

RESEARCH ARTICLE

Open Access



Paleozoic diversification of terrestrial chitin-degrading bacterial lineages

Danielle S. Gruen^{1,2*} , Joanna M. Wolfe¹ and Gregory P. Fournier¹

Abstract

Background: Establishing the divergence times of groups of organisms is a major goal of evolutionary biology. This is especially challenging for microbial lineages due to the near-absence of preserved physical evidence (diagnostic body fossils or geochemical biomarkers). Horizontal gene transfer (HGT) can serve as a temporal scaffold between microbial groups and other fossil-calibrated clades, potentially improving these estimates. Specifically, HGT to or from organisms with fossil-calibrated age estimates can propagate these constraints to additional groups that lack fossils. While HGT is common between lineages, only a small subset of HGT events are potentially informative for dating microbial groups.

Results: Constrained by published fossil-calibrated studies of fungal evolution, molecular clock analyses show that multiple clades of Bacteria likely acquired chitinase homologs via HGT during the very late Neoproterozoic into the early Paleozoic. These results also show that, following these HGT events, recipient terrestrial bacterial clades likely diversified ~300–500 million years ago, consistent with established timescales of arthropod and plant terrestrialization.

Conclusions: We conclude that these age estimates are broadly consistent with the dispersal of chitinase genes throughout the microbial world in direct response to the evolution and ecological expansion of detrital-chitin producing groups. The convergence of multiple lines of evidence demonstrates the utility of HGT-based dating methods in microbial evolution. The pattern of inheritance of chitinase genes in multiple terrestrial bacterial lineages via HGT processes suggests that these genes, and possibly other genes encoding substrate-specific enzymes, can serve as a “standard candle” for dating microbial lineages across the Tree of Life.

Keywords: Horizontal gene transfer, Chitinase, Chitin, Bacteria, Fungi, Arthropods

Background

Dating when new metabolisms evolved and when major clades of Bacteria arose, particularly on the order of hundreds of millions of years, remains a key challenge in biology [1]. Despite progress in understanding the molecular record of extant bacterial genomes, the timing of the evolution of major clades of Bacteria is especially problematic to resolve due to complex gene histories and a lack of clear phenotypic traits that can be correlated with a diagnostic fossil record [2]. In the near-absence of physical (geochemical or fossil) records

of microbial evolution, it is difficult to determine and date the evolutionary history of bacterial lineages [3].

Leveraging the information contained in horizontal gene transfer (HGT) events can substantially improve estimates of the timing of events within microbial evolution [4–9]. Vertical inheritance passes genetic information from parent to offspring, but HGT passes genetic information between organismal lineages, across all degrees of evolutionary distance. This can be particularly useful for molecular clock dating, as HGTs establish cross-cutting relationships between lineages and serve as a “temporal scaffold” upon which fossil calibrations or other date information from even distantly related taxa may be placed [5, 8, 10]. While HGT is a major process in microbial evolution [7, 11], HGT events between microbes and eukaryotes with a fossil record are less frequently identified [12]. Furthermore, the donor-recipient

* Correspondence: dgruen@mit.edu

¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA



relationships are often difficult to infer for many gene histories due to multiple HGT events and gene losses or the lack of a strong phylogenetic signal [13]. The function of a gene is not necessarily relevant to its utility in propagating time constraints (e.g. [4]); however, in some cases, this gene function may be additionally informative, and provide independent support for age estimates. This is the case, for example, if the protein encoded by the transferred gene is specific for a substrate that can, itself, be temporally constrained. Given all of these criteria, a very small number of HGT events may be especially valuable for dating microbial lineages; these “index transfers” [9] can be even more valuable if multiple HGT recipients are present, closely correlating the ages of the recipients in time, a “standard candle” (a term used in astronomy to describe an object with known luminosity used to infer the cosmic distances to other objects of interest) [14].

Environmental distribution of chitin

Chitin is one of the most abundant structural polysaccharides in nature [15, 16], and chitin degradation by chitinase enzymes is a critical process in the biogeochemical cycling of carbon and nitrogen in terrestrial and aquatic ecosystems [15]. There are two dominant biogenic sources of chitin: arthropods [16] and fungi [16]. Chitin may therefore have increased in abundance in terrestrial systems following the terrestrialization of arthropods, sometime after the Cambrian [17]. In modern aquatic systems, arthropods are the dominant chitin-producing organisms. While there is a great deal of uncertainty in these estimates, the chitin sourced from arthropods is roughly 2.8×10^7 Mg yr⁻¹ in freshwater ecosystems and 1.3×10^9 Mg yr⁻¹ in marine ecosystems [18]. The majority of chitin in terrestrial ecosystems is produced by fungi [19] largely due to their contribution of biomass to the soil environment [20]. While global estimates for the contribution of arthropod biomass, and thus chitin, to the environment over time are lacking, arthropods nonetheless make up the largest pool of animal biomass today [21].

Chitin production and the evolution of Fungi

The evolution of chitin producers is anchored to the fossil record through diagnostic morphological characters [22–26]. In the case of Fungi, Cryptomycota form the most deeply branching fungal clade, and contain the most deeply branching chitinous Fungi (e.g., *Rozella*) [24, 25]. Fossil-calibrated molecular clock studies generally agree that early Fungi diverged around 1145–738 million years ago (Ma) [27]. Fossil and molecular clock evidence also indicates that divergence of Ascomycota and Basidiomycota within the major fungal group Dikarya occurred around 830–518 Ma [24] with a fossil

minimum around 405 Ma [23, 28, 29]. Posterior age estimates from molecular clock studies suggest that crown Ascomycota diversified 715–408 Ma [30] and crown Basidiomycota diversified 655–400 Ma [28]. Therefore, studies of fungal evolution can inform the timing of chitinase gene evolution.

Based on fossil and molecular clock dating methods, marine crown-group euarthropods appeared around 521–514 Ma, shortly after the start of the Cambrian, and radiated into the lower and middle Cambrian [29, 31]. Molecular clock and fossil evidence suggests that terrestrialization of major arthropod groups occurred from the Cambrian into the Silurian [32]. The oldest terrestrial myriapod body fossil (the oldest undisputedly terrestrial animal) is the 416 Ma *Crussolum sp.* [29]. However, the radiation of terrestrial arthropods (including insects) likely continued into the Devonian [17, 33, 34].

The evolution of chitinase gene families

Chitinases are proteins that catalyze the breakdown of glycosidic linkages in polymers of chitin [16]. Chitinases are a type of glycoside hydrolase (GH) specific to chitin [16, 35]. There are two main families of chitinases: glycoside hydrolase family 18 (GH18) and glycoside hydrolase 19 (GH19) [16]. GH18 chitinases are distributed across the three domains of life [16, 36], whereas GH19 chitinases are restricted to mostly plants and are rarely associated with bacteria [36]. In one well-studied bacterial model, *Streptomyces*, there were ten genes associated with the GH18 family of chitinases (homologs chiA-E, and H-L) and two genes associated with GH19 (chiF, G) [37]. It has been suggested that some of these genes may have evolved under selective pressures related to the host environment or to the presence and proximity to other organisms, which may have even precipitated HGT events [37–39]. Myxobacterial chitinases have been hypothesized to have evolved via HGT [40], and other bacterial lineages within Actinobacteria are hypothesized to have co-opted a fungal chitinase for self-defense [37]. Because of the specific associations between substrate and gene, it stands to reason that there may be an evolutionary link between the major producers of environmental chitin (fungi and arthropods) and chitin-degrading genes in bacteria. It has been shown that some bacterial chitin degradation systems are even adapted to the environments (aquatic vs. terrestrial) and most abundant chitin producers (exoskeletons of crustaceans vs. fungal cell walls) that they encounter [15]. Nonetheless, it remains to be tested whether chitinase genes also reflect widespread environmental adaptations over geological time.

It has been shown that chitinases may retain a molecular record of evolutionary events hundreds of millions of years ago [41]. While some of the phylogenetic

distribution of these genes may indicate a pattern of vertical inheritance, other chitinase genes may have evolved via horizontal gene transfer [37]. For these reasons, and the criteria described above, chitinase genes are an attractive potential source of temporal information for microbial evolution. Therefore, we sought to test the hypothesis that specific bacterial chitinases evolved via HGT, and if so, if these HGT events could be leveraged to propagate known fossil calibrations between donor and recipient lineages. Bacterial chitinases are especially useful because they metabolize chitin, a specific biopolymer only produced in abundance by arthropods and fungi, two groups with fossil records, and thus likely age estimates, much more precise than those of most microbial groups. Previous work has also suggested that some chitinases are distributed between the domains of life via HGT, for example, postulating that some chitinase genes were transferred from plants to Actinobacteria and then to arthropods [42]. However, the evolutionary history of the many disparate chitinase gene families in microbes has not been fully investigated.

Bayesian molecular dating

Fossil-calibrated molecular clock models are applied to estimate divergence times of organisms (e.g. [3, 43]). Many divergence time analysis parameters have only been recently developed, and few have been applied to microbes with divergence time estimates that span geologic time or have undergone rampant horizontal gene transfer events (e.g. [8, 44]). For a more detailed review of these parameters and challenges see, for example, [43, 45–53]. The issues inherent to assessing microbial evolution present a challenge for this work, but also an opportunity to explicitly test these model parameters and assumptions in order to determine those that are valid for this specific set of evolutionary conditions.

Molecular clock dating is based on a Bayesian framework, reviewed in greater detail by others [51, 52, 54]. There are a few major components used to determine posterior probabilities or date distributions such as data selection, calibrations, the molecular clock model, the tree process prior, and the rate distribution model. The sequence data assessed in this work are the chitinase genes present in bacterial and eukaryotic lineages. Tree process priors include birth-death and uniform. Rate distribution models include lognormal autocorrelated and uncorrelated gamma.

We tested the uniform prior and the birth-death tree process priors. The uniform prior considers every possible topology to be equal and favors divergences that are evenly spaced across the tree from the root to tip [55, 56]. The birth-death model is defined by speciation (“birth”) and extinction (“death”). In contrast to the uniform prior, this tree process ascribes more weight to tree topologies with certain branching patterns [57]. The

birth-death process generally biases the model such that deeper branches are longer and the more shallow branches are shorter, because it is assumed the “older” lineages more often end in extinction [52]. Biases such as this can have large effects on the posterior age estimates and inappropriate model selection can result in less precise dates.

All models in this study assume a relaxed molecular clock model for a prior on the branch rate. However, two relaxed clock models for the branch rates are assessed: autocorrelated and uncorrelated. Uncorrelated clocks make no assumption that branches next to each other on the tree should share similar rates. In other words, the rate on each branch of the tree is independent. Conversely, autocorrelated clocks assume that more closely related branches on the tree should also have more similar rates [46, 56, 58, 59]. The assumption that neighboring branches should share more similar rates makes sense when we consider that the evolution of genetic information between related lineages is often affected by many of the same processes that affect the rates of evolution (e.g. environment, population) [52]. Biological events such as horizontal gene transfer may invalidate model assumptions, but the mechanisms of rate variation and quantifying the relative importance of various biological events are still debated [1]. Choosing between these models is a matter of ongoing debate in the field, and is often dependent on the data [52, 56, 60]. Thus, we detail the effects of model selection in our analyses.

