An Age Dependent Branching Model for Macroevolution

Stephanie Keller-Schmidt[1*], Murat Tuğrul[2*], Víctor M. Eguíluz[2],
Emilio Hernández-García[2] and Konstantin Klemm[1]
[1]Bioinformatics, Institute of Computer Science, University Leipzig
Härtelstr. 16-18, 04107 Leipzig, Germany
[2]IFISC (CSIC-UIB) Instituto de Física Interdisciplinar y Sistemas Complejos,
E-07122 Palma de Mallorca, Spain

December 14, 2010

Abstract

The imbalance of phylogenetic trees exhibits a systematic deviation from the expectation of a purely random tree growth process as in the Yule or ERM models. Here we introduce an age dependent growth model based on the hypothesis that speciation rate is a decreasing function of the waiting time since the last speciation. We find that the imbalance in terms of the mean distance of tips from root (Sackin index) grows as $(\log n)^2$ in leading order with tree size n. This result is in good agreement with the trend observed by exhaustive analysis of the phylogenetic databases TreeBASE and PANDIT. Exact likelihood computation of the model on the trees up to 20 tips contained in the databases is performed. Higher likelihoods values are found when compared with a previously suggested model (Blum and François, 2006).

keywords: macroevolution, phylogenetic trees, branching models, imbalance

Reconstruction of phylogenies is instrumental for an understanding of the driving forces of evolution that have led to the diversity of living organisms. Hypotheses about the dynamical rules of speciation and extinction governing all evolutionary processes may come under scrutiny with large collections of phylogenetic trees available nowadays (Sanderson et al., 1994; Whelan et al., 2006). A suitable starting point and null hypothesis is the Equal Rate Markov (ERM) process suggesting that species undergo further speciation at a constant homogeneous rate, independently of previous events and other species present. The resulting topology of the growing tree, which is equivalent to the one produced by the Yule model (Yule, 1925), tends to generate compact and nearly *balanced* tree shapes, where *shape* denotes the simple rooted tree without branch lengths nor node labels (Mooers and Heard, 1997). When comparing with the shape of observed trees above a certain moderate size, however, the ERM hypothesis can be rejected, as most real phylogenetic trees are significantly less balanced than those generated by the ERM and Yule models (Herrada et al., 2008). *Imbalance* refers to an uneven distribution of the number of taxa (tips) between left and right branches of a tree or subtree.

The so-called beta-splitting (Aldous, 1996) is a class of models for stochastic tree generation with expected imbalance tunable by a parameter $\beta \in [-3/2; +\infty[$. The ERM model is reproduced at $\beta = 0$. The model of Proportional to Distinguishable Arrangements (PDA, $\beta = -3/2$) assigns to all tree shapes of a given size the same probability. The trees in TreeBASE have been found to match best with the intermediate parameter value $\beta = -1.0$ (Blum

*These authors contributed equally this work. Present address of M. Tuğrul: IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria

and François, 2006) a case called Aldous' Branching (AB) model (Aldous, 1996). The AB model and others (Ford, 2006) introduced to account for tree imbalance, however, assign probabilities to tree shapes in a way which is not based on any evolutionary principles. While the model can statistically reproduce features of the trees in the databases, it does not hint at any biological explanation of these features, as Blum and François (2006) remark.

Here we define a stochastic procedure to grow trees by iterative attachment of tips, similarly to the ERM model. At difference with the latter, the rate of speciation of a given species is assumed to decrease with the amount of time since last speciation of that species. We show below that this *age model* yields larger or equal likelihoods on small and medium-sized trees in the databases (where likelihood computation is feasible) when compared with the AB model. Furthermore, we use the depth (Sackin index) to quantify tree imbalance. For the age model, there is evidence that the expected depth increases as $(\log n)^2$ with the number of tips n. This growth law, identical to the AB model, is in good agreement with the depth values obtained from the databases.

Tree Balance

Several indices for balance measurement have been proposed, used and compared in the literature (see Mooers and Heard (1997); Matsen (2006); Agapow and Purvis (2002) for detailed discussion). Here we consider the depth (Sackin, 1972)

$$d = n^{-1} \sum_{i=1}^{n} d_i \tag{1}$$

as a measure of imbalance. In a tree with n tips, d_i denotes the number of edges to be traversed to reach the root from node $i \in \{1, \ldots, n\}$. This measure may be applied to non-binary trees (including polytomies and monotomies) and favors analytical treatment. Comparisons with other measures have been made by Matsen (2006); Agapow and Purvis (2002).

