test is not conclusive because of the uncertainty; however, if the estimated \( P \)-value coming from our index is much smaller than the level required, then it is safe to reject the hypothesis of lack of congruence between the considered trees.

From our two figures, we see that our test performs its function very satisfactorily in most practical situations, rapidly providing a good idea of the level of similarity between two trees, before going into more detailed analyses. Furthermore, we have added to our web page which computes estimated \( P \)-values a quote of the exact \( P \)-value when the size of the MAST is critical (i.e. when the estimated and exact \( P \)-values are on different sides of the significance level of 5%) for trees with 100 leaves or less. Thus, users working at this significance level of 5% (a frequently used value) will know rigorously whether they can reject the hypothesis that two trees are no more congruent than random. The online calculation of the congruence index and the associated \( P \)-value is available at http://www.esf-upsud.fr/bases/upresa/pages/devienne/index.html. Note that for trees having between 78 and 100 leaves, when \( \alpha = 5\% \), the exact critical MAST and the one obtained from estimated \( P \)-values are identical. Furthermore, our method does not only apply to trees with less than...
Kupczok and Haeseler (2008) further objected that our method suffered from two additional pitfalls. First, by using only the size of the MAST, the positions of the taxa pruned are ignored. This is true, but as stated above and in de Vienne et al. (2007), the goal of our test is to provide a rapid and first estimation of the degree of congruence between two trees, not to replace more detailed analyses, such as those developed by Legendre et al. (2002) or Huelsenbeck et al. (1997, 2000), that are useful to investigate evolutionary scenarios and test for temporal congruence. If the use of our (approximate) test leads to the rejection of the null hypothesis, it simply indicates that the evolution in one group has most probably been dependent on the evolution in the other. However, other factors are susceptible to impact on the significance of the test, such as biases in tree reconstruction or an incorrect null distribution for trees if the model of equally likely trees is incorrect (Blum and François, 2006); by looking at more detailed information than the MAST size, one may get more stringent tests, but ours remains perfectly legitimate nevertheless. Another criticism of Kupczok and Haeseler (2008) was that a high number of taxa can be pruned to obtain identical trees, while the trees are still considered more similar than expected by chance. This is true, but means that the number of leaves to be pruned is still small enough to be very unlikely to have arisen by chance alone, and is thus an indication of some inter-dependent evolution of the two groups analyzed.

In conclusion, we believe that our index can be very useful for deciphering evolutionary associations among organisms, to get a first and rapid (computationally cheap) insight on the topological congruence between two trees. Also, with the particular treatment of the 5% significance level as implemented on the web page for the index, the exact critical MAST values and the associated exact P-values are available so users can reject the null hypothesis rigorously when appropriate.

REFERENCES


