

# Episodic Evolution and Adaptation of Chloroplast Genomes in Ancestral Grasses

Bojian Zhong, Takahiro Yonezawa, Yang Zhong, Masami Hasegawa\*

School of Life Sciences, Fudan University, Shanghai, China

## Abstract

**Background:** It has been suggested that the chloroplast genomes of the grass family, Poaceae, have undergone an elevated evolutionary rate compared to most other angiosperms, yet the details of this phenomenon have remained obscure. To know how the rate change occurred during evolution, estimation of the time-scale with reliable calibrations is needed. The recent finding of 65 Ma grass phytoliths in Cretaceous dinosaur coprolites places the diversification of the grasses to the Cretaceous period, and provides a reliable calibration in studying the tempo and mode of grass chloroplast evolution.

**Methodology/Principal Findings:** By using chloroplast genome data from angiosperms and by taking account of new paleontological evidence, we now show that episodic rate acceleration both in terms of non-synonymous and synonymous substitutions occurred in the common ancestral branch of the core Poaceae (a group formed by rice, wheat, maize, and their allies) accompanied by adaptive evolution in several chloroplast proteins, while the rate reverted to the slow rate typical of most monocot species in the terminal branches.

**Conclusions/Significance:** Our finding of episodic rate acceleration in the ancestral grasses accompanied by adaptive molecular evolution has a profound bearing on the evolution of grasses, which form a highly successful group of plants. The widely used model for estimating divergence times was based on the assumption of correlated rates between ancestral and descendant lineages. However, the assumption is proved to be inadequate in approximating the episodic rate acceleration in the ancestral grasses, and the assumption of independent rates is more appropriate. This finding has implications for studies of molecular evolutionary rates and time-scale of evolution in other groups of organisms.

**Citation:** Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2009) Episodic Evolution and Adaptation of Chloroplast Genomes in Ancestral Grasses. PLoS ONE 4(4): e5297. doi:10.1371/journal.pone.0005297

**Editor:** Simon Joly, McGill University, Canada

**Received:** January 29, 2009; **Accepted:** March 19, 2009; **Published:** April 24, 2009

**Copyright:** © 2009 Zhong et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The study was financially supported by Shanghai Leading Academic Discipline Project (B111) and National Infrastructure of Natural Resources for Science and Technology (2005DKA21403) to Yang Zhong. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: hasegawa@fudan.edu.cn

## Introduction

The grass family, Poaceae, is one of the largest plant families, comprising about 10,000 species including the most important agricultural plants, rice, wheat and maize, and grass-dominated ecosystems comprise about one-third of Earth's vegetative cover and support a vast number of animals [1]. It has long been suggested that the chloroplast (chl) genomes of the grass family have undergone an elevated evolutionary rate compared to other angiosperms [2–5], yet little is known when, why and how this rate change occurred.