For a complete binary tree, $d = \log_2 n$ since all $n = 2^k$ tips are at level k. As the other extreme, a comb (or pectinate) tree has $nd = 1 + 2 + \cdots + (n-2) + 2(n-1)$ resulting in asymptotically linear scaling $d \sim n$ (we use the notation $f \sim g$ to indicate similar asymptotic behavior, i.e. to indicate that for large values of n, the ratio $f(n)/g(n)$ tends to a constant; an alternative notation is $f \in \Theta(g)$).

In the present work we calculated the depth d for all trees (and subtrees) in the phylogenetic databases TreeBASE (containing species phylogenies, Sanderson et al. (1994)) and PANDIT (protein phylogenies, Whelan et al. (2006)). The results in Figure 1 suggest that the average depth of the trees grows with the number of tips as

$$d \sim (\log n)^2 \tag{2}$$

in good approximation. Alternative analytic expressions for the growth can be fitted, in particular a power law $d \sim n^\alpha$, with $\alpha \approx 0.4$ describes TreeBASE data equally well (Herrada et al., 2008). But for the larger tree sizes contained in PANDIT, the $(\log\ n)^2$ form is more accurate.

Previous models

A simple mechanism to generate a binary rooted tree is given by the ERM model. It departs from a root node counting as a single tip ($n = 1$). Then in each iteration, a node i is drawn from the flat distribution on the set of tips, and two tips are attached to node i, increasing the number of tips by 1.

The ERM model provides a mechanism of a growing tree. This mechanism may be interpreted as a hypothesis for evolutionary dynamics. Here the hypothesis is that all species are involved in further speciation at the equal rates, irrespective of evolutionary history. The Yule model (Yule, 1925) is another implementation of this idea leading to trees with the same shape.

Abstracting from the dynamics behind tree generation, one may formulate a model directly in terms of a probability distribution on a set of trees. More precisely, a probability distribution is given separately for each set of all eligible trees of the same given tree size n. Here eligible trees are oriented binary rooted trees. *Oriented* is to say that left and right subtrees

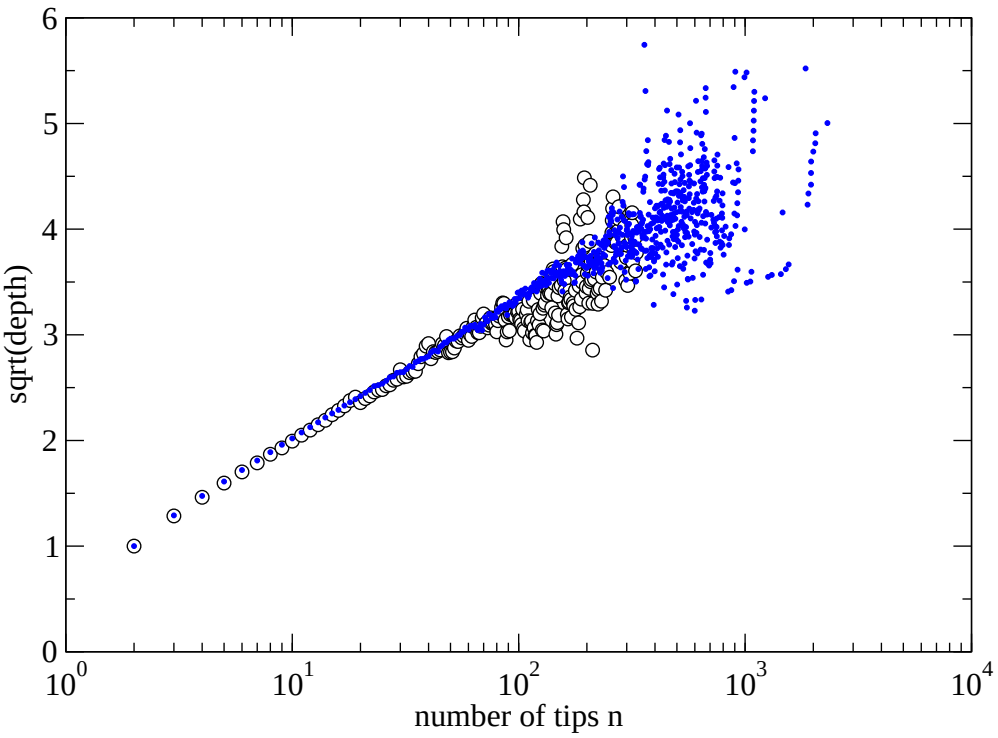

Figure 1: The square-root of the mean depth vs. size of phylogenetic trees contained in databases for species (TreeBASE; empty circles) and proteins (PANDIT; filled circles). The mean depth is averaged for all trees having the same number of tips. In this scale (log-linear), the behavior $\langle d \rangle \sim (\log n)^2$ is a straight line. Data from TreeBASE has been downloaded from `http://www.treebase.org` on June, 2007 containing 5,212 phylogenetic trees; data from PANDIT has been downloaded from `http://www.ebi.ac.uk/goldman-srv/pandit` on May 2008 and contains 7,738 protein families.