The primary objective of this work is to test whether fossil-calibrated age estimates within fungi can be propagated to bacterial lineages through the use of HGT events between these lineages under different model assumptions. Secondly, we seek to understand possible ecological implications of the evolution of chitinases in fungi and bacteria. If bacterial chitinase genes were acquired in response to environmental chitin availability, then arthropod evolutionary history provides a prediction for the timing of these events within bacterial lineages. We hypothesize that terrestrial bacterial chitinases diversified from the Cambrian into the Devonian following the distribution of environmental chitin. We independently date chitinase evolution in microbial lineages by first testing and then applying molecular clock models to chitinase gene trees, constrained by fungal date calibrations tethered via HGT. We show that certain model parameters seem to outperform others. Moreover, our posterior date distributions for bacterial lineages support the utility of HGT-propagated fossil calibrations in accurately estimating the ages of microbial lineages as an avenue for future work.

Methods

Taxon sampling

We queried The National Center for Biotechnology Information (NCBI) nonredundant (nr) database using the

protein Basic Local Alignment Search Tool (BLASTp) for sequences homologous to the *Myxococcus fulvus* ChiD protein (WP_046715376.1). Complete protein sequences of the top 5000 hits from NCBI were downloaded (E-value $< 10^{-5}$). Sequences were subsampled from this list to include a single representative of each species as annotated in NCBI, to avoid an overabundance of terminal taxa representing multiple strain isolates of the same species. We further used BLASTp to more exhaustively identify potential homologs within Fungi, repeating this method for specific searches within Ascomycota, Basidiomycota, and more deeply-rooting Fungi (e.g. Blastocladiomycota, Chytridiomycota, Zoopagomycota, and Mucorales).

Sequences and alignments

Sequences were aligned using the program MUSCLE [61]. Poorly aligning regions that contained misaligned gaps in the deeply rooting fungi were identified via manual inspection and removed using Jalview [62]. The resulting alignment was then manually edited to correct obvious misalignments in generally well-aligned regions adjacent to indels in *Fonsecaea multimorphosa* and *Phialophora americana* (sites 2390–2470). We also removed the misaligned C-terminal region from *Phelbia centrifuga*, and the misaligned C-terminal regions from *Rhizopus*, *Mucor*, *Syncephalastrum*, *Absidia*, and *Lictheimia* (sites 2393 onward). Both datasets (before and after trimming) are publicly available [63].

A profile alignment of bacterial and fungal sequences was made [61]. This revealed a highly conserved alignment region shared across bacterial and fungal sequences (sites 1844–2470) and another well-aligned N-terminal region conserved across Bacteria, but absent or poorly aligned in Fungi. In order to maximize the sequence information used for phylogenetic reconstruction and molecular clocks without introducing misalignments between bacterial and fungal sequences, a composite alignment was generated. This involved concatenating the conserved region for both Fungi and Bacteria with the N-terminal region aligned for just Bacteria. From this alignment, a single gene tree was generated for determining the relationship between Fungi and Bacteria, and for maximally resolving splits within the bacterial tree.

Phylogenetic analyses

Gene tree

The gene trees were inferred using RaxML v1.8.9 using the PROTGAMMALGF substitution model [64] as fit by PROTTEST [65], and 100 bootstrap replicates. The resulting tree showed relationships between fungal taxa that are congruent with published phylogenies [22–26]. We rooted the gene tree on the branch leading to *Rozella*, which is considered to be either part of a sister

group to the most deeply-rooting fungal clades, or a member of Chytridiomycota, one of the most deeply-rooting Fungi [24]. This root resulted in bacterial chitinases as a clade diverging within crown Fungi, polarizing the origin of the bacterial homologs as originating via an HGT from a fungal donor. A consensus tree was generated from Bayesian Inference using PhyloBayes v3.3 (CAT20 set of substitution models [60], effective size > 50 , and variable discrepancies < 0.30).

Divergence time estimation

Divergence times were estimated using PhyloBayes v3.3 under the CAT20 set of substitution models [60]. Divergence time estimates were generated under several sets of model priors. Specific model parameters are described in Tables 1 and 2. After chain convergence (effective size > 50 , variable discrepancies < 0.30), trees and posterior probability support values were generated from completed chains after the initial 20% of sampled generations were discarded as burn-in.

Date constraints

Secondary calibrations were applied to the divergence times of major fungal groups within the gene tree. For all analyses, we applied a root prior and one internal date constraint to the split of Ascomycota and Basidiomycota consistent with reported molecular clock and fossil evidence within Fungi [22–30, 66]. In order to avoid false precision, uniform priors were used in both cases, 1145–739 Ma for the fungal root [27] and 830–518 Ma for the Ascomycota-Basidiomycota split [28]. We also tested the addition of secondary calibrations on the nodes leading to the Ascomycota (715–408 Ma) and Basidiomycota (655–400 Ma) clades [28, 30]. Finally we tested the application of a primary fossil minimum calibration on the split of Ascomycota and Basidiomycota (830–405 Ma) [23, 28, 29]. All calibration structures are listed in Table 2.

Results

Phylogeny of ChiD and ChiC homologs

Figure 1 illustrates the relationships between sequences in this study (Additional file 1: Table S1) as a maximum-likelihood gene tree generated with RAXML. The tree is rooted with the most deeply-branching fungal taxon, *Rozella* (Cryptomycota). The group of deeply-rooting Fungi include members of Cryptomycota, Blastocladiomycota, Chytridiomycota, Blastocladiomycota, Chytridiomycota, Mucormycotina, and Zoopagomycota (in order of branching from the root), are generally consistent with recent results of phylogenomic analyses of the divergence of basal Fungi [67]. Bootstrap supports are low for many bipartitions within this deeply-rooting group. Support for the bipartitions placing bacterial

Table 1 PhyloBayes model parameters tested in this study. For each model, a +/- indicates the presence or absence of a condition. Sequence data were used to generate posterior probability distributions for all models, which were also tested under the prior by removing sequence data (-prior flag in PhyloBayes as indicated by a "p"). BD refers to birth-death. LN stands for lognormal autocorrelated. UGAM stands for uncorrelated gamma multipliers. The AB split refers to the split between Ascomycota and Basidiomycota. Fossil refers to the fossil minimum referenced in the Calibration Table (Table 2)

Model	Prior		Relaxed Clock Model		Calibrations				-prior
	Uniform	BD	LN	UGAM	Root	AB split	A crown, B crown	Fossil	
1	+	-	+	-	+	+	-	-	-
2	+	-	+	-	+	+	+	-	-
3	+	-	+	-	+	-	-	+	-
4	+	-	+	-	+	-	+	+	-
1p	+	-	+	-	+	+	-	-	+
2p	+	-	+	-	+	+	+	-	+
3p	+	-	+	-	+	-	+	+	+
4p	+	-	+	-	+	-	-	+	+
5	+	-	-	+	+	+	-	-	-
6	+	-	-	+	+	+	+	-	-
7	+	-	-	+	+	-	+	+	-
8	+	-	-	+	+	-	-	+	-
5p	+	-	-	+	+	+	-	-	+
6p	+	-	-	+	+	+	+	-	+
7p	+	-	-	+	+	-	+	+	+
8p	+	-	-	+	+	-	-	+	+
9	-	+	-	+	+	+	-	-	-
10	-	+	-	+	+	+	+	-	-
11	-	+	-	+	+	-	+	+	-
12	-	+	-	+	+	-	-	+	-
9p	-	+	-	+	+	+	-	-	+
10p	-	+	-	+	+	+	+	-	+
11p	-	+	-	+	+	-	+	+	+
12p	-	+	-	+	+	-	-	+	+
13	-	+	+	-	+	+	-	-	-
14	-	+	+	-	+	+	+	-	-
15	-	+	+	-	+	-	+	+	-
16	-	+	+	-	+	-	-	+	-
13p	-	+	+	-	+	+	-	-	+
14p	-	+	+	-	+	+	+	-	+
15p	-	+	+	-	+	-	+	+	+
16p	-	+	+	-	+	-	-	+	+

sequences within Fungi are higher (74, 71). Support for the monophyly of Ascomycota and Basidiomycota is high (100). Support is also high for the monophyly of bacterial sequences (99). While the deeper branches in the fungal tree have weak bootstrap support, the relatively short branches relating these groups and the lack of any calibrations sensitive to their specific crown-group topology suggest the observed phylogenetic

uncertainty has little impact on divergence times for more distal clades within the tree.

Within Bacteria are the generally well-supported and often monophyletic bacterial clades including groups within Betaproteobacteria (Burkholderiales, Chromobacteriaceae), Deinococcus, Actinobacteria, Bacteroidetes (Cytophagia, Flavobacteriaceae, Chitinophagia), Firmicutes (Bacillales), and Deltaproteobacteria (Myxococcales). Gammaproteobacteria

Table 2 Calibrations used in molecular clock models. All calibrations listed in Ma. Taxon 1 and Taxon 2 refer to the taxa used in PhyloBayes commands

Node	Taxon 1	Taxon 2	Calibration	Reference
Root	RozeAll205	Paeniba168	1145–738	Sharpe et al., 2015
AB Split	SerpLac210	TricVir244	830–518	Floudas et al., 2012
Ascomycota Crown	RoseNec204	MetaBru136	715–408	Prieto et al., 2013
Basidiomycota Crown	HessVes103	AgarBisp08	655–400	Floudas et al., 2012
Fossil minimum on AB Split	SerpLac210	TricVir244	830–405	Berbee and Taylor, 2010; Floudas et al., 2012

are polyphyletic, including Vibrionales, Xanthomonadales, and one Gammaproteobacteria taxon in Actinobacteria (*Cellvibrio*, WP_049631752.1), a cellulolytic bacterium in the order Pseudomonadales [68], suggesting multiple independent acquisitions of ChiD. Actinobacteria (bootstrap support 99%), Bacteroidetes (bootstrap support 42%), Firmicutes (bootstrap support 79%), and Deltaproteobacteria (bootstrap support 88%) are also monophyletic. Deltaproteobacteria sit on a reticulating branch within Firmicutes. The tree generated with PhyloBayes recovered a similar topology, further supporting the placement of the root within this group of deeply branching fungi (Additional file 2: Figure S1). Additional annotations for node numbers and clades are included in Additional file 3: Figure S2.

Divergence time estimates of bacterial chitinases

Divergence time estimates were tested under several models, with the impacts of taxon sampling (inclusion or exclusion of bacterial sequences), tree priors (uniform vs. birth-death), and relaxed clock models (autocorrelated lognormal vs. uncorrelated gamma rate distributions) subsequently evaluated. Our preferred model is uncorrelated gamma distribution under a uniform prior with calibrations on the root (1145–738), Dikarya (830–518) and crown Ascomycota (715–408) and crown Basidiomycota (655–400).