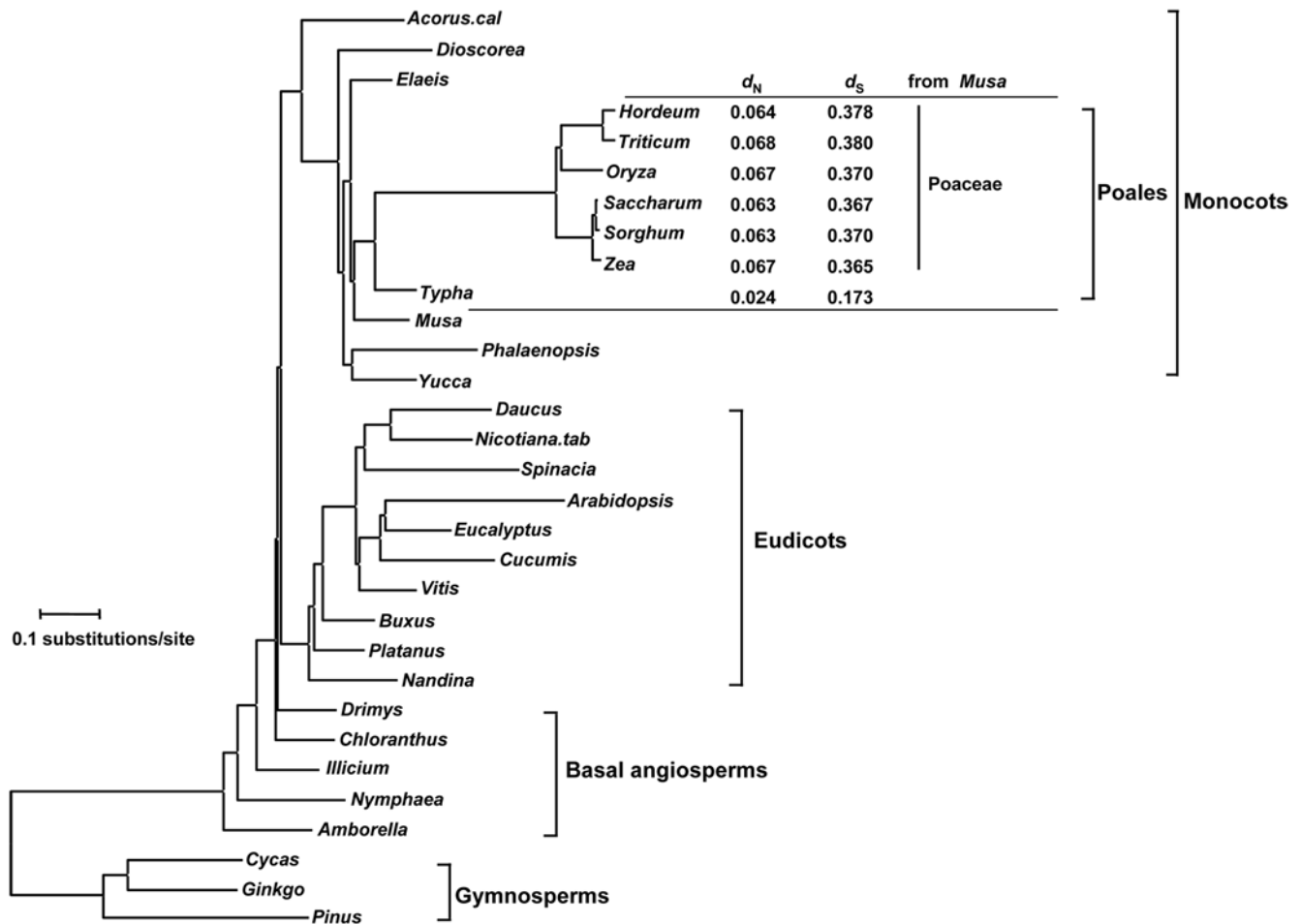
To examine how the rate change occurred during evolution, it is prerequisite to know the time-scale of evolution. It has become increasingly feasible to estimate the phylogenetic tree and the time-scale of Angiosperm evolution by using chl genome sequences [5–9]. A reliable calibration is necessary to obtain reliable time estimates, but lack of good fossil evidence of the ancestral grasses has prevented us from addressing this issue. The recent finding of 65 Ma grass phytoliths in Cretaceous dinosaur coprolites [10,11] places the diversification of the grasses to the Cretaceous period, and provides a reliable calibration in studying the tempo and mode of grass chl evolution. By using this

calibration, we here find that episodic rate acceleration occurred in the common ancestral branch of the core Poaceae (a clade formed by rice *Oryza*, wheat *Triticum*, maize *Zea*, and their allies) accompanied by adaptive evolution in several chl proteins, while the rate reverted to the slow rate typical of most monocot species in the terminal branches. We also find that the widely used method for estimating divergence times based on the assumption of correlated rates between ancestral and descendant lineages [12–14] proved to be inadequate in approximating the process of grass chl evolution, and the assumption of independent rates [15] is more appropriate to studies of rate change over time. These results have implications for studies of molecular evolutionary rates and time-scale of evolution in other groups of organisms.