are explicitly distinguishable (by a left-right labeling). By this choice we avoid the need to consider isomorphy classes w.r.t. left-right symmetry.

In a particular class of models, the probability $L(T)$ of a tree T is defined by a product over its inner nodes $I = \{1, 2, \ldots, n-1\}$ according to

$$L(T) = \prod_{i \in I} p_{\text{model}}(l_i|n_i) \ . \tag{3}$$

The model-specific probability factor p_{model} depends on the total number n_i of tips in the subtree with root node i and the number l_i of tips in the left subtree of i. Arguments naturally fulfill $1 < l_i < n_i$. Left-right symmetry is ensured by

$$p_{\text{model}}(i|n) = p_{\text{model}}(n-i|n) \tag{4}$$

such that $L(T_1) = L(T_2)$ when T_1 is isomorphic to T_2. The choice of the functional form of p_{model} determines the expected balance of the trees. By concentrating probability mass at values l close to 2 and close to $n-1$, imbalance is enforced.

The ERM model is recovered as the case where all $n-1$ possibilities of left-right splitting are equally likely, so

$$p_{\text{Yule}}(l|n) = \frac{1}{n-1} \tag{5}$$

with equal probability for all $n-1$ possibilities for splitting n tips between left and right subtree. The expected depth grows as the logarithm of the number of tips,

$$\langle d \rangle(n) \sim \log n \ . \tag{6}$$

The ERM model is a particular case ($\beta = 0$) of *beta-splitting*. This is a one-parametric class of models with imbalance tunable by the parameter $\beta \in [-3/2, \infty[$.

$$p_\beta(l|n) = \frac{1}{a_\beta(n)} \frac{\Gamma(\beta + l + 1)\Gamma(\beta + n - l + 1)}{\Gamma(l+1)\Gamma(n-l+1)} \tag{7}$$

with suitably chosen normalization $a_\beta(n)$. Taking $\beta \to \infty$ produces maximally balanced trees. As the opposite extreme, the *Proportional to Distinguishable arrangements* (PDA) model is obtained at $\beta = -3/2$. Here the depth grows algebraically with n as

$$\langle d \rangle(n) \sim \sqrt{n} \ . \tag{8}$$

This *square-root* scaling is obtained also for a different model of tree growth, the *activity model* (Hernández-García et al., 2010).

The parameter value $\beta = -1$ is of particular interest because it has been demonstrated to maximize the agreement of beta-splitting with observed phylogenetic trees (Blum and François, 2006) in terms of imbalance. This choice $\beta = -1$ is also called Aldous' branching (AB) model with probabilities

$$p_{-1}(l|n) = \frac{1}{a_{-1}(n)} \frac{n}{l(n-l)} \tag{9}$$

The expected mean depth increases as

$$\langle d \rangle(n) \sim (\ln n)^2 \ . \tag{10}$$

The age model

Similar to the ERM model, the age model describes the growth of a binary tree by iterative stochastic addition of tips. Each tip i is assigned an age $\tau_i(t)$ being the time that passed from the birth of the tip, t_i, to present time t, i.e. $\tau_i(t) = t - t_i$. The growth proceeds by iterating the following three steps. (i) A tip i is chosen with probability $p_i(t)$ inversely proportional to age

$$p_i(t) = \frac{\tau_i^{-1}}{c(t)}, \tag{11}$$

where $c(t)$ is chosen such that probabilities of all tips sum up to 1; (ii) Two new tips j and k with creation times $t_j = t_k = t$ are attached to node i; (iii) time t is increased by Δt and the process resumes at (i). Here we consider a constant time increment $\Delta t = 1$ unless indicated otherwise. With this choice, time t is equivalent to number of branching events, and $t = n - 1$. A different choice is briefly considered below.

