

Estimating TimeTrees with MEGA and the TimeTree Resource

Beatriz Mello^{*1}

¹Department of Genetics, Federal University of Rio de Janeiro, RJ, Brazil

***Corresponding author:** E-mail: biau_mello@gmail.com.

Associate editor: Joel Dudley

Abstract

The main outcome of molecular dating, the timetree, provides crucial information for understanding the evolutionary history of lineages and is a requirement of several evolutionary analyses. Although essential, the estimation of divergence times from molecular data is frequently regarded as a complicated task. However, establishing biological timescales can be performed in a straightforward manner, even with large, genome-wide data sets. This protocol presents all the necessary steps to estimate a timetree in the program MEGA X. It also illustrates how the TimeTree resource can be a useful tool to obtain chronological information based on previous studies, therefore yielding calibration boundaries.

Key words: RelTime, divergence times, molecular dating, calibration, MEGA X.

Protocol

In a phylogenetic tree, internal nodes represent the most recent common ancestors (MRCA) of at least two evolutionary lineages. These lineages may be species, genes, or alleles within populations (Nei and Kumar 2000). The use of biological sequences (nucleotides or amino acids) has become the most frequent way to infer the relationship among taxa across the tree of life, from microorganisms to plants and animals (Marin et al. 2016; Morris et al. 2018; Shin et al. 2018). Likewise, molecular dating, that is, estimating the age of internal nodes based on nucleotides or amino acid sequences, is now the standard method to infer timetrees. Establishing a chronological scale for the evolution of clades/lineages is a fundamental step in order to draw biological conclusions from molecular sequence data (Ho and Duchêne 2014). The significance of timetrees in biology ranges across a wide temporal spectrum that goes from deep time studies, such as the diversification of major taxa and their association with Earth's history (Misof et al. 2014), to contemporaneous epidemiological problems, such as the inference of evolutionary dynamics of pathogens and the spread of diseases (Faria et al. 2017).

Numerous methodological approaches were devised to estimate timetrees, and molecular dating underwent distinct phases since it was first proposed in the 1960s by Zuckerkandl and Pauling (Zuckerkandl and Pauling 1962, 1965; Kumar 2005; Kumar and Hedges 2016). For instance, modern methods are free from the assumption of constancy of evolutionary rates (i.e., a strict molecular clock). Sophisticated approaches, framed on Bayesian statistics, have been demonstrated to be very valuable in estimating divergence times by accommodating rate variation among lineages and incorporating uncertainty from calibration information from the fossil record (dos Reis et al. 2016). Among the several software implementations of Bayesian molecular dating methods, MCMCTree (Yang 2007) and BEAST (Bouckaert et al. 2014)

are the most widely used. However, with the tremendous growth in genomic data over the last several decades, the use of Bayesian methods is frequently unfeasible, as they require a large computational time (Akerborg et al. 2008; Battistuzzi et al. 2011; Ho 2014; Mello et al. 2017). Thus, fast-dating methods stand out as a powerful strategy to accelerate timetree inference and, consequently, all the scientific knowledge that emerges from it (Tamura et al. 2012; Ho 2014; To et al. 2016).

In this protocol, I will present a step-by-step approach to carrying out molecular dating with the RelTime method, a fast-dating and high-performance algorithm that is implemented in MEGA X (Kumar et al. 2016). MEGA is a software that performs a wide range of molecular evolutionary analyses, and it has been extensively used and continuously updated in the last 25 years (Kumar et al. 1994, 2018). MEGA has been established as a unified platform that allows users to carry out molecular evolution and phylogenetic analyses in one single environment. This makes MEGA very user-friendly and encourages researchers, even those new to the field, to perform complex evolutionary analyses. Additionally, as the dating of species divergences is generally meaningful only when based on time units measured in calendar years, the TimeTree web resource (Hedges et al. 2006; Kumar et al. 2017) will be used to inform the RelTime method about the ages of calibrations, transforming relative node ages into absolute times.

Estimating a Phylogenetic Tree

A previous MBE protocol already provides guidelines to estimate phylogenetic trees from molecular sequence data in MEGA (Hall 2013). This very didactic and useful protocol instructs researchers from the first steps (acquiring and aligning molecular sequences) to selecting the best substitution model and inferring maximum likelihood (ML) trees. In case the user is not familiar with phylogenetic tree inference, please refer to that protocol before going into molecular

dating analyses. Here, it is assumed that those preliminary steps were successfully completed, and a phylogenetic tree is available to estimate a timetree. Apart from MEGA, many other software has been written to estimate phylogenetic trees based on distance, ML and Bayesian methods (Guindon et al. 2010; Ronquist et al. 2012; Stamatakis 2014; Lefort et al. 2015; Nguyen et al. 2015). The phylogenetic trees output by these programs can be readily imported into MEGA to estimate divergence times. Note that the phylogenetic tree input format read by MEGA is the widely used Newick format. If your estimated phylogeny was not saved in this file format, guidelines to convert it to Newick are provided on the section “Estimating Divergence Times with MEGA X” (on “Step 2: Loading the Tree File”).