## Results

### Estimation of time-scale and pattern of rate change

Fig. 1 shows the ML phylogenetic tree of Angiosperm chl with Gymnosperm as an outgroup. The elongated branches of Poaceae are in accord with their widely accepted rate acceleration [2–5]. The global clock model in Poales (including Poaceae and *Typha*)+*Musa* was rejected when compared with the 2-local-clocks



**Figure 1. The phylogenetic tree of chloroplast genomes for the 31 species.** The tree topology in Fig. 3 of ref.[6] was used, and the branch lengths were estimated by the ML with the codon-substitution model [21,22] (CODEML in PAML [19]). The root was arbitrarily placed between Gymnosperm and Angiosperm. Non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) distances of Poales from *Musa* were estimated by CODEML. doi:10.1371/journal.pone.0005297.g001

model (Poaceae lineages have a different rate from basal lineages such as *Typha* and *Musa*) by the likelihood ratio test (LRT) ( $\chi^2 = 2903.89$ ,  $P < 10^{-300}$  with the codon-substitution model). Moreover, longer distances of the Poaceae species from *Musa* than the *Typha/Musa* distance both in terms of non-synonymous and synonymous substitutions (Fig. 1) indicate that both types of substitutions have undergone rate acceleration along the line leading to Poaceae.

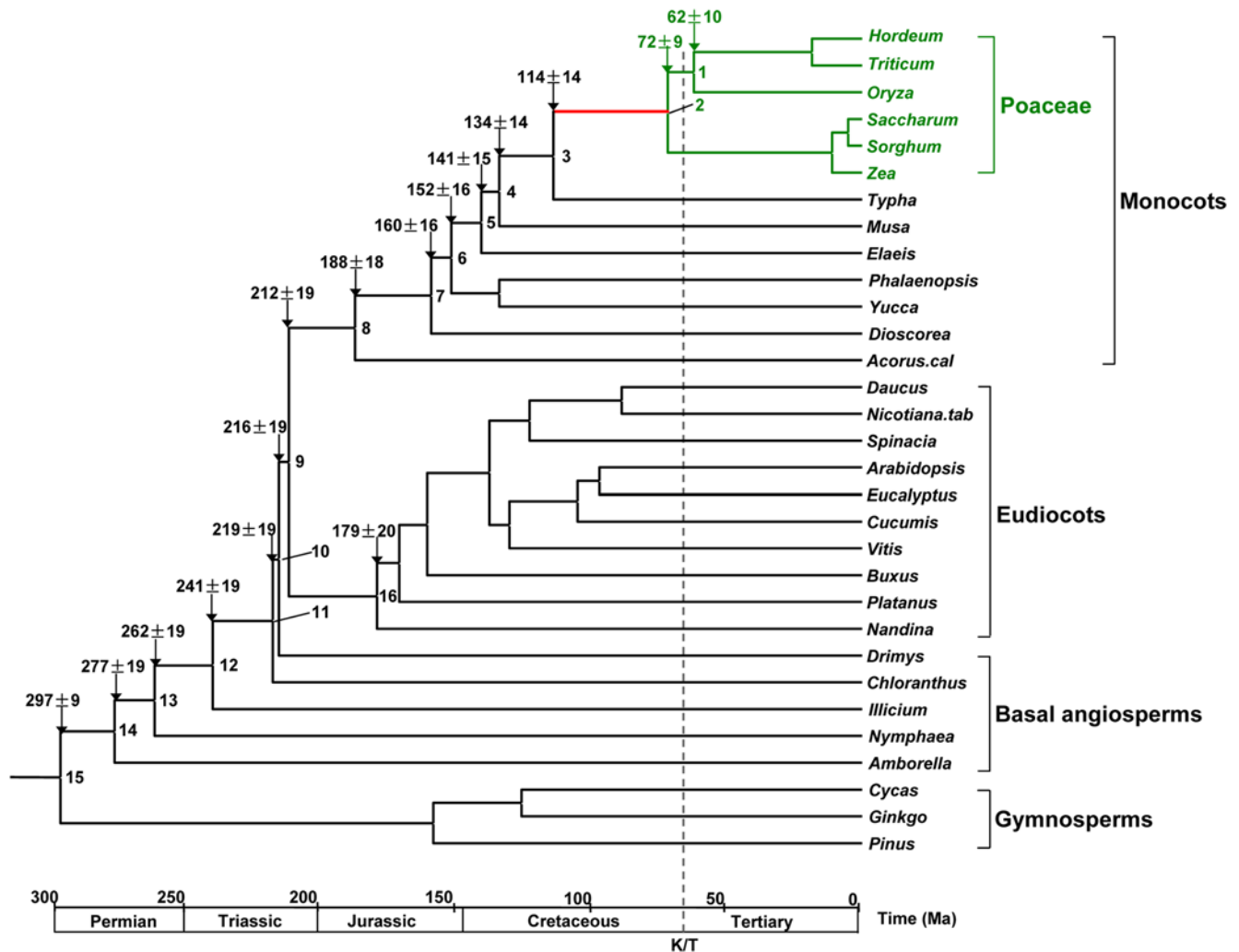
To explore the pattern of rate change during the course of grass evolution in more detail, we estimated the time-scale of Angiosperm phylogeny, particularly focusing on monocots. Although several powerful methods have been developed for molecular time estimation allowing the rate change (a relaxed clock) [12–17], the poor quality of the fossil record for early grasses has prevented us from addressing this issue. Previously, the divergence among major groups of Poaceae was thought to have occurred in early Cenozoic, and the 60 Ma [18] and 50–60 Ma [5] date calibrations for the maize/wheat divergence were used in estimating the monocots/eudicots divergence time with chl DNA sequences. However, recent findings of grass phytoliths in Cretaceous dinosaur coprolites [10,11] provided evidence that the major groups of core Poaceae had already diversified before Cretaceous/Tertiary (K/T) boundary of 65 Ma. Fig. 2A gives time estimates of the monocots evolution (Fig. 3 and Table 1 for