Depth asymptotics

Now we analyze the n-dependence of the expected depth of trees stochastically generated by the age model with $\Delta t = 1$. Numerical and heuristic arguments strongly suggest that $d \sim (\log n)^2$ is the

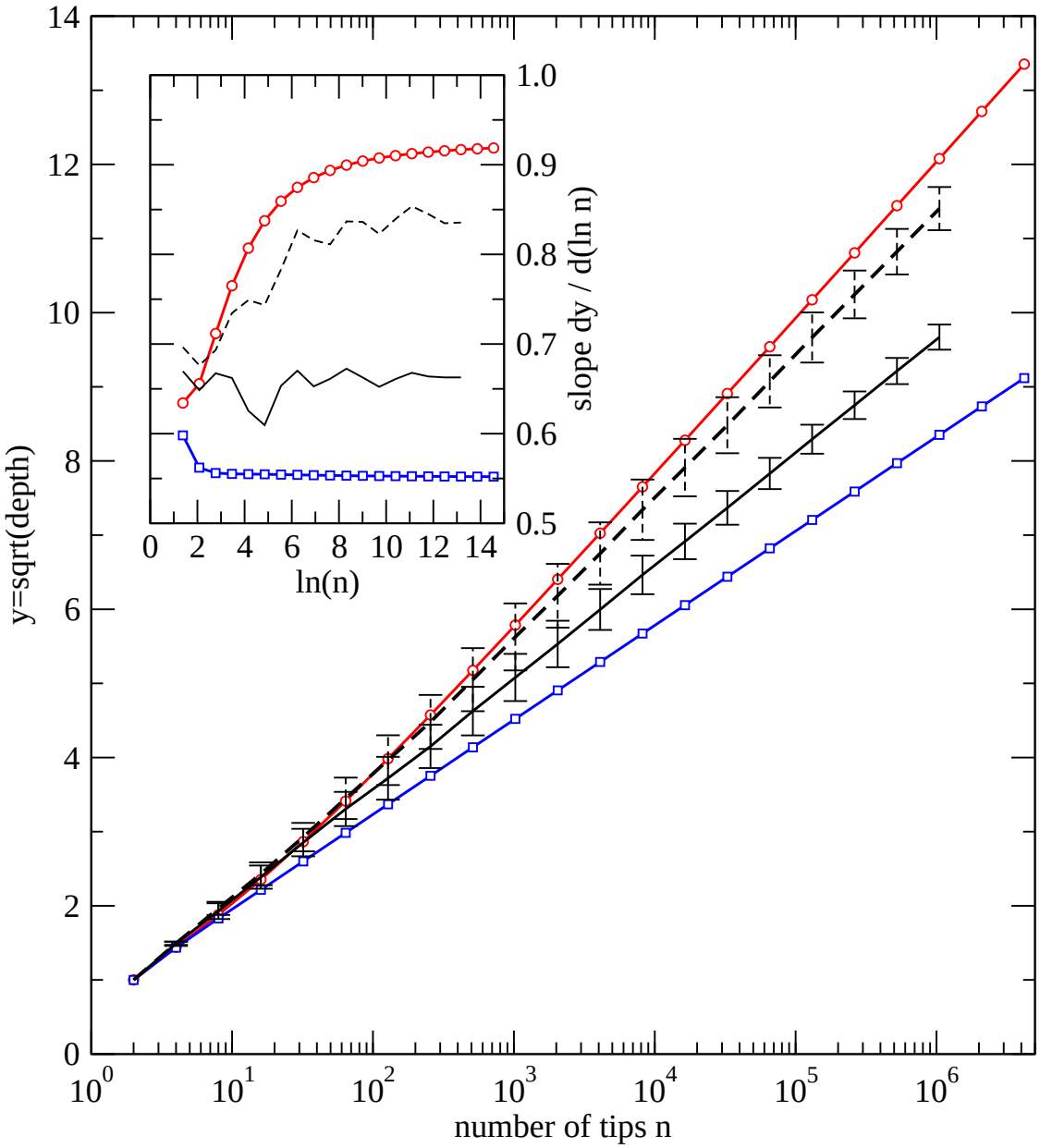

Figure 2: The dependence of depth d on the number of tips n. Curves with symbols are the lower (squares) and upper (circles) bounds obtained by the recursion Eqs. (18) and (21) inserted into Eq. (23). Stochastic simulations yield an average depth plotted as the solid line with error bars indicating standard deviation over the 30 independent realizations with $\Delta t = 1$. Analogously, the dashed line is for stochastic simulations but using a time increment $\Delta t = 1/n$. Note that $\sqrt{d}$ is plotted over a logarithmic n-axis, so the dependence $d \sim (\log n)^2$ results in a straight line. The inset shows the slopes of the curves in the main panel, which display better the asymptotic approach to a constant slope, i.e. the approach to a $(\log n)^2$ growth.

asymptotic growth law for this model, but we can not provide a fully rigorous demonstration of that. Instead we establish here upper and lower bounds for the depth in the model, and provide convincing numerical evidence for the $(\log n)^2$ scaling of them, from which the same behaviour would hold for $\langle d\rangle(n)$.