Calibrations and the TimeTree Database

Before carrying out the estimation of divergence times, temporal information should be obtained in order to calibrate the relative node ages. With this information, molecular dating methods are capable of converting time estimates to an absolute scale in calendar years. Thus, the ages of the internal nodes in the phylogenetic tree become meaningful and comparable between studies. Calibration information relies mostly on the fossil record, although other sources, such as the age geological events, are also used (Ho et al. 2015). Therefore, a detailed evaluation of the literature is required to derive calibrations. It mostly consists of compiling paleontological information that justifies minimum and/or maximum values that will be used to calibrate a node, that is, a divergence between lineages, in the phylogenetic tree. This is a crucial step in molecular dating, since the final timetree will depend on the calibration information provided. Because of this, several studies discuss how to carefully choose and derive boundaries for calibration points (Ho and Phillips 2009; Parham et al. 2012).

Although complex, the task of choosing suitable time boundaries is greatly facilitated by the TimeTree web resource (<http://www.timetree.org/>) (Kumar et al. 2017). TimeTree is a database that compiles divergence times estimated by distinct molecular dating studies across the tree of life. It also provides users with timelines for a particular taxon and subtrees of the timetree of life estimated in Hedges et al. (2015), which has been further improved (Kumar et al. 2017). Although TimeTree is not a fossil database, it is extremely useful to begin a search for calibrations and to collect the estimated molecular times for a divergence of interest.

To obtain previous estimates of divergence times for a pair of taxa, you should only provide the taxon names. For instance, suppose you need to establish a timetree for great apes. To get the pairwise divergence times between gorillas and humans, just type their names on the **Get Divergence Time for a Pair of Taxa** field. The TimeTree resource will first try to resolve any ambiguities. You should then click on the **Show Time** button. You will recover a list of molecular times along with information of the reference studies, as well as physical and geological information (similar to figure 6 from Kumar et al. 2017). For gorillas and humans, the median time

among all retrieved studies is 8.61 Ma, and the estimated time based on the study of Hedges et al. (2015) (updated in Kumar et al. 2017) is 9.06 Ma, with a confidence interval ranging from 8.44 to 9.69 Ma. By replicating this same procedure for the split between both living species of chimpanzees, the bonobo and the common chimpanzee, you will get a median time of 2.375 Ma, and estimated time of 2.820 Ma (confidence interval from 2.399 to 3.241 Ma).

Since TimeTree is based on many published works, the time information provided by this resource is a good starting point to obtain calibration information. In particular, the confidence intervals may provide reasonable ages for minimum and maximum boundaries to calibrate the divergence time estimation and, for didactic purposes, we will use it in this protocol. It is worth mentioning though that there are other very useful web-resources that provide fossil information, such as the Fossil Calibration Database (Ksepka et al. 2015) and the Paleobiology Database (<https://paleobiodb.org/>). However, regardless of the source of calibration you are using, make sure you have adequately selected it before performing a molecular dating analysis. Finally, once you have the phylogeny and the calibration boundaries, you are ready to estimate divergence times with RelTime in MEGA X.

Estimating Divergence Times with MEGA X

MEGA was recently updated to a new version, MEGA X. As a major improvement, in addition to Windows, MEGA X now runs natively on Linux machines, maximizing the use of computer resources when running on this operating system. MEGA X version for macOS is currently being developed and will be released in the near future (Kumar et al. 2018). The program can be downloaded at <https://www.megasoftware.net>. Once the download is complete, double-click the executable and follow the steps for installation. The default installation process will generate a Desktop icon, which launches MEGA X once you double-click it. It also creates a folder called MEGA X in your Documents directory, which contains several example files. Within this folder, you will find the files named mtCDNA.meg and mtCDNA.nwk, which contain mitochondrial alignment data and the phylogenetic tree of the great apes' example for which molecular times were previously retrieved from the TimeTree resource (example files are also provided as [Supplementary Material](#)). With these two example files, all the following steps can be performed in MEGA X.

When you start MEGA, the main MEGA window will be opened. To perform a molecular dating analysis, from the **Clocks** menu, choose **Compute TimeTree**. Then, the next three options will be available to users: **RelTime-ML**, **RelTime-OLS**, and **RelTime-Branch Lengths**. The **RelTime-ML** option is the original method proposed by Tamura et al. (2012). In most cases, this is the option you want. Time estimates are computed based on branch lengths optimized by ML, which is a robust statistical method.

RelTime-OLS relies on the estimation of branch lengths by the ordinary least-squares approach, which is a distance method (Rzhetsky and Nei 1993). Such an option can be

useful when dealing with big data sets that have a good coverage (i.e., small amounts of missing data), because it is faster than the ML-based option. Lastly, **RelTime-Branch Lengths** allows the user to provide the phylogenetic tree with previously estimated branch lengths. In this case, MEGA will not provide confidence intervals for node ages. This option is useful when the user has already optimized branch lengths under approaches different from ML or OLS, or under models of evolution that are not yet implemented in MEGA (e.g., the Lewis 2001 model for morphological data). In this case, you should not provide your alignment data and go directly to step 2.