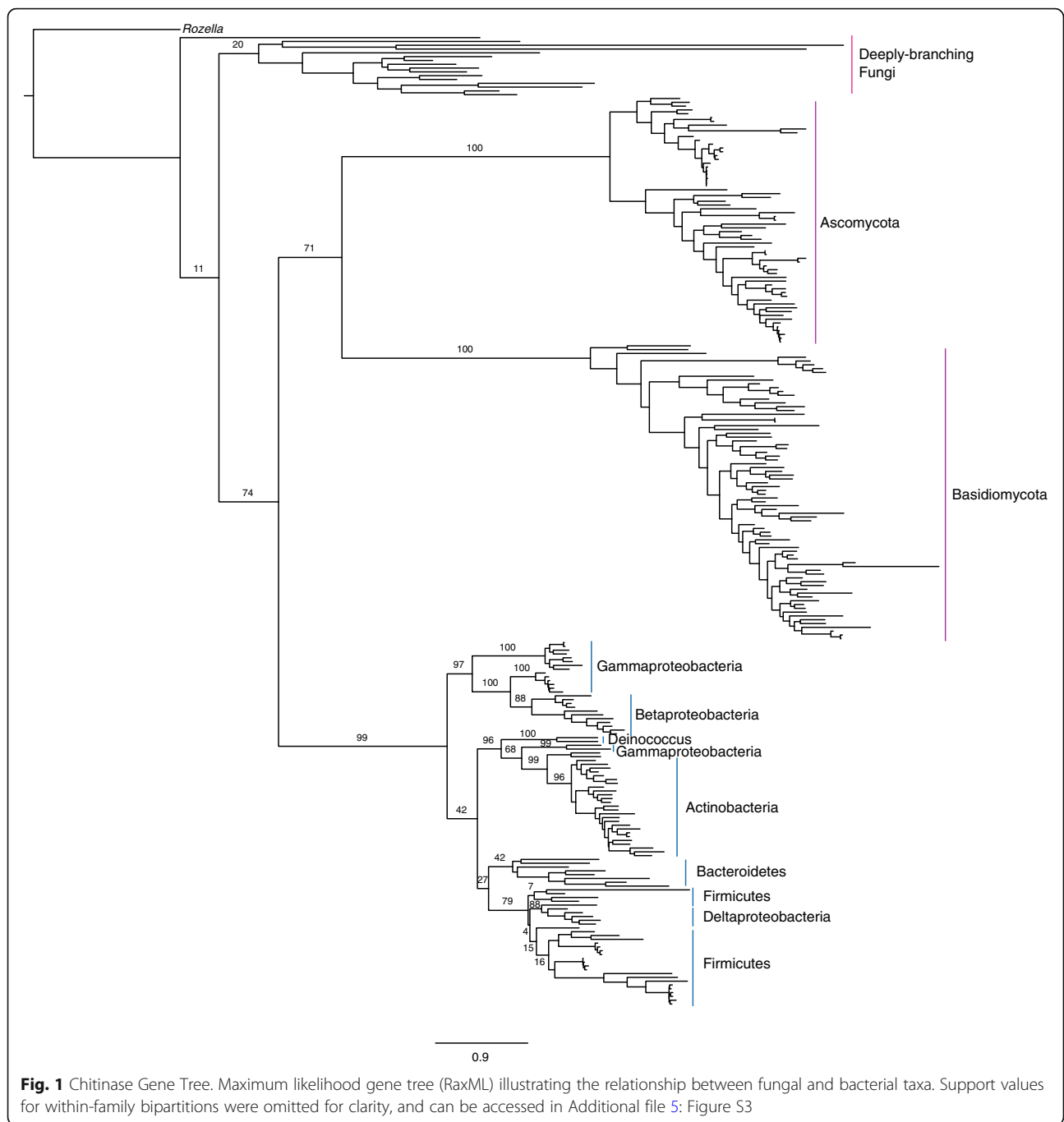
Few published age estimates exist for the bacterial clades present in our tree. For example, the only other published divergence time estimate for Vibrionales (the last common ancestor of *Vibrio* and *Photobacterium*) was an uncalibrated RelTime clock built by using 16S rRNA and protein datasets [69]. The result for this clade was 124 Ma. Based on the chitinase HGT from a time-calibrated Fungi tree with a uniform prior and uncorrelated gamma clock model, the posterior age estimate for crown-group Vibrionales is ~188 Ma with an uncertainty spanning ~278–113 Ma.

The chronogram depicted in Fig. 2 shows that bacterial chitinases have a common ancestor ~780 Ma (Node 3, Table 3) and were acquired from Fungi prior to the evolution of marine arthropods in the Cambrian. Subsequent HGT events between bacterial groups distributed this gene, with the major bacterial clades in the tree acquiring chitinase ~505–188 Ma. This age range is consistent with the

ecological and taxonomic dispersal of bacterial chitinases being correlated with the origin and diversification of crown group euarthropods around 521–514 Ma [29, 31]. Interestingly, four major clades of terrestrial Bacteria in the tree, Gammaproteobacteria (Xanthomonadales), Betaproteobacteria, Actinobacteria, Firmicutes, all diversify ~408–365 Ma, temporally consistent with the terrestrialization of arthropod groups, as terrestrial myriapods were present by 416 Ma [17, 29, 32, 33] (Fig. 1). This timing is also consistent with molecular clock evidence for the early terrestrialization of land plants (middle Cambrian – Early Ordovician) and vascular plants (Late Ordovician – Silurian) [70], and alternatively, may represent the establishment of plant-degrading Fungi in soils by 300 Ma [28].

Testing molecular clock models

Molecular clock models as listed in Table 1 were tested to assess model parameter sensitivities. The results for Model 6 (selected for further analysis) are presented in Table 3. The results of all model outputs are listed in Additional file 4: Table S2. An analysis of these models is presented in Table 4 and further elaborated upon in the following sections. Table 4 illustrates the models excluding calibrations on Ascomycota and Basidiomycota crown groups that recover the expected age ranges for these nodes in the literature. For this analysis, Calibrations 1 and 3 were used, as these do not impose dates on crown Ascomycota or Basidiomycota clades, enabling comparison between estimated and expected model output for these clades. Table 4 shows that the 95% CI posterior ages fall within expected ranges for the uniform prior and uncorrelated gamma relaxed clock model for Ascomycota under Calibrations 1 and 3. The model ages also fall within expected ranges the uniform prior and uncorrelated gamma clock model for Basidiomycota Calibration 1; uniform prior and lognormal autocorrelated clock model for Ascomycota, Calibration 3; and birth-death prior and uniform gamma distributed model for Basidiomycota, Calibration 3. Mean ages for the birth-death prior and uncorrelated gamma model and for the uniform prior with lognormal model fall outside of expected age ranges under Calibrations 1 and 3 for Ascomycota and under Calibration 1 for Basidiomycota.



Impact of the tree process prior and rate distribution model

The effects of the tree process prior (birth-death vs. uniform) and the rate distribution model (lognormal correlated vs. uncorrelated gamma) were evaluated (Table 4, Additional file 4: Table S2). Prior and posterior age estimates for the chitinase tree using a uniform vs. birth-death prior and lognormal vs. gamma rate distribution return different date distributions across nodes, in both bacterial and fungal groups. Across the Bacterial nodes, the uniform prior with lognormal autocorrelated clock

model corresponded to the oldest date estimates across nodes, followed by the uniform prior and uncorrelated gamma model, birth-death prior and lognormal autocorrelated model, and finally the youngest birth-death prior and uncorrelated gamma relaxed clock model (Fig. 3). The birth-death prior resulted in the youngest age estimates as compared to the uniform prior (Fig. 3). The same pattern holds for the Ascomycota and Basidiomycota within the fungal nodes. However, a slightly different result is observed for the deeply-rooting fungal nodes (root,

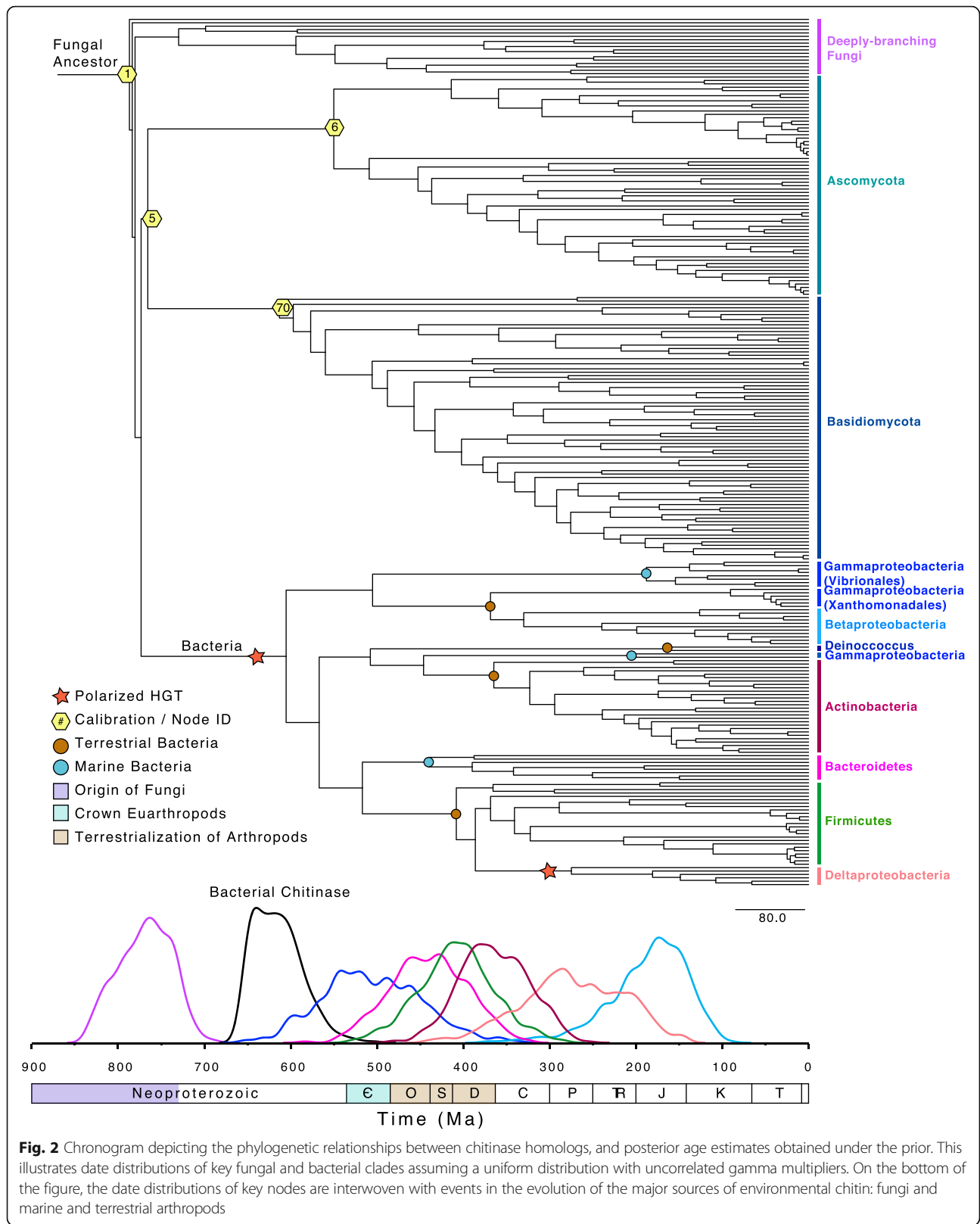


Fig. 2 Chronogram depicting the phylogenetic relationships between chitinase homologs, and posterior age estimates obtained under the prior. This illustrates date distributions of key fungal and bacterial clades assuming a uniform distribution with uncorrelated gamma multipliers. On the bottom of the figure, the date distributions of key nodes are interwoven with events in the evolution of the major sources of environmental chitin: fungi and marine and terrestrial arthropods