Let us first consider a single realization of the stochastic process. For each integer time $t > 0$, let $\delta(t)$ be the distance from root of the two new tips added at time t. This means that $\delta(t) - 1$ is the distance from root of the tip chosen to speciate. Let $\tau(t)$ be the age of the tip chosen at time t. Then $\delta(t)$ obeys the recursion

$$\delta(t) = \delta(t - \tau(t)) + 1 \ . \tag{12}$$

for $t > 1$ with $\delta(1) = 1$ as initial condition.

Let us now consider the case that the process has generated the sequence $\delta(1), \delta(2), \ldots, \delta(t-1)$ and we would like to know the expectation value η of $\delta(t)$. In the calculation of η also the distribution $f(\tau, t)$ of ages of the tips of the tree enters as

$$\eta = 1 + \frac{\sum_{\tau=1}^{t-1} f(\tau,t)\tau^{-1}\tau^{-1}\delta(t-\tau)}{\sum_{\sigma=1}^{t-1} f(\sigma,t)\sigma^{-1}} \ . \tag{13}$$

Let us now establish an f-independent lower-bound on η. To this end, we define a particular age distribution as

$$f_{\leq}(\tau,t) = \begin{cases} 2t^{-1}, & \text{if } \tau = 1 \\ t^{-1}, & \text{if } 2 \leq \tau < t \\ 0, & \text{if } t \leq \tau \end{cases} \tag{14}$$

Dynamically, this age distribution is obtained when one of the youngest tips ($\tau = 1$) is chosen in each step. One can show that replacing the actual age distribution f by $f_{\leq}$, the expected level does not increase. Therefore

$$\eta \geq 1 + \frac{\sum_{\tau=1}^{t-1} f_{\leq}(\tau,t)\tau^{-1}\tau^{-1}\delta(t-\tau)}{\sum_{\sigma=1}^{t-1} f_{\leq}(\sigma,t)\sigma^{-1}} \ . \tag{15}$$

The expectation value $\langle\delta\rangle(t)$ over the whole stochastic process is obtained formally by an average over all histories as follows. Call $\mathcal{D}_t$ the set of all eligible distance sequences of length $t-1$ and $\mathcal{F}_t$ the set of all eligible age distributions at time t. Then we may write

$$\langle\delta\rangle(t) = 1 + \sum_{\delta\in\mathcal{D}_t}\sum_{f\in\mathcal{F}_t} p(\delta,f,t)\frac{\sum_{\tau=1}^{t-1} f(\tau,t)\tau^{-1}\delta(t-\tau)}{\sum_{\sigma=1}^{t-1} f(\sigma,t)\sigma^{-1}} \tag{16}$$

with p being the joint distribution of distance sequence and age distribution at a given time. An exact solution for $\langle\delta\rangle(t)$ would thus involve a recursion for p, which is difficult to treat. The lower bound on η in Eq. (15), however, is valid for each possible realization of the process. Therefore

$$\langle\delta\rangle(t) \geq 1 + \frac{\sum_{\tau=1}^{t-1} f_{\leq}(\tau,t)\tau^{-1}\sum_{\delta\in\mathcal{D}_t}\delta(t-\tau)p'(\delta,t)}{\sum_{\sigma=1}^{t-1} f_{\leq}(\sigma,t)\sigma^{-1}} \tag{17}$$

where p' is the marginal of p after summation over $\mathcal{F}_t$. Performing the sum over $\mathcal{D}_t$ yields

$$\langle\delta\rangle(t) \geq 1 + \frac{\sum_{\tau=1}^{t-1} f_{\leq}(\tau,t)\tau^{-1}\langle\delta(t-\tau)\rangle}{\sum_{\sigma=1}^{t-1} f_{\leq}(\sigma,t)\sigma^{-1}} \tag{18}$$

Thus we have established a recursion for a lower bound on $\langle\delta\rangle$.