It is important to mention that in case you have just estimated a phylogenetic tree with MEGA, please note the section entitled “Performing a Phylogenetic Reconstruction Followed by a TimeTree Estimation in MEGA X” near the end of this paper.

Step 1: Loading Sequence Data

Choose between the ML or OLS optimization of branch lengths (ML is most suitable except in the case of very large data sets), and then the TimeTree wizard similar to that of figure 1 will open. In this wizard, you will provide all the relevant information to estimate the timetree. The first step consists of loading the sequence data, which can be nucleotide or amino acid sequences. This is the same alignment file that you have used to infer the phylogenetic tree to be dated. Click **Browse** and select the file in your computer that contains your alignment. If you are using the MEGA example files, select the “mtCDNA.meg” file that is within MEGA X folder. Two alignment formats are accepted, MEGA (.meg extension) or fasta (.fas or .fasta extensions). The default behavior of MEGA is to show MEGA files. If your alignment is in fasta format, change to this format on the bottom right panel and select your alignment file accordingly.

Step 2: Loading the Tree File

Once the alignment is loaded, you will need to load the phylogenetic tree that you have previously estimated based on that same sequence alignment (or the “mtCDNA.nwk” file, in case you are using the MEGA example files). Notice that if you have chosen the **RelTime-Branch Lengths** option, this is your first step, because in this case, no information from the alignment is required. The tree should be in the Newick format (.nwk extension), which is the parenthetical notation used by most phylogenetic software. Another common file format that is used to save phylogenetic trees is Nexus (.nex or .nexus extension). Some phylogenetic programs output the estimated tree in this format by default, such as MrBayes (Ronquist et al. 2012). In this case, you can open your tree in Nexus format in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) and export it to Newick. To do this, open FigTree, click on **File** and then **Open**. Select the Nexus file you want to convert to Newick format and click **Open**. The phylogenetic tree will be displayed. Then, to save it in Newick, click **File** and **Export Trees...** In the **Tree file format**, choose **NEWICK**, and then click **OK**. Lastly, provide a name for your Newick tree file. The easiest way is to provide the

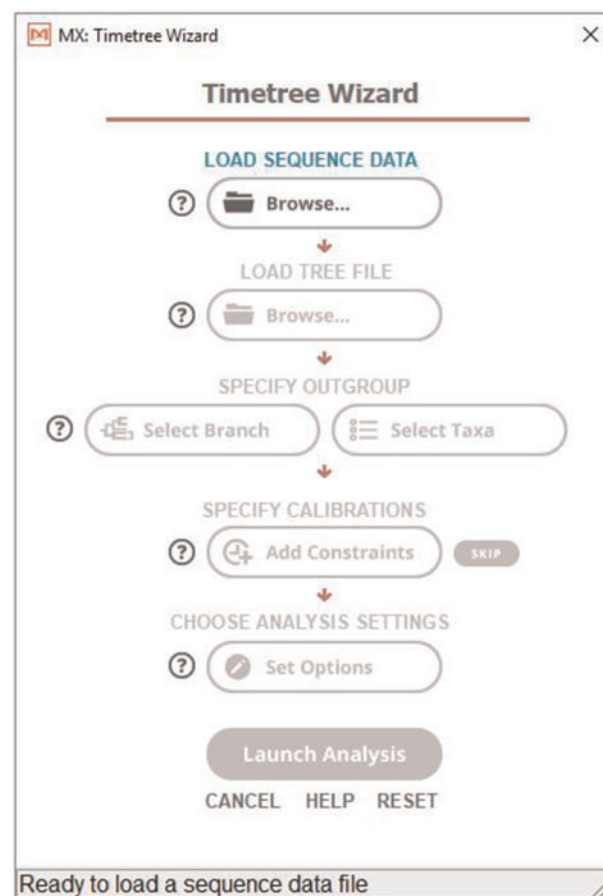


Fig. 1. Timetree wizard.

same name of your Nexus tree file, but add the extension “.nwk,” so that MEGA will recognize it automatically. Once the phylogenetic tree in Newick format is ready, in the MEGA TimeTree wizard, click **Browse** and select it. Make sure the tip names match between your alignment and phylogeny; if they do not, it is impossible to relate data information with the phylogenetic relationships.

Step 3: Specifying Outgroups

Next, you need to specify the outgroups of your phylogeny. This is because phylogenetic building approaches, such as Neighbor Joining, ML, and Bayesian methods, estimate unrooted trees. However, to estimate timetrees, it is necessary to provide direction to the evolutionary process, which is commonly done by the use of outgroups. Outgroup sequences or taxa contain information external to the *ingroup*, which is the set of sequences/taxa in which you are interested. For this reason, outgroup information is able to root the phylogeny and place a temporal direction in its internal divergences. See step 4.1 and 4.2 of Hall (2013) for more details about how to choose outgroups.