the whole angiosperms) by a relaxed clock based on the Bayesian method implemented in MCMCTREE (in PAML [19]) with a constraint of >65 Ma for the *Zea/Oryza* divergence and with the independent-rates model for the rate change along lineages [15,17]. In order to illustrate the rate change during evolution, a single instance of estimated rates along the lineage from the root to *Oryza* is also shown in Fig. 2A, where elevation of the rate only occurred on the common ancestral branch of Poaceae after they diverged from *Typha*.

Although the fossil evidence for the >65 Ma constraint of the *Zea/Oryza* divergence is important in demonstrating the rate acceleration in ancestral grasses with subsequent slow-down, it is not a prerequisite. Even when the constraint was removed, almost the same pattern of rate change as that with the >65 Ma constraint was obtained when the IR model was used (Fig. 2B and Table 2), although the time estimate of the *Zea/Oryza* separation became younger (55.0 Ma). This time estimate is consistent with a conservative date of >50 Ma presented in refs. [5,20], and our conclusion of the reverted slow rate in contemporary Poaceae can be regarded as robust to the calibration points used.

In the first model of the relaxed clock implemented by Thorne and colleagues [12,13], rates are auto-correlated between ancestral and descendant lineages on the tree, and the model is called the correlated-rates (CR) model. Sanderson's method of nonparamet-





**Figure 3. Posterior estimates of divergence times of a whole Angiosperm tree.** Estimations were done by using MCMCTREE [19] with the IR model [15]. The >65 Ma constraint to the *Zea/Oryza* separation was applied. doi:10.1371/journal.pone.0005297.g003

value even if positive selection operated in some parts of some proteins. Therefore, the branch-site model [25,26] was applied.

To identify positively selected sites, among 61 protein-encoding “genes”, we at first selected 16 “genes”, for which the reverted 2 $\omega$ -model (with the rate ratio  $\omega_1$  of the common ancestral branch of Poaceae larger than the rate ratio  $\omega_0$  of other branches) is significantly better than the 1 $\omega$ -model ( $P < 0.05$ ) (Table 5), and by using the branch-site model [25,26], we identified 5 genes (*atpE*, *cemA*, *clpP*, *rpoB*, and *rps11*) which have  $P$  value of the branch-site LRT less than 0.05 and contain positively selected sites (Table 6).