Likewise, the age distribution

$$f_{\geq}(\tau,t) = \begin{cases} 2t^{-1}, & \text{if } \tau \leq \lfloor t/2 \rfloor \\ t^{-1}, & \text{if } \tau = (t+1)/2 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

can be used to establish an upper-bound recursion. Dynamically, this age distribution is obtained when an oldest tip is chosen in each step. One can show that

$$\eta \leq 1 + \frac{\sum_{\tau=1}^{t-1} f_{\leq}(\tau,t)\tau^{-1}\tau^{-1}\delta(t-\tau)}{\sum_{\sigma=1}^{t-1} f_{\leq}(\sigma,t)\sigma^{-1}} \ . \tag{20}$$

By arguments analogous to the above, we arrive at the upper-bound recursion

$$\langle\delta\rangle(t) \leq 1 + \frac{\sum_{\tau=1}^{t-1} f_{\leq}(\tau,t)\tau^{-1}\langle\delta(t-\tau)\rangle}{\sum_{\sigma=1}^{t-1} f_{\leq}(\sigma,t)\sigma^{-1}} \ . \tag{21}$$

For transforming $\langle\delta\rangle$ into expected depth d, consider the sum of distances of tips from root, $D(t) =$

$td(t)$. Addition of two tips at distance x from root increases D by $2x - (x-1) = x+1$. Thus

$$D(t) = \sum_{s=2}^{t} (\delta(s) + 1) \tag{22}$$

for a realization of the stochastic process with level sequence δ. By linearity of expectation values the expected depth is

$$\langle d(t) \rangle = \sum_{s=2}^{t} [\langle \delta(s) \rangle + 1]/t \ . \tag{23}$$

Figure 2 shows upper and lower bounds on the expected depth $\langle d \rangle$ obtained as numerical solutions of the recursion equations (18) and (21). In the same diagram we plot the results of direct simulations of the model. In one set of simulations we use the usual time increment $\Delta t = 1$, so that $t \sim n$. Another set of simulations is performed with $\Delta t = 1/n$ to check for robustness under different evolution of overall speciation rates. Upper and lower bounds as well as the two sets of simulations strongly suggest that the asymptotic growth behavior for the depth is $(\log n)^2$.

Likelihood Analysis

Abstracting from the algorithmic formulation of tree generation, a branching model A can be characterized by the probability $L_A(T)$ of obtaining a given tree T. The quantity $L_A(T)$ is also called the *likelihood* of model A under the data (tree) T. When aiming at modeling empirical data, we would say that one model A is better than a different model B, if $L_A(T) > L_B(T)$ for an observed tree T.

For the ERM model and the AB models defined in the previous section, the calculation of likelihoods is straight-forward:

$$L_A(T) = \sum_{x \in I(T)} p_{\mathrm{A}}(s(\mathrm{left}(x)) | s(x)) \tag{24}$$

where A is the model under consideration, $I(T)$ is the set of inner nodes of tree T, $s(x)$ is the number of tips in the subtree with the root x and left(x) is the left child node of node x. For the age model it is not clear if a simple method of exact likelihood computation exists. Here we calculate the exact value of $L_{\mathrm{age}}(T)$ by adding up probabilities of *all* branching orders leading to the observed tree T. Details are described in the Appendix.

Figure 3 shows that the likelihoods of the age model and AB model are clearly correlated under the trees in the databases. The variation of likelihoods across trees of the same size n is smaller in the age model compared to that in the AB model. Notably, the age model has larger likelihood than the AB model under more than half of the trees under consideration, so that it can be considered a better description of the evolutionary process.

Concluding remarks

The proposed *age model* compares with observed phylogenetic trees better than previous models. In addition, it describes the tree generation process in a way which is easy to interpret biologically: it assumes that lineages which have not speciated for a long time would display in the future a still more reduced speciation rate.

Future work should provide a more detailed analysis of the model itself and further comparison to real phylogenetic trees. For the former, an analytic solution for the expected depth of for its bounds is lacking. It would be also desirable to obtain expressions or at least numerical results for the second and perhaps higher moments. For the likelihood expressions, a factorization or other kind of decomposition would allow for faster exact computation. Instead of exact computation, estimation by a Monte-Carlo sampling method may circumvent the present size limitation of trees in the likelihood analysis.