MEGA provides two distinct ways to specify outgroups. The distinction between them is only a matter of preference, because in the end, regardless of the way you choose, the definition of outgroups is the same. If you choose **Select Branch**, a tree explorer window will open in which you have to select the branch of the phylogeny where all outgroups are contained. Once you select it, all the descendent

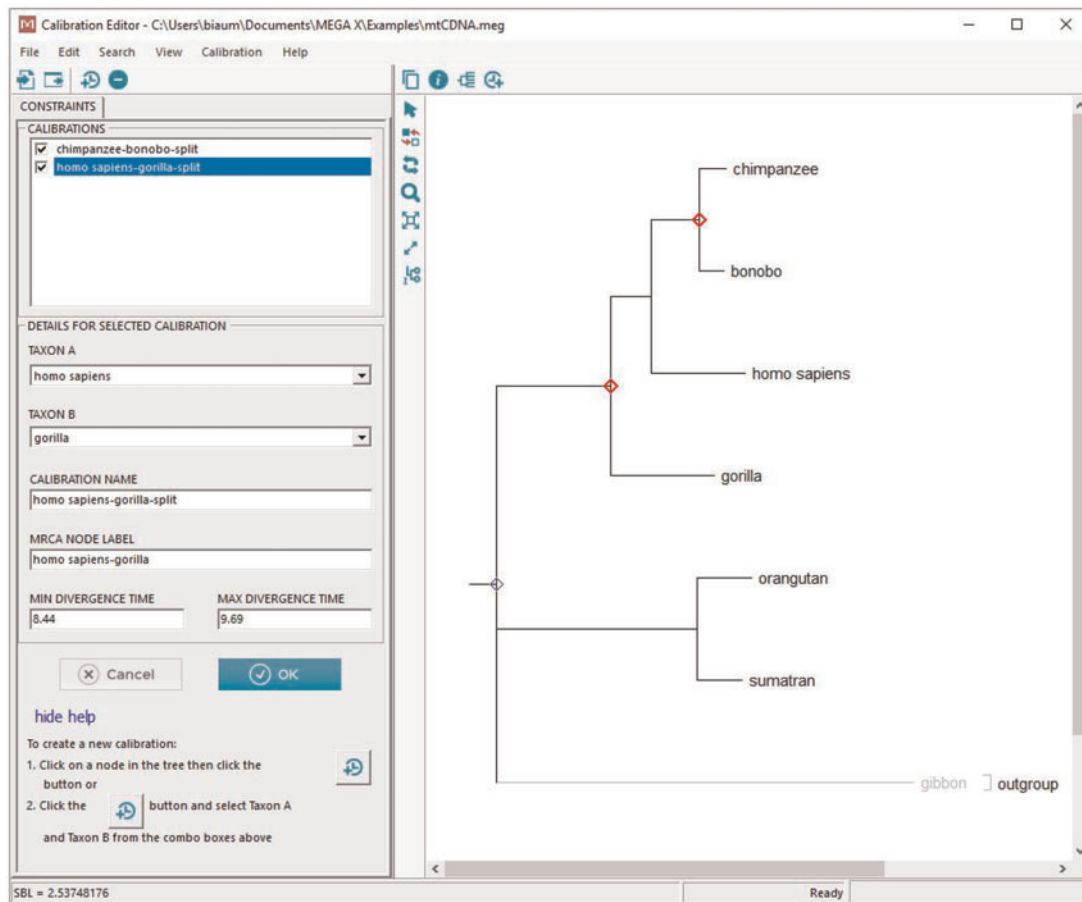


FIG. 2. Calibration editor. MEGA X highlights the nodes with assigned calibrations.

lineages and tip names will be highlighted in red and the word “outgroup” will be displayed on the right (in the great apes’ example, you should select the gibbon branch). If the tree is too large, some features of the tree explorer window can be very useful. For instance, in the case of big phylogenies, overlapping tip names will possibly be hidden. If you want to see all of them, you should deselect the “Hide overlapping taxa” option from the **View** menu. Additionally, you can search a tip name by clicking on the “magnifying glass” button and choosing whether the name should contain or starts with the text you entered. When you click the “search” button, the sequences for which you are looking will be displayed in green.

Double check to make sure you defined the outgroups correctly. This is very important, since an incorrect specification of outgroups will compromise the whole analysis. Once you are sure the selection is correct, click **OK** at the top. A window will prompt you for the confirmation of the outgroups that were selected. If, for some reason, they were not specified properly, click **NO** and correct the misspecification. If the outgroups are right, click **Yes**.

The alternative way of specifying the outgroups is by **Select taxa**. With this option, a window will be opened containing the list of all ungrouped taxa of the phylogenetic tree in the tab **Select Outgroup Taxa**. The left panel lists the tips that will be specified as outgroups, where you can include or

exclude them by using the arrows. Include the gibbon in case you are using the example files. Again, make sure the outgroup sequences/taxa were specified properly and click **OK**.