Table 3 Posterior divergence time estimates. Divergence time estimates calculated under the setup described for Model 6 in Table 1. Node ID represents the node identification number as depicted in Fig. S2. Cal represents the calibration used. Posterior divergence time estimates (Mean Age) and 95% confidence intervals (CI) are reported in Ma. This table represents our preferred model assumptions and results for bacterial divergence time estimates

Node	Node ID	Cal	Mean Age	95% CI
Root	1	1145–738	787	(738–837)
Ancestral Fungi	3		780	(729–833)
Dikarya (AB Split)	5	830–518	766	(720–828)
Ascomycota	6	715–408	550	(480–618)
Basidiomycota	70	655–400	613	(566–655)
Bacteria	147		605	(537–672)
Gammaproteobacteria	148		505	(393–605)
Vibrionales	165		188	(113–278)
Betaproteobacteria	155		330	(223–442)
Bacteroidetes	174		440	(359–525)
Firmicutes	181		408	(323–498)
Deltaproteobacteria	186		275	(174–389)
Actinobacteria	215		365	(290–435)

Fungi, and Dikarya). For these fungal nodes, the opposite pattern is seen with the oldest date distributions resulting from the birth-death prior and uncorrelated gamma clock model, followed by the birth-death prior and lognormal autocorrelated model, the uniform prior and uncorrelated gamma model, and finally the youngest uniform prior and lognormal autocorrelated clock model (Fig. 3). This empirical control on predicting fungal age estimates for nodes that have had their calibrations removed suggests that the uniform tree process and uncorrelated gamma rate distribution provide the most accurate age estimates for this gene family.

This model selection is also theoretically justifiable. A birth-death prior is a tree process prior that assumes a tree generated by speciation and extinction events across a lineage [49]. This assumption is violated for trees that include HGT events, especially if several such events are present. Birth-death priors are therefore not appropriate for gene trees that show histories of extensive HGT, since the underlying assumption, that nodes are distributed across a continuity of lineage speciation and extinction, is invalid. This is especially true for HGTs between microbes and eukaryotes, which sometimes have very different patterns of speciation and extinction occurring over very different timescales and sampling densities. The chitinase tree is an especially good test of these

Table 4 Posterior date distributions for model parameters. The calibration refers to the node, model number (in parentheses), calibration age (in Ma). The node refers to either Ascomycota (A) or Basidiomycota (B). The prior refers to the tree process prior, either uniform or birth-death. The clock model refers to the relaxed clock model, either lognormal autocorrelated (LN) or uncorrelated gamma (ugam). The model ID column refers to the model number as delineated in Table 1. The posterior mean age estimates and 95% confidence intervals (CI) are in Ma as calculated in the output of each PhyloBayes model. The expected age ranges (Ma) are listed based on literature values (user priors) as noted in Table 2. The Outside Expected (OE) column lists the percent of the 95% CI outside of the expected range

Cal Node, Number, Age	Node	Prior	Clock Model	Model ID	Mean Age	95% CI	Expected	OE (%)
AB-Split (Cal 1) 830–518	Ascomycota	uniform	LN	1	651	728 580	715 408	8.8
		uniform	ugam	5	559	630 481	715 408	0.0
		BD	LN	13	494	592 391	715 408	8.5
		BD	ugam	9	366	457 267	715 408	74.2
	Basidiomycota	uniform	LN	1	674	749 602	655 400	63.9
		uniform	ugam	5	628	698 559	655 400	30.9
		BD	LN	13	550	650 443	655 400	0.0
		BD	ugam	9	456	572 349	655 400	22.9
AB-Split (Cal 3) 830–405	Ascomycota	uniform	LN	3	656	731 590	715 408	11.3
		uniform	ugam	7	560	637 484	715 408	0.0
		BD	LN	15	510	616 397	715 408	5.0
		BD	ugam	11	366	447 276	715 408	77.2
	Basidiomycota	uniform	LN	3	677	754 607	655 400	67.3
		uniform	ugam	7	626	687 554	655 400	24.1
		BD	LN	15	560	674 430	655 400	7.8
		BD	ugam	11	462	570 369	655 400	15.4

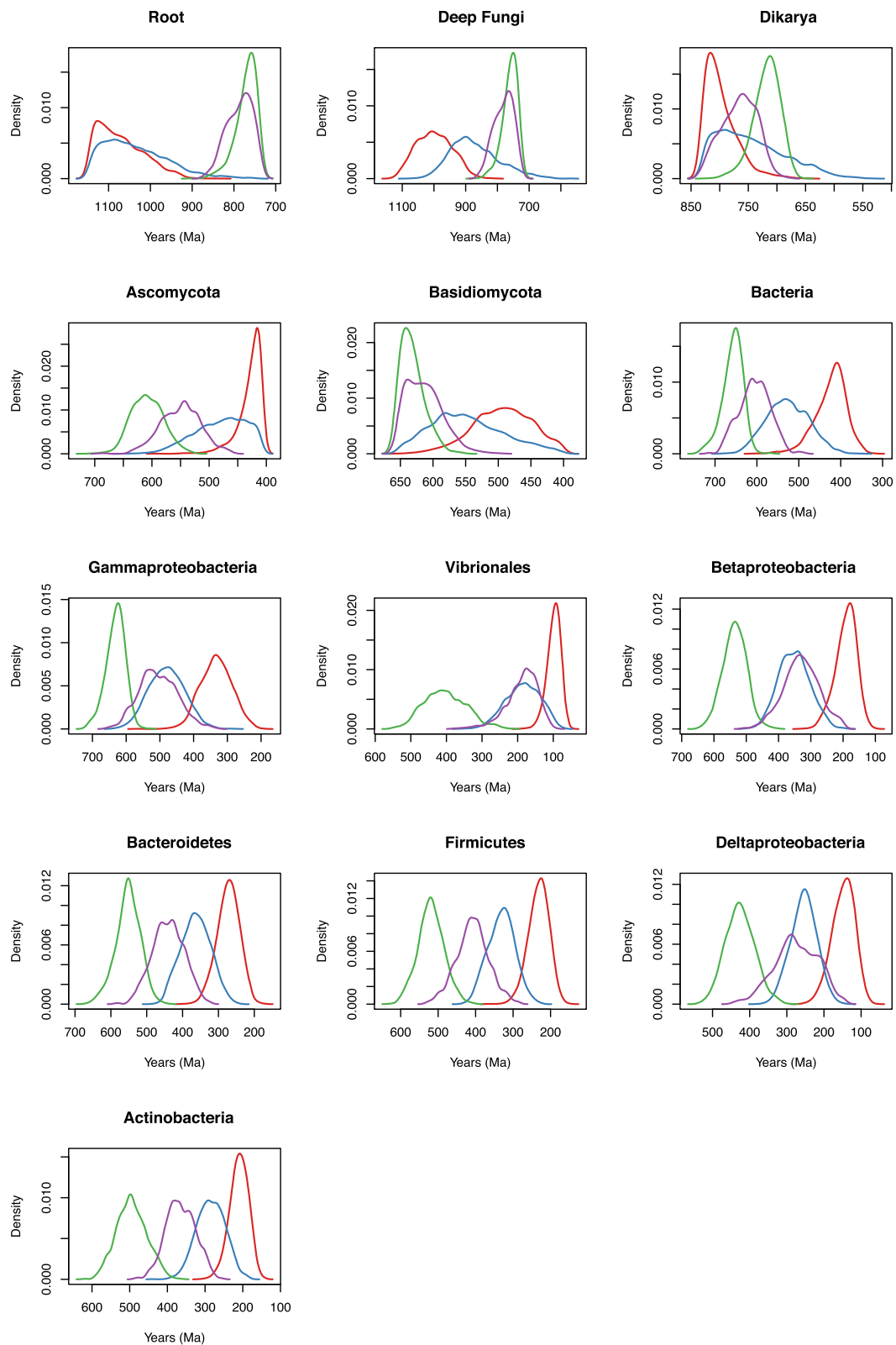


Fig. 3 Posterior chitinase date distribution across nodes with varied model selection. Models: 10, birth-death prior and uncorrelated gamma relaxed clock model (red); 14, birth-death prior and lognormal autocorrelated relaxed clock model (blue); 2, uniform prior and lognormal autocorrelated relaxed clock model (green); and 6, and uniform prior and uncorrelated gamma relaxed clock model (purple)

hypotheses, as in this dataset we infer multiple HGTs between Bacteria after a primary HGT from Fungi. There are many nodes that are clearly not the consequence of birth-death processes. In fact, the ecological dispersal of genes via HGT should be expected to locally increase node densities in the tree entirely independent of any underlying assumptions of speciation or extinction. In the absence of a different model sensitive to nodes mapping as transfers vs. speciation events, it is important to avoid assumptions made in the birth-death model. In addition, for many of the bacterial nodes, the uniform tree process prior results in broader prior ages than the birth-death prior. Therefore, the violation of the assumptions of a birth-death process in the bacterial chitinase tree may result in overly narrow priors that are too informative. Additionally, autocorrelated rate distribution models generally perform poorly for large evolutionary distances [46], and inspection of the gene tree does not readily reveal any lineage-specific branch length effects that suggest rate biases that would be poorly accounted for under an uncorrelated model.

Discrimination between these priors and evolutionary models would be substantially aided by the presence of crown-group calibrations within the bacterial clades recovered within the HGT recipient subtree. While diagnostic body fossils representing these microbial clades are almost certain not to be found, future studies may provide such calibrations in the form of additional HGT events, or inferred cospeciations with fossil-calibrated metazoan host lineages [71]. Such additional calibrations would also permit sensitivity analyses to be performed for the HGT-based calibrations used in this study.