An additional interesting point of analysis and comparison of phylogenetic trees is the distribution of branch length. Branch length data, however, are not as reliable as the topological structure of phylogenetic trees (Barraclough and Nee, 2001). This argument is supported by Pigolotti et al. (2005), summarizing the variety of behaviors of distributions found in the literature. We believe that future studies in the line of Venditti et al. (2010) will accumulate sufficiently

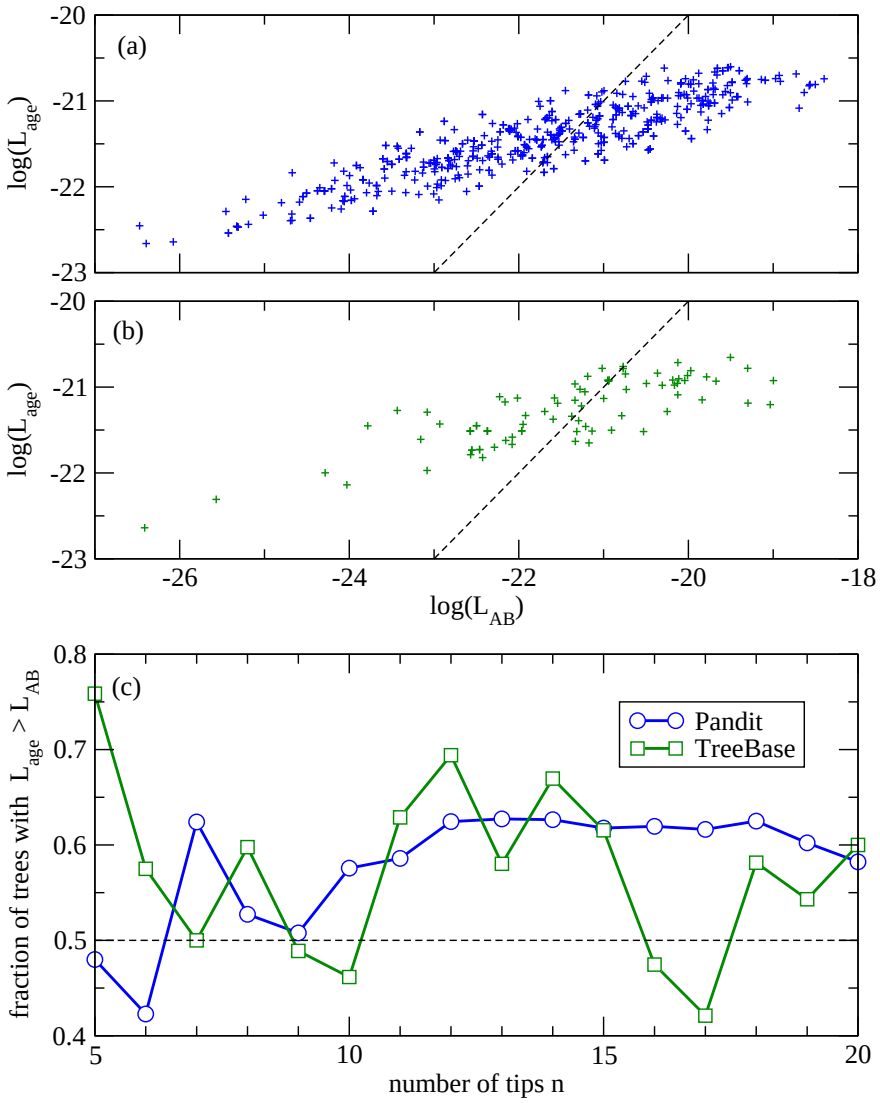

Figure 3: Comparison between age and AB models by likelihoods under tree shapes from databases. (a) $\log(L_{\text{age}}(T))$ versus $\log(L_{\text{AB}}(T))$ for each of the 541 tree shapes T with $n = 20$ tips in the database PANDIT. The dashed line is the identity. (b) Same as (a) for the 85 tree shapes with $n = 20$ tips in the database TreeBASE. (c) Fraction of trees T with $L_{\text{age}}(T) > L_{\text{AB}}(T)$, separately for each $n \in \{5, \ldots, 20\}$. The overall fraction is $0.554 = 14088/25441$ for PANDIT and $0.562 = 842/1499$ for TreeBASE. The number of available tree instances is one order of magnitude smaller in TreeBASE than in PANDIT leading to larger fluctuations in the TreeBASE results.

reliable branch-length data to allow for comparison to models such as the present one.

Finally, timing in the model is worth further clarification. The model describes tree growth as a Markov chain where exactly one speciation event occurs at each time step. A more realistic version would formulate a Markov process that assigns a speciation rate to each species at any moment in continuous time. The choice $\Delta t = 1/n$ in the results of Fig. 2 is a first step in that direction.