Step 4: Specifying Calibrations

The RelTime method does not require any calibration information to estimate divergence times. If no calibrations are provided, it will compute relative times without assuming a strict molecular clock. Therefore, the order of divergences is estimated, although the absolute scale in calendar years is absent. Nevertheless, relative times can still be very useful to date duplication events of gene families and to estimate timetrees for biological groups without calibration information. In the calibration specification stage, if you select **SKIP**, RelTime will then provide you a timetree on a relative time scale. Despite being valuable, a relative timescale may not be meaningful in most cases. Studies mostly rely on absolute divergence times to draw biological conclusions from timetrees. The absolute chronological information contained in the timetree can be compared with many other biotic and abiotic factors. Therefore, in this protocol, we will focus on estimating absolute node ages.

Consequently, to estimate absolute times, click **Add Constraints**. A window called “Calibration Editor” will be opened, where all the temporal constraints derived from the calibration information should be provided (fig. 2). To

Option	Setting
ANALYSIS	
Tree to Use →	C:\Users\biaum\Documents\MEGA X\Examples\mtCDNA.nwk
User Tree File →	C:\Users\biaum\Documents\MEGA X\Examples\mtCDNA.nwk
CLOCK SETTINGS	
Clock Type →	Local clocks
Variance Estimation Method →	Analytical method
Statistical Method →	Maximum Likelihood
SUBSTITUTION MODEL	
Substitutions Type →	Nucleotide
Model/Method →	General Time Reversible model
RATES AND PATTERNS	
Rates among Sites →	Gamma Distributed (G)
No of Discrete Gamma Categories →	5
DATA SUBSET TO USE	
Gaps/Missing Data Treatment →	Use all sites
Site Coverage Cutoff (%) →	Not Applicable
Branch Swap Filter →	None
SYSTEM RESOURCE USAGE	
Number of Threads →	1

FIG. 3. RelTime-ML analysis preferences.

assign temporal information to a node, click on the button “Add a new calibration point” (represented by a clock and a plus sign). Then, choose two taxa (“TAXON A” and “TAXON B”) for which the MRCA (most recent common ancestor) corresponds to the node you will provide a calibration information. Set the minimum and maximum divergence times that define the time boundaries for the calibrated split. For each added calibration, MEGA will highlight the corresponding node on the phylogeny you provided. Make sure you correctly assigned the calibration points to the desired nodes. Figure 2 shows the time range used to calibrate the divergence between humans (“TAXON A”) and gorillas (“TAXON B”) (8.44–9.69 Ma), which was derived from the confidence interval provided by the TimeTree database, as previously discussed.

Step 5: Choosing Analysis Settings

After setting up calibration constraints, you will need to specify the analysis settings. In the “Analysis Preferences” window (fig. 3), first, you should select the substitution model/method that will be used to estimate branch lengths. If you provided a nucleotide sequence alignment, the models that you may choose are distinct from the ones you may choose for amino acid sequences. This is the model you used previously to estimate the phylogenetic tree. To choose among substitution models, see step 3.2 in Hall (2013).

Specifically, if you choose to optimize branches by OLS (RelTime-OLS), you may select which substitutions should be used to calculate the distance matrix, as it is a distance method. You can select transitions, transversions or both in the “Substitutions to Include” field. Generally, choose “Transitions+Transversions” unless you have very specific questions to address regarding the substitution process.

If the substitution model does not contemplate rate heterogeneity among sites, select “Uniform Rates” in the “Rates among Sites” field (“Rates and Patterns” panel). Conversely, if the rates should vary among sites, select one of the other options. The best option to choose will depend on the substitution model selected as the best one in the model selection procedure (Hall 2013). The options are gamma distributed (G), invariant sites (I) or both of them (G + I) in the case of RelTime-ML. The number of discrete gamma categories is set to 5 by default in MEGA X, which is a reasonable value. RelTime-OLS allows you to use only a gamma-distributed pattern of rate variation, and the gamma parameter should be provided. This is because distance methods require a continuous gamma distribution to take rate variation into account, so that you do not need to specify a “number of discrete gamma categories.” Therefore, if you decide to contemplate rate variation among sites while using the optimization of branch lengths by OLS, you should provide a gamma parameter (if it is known from the literature) or

estimate it previously in programs such as baseml/codeml (Yang 2007).

Next, select whether a subset of your alignment data should be used or not to estimate the branch lengths. You can choose a distinct treatment for gaps and missing data of your alignment. “Complete deletion” means that all alignment positions where gaps and/or missing data are present will not be considered to estimate all the branch lengths. Specifically, for **RelTime-OLS**, you may choose “Pairwise deletion.” This option means that, considering each pair of sequences, gaps and or/missing data will be discarded from the distance calculation. Another option is “Partial deletion,” in which a “Site Coverage Cutoff (%)” value should be provided. In this field, you should provide the cutoff value (in percentage) for an alignment position to be included or not during branch length optimization. For instance, if you provide a cutoff value of 95, it means that only sites present in at least 95% of the sequences/taxa will be used to calculate branch lengths. Finally, in the **RelTime-ML** approach, there is the option to consider all gaps and missing data during branch length calculations (“Use all sites”). Generally, this is the option you want.