Among the 16 genes with significantly higher  $\omega_1$  than  $\omega_0$  in Table 5 and among the 5 genes with positively selected sites in Table 6, only *atpE* is among the 14 genes with significant heterogeneity of nucleotide substitution rates for maize vs. rice, maize vs. wheat, or rice vs. wheat comparisons listed in Table 5 of ref. [27]. Four “genes”, *psaC*, *rbcL*, *rpl6*, and PS13, have significantly lower  $\omega_1$  than  $\omega_0$  (stronger purifying selection in the ancestral branch of Poaceae than in other branches). In Table 5, we carried out multiple tests for 61 “genes”. The Bonferroni correction is a safeguard against multiple tests falsely giving the appearance of significance, since 1 out of every 20 hypotheses tests is expected to be significant at the 5% level purely by chance. After

performing the Bonferroni correction, 7 genes with \* in Table 5 remained significant, that means, all the genes listed in Table 6 remained significant even by the conservative test of Bonferroni. On the other hand, the 4 “genes” with lower  $\omega_1$  than  $\omega_0$  in Table 5 were not significant after the Bonferroni correction.

## Discussion

In our study, the IR model gives more consistent results than the CR model, which has been widely used in estimating divergence times [9,12–14,28–34]. A basic assumption of the CR model is that rates change gradually over the tree. Our results suggest that the magnitude of the rate acceleration is underestimated by the CR model and that the IR model is more appropriate in approximating the rate change in the grass chl evolution. Although there exists a case in which the CR model outperforms the IR model [34], a number of authors have recently begun to notice that the IR model is superior to the CR model in approximating the evolution of evolutionary rates in several cases [17,35–37].

Our analysis has revealed an episodic acceleration of the evolutionary rate of chl genomes during the emergence of core Poaceae, accompanied by adaptive evolution in several protein-

**Table 1.** Posterior estimates of divergence times with the >65 Ma constraint to the *Zea/Oryza* separation.

Node	Independent-rates (IR) model		Correlated-rates (CR) model	
	Time (Ma)	Rate	Time (Ma)	Rate
1 ( <i>Triticum</i> )	62.3 (44.4, 84.4)	0.033 (0.009, 0.090)	58.4 (45.5, 76.2)	0.030 (0.014, 0.058)
2 ( <i>Zea</i> )	<b>72.2 (59.1, 93.7)</b>	<b>0.253 (0.137, 0.460)</b>	<b>67.7 (54.3, 86.9)</b>	<b>0.119 (0.087, 0.167)</b>
3 ( <i>Typha</i> )	114.1 (91.4, 144.5)	0.069 (0.029, 0.157)	150.7 (123.2, 179.5)	0.062 (0.033, 0.104)
4 ( <i>Musa</i> )	134.3 (111.9, 165.8)	0.040 (0.012, 0.106)	171.0 (145.1, 197.7)	0.048 (0.027, 0.079)
5 ( <i>Elaeis</i> )	141.0 (117.0, 173.4)	0.046 (0.015, 0.116)	175.6 (149.3, 202.1)	0.065 (0.038, 0.104)
6 ( <i>Phalaenopsis/Yucca</i> )	152.2 (125.7, 185.4)	0.041 (0.012, 0.109)	182.0 (156.0, 208.0)	0.047 (0.028, 0.076)
7 ( <i>Dioscorea</i> )	159.6 (131.8, 193.6)	0.083 (0.035, 0.185)	187.0 (160.7, 212.9)	0.070 (0.043, 0.107)
8 ( <i>Acorus</i> )	187.8 (153.1, 224.4)	0.057 (0.023, 0.136)	216.3 (188.8, 241.5)	0.053 (0.032, 0.085)
9 (Eudicots)	212.5 (175.0, 247.5)	0.035 (0.009, 0.096)	239.1 (211.5, 262.3)	0.055 (0.033, 0.086)
10 ( <i>Drimys</i> )	216.2 (178.0, 251.4)	0.033 (0.008, 0.091)	240.9 (213.4, 263.8)	0.050 (0.029, 0.079)
11 ( <i>Chloranthus</i> )	218.6 (180.1, 254.0)	0.061 (0.024, 0.145)	242.1 (214.6, 264.9)	0.055 (0.034, 0.086)
12 ( <i>Illicium</i> )	240.8 (199.7, 274.7)	0.053 (0.020, 0.125)	263.0 (235.3, 283.9)	0.056 (0.032, 0.091)
13 ( <i>Nymphaea</i> )	262.4 (219.8, 293.3)	0.043 (0.014, 0.108)	280.8 (253.3, 299.5)	0.057 (0.032, 0.092)
14 ( <i>Amborella</i> )	276.9 (233.3, 305.1)	0.054 (0.011, 0.174)	289.8 (262.2, 307.8)	0.059 (0.035, 0.096)
15 (Gymnosperm)	297.2 (280.5, 310.3)	–	299.5 (281.4, 310.6)	–
16	179.4 (142.1, 218.8)	0.057 (0.024, 0.135)	196.7 (166.5, 230.0)	0.039 (0.025, 0.071)
Terminal branch to <i>Oryza</i>	0	0.038 (0.027, 0.052)	0	0.040 (0.030, 0.051)