Impact of taxon sampling and fungal divergence times

The impact of taxon sampling was evaluated (Additional file 1: Table S1, Additional file 4: Table S2). Within Fungi, the chitinase gene appears to follow a history of vertical descent, and therefore better modeled under a birth-death tree process prior. Therefore, one test of the appropriateness of a birth-death process prior is if the presence of bacterial sequences within the tree impacts the effective prior ages within Fungi. Ascomycota and Basidiomycota groups each have prior ages ~ 100 Ma younger under the birth-death model when Bacteria are removed. Under the uniform model, Ascomycota is the same age whether or not Bacteria are included, while Basidiomycota is also ~ 100 Ma younger. In general, the birth-death model gives much younger prior ages, ~ 150 Ma for Bacteria and Basidiomycota, whether or not Bacteria are in the tree, and ~ 150 Ma for Ascomycota in the presence of Bacteria, and ~ 250 Ma in the absence of Bacteria. Ascomycota and Basidiomycota crown group age priors are very sensitive to the tree process prior. Therefore, we chose to use additional

secondary calibrations within Dikarya to constrain the prior on the Ascomycota and Basidiomycota nodes.

Impact of calibrations

In general, the date distributions across all nodes do not appear to be very sensitive to the calibrations applied under the uniform distribution and uncorrelated gamma relaxed clock model. Because the calibrations are all roughly in the same range, it appears that all calibration results lead to similar date distributions (Fig. 4). However, Calibration 2 (calibrations on the root, split of Ascomycota and Basidiomycota, and Ascomycota and Basidiomycota crown lineages, not including the fossil minima) lead to slightly more precise peaks (Fig. 4). There are two potential problems with using single gene alignments to generate a posterior age estimate for an HGT: (1) a single gene has limited rate information from aligned sites for an informative molecular clock, and (2) if HGT increases the rate of evolution along reticulate branches due to genes evolving faster once in a recipient genome, then the posteriors will bias results towards under-estimating the ages of these groups. Therefore, we assessed whether younger posterior dates generated by the birth-death prior as compared to the uniform prior were due to the long branch separating Bacteria from Fungi in the tree. It is possible that this long branch may either be representative of a longer time interval (and thus younger crown ages) or of a faster evolutionary rate (and thus older crown ages). The maximum likelihood tree (Fig. 1 and Additional file 5: Figure S3) illustrates that when rooted, Ascomycota and Basidiomycota actually have slightly longer distances to the root, suggesting that the relative rates of evolution in this gene tree are not accelerated in the bacterial group. Consequently, the limited sequence information contained in this dataset may be used to calculate posterior age estimates that are unlikely to be biased by HGT-induced rate effects. Including additional internal constraints on the fungal clades push the priors under the uniform and birth-death models closer together for bacterial nodes. These additional secondary calibrations are thus important for constraining the tree process prior, and this type of approach may be important for using single gene HGTs to improve age estimates in general.

Informativeness of sequence data

We assessed the informative of the sequence data by running PhyloBayes under the prior (effective prior, including calibrations (Additional file 6: Figure S4)). Posterior age distributions for bacterial chitinase nodes substantially differed from prior age distributions, showing that sequence data is meaningfully informing age estimates via the relaxed molecular clock (Figs. 5 and 6).

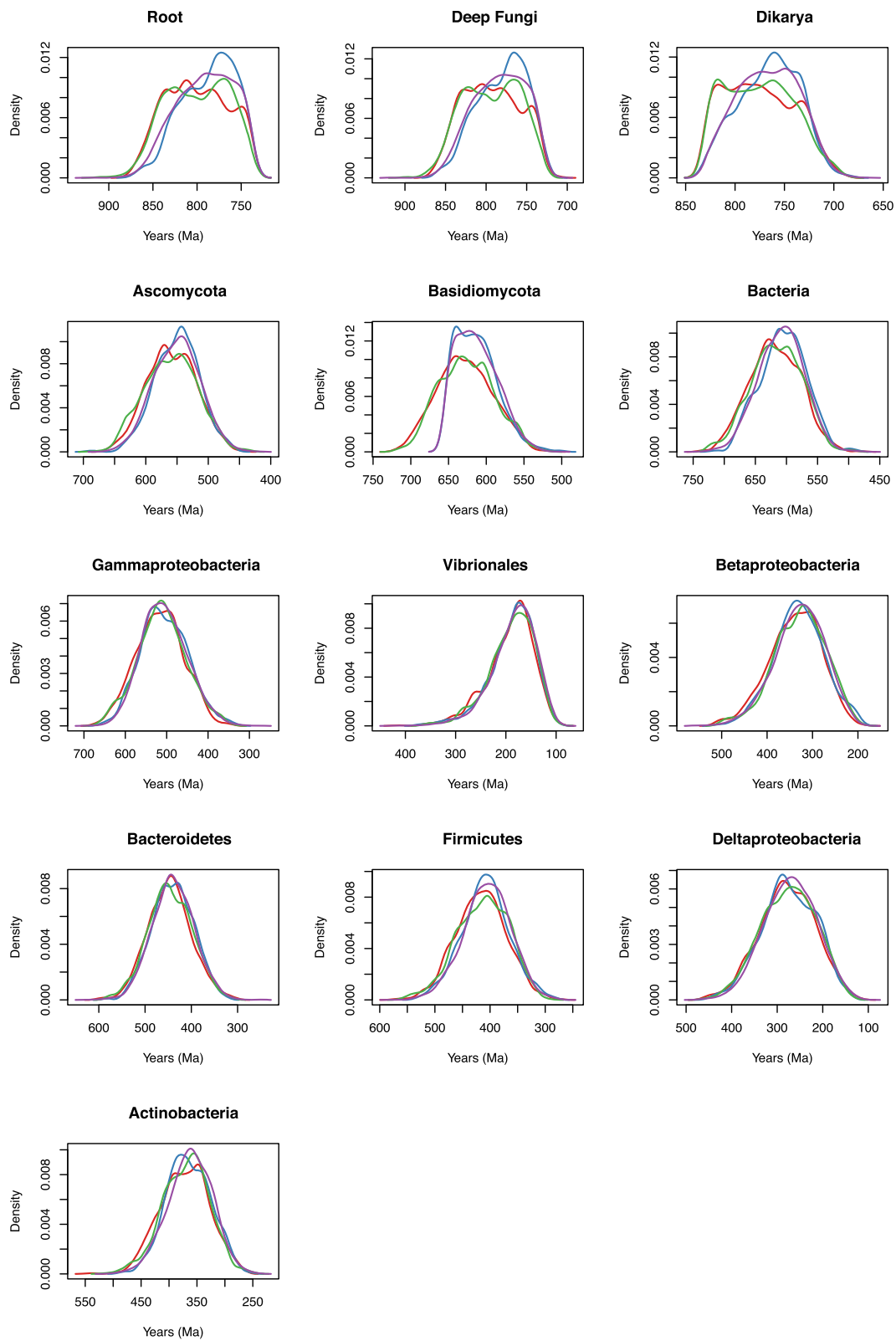


Fig. 4 Posterior chitinase date distribution across nodes with varied calibration. Illustrates the posterior age distributions across nodes under the four calibration setups, 1 (red), 2 (blue), 3 (green) and 4 (purple). These correspond to Models 5, 6, 7, and 8

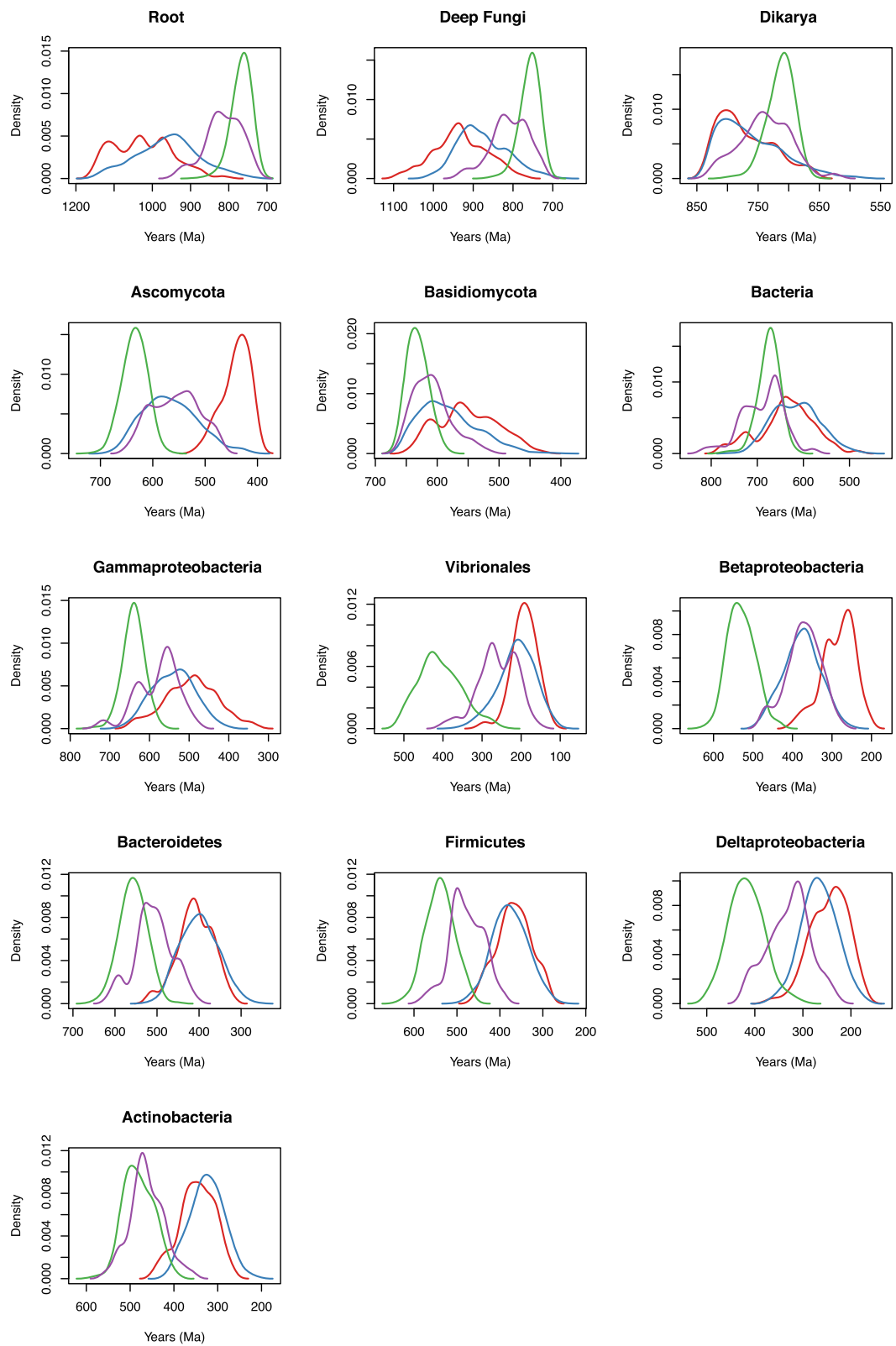


Fig. 5 Prior date distributions across nodes. Models: 10p, birth-death prior and uncorrelated gamma relaxed clock model (red); 14p, birth-death prior and lognormal autocorrelated relaxed clock model (blue); 2p, uniform prior and lognormal autocorrelated relaxed clock model (green); and 6p, and uniform prior and uncorrelated gamma relaxed clock model (purple)