There are two additional options you may set only for **RelTime-ML** analyses. First, you may select a “Branch Swap Filter,” which sets the level of optimization stringency for a ML analysis (branch lengths and likelihood improvements). Choose “None,” which indicates that no heuristic will be used, and a full optimization process will be performed. The other options allow for weaker or stronger filters, leading to more or less exhaustive searches, which results in slower/faster computational times. Second, you may choose the number of threads MEGA will be able to access from your computer. Obviously, more threads result in faster computation, but in general, one or two threads is enough for most analyses.

In case you have begun the analysis from step 2, because you provided a phylogenetic tree with branch lengths already optimized, you do not need to specify a substitution model and any analysis settings.

Remember that in case you got confused, the “Analysis Preferences” window has a “Help” button that takes you to the MEGA X help and explains each field that should be filled with analysis information. In addition, it provides many other explanations, from the basic understanding of phylogenetic representations to much more advanced topics.

Step 6: Launching Analysis

Once you finish configuring the analysis, click **OK** and select **Launch Analysis** in the TimeTree wizard. A “Progress” window will open displaying a progress bar and informing the stages of the timetree estimation process. It also shows relevant information, such as the starting time of the analysis and the analysis options you chose. Do not close this window, since it will interrupt the analysis.

Depending on the size of your data set, the analysis will take more or less time. Regular-sized data sets usually take just a few minutes. For instance, a mitochondrial mammalian data set with 274 species and ~7, 300 nucleotide sites (dos

Reis et al. 2012) takes less than 14 minutes to complete the calculations (with a GTR+G5 substitution model, using all alignment sites and one thread). Importantly, at any time, you can click the **Stop** button to interrupt the process. For instance, in case you noticed a misspecification of the substitution model, stop the analysis, and start from step 1 again.

Saving and Visualizing the Timetree

The tree explorer window will be opened once the analysis has finished, containing two tabs: one with the original tree topology you provided, and the other with the timetree estimated by the RelTime method (fig. 4). The calibrated nodes are highlighted, and the estimated divergence times are displayed close to the nodes to which they refer. Outgroup taxa/sequences are shown in gray with a blue mark. An axis is shown below the timetree, indicating the timescale. In the bottom of the tree explorer is the figure legend describing detailed information about your analysis and the references to the models and methods that were employed. The caption can be saved as a text file if you select **Caption** and **Save to Text File**.

MEGA X allows you to export the timetree in three distinct formats: Newick, tabular, and Nexus. You can do this by clicking **File** and choosing the format you want. It is advisable to save your timetree in all of these formats, as each of them can be useful for distinct purposes in the future. The parenthetical notation of the Newick format is the standard way to store information from phylogenies in general. After choosing **Export Current Tree (Newick)**, click on the TimeTree options and select **Divergence Times**; then click **OK**. Your timetree will be displayed in Newick format in the Text File Editor window. Click **File** and **Save**, type the name of the file and **Save**.

The tabular format allows you to manipulate the results in Excel, for instance. The Nexus format is useful to store data and visualize your timetree in other tree viewer software, such as FigTree. Repeat the procedure described above to save the tree in tabular and Nexus formats, but instead of choosing **Export Current Tree (Newick)**, choose **Export Current Tree (Tabular)**, and **Export Current Tree (Nexus)**. You should also save the analysis summary, by clicking **File** and **Exporting Analysis Summary**, in order to store estimated parameter values and the settings you used to generate the timetree. The output files resulting from the great apes’ analysis are provided as [Supplementary Material](#).

The tree explorer contains several features in the left panel that allows you to swap, flip, and compress a subtree (in addition to showing a subtree separately); add taxon/subtree markers; edit the branch lines; display images; and search for taxon names, among many others. Specifically, for timetrees, you can choose to hide the timescale and the estimated divergence times and show the estimated error bars (confidence intervals) for each node age, as well as the percentage of data coverage for each internal node (the amount of gaps/missing present while estimating a node