MCMCTREE in PAML [19] was used with the GTR+ $\Gamma_5$  model. Shape and scale parameters,  $\alpha$  and  $\beta$ , in the gamma prior for parameter  $\sigma^2$  were 1.0 and 10.0, respectively. 95% highest posterior density (HPD) is shown in parentheses. Rate ( $\times 10^{-8}$  substitutions/ nucleotide/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*. doi:10.1371/journal.pone.0005297.t001

**Table 2.** Posterior estimates of divergence times without constraint to the *Zea/Oryza* separation.

Node	Independent-rates (IR) model		Correlated-rates (CR) model	
	Time (Ma)	Rate	Time (Ma)	Rate
1 ( <i>Triticum</i> )	47.2 (28.4, 71.5)	0.040 (0.012, 0.105)	32.4 (21.6, 49.6)	0.059 (0.028, 0.103)
2 ( <i>Zea</i> )	55.0 (35.2, 80.4)	0.212 (0.124, 0.377)	36.9 (24.6, 56.7)	0.119 (0.087, 0.159)
3 ( <i>Typha</i> )	104.2 (78.7, 135.5)	0.063 (0.026, 0.144)	119.8 (90.9, 156.5)	0.059 (0.034, 0.096)
4 ( <i>Musa</i> )	126.8 (105.7, 158.6)	0.040 (0.012, 0.105)	140.8 (113.0, 176.2)	0.043 (0.025, 0.069)
5 ( <i>Elaeis</i> )	133.6 (111.1, 166.1)	0.044 (0.014, 0.112)	145.8 (117.4, 181.1)	0.057 (0.033, 0.091)
6 ( <i>Phalaenopsis/Yucca</i> )	145.3 (120.2, 179.3)	0.041 (0.012, 0.108)	153.2 (124.5, 187.7)	0.043 (0.026, 0.066)
7 ( <i>Dioscorea</i> )	152.6 (125.8, 187.8)	0.080 (0.034, 0.179)	158.7 (129.4, 193.2)	0.059 (0.035, 0.093)
8 ( <i>Acorus</i> )	181.7 (147.1, 220.3)	0.054 (0.022, 0.128)	193.7 (159.4, 226.9)	0.046 (0.027, 0.073)
9 (Eudicots)	207.7 (169.2, 245.1)	0.036 (0.010, 0.098)	220.4 (184.1, 251.4)	0.047 (0.029, 0.072)
10 ( <i>Drimys</i> )	211.3 (172.3, 249.1)	0.034 (0.009, 0.094)	222.6 (186.2, 253.2)	0.044 (0.026, 0.070)
11 ( <i>Chloranthus</i> )	213.6 (174.3, 251.5)	0.060 (0.023, 0.141)	223.9 (187.4, 254.5)	0.047 (0.029, 0.072)
12 ( <i>Illicium</i> )	236.2 (193.9, 272.9)	0.053 (0.020, 0.128)	248.6 (210.3, 277.3)	0.048 (0.027, 0.079)
13 ( <i>Nymphaea</i> )	258.2 (213.6, 291.8)	0.042 (0.014, 0.107)	269.2 (229.4, 296.1)	0.048 (0.027, 0.078)
14 ( <i>Amborella</i> )	272.8 (227.0, 304.2)	0.056 (0.012, 0.177)	279.7 (239.6, 305.8)	0.053 (0.031, 0.086)
15 (Gymnosperm)	296.6 (280.3, 310.2)	–	297.2 (280.4, 310.3)	–
16	176.9 (138.5, 217.4)	0.061 (0.026, 0.139)	178.5 (141.1, 215.1)	0.040 (0.024, 0.068)
Terminal branch to <i>Oryza</i>	0	0.051 (0.032, 0.080)	0	0.074 (0.046, 0.106)