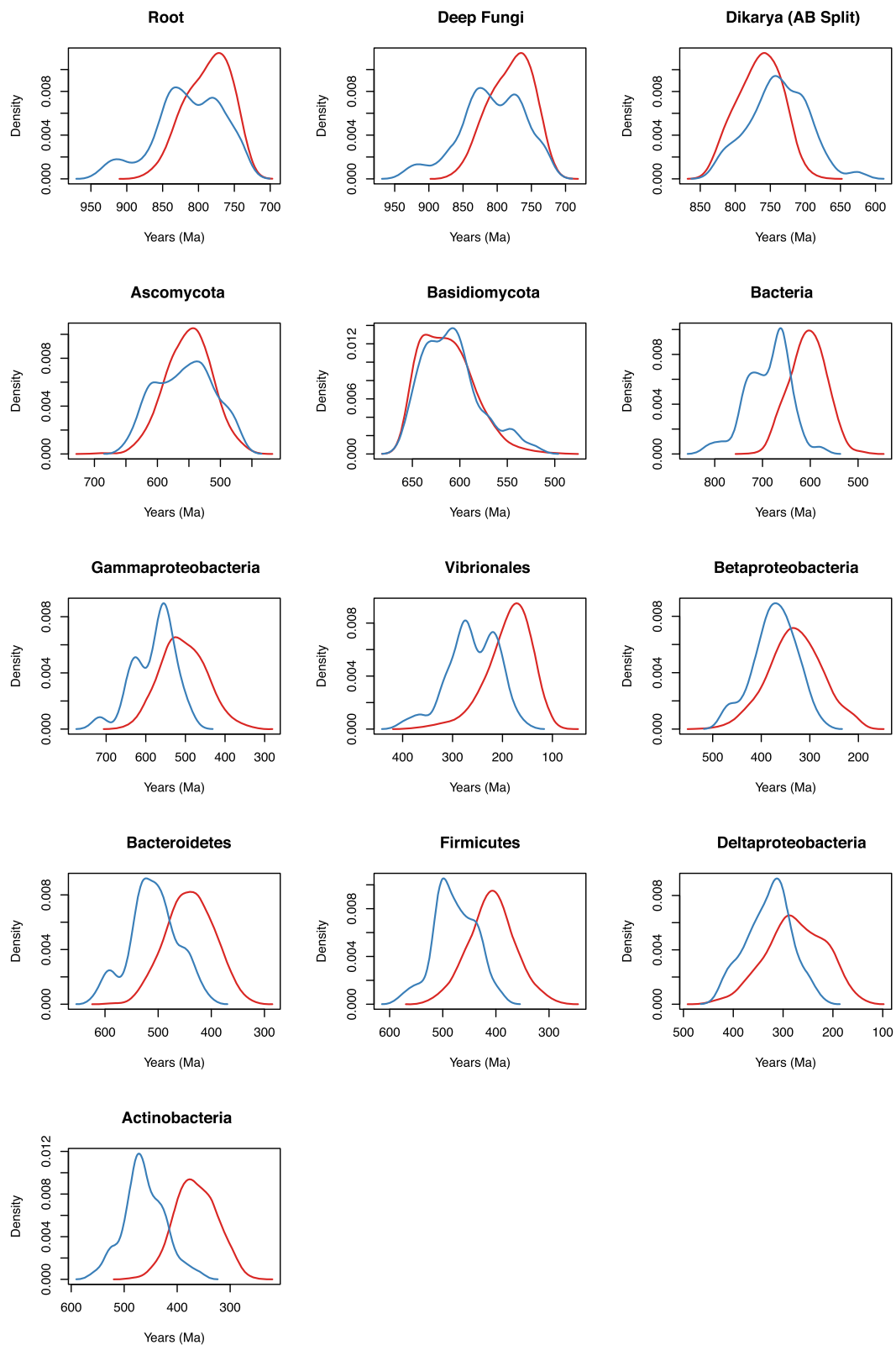


Fig. 6 Prior vs. posterior chitinase date distributions across nodes run under the prior (blue) and including sequence data (red)

Discussion

Fungal origin and distribution of bacterial Chitinases

The gene tree topology for ChiC/D and its inferred rooting within Fungi show that bacterial chitinase was acquired via HGT from a fungal donor lineage. By including secondary age calibrations on nodes within Fungi, molecular clock estimates show that this gene was acquired by bacteria by 605 Ma (range of 655–566 Ma), slightly predating estimates for the evolution of crown marine euarthropods [31, 32]. While the environment of the first Fungi is uncertain, the earliest Fungi likely evolved from aquatic ancestors, and colonized land by moving from shallow marine or freshwater environments to terrestrial environments [22, 24, 72, 73]. This is consistent with the hypothesis that bacterial chitinases evolved from an aquatic ancestor. Nevertheless, the initial acquisition of bacterial chitinase is unlikely to be in response to increases in marine arthropod chitin, which is unlikely to have been widespread at that time.

The HGT between Fungi and Bacteria also seems plausible from environmental and mechanistic perspectives. Bacteria and Fungi occupy similar environments, and other bacterial chitinases within the GH18 family (e.g. ChiI) have been hypothesized to have evolved via HGT, possibly from Fungi [37]. Following the initial transfer into a bacterial lineage, bacterial groups have all acquired chitinases from one another via subsequent HGT events, although the donors of these HGTs cannot be directly inferred from the tree topology, except in the case of Firmicutes to Deltaproteobacteria.

Importance of chitinase evolution for dating microbial metabolisms

Several types of calibrations exist for constraining divergence time estimates for clades of organisms. Fossil node calibrations with well-defined phylogenetic histories, morphological, and age information provide some of the strongest constraints [74]. Tip-dating or Total Evidence Dating expands the utility of fossil age constraints by merging fossil species information with extant species information [51]. HGT events also constrain the relative age of donor and recipient clades [4, 5, 9]. Nearly all bacterial groups lack fossil evidence that could potentially constrain crown-group clades. There are some fossil constraints within Cyanobacteria [3, 53], and other bacterial lineages contain proxy eukaryote fossil calibrations, such as mitochondrial lineages within Alphaproteobacteria [9]. Nonetheless, major lineages such as Firmicutes, which are distant relatives to these better-calibrated groups, are difficult to date, and because they are so distant, calibrations for other regions of the tree, even if they exist, are essentially not informative. HGT-propagated calibrations are thus especially valuable for dating microbial lineages.

Substrate-specific genes, such as chitinases, are also valuable for placing absolute older-bound ages on

microbial lineages, as they can be inferred to have evolved in direct response to a derived character (e.g., chitin synthesis) found within another, better-calibrated part of the Tree of Life. In the special case that these genes were acquired by multiple HGTs across diverse recipient lineages, as we observe for ChiD, a further inference can be made: HGT acquisitions in these groups are likely the result of substrate availability increasing or expanding across multiple microbial niches. This suggests that recipient clades are all of similar ages, regardless of their taxonomic diversity. In effect, the substrate-dependent dispersal of these genes act as “Standard Candles” for dating microbial groups. The concept of Standard Candles is taken from astronomy, where absolute distances of objects can be as difficult to infer as absolute dates for species divergence times. In order to help solve this problem, the known absolute luminosity of some sets of objects can be inferred from their physical properties, such as Cepheid variable stars [75] and Type IA supernovae [14]. Given their observed (apparent) luminosities, a distance calculation can then be made, and extrapolated to other objects. Similarly, substrate-dependent HGTs may permit multiple clades to be established to be within a specific absolute age range, improving divergence time estimates across the Tree of Life. Future dating efforts will likely benefit from exploring a broader set of temporally-constrained, substrate-specific HGTs. Depending upon their ubiquity, this may be a robust means of proxy dating microbial lineages, at least within the time horizon of eukaryal life for which a diagnostic fossil record exists.

Divergence time age estimates from this study can also be useful for future investigations. While a single gene, such as ChiD, contains limited sequence data for informing posterior age distributions, posterior dates from HGT-calibrated gene trees can be used as constraints that may improve accuracy in molecular clock studies using larger alignments [57, 76]. Furthermore, in at least one case, our results suggest the likely transfer of chitinase genes from within one phylogenetically distant [77] microbial clade to another: Bacilli (Firmicutes) to Myxococcales (Deltaproteobacteria) (Fig. 1). Because these clades are “nested” one could therefore polarize the direction of the transfer and apply a relative age constraint between these two groups on a species tree, independent of any propagated absolute constraints. Future work should assess how the application of such constraints affect divergence time estimates on species trees containing additional sequence information and increased taxon sampling.

Ecological implications of chitinase evolution

Our results show that numerous clades of bacteria acquired chitinase genes during the early Paleozoic (Fig. 7), suggesting that their dispersal throughout the microbial

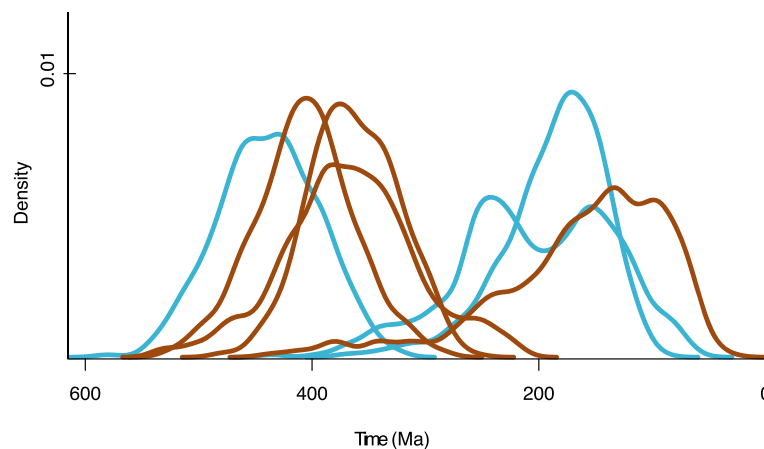


Fig. 7 Posterior date distributions for aquatic (blue) and terrestrial (brown) bacterial chitinases. Posterior date distributions correspond to the uniform prior and uncorrelated gamma relaxed clock model with calibration 4 (Model 6). Aquatic nodes represent ancestors of groups within Vibrionales (165), Gammaproteobacteria (214), and Bacteroidetes (174). Terrestrial nodes represent ancestors of groups within Betaproteobacteria (149), Deinococcus (212), Actinobacteria (215) and Firmicutes (181)

world was in direct response to the evolutionary and ecological expansion of detrital-chitin producing arthropod groups. Nonetheless, it is uncertain how the primary origin of environmentally-relevant amounts of chitin has evolved through time; did this originate from fungal cell walls or detrital chitin from the molted exoskeletons of arthropods? The genomic record may aid in distinguishing these sources. Robust across our model parameters and assumptions, bacterial chitinase (ChiD) appears to have evolved from fungi, likely in response to the availability of chitin as a major structural component of fungal cell walls in the Proterozoic, prior to early arthropod evolution. Subsequent HGT and inheritance of chitinase within terrestrial bacterial clades appears to be a much more recent series of evolutionary events within the early Paleozoic, consistent with evidence for plant and arthropod terrestrialization during this time [32, 78]. The taxonomic distribution of ChiD within marine microbial groups is too sparse to infer the timing of their acquisition, or to polarize the deep HGT events between microbial lineages, which, presumably, progressed from marine to terrestrial clades (Fig. 7). The long reticulating branch leading from Fungi to the bacterial ChiD sequences suggests that the direct fungal donor clade is not represented in the current tree; this may be due to unsampled extant fungal diversity, or patterns of extinction among ancient marine fungal groups.