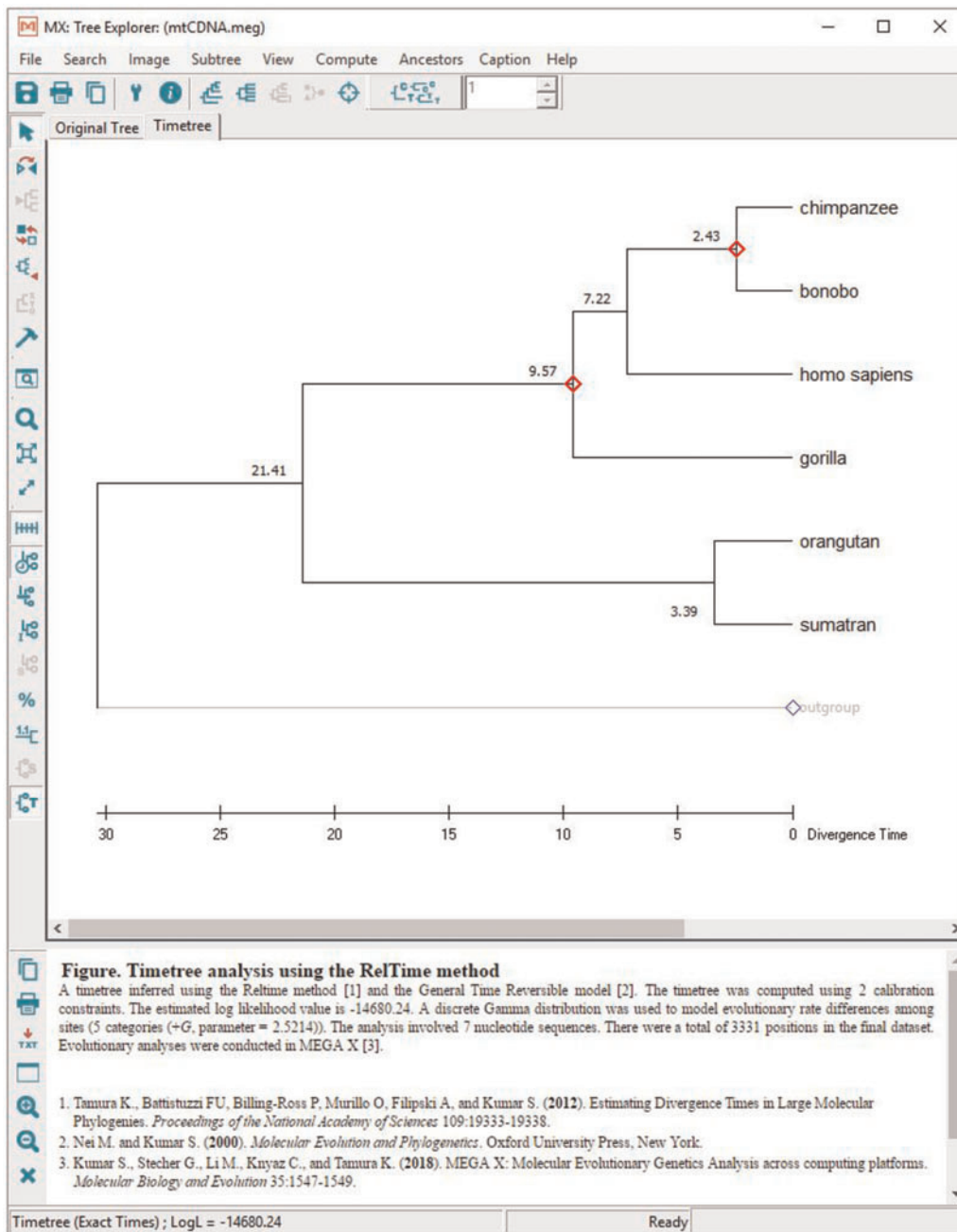


Fig. 4. Tree explorer displaying the estimated timetree in RelTime.

height). With all these features, you can format the timetree as you want, depending on your purpose. Then, save it in a graphic file format by selecting **Save as Image File** from the **Image** menu. Among all the file format options, PDF (portable document format) is the most universal, as it is accepted by several image editing programs. Therefore, you will be able to manipulate the timetree image by increasing and changing the font style, as well as adding drawings. This way, you are able to present a nice timetree figure while publishing your results. Less versatile, but still useful, the other image file formats in which MEGA X allows you to save the timetree are BMP (Windows Bitmap), PNG (Portable Networks Graphic), and TIF (Tagged Image Format).

Performing a Phylogenetic Reconstruction Followed by a Timetree Estimation in MEGA X

In case you have just estimated a phylogenetic tree with MEGA, as in step 3.3 of Hall (2013), you can directly open the TimeTree wizard from the tree explorer window. When performing a molecular dating analysis from the tree explorer window, click on **Compute** and then **TimeTree** (or by clicking on the “clock” button). By doing this, you go directly to step 3 and do not need to load alignment and tree files (because you have already done it to estimate the phylogeny). Additionally, you do not need to specify any analysis settings

as you did in step 5, and the timetree will be estimated based on the branch lengths present in your current tree. The final timetree will be displayed in an additional tab, so that you can access both the phylogeny and the timetree and save them separately in distinct Newick and/or Nexus files. It is important to mention that even if you opt for doing a timetree reconstruction directly from the tree explorer window, it is important to save the estimated phylogeny and the timetree. This way, you have both the phylogenetic tree with branch lengths and the timetree so that they can be modified and viewed in other tree viewer software.

More Advanced Analyses with MEGA-CC

For more advanced analyses, such as simulations or estimating divergence times for genomic data and huge phylogenetic trees (up to thousands of taxa), MEGA's computational core (MEGA-CC) is an ideal framework. MEGA-CC is the command-line version of MEGA, and therefore, it allows the implementation of pipelines to perform MEGA analyses. It can be downloaded from the MEGA website and installed in the same way as the GUI version, which is the graphical version used in this protocol. Importantly, MEGA-CC runs natively on Microsoft Windows, Linux, and macOS.

Additionally, for the sequence alignment (in fasta or MEGA format) and the phylogenetic tree (in Newick format), estimating timetrees with MEGA-CC requires input files with the calibrations, outgroups and analysis settings. An easy way to get the additional files required to launch a MEGA-CC analysis is through the GUI version of MEGA X. First, the user will need to follow all the stages until step 3, and then, calibrations can be exported on the calibration editor window through the button "Export all calibrations to a text file."