MCMCTREE in PAML [19] was used with the GTR+ $\Gamma_5$  model. Shape and scale parameters,  $\alpha$  and  $\beta$ , in the gamma prior for parameter  $\sigma^2$  were 1.0 and 10.0, respectively. 95% HPD is shown in parentheses. Rate ( $\times 10^{-8}$  substitutions/ nucleotide/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*. doi:10.1371/journal.pone.0005297.t002

**Table 3.** Posterior estimates of divergence times excluding Poaceae.

Node	Independent-rates (IR) model		Correlated-rates (CR) model	
	Time (Ma)	Rate	Time (Ma)	Rate
4 ( <i>Musa/Typha</i> )	115.5 (96.3, 136.7)	0.033 (0.011, 0.083)	110.6 (91.6, 128.6)	0.031 (0.016, 0.052)
5 ( <i>Elaeis</i> )	125.1 (103.9, 149.2)	0.033 (0.012, 0.081)	119.2 (99.0, 138.7)	0.034 (0.022, 0.053)
6 ( <i>Yucca/Phalaenopsis</i> )	139.5 (115.5, 167.3)	0.034 (0.010, 0.086)	130.8 (109.9, 150.3)	0.034 (0.021, 0.052)
7 ( <i>Dioscorea</i> )	147.7 (122.0, 177.2)	0.060 (0.029, 0.128)	137.3 (115.4, 158.1)	0.037 (0.026, 0.056)
8 ( <i>Acorus</i> )	184.6 (151.4, 219.8)	0.041 (0.019, 0.090)	190.8 (162.0, 218.1)	0.037 (0.022, 0.060)
9 (Eudicots)	216.5 (181.5, 248.9)	0.033 (0.010, 0.085)	223.2 (194.0, 246.9)	0.041 (0.027, 0.060)
10 ( <i>Drimys</i> )	220.1 (184.7, 252.6)	0.031 (0.009, 0.081)	225.5 (196.2, 248.9)	0.041 (0.026, 0.062)
11 ( <i>Chloranthus</i> )	222.2 (186.6, 254.9)	0.057 (0.024, 0.128)	226.7 (197.4, 250.2)	0.041 (0.028, 0.060)
12 ( <i>Illicium</i> )	244.7 (206.9, 275.8)	0.050 (0.020, 0.114)	253.8 (222.5, 276.8)	0.045 (0.028, 0.070)
13 ( <i>Nymphaea</i> )	266.3 (227.3, 294.5)	0.040 (0.014, 0.096)	274.8 (241.9, 297.0)	0.044 (0.028, 0.068)
14 ( <i>Amborella</i> )	280.2 (240.4, 305.9)	0.047 (0.011, 0.139)	285.2 (251.9, 307.0)	0.049 (0.031, 0.076)
15 (Gymnosperms)	298.3 (280.8, 310.4)	–	300.3 (282.0, 310.7)	–
16	190.4 (157.5, 223.6)	0.067 (0.031, 0.141)	188.6 (161.2, 213.2)	0.045 (0.030, 0.066)

MCMCTREE in PAML [19] was used with the GTR+ $\Gamma_5$  model. Shape