Conclusion

The evolution of the metabolic degradation of detrital arthropod and fungal chitin may provide important temporal clues for dating early bacterial diversification. If bacterial chitinase genes were acquired in response to environmental chitin availability, then arthropod evolutionary history provides a

prediction for the timing of these events within bacterial lineages: terrestrial bacterial chitinases should have diversified from the Cambrian into the Devonian. We independently date chitinase evolution in microbial lineages by applying molecular clocks to chitinase gene trees, constrained by fungal date calibrations tethered via HGT. We show that concordance with terrestrial arthropod evolution indeed appears to be the case, further supporting the utility of HGT-propagated fossil calibrations in accurately estimating the ages of microbial lineages.

Bacterial chitinases appear to have diversified from the time of their acquisition, roughly 600 Ma, into the last 200 Ma of Earth history (Table 3). This is consistent with the hypothesis that bacterial chitinases evolved in response to the seeding of marine and terrestrial environments with globally-significant amounts of chitin, first from fungi, then later from marine and then terrestrial arthropods. There is later evidence of at least one HGT event within bacterial lineages, from within Firmicutes to Deltaproteobacteria. Although we only assessed one chitinase gene tree in this study, future work evaluating the phylogenetic distribution of other chitinase genes will be important for quantifying chitinase evolution in marine and terrestrial environments to further test the hypothesis that the phylogenetic distribution of chitinase genes mirrors the evolution and terrestrialization of environmental chitin sources.

Further, we show the importance of prior choice, highlighting that this dataset, which includes at least one deeply-rooted HGT, violates the birth-death prior. Moreover, we argue for the use of a uniform prior, uncorrelated gamma multipliers model, and three internal secondary calibrations propagating fossil calibrations from within Fungi to Bacteria.

Finally, we suggest that our dataset does not demonstrate HGT-associated heterotachy. Thus, our fungal

priors and perhaps even posterior bacterial date distributions, may be more broadly applicable for future molecular clock studies assessing the divergence times of these major clades of Bacteria.

Additional files

Additional file 1: Table S1. Taxa in this study and their environment. Environment refers to the environment that these substrates operate in¹ or where the organism was sampled from according to NCBI. Aquatic represents marine environments. Terrestrial refers to land which includes shallow freshwater ponds/swamp type sample locations. (DOCX 42 kb)

Additional file 2: Figure S1. PhyloBayes chitinase gene tree with posterior probabilities supporting nodes. (PDF 46 kb)

Additional file 3: Figure S2. Chronogram with corresponding node numbers used in analysis. Clades are annotated on corresponding nodes. (PDF 374 kb)

Additional file 4: Table S2. Extended Table 2 of all results and parameters tested. Prior and posterior divergence time estimates calculated under all model assumptions described in Table 1. Divergence date ranges are given in Ma. (XLSX 12 kb)

Additional file 5: Figure S3. Expanded RAXML gene tree with tip labels and bootstrap support values. (PDF 16 kb)

Additional file 6: Figure S4. Prior date distributions across nodes under the four calibration setups. Models: Calibration 1 (red), Calibration 2 (blue), Calibration 3 (green), Calibration 4 (purple). (PDF 492 kb)

Abbreviations

BLAST: Basic Local Alignment Search Tool; Chi (A-J): Chitinase (A-J); GH: Glycoside hydrolase; HGT: Horizontal gene transfer; Ma: Million years ago; NCBI: The National Center for Biotechnology Information; nr: Nonredundant

Acknowledgements

The authors thank Cara Magnabosco for valuable discussions and Abigail Caron for developing sequence analysis pipelines. We also thank the three anonymous reviewers for their time and constructive comments on this manuscript.

Funding

This work was supported by a National Science Foundation (NSF) Graduate Research Fellowship Program Award to DSG, and Simons Collaboration on the Origins of Life Award #339603 and NSF Integrated Earth Systems Program Award #1615426 to GPF. The funding agencies for this study had no role in study design, data collection, data analysis and interpretation, or in writing the manuscript.

Availability of data and materials

The data set supporting the results of this article is available in the Dryad repository, <https://doi.org/10.5061/dryad.np0s26r>.

Authors' contributions

DSG and GPF conceived of this project. DSG performed the analyses, and drafted the manuscript. Calibrations, molecular clock evaluations, and final preparation of the manuscript were supervised by GPF and JMW. All authors have read and approved this manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 17 October 2018 Accepted: 9 January 2019

Published online: 28 January 2019

References

- dos Reis M, Donoghue PCJ, Yang Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet.* 2015;17:1–10. <https://doi.org/10.1038/nrg.2015.8>.
- Shih PM. Photosynthesis and early Earth. *Curr Biol.* 2015;25:R855–9. <https://doi.org/10.1016/j.cub.2015.04.046>.
- Schirmer BE, Gugger M, Donoghue PC. Cyanobacteria and the great oxidation event: evidence from genes and fossils. *Palaeontology.* 2015;58:935–6. <https://doi.org/10.1111/pala.12193>.
- Wolfe JM, Fournier GP. Horizontal gene transfer constrains the timing of methanogen evolution. *Nat Ecol Evol.* 2018;129494. <https://doi.org/10.1038/s41559-018-0513-7>.
- Davin AA, Tannier E, Williams TA, Boussau B, Daubin V, Szöllösi GJ. Gene transfers can date the tree of life. *Nat Ecol Evol.* 2018;2:904–9. <https://doi.org/10.1038/s41559-018-0525-3>.
- Gold DA, Caron A, Fournier GP, Summons RE. Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature.* 2017;543:420–3. <https://doi.org/10.1038/nature21412>.
- Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet.* 2015;16:472–82. <https://doi.org/10.1038/nrg3962>.
- dos Reis M. Fossil-free dating. *Nat Ecol Evol.* 2018;2:771–2. <https://doi.org/10.1038/s41559-018-0532-4>.
- Magnabosco C, Moore KR, Wolfe JM, Fournier GP. Dating phototrophic microbial lineages with reticulate gene histories. *Geobiology.* 2018;16:179–89. <https://doi.org/10.1111/gbi.12273>.
- Wolfe JM, Fournier GP. Reply to 'Molecular clocks provide little information to date methanogenic Archaea.' *Nat Ecol Evol.* 2018;2:1678–1678. <https://doi.org/10.1038/s41559-018-0685-1>.
- Andam CP, Williams D, Gogarten JP. Natural taxonomy in light of horizontal gene transfer. *Biol Philos.* 2010;25:589–602. <https://doi.org/10.1007/s10539-010-9212-8>.
- Husnik F, McCutcheon JP. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol.* 2017;16:67–79. <https://doi.org/10.1038/nrmicro.2017.137>.
- Fournier GP, Huang J, Gogarten JP. Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Philos Trans R Soc B Biol Sci.* 2009;364:2229–39. <https://doi.org/10.1098/rstb.2009.0033>.
- Colgate SA. Supernovae as a standard candle for cosmology. *Astrophys J.* 1979;232:404. <https://doi.org/10.1086/157300>.
- Bai Y, Eijssink VGH, Kielak AM, van Veen JA, de Boer W. Genomic comparison of chitinolytic enzyme systems from terrestrial and aquatic bacteria. *Environ Microbiol.* 2016;18:38–49. <https://doi.org/10.1111/1462-2920.12545>.
- Talamantes D, Biabini N, Dang H, Abdoun K, Berlemont R. Natural diversity of cellulases, xylanases, and chitinases in bacteria. *Biotechnol Biofuels.* 2016;9:133. <https://doi.org/10.1186/s13068-016-0538-6>.
- Schwentner M, Combosch DJ, Pakes Nelson J, Giribet G. A Phylogenomic Solution to the Origin of Insects by Resolving Crustacean-Hexapod Relationships. *Curr Biol.* 2017;27:1818–1824.e5. <https://doi.org/10.1016/j.cub.2017.05.040>.
- Cauchie H-M. Chitin production by arthropods in the hydrosphere. *Hydrobiologia.* 2002;470 1/3:63–95. <https://doi.org/10.1023/A:1015615819301>.
- Smrž J, Čatská V. Mycophagous mites and their internal associated bacteria cooperate to digest chitin in soil. *Symbiosis.* 2010;52:33–40. <https://doi.org/10.1007/s13199-010-0099-6>.
- Holtkamp R, Kardol P, van der Wal A, Dekker SC, van der Putten WH, de Ruiter PC. Soil food web structure during ecosystem development after land abandonment. *Appl Soil Ecol.* 2008;39:23–34. <https://doi.org/10.1016/j.apsoil.2007.11.002>.
- Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. *Proc Natl Acad Sci.* 2018;115:6506–11. <https://doi.org/10.1073/pnas.1711842115>.
- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature.* 2006;443:818–22. <https://doi.org/10.1038/nature05110>.