Funding

This work has been supported by the European Commission NEST Pathfinder initiative on Complexity through project EDEN (Contract 043251), by MICINN and FEDER through project FISICOS (FIS2007-60327) and by VolkswagenStiftung through contract I / 82 719.

Acknowledgements

MT would like to thank Alejandro Herrada, Stephan Steigele, Joan Pons for their valuable discussions with him regarding biological evolution.

References

Agapow, P. M. and A. Purvis. 2002. Power of eight tree shape statistics to detect non-random diversification: A comparison by simulation of two models of cladogenesis. Syst. Biol. 51:866–872.

Aldous, D. 1996. Probability distributions on cladograms. Pages 1–18 *in* Random Discrete Structures (D. Aldous and R. Pemantle, eds.) Springer.

Barraclough, T. and S. Nee. 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16:391–399.

Blum, M. G. and O. François. 2006. Which random processes describe the tree of life? a large-scale study of phylogenetic tree imbalance. Syst. Biol. 55:685–691.

Ford, D. J. 2006. Probabilities on cladograms: introduction to the alpha model. Ph.D. thesis Stanford University available from `arXiv:math.PR/0511246`.

Hernández-García, E., M. Tugrul, E. Herrada, V. Eguíluz, and K. Klemm. 2010. Simple models for scaling in phylogenetic trees. Int. J. Bif. Chaos 20:805–811.

Herrada, E. A., C. J. Tessone, K. Klemm, V. M. Eguíluz, E. Hernández-García, and C. M. Duarte. 2008. Universal scaling in the branching of the tree of life. PLoS ONE 3:e2757.

Matsen, F. A. 2006. A geometrical approach to tree shape statistics. Syst. Biol. 55:652–661.

Mooers, A. and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:31–54.

Pigolotti, S., A. Flammini, M. Marsili, and A. Maritan. 2005. Species lifetime distribution for simple models of ecologies. Proc Natl Acad Sci U S A 102:15747–15751.

Sackin, M. 1972. Good and bad phenograms. Syst. Zool. 21:225–226.

Sanderson, M. J., M. J. Donoghue, W. Piel, and T. Eriksson. 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. American Journal of Botany 81:183 http://www.treebase.org.

Venditti, C., A. Meade, and M. Pagel. 2010. Phylogenies reveal new interpretation of speciation and the red queen. Nature 463:349–352.

Whelan, S., P. I. W. de Bakker, E. Quevillon, N. Rodriguez, and N. Goldman. 2006. PANDIT: an evolution-centric database of protein and associated nucleotide domains with inferred trees. Nucleic Acids Res 34:D327–D331 http://www.ebi.ac.uk/goldman-srv/pandit/.

Yule, G. U. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Phil. Trans. R. Soc. Lond. B 213:21–87.

APPENDIX: Likelihood computation

The probability $L_{\text{age}}(T)$ of the age model generating a given rooted binary tree shape T is calculated as follows. The nodes of the tree are assigned unique labels in $\{1,\ldots,2n-1\} := A$, where the *inner* nodes have the labels $\{1,\ldots,n-1\} := I$. The root has label 1. For a non-root node $i > 1$, we denote the unique parent node by $m(i)$. We call S the set of all permutations of I, so each element of S is a bijection $s : I \to I$. Such a permutation is to encode a branching order of a tree: $s(i)$ is the time step at which node i branches. In a valid branching order, children cannot branch earlier than the parent. Thus we say that $s \in S$ is *compatible* with T, if $s(i) > s(m(i))$ for each $i \in I \setminus \{1\}$. We call $S_c(T) \subseteq S$ the set of compatible permutations. When branching according to $s \in S_c(T)$, the set of tips at time $t > 1$ is

$$\begin{aligned} B(s,t) &= \{j \in I \setminus \{1\} \mid s(m(j)) < t < s(j)\} \\ &\quad \cup \{j \in A \setminus I \mid s(m(j)) < t\} \ . \end{aligned} \tag{25}$$

The age of a tip j at time $t > 1$ is $t - m(j)$. Thus the age model generates the tree T with the branching order given by $s \in S_c$ with probability

$$p(s,T) = \prod_{i=2}^{n-1} \frac{(s(i) - s(m(i))^{-1}}{\sum_{j \in B(s,s(i))} (s(i) - s(m(j))^{-1}} \tag{26}$$

The overall probability of generating T with the age model is obtained by summing over all branching orders generating T,

$$L_{\text{age}}(T) = \sum_{s \in S_c(t)} p(s,T) \ . \tag{27}$$