To obtain a file with the analysis settings, chose mode "Prototype" in the bottom right of MEGA X main window. Subsequently, specify the data input file type. It depends on which kind of data you are using to estimate the divergence times, which is generally nucleotide or amino acid. After selecting the proper data type, on the **Clocks** menu, choose **Compute TimeTree**, and then **RelTime-ML** or **RelTime-OLS**. An "Analysis Preferences" window (similar to fig. 3) will be opened. Set all the required information as in step 5 (*Choosing analysis settings*) and click on the "Save settings" button. This will generate a file with a ".mao" extension that specifies the settings to run your analysis in the command line.

Then, it is possible to estimate a timetree by using the following command line: `megacc -a "analysis preferences file" -d "sequence alignment file" -t "phylogenetic tree file" -g "outgroup file" -o "output file name."` The outgroup file should contain a list specifying the outgroup sequences/taxa of your analysis. This is a simple text file in which each line contains the following text: "outgroup name"=outgroup. As well as in the GUI version, MEGA-CC outputs the estimated timetree in Newick format, both in relative and absolute (exact) times. The timetree can be visualized in MEGA X (GUI version) or another tree viewer program. Nexus

(absolute times) and tabular formats are also generated, as well as an analysis summary file.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

References

- Akerborg O, Sennblad B, Lagergren J. 2008. Birth-death prior on phylogeny and speed dating. *BMC Evol Biol.* 8:77.
- Battistuzzi FU, Billing-Ross P, Paliwal A, Kumar S. 2011. Fast and slow implementations of relaxed-clock methods show similar patterns of accuracy in estimating divergence times. *Mol Biol Evol.* 28(9):2439–2442.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 10(4):e1003537.
- dos Reis M, Donoghue PCJ, Yang Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet.* 17(2):71–80.
- dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B Biol Sci.* 279(1742):3491–3500.
- Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, da Costa AC, et al. 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546(7658):406–410.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hall BG. 2013. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol.* 30(5):1229–1235.
- Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.
- Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol.* 32(4):835–845.
- Ho SYW. 2014. The changing face of the molecular evolutionary clock. *Trends Ecol Evol.* 29(9):496–503.
- Ho SYW, Duchêne S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol.* 23(24):5947–5965.
- Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst Biol.* 58(3):367–380.
- Ho SYW, Tong KJ, Foster CSP, Ritchie AM, Lo N, Crisp MD. 2015. Biogeographic calibrations for the molecular clock. *Biol Lett.* 11(9):20150194.
- Ksepka DT, Parham JF, Allman JF, Benton MJ, Carrano MT, Cranston KA, Donoghue PCJ, Head JJ, Hermesen EJ, Irmis RB, et al. 2015. The fossil calibration database – a new resource for divergence dating. *Syst Biol.* 64(5):853–859.
- Kumar S. 2005. Molecular clocks: four decades of evolution. *Nat Rev Genet.* 6(8):654–662.
- Kumar S, Hedges SB. 2016. Advances in time estimation methods for molecular data. *Mol Biol Evol.* 33(4):863–869.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 35(6):1547–1549.
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 34(7):1812–1819.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.

- Kumar S, Tamura K, Nei M. 1994. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics* 10(2):189–191.
- Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program: table 1. *Mol Biol Evol.* 32(10):2798–2800.
- Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol.* 50(6):913–925.
- Marin J, Battistuzzi FU, Brown AC, Hedges SB. 2017. The timetree of prokaryotes: new insights into their evolution and speciation. *Mol Biol Evol.* 34(2):437–446.
- Mello B, Tao Q, Tamura K, Kumar S. 2017. Fast and accurate estimates of divergence times from big data. *Mol Biol Evol.* 34(1):45–50.
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346(6210):763–767.
- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The time-scale of early land plant evolution. *Proc Natl Acad Sci USA.* 115(10):E2274–E2283.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. New York: Oxford University Press.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, Inoue JG, Irmis RB, Joyce WG, Ksepka DT, et al. 2012. Best practices for justifying fossil calibrations. *Syst Biol.* 61(2):346–359.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
- Rzhetsky A, Nei M. 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol.* 10(5):1073–1095.
- Shin S, Clarke DJ, Lemmon AR, Moriarty Lemmon E, Aitken AL, Haddad S, Farrell BD, Marvaldi AE, Oberprieler RG, McKenna DD. 2018. Phylogenomic data yield new and robust insights into the phylogeny and evolution of weevils. *Mol Biol Evol.* 35(4):823–836.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci USA.* 109(47):19333–19338.
- To T-H, Jung M, Lycett S, Gascuel O. 2016. Fast dating using least-squares criteria and algorithms. *Syst Biol.* 65(1):82–97.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Zuckerandl E, Pauling L. 1962. Molecular disease, evolution, and genic heterogeneity. In: Kasha M, Pullman B, editors. *Horizons in biochemistry*. New York: Academic Press. p. 189–225.
- Zuckerandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V & Vogel HJ, editors. *Evolving Genes and Proteins*. New York: Academic Press. p. 97–166.