23. Berbee ML, Taylor JW. Dating the molecular clock in fungi – how close are we? *Fungal Biol Rev.* 2010;24:1–16. <https://doi.org/10.1016/j.fbr.2010.03.001>.
24. Berbee ML, James TY, Strullu-Derrien C. Early Diverging Fungi: Diversity and Impact at the Dawn of Terrestrial Life. *Annu Rev Microbiol.* 2017;71:41–60. <https://doi.org/10.1146/annurev-micro-030117-020324>.
25. Strullu-Derrien C, Spencer ART, Goral T, Dee J, Honegger R, Kenrick P, et al. New insights into the evolutionary history of Fungi from a 407 Ma Blastocladiomycota fossil showing a complex hyphal thallus. *Philos Trans R Soc Lond B Biol Sci.* 2018;373:20160502. <https://doi.org/10.1098/rstb.2016.0502>.
26. Taylor JW, Berbee ML. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia.* 2006;98:838–49. doi:10.1080/15572536.2006.11832614.
27. Sharpe SC, Eme L, Brown MW, Roger AJ. Timing the Origins of Multicellular Eukaryotes Through Phylogenomics and Relaxed Molecular Clock Analyses. In: Ruiz-Trillo I, Nedelcu AM, editors. *Evolutionary Transitions to Multicellular Life.* Springer; 2015. p. 3–29. https://doi.org/10.1007/978-94-017-9642-2_1.
28. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science (80-).* 2012;336:1715–9. <https://doi.org/10.1126/science.1221748>.
29. Wolfe JM, Daley AC, Legg DA, Edgecombe GD. Fossil calibrations for the arthropod Tree of Life. *Earth-Science Rev.* 2016;160:43–110. <https://doi.org/10.1016/j.earscirev.2016.06.008>.
30. Prieto M, Wedin M. Dating the Diversification of the Major Lineages of Ascomycota (Fungi). *PLoS One.* 2013;8:e65576. <https://doi.org/10.1371/journal.pone.0065576>.
31. Daley AC, Antcliffe JB, Drage HB, Pates S. Early fossil record of Euarthropoda and the Cambrian Explosion. *Proc Natl Acad Sci.* 2018;115:5323–31. <https://doi.org/10.1073/pnas.1719962115>.
32. Lozano-Fernandez J, Carton R, Tanner AR, Puttick MN, Blaxter M, Vinther J, et al. A molecular palaeobiological exploration of arthropod terrestrialization. *Philos Trans R Soc B Biol Sci.* 2016;371:20150133. <https://doi.org/10.1098/rstb.2015.0133>.
33. Oakley TH, Wolfe JM, Lindgren AR, Zaharoff AK. Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny. *Mol Biol Evol.* 2013;30:215–33. <https://doi.org/10.1093/molbev/mss216>.
34. Glenner H, Thomsen PF, Hebsgaard MB, Sørensen MV, Willerslev E. The Origin of Insects. *Science (80-).* 2006;314:1883–4. <https://doi.org/10.1126/science.1129844>.
35. Berlemont R, Martiny AC. Genomic Potential for Polysaccharide Deconstruction in Bacteria. *Appl Environ Microbiol.* 2015;81:1513–9. <https://doi.org/10.1128/AEM.03718-14>.
36. Chang T-C, Stergiopoulos I. Evolutionary analysis of the global landscape of protein domain types and domain architectures associated with family 14 carbohydrate-binding modules. *FEBS Lett.* 2015;589:1813–8. <https://doi.org/10.1016/j.febslet.2015.05.048>.
37. Ubhayasekera W, Karlsson M. Bacterial and fungal chitinase chiJ orthologs evolve under different selective constraints following horizontal gene transfer. *BMC Res Notes.* 2012;5:581. <https://doi.org/10.1186/1756-0500-5-581>.
38. Ihrmark K, Asmail N, Ubhayasekera W, Melin P, Stenlid J, Karlsson M. Comparative Molecular Evolution of Trichoderma Chitinases in Response to Mycoparasitic Interactions. *Evol Bioinforma.* 2010;6:EBO.S4198. <https://doi.org/10.4137/EBO.S4198>.
39. Mamarabadi M, Jensen B, Lübeck M. Three endochitinase-encoding genes identified in the biocontrol fungus *Clonostachys rosea* are differentially expressed. *Curr Genet.* 2008;54:57–70. <https://doi.org/10.1007/s00294-008-0199-5>.
40. Sharma G, Subramanian S. Unravelling the Complete Genome of Archangium gephyra DSM 2261T and Evolutionary Insights into Myxobacterial Chitinases. *Genome Biol Evol.* 2017;9:1304–11. <https://doi.org/10.1093/gbe/evx066>.
41. Emerling CA, Delsuc F, Nachman MW. Chitinase genes (CHIA s) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Sci Adv.* 2018;4:ear6478. <https://doi.org/10.1126/sciadv.aar6478>.
42. Lacombe-Harvey M-E, Brzezinski R, Beaulieu C. Chitinolytic functions in actinobacteria: ecology, enzymes, and evolution. *Appl Microbiol Biotechnol.* 2018;102:7219–30. <https://doi.org/10.1007/s00253-018-9149-4>.
43. Donoghue PCJ, Yang Z. The evolution of methods for establishing evolutionary timescales. *Philos Trans R Soc B Biol Sci.* 2016;371:20160020. <https://doi.org/10.1098/rstb.2016.0020>.
44. Louca S, Shih PM, Pennell MW, Fischer WW, Parfrey LW, Doebeli M. Bacterial diversification through geological time. *Nat Ecol Evol.* 2018;1. <https://doi.org/10.1038/s41559-018-0625-0>.
45. Hillis DM, Heath TA, John KS. Analysis and Visualization of Tree Space. *Syst Biol.* 2005;54:471–82. <https://doi.org/10.1080/10635150590946961>.
46. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* 2006;4:e88. <https://doi.org/10.1371/journal.pbio.0040088>.
47. Donoghue PCJ, Benton MJ. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol Evol.* 2007;22:424–31. <https://doi.org/10.1016/j.tree.2007.05.005>.
48. Edwards S V. IS A NEW AND GENERAL THEORY OF MOLECULAR SYSTEMATICS EMERGING? *Evolution (N Y).* 2009;63:1–19. <https://doi.org/10.1111/j.1558-5646.2008.00549.x>.
49. Heled J, Drummond AJ. Calibrated Birth–Death Phylogenetic Time-Tree Priors for Bayesian Inference. *Syst Biol.* 2015;64:369–83. <https://doi.org/10.1093/sysbio/syu089>.
50. Guyader J, Ungerfeld EM, Beauchemin KA. Redirection of Metabolic Hydrogen by Inhibiting Methanogenesis in the Rumen Simulation Technique (RUSITEC). *Front Microbiol.* 2017;8 MAR:393. <https://doi.org/10.3389/fmicb.2017.00393>.
51. O'Reilly JE, dos Reis M, Donoghue PCJ. Dating Tips for Divergence-Time Estimation. *Trends Genet.* 2015;31:637–50. <https://doi.org/10.1016/j.tig.2015.08.001>.
52. Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW. Bayesian molecular dating: opening up the black box. *Biol Rev.* 2017;93:1165–91. <https://doi.org/10.1111/brv.12390>.
53. Knoll AH, Nowak MA. The timetable of evolution. *Sci Adv.* 2017;3:e1603076. <https://doi.org/10.1126/sciadv.1603076>.
54. Nascimento FF, dos Reis M, Yang Z. A biologist's guide to Bayesian phylogenetic analysis. *Nat Ecol Evol.* 2017;1:1446–54. <https://doi.org/10.1038/s41559-017-0280-x>.
55. Huelsenbeck JP, Larget B, Miller RE, Ronquist F. Potential Applications and Pitfalls of Bayesian Inference of Phylogeny. *Syst Biol.* 2002;51:673–88. <https://doi.org/10.1080/10635150290102366>.
56. Lepage T, Bryant D, Philippe H, Lartillot N. A General Comparison of Relaxed Molecular Clock Models. *Mol Biol Evol.* 2007;24:2669–80. <https://doi.org/10.1093/molbev/msm193>.
57. Rannala B, Yang Z. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *J Mol Evol.* 1996;43:304–11. <https://doi.org/10.1007/BF02338839>.
58. Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 1998;15:1647–57. <https://doi.org/10.1093/oxfordjournals.molbev.a025892>.
59. Ho SYW, Duchêne S. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol.* 2014;23:5947–65. <https://doi.org/10.1111/mec.12953>.
60. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics.* 2009;25:2286–8.
61. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7. <https://doi.org/10.1093/nar/gkh340>.
62. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbook. *Bioinformatics.* 2009;25:1189–91. <https://doi.org/10.1093/bioinformatics/btp033>.
63. Gruen DS, Wolfe JM, Fournier GP. Data from: Paleozoic diversification of terrestrial chitin-degrading bacterial lineages. In: Dryad Digit Repos. <https://doi.org/105061/dryad.np0s26r>. 2018.
64. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90. <https://doi.org/10.1093/bioinformatics/btl446>.
65. Darriba B, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011;27:1164–5. <https://doi.org/10.1093/bioinformatics/btr088>.
66. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A.* 2011;108:13624–9. <https://doi.org/10.1073/pnas.1110633108>.
67. Torruella G, Grau-Bové X, Moreira D, Karpov SA, Burns JA, Sebé-Pedrós A, et al. Global transcriptome analysis of the aphelid Paraphelidium tribonemae supports the phagotrophic origin of fungi. *Commun Biol.* 2018;1:231. <https://doi.org/10.1038/s42003-018-0235-z>.
68. Mergaert J, Lednická D, Goris J, Cnockaert MC, De Vos P, Swings J. Taxonomic study of *Cellvibrio* strains and description of *Cellvibrio ostraviensis* sp. nov., *Cellvibrio fibrivorans* sp. nov. and *Cellvibrio gandavensis* sp. nov. *Int J Syst Evol Microbiol.* 2003;53:465–71. <https://doi.org/10.1099/ijss.0.02316-0>.

69. Marin J, Battistuzzi FU, Brown AC, Hedges SB. The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation. *Mol Biol Evol.* 2016;34:msw245. doi:10.1093/molbev/msw245.
70. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, et al. The timescale of early land plant evolution. *Proc Natl Acad Sci.* 2018;115:E2274–83. <https://doi.org/10.1073/pnas.1719588115>.
71. Hafner MS, Nadler SA. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature.* 1988;332:258–9. <https://doi.org/10.1038/332258a0>.
72. Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, et al. Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. *Genome Biol Evol.* 2015;7:1590–601. <https://doi.org/10.1093/gbe/evv090>.
73. Beimforde C, Feldberg K, Nylinder S, Rikkinen J, Tuovila H, Dörfelt H, et al. Estimating the Phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data. *Mol Phylogenet Evol.* 2014;78:386–98. <https://doi.org/10.1016/j.ympev.2014.04.024>.
74. Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, et al. Best Practices for Justifying Fossil Calibrations. *Syst Biol.* 2012;61:346–59. <https://doi.org/10.1093/sysbio/syr107>.
75. Udalski A, Soszynski I, Szymanski M, Kubiak M, Pietrzynski G, Wozniak P, et al. The Optical Gravitational Lensing Experiment. Cepheids in the Magellanic Clouds. IV. Catalog of Cepheids from the Large Magellanic Cloud. *Acta Astron.* 1999;49:437–520. <http://arxiv.org/abs/astro-ph/9908317>.
76. Yang Z, Rannala B. Branch-Length Prior Influences Bayesian Posterior Probability of Phylogeny. *Syst Biol.* 2005;54:455–70. <https://doi.org/10.1080/10635150590945313>.
77. Raymann K, Brochier-Armanet C, Gribaldo S. The two-domain tree of life is linked to a new root for the Archaea. *Proc Natl Acad Sci.* 2015;112:6670–5. <https://doi.org/10.1073/pnas.1420858112>.
78. Kenrick P, Wellman CH, Schneider H, Edgecombe GD. A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Philos Trans R Soc B Biol Sci.* 2012;367:519–36. <https://doi.org/10.1098/rstb.2011.0271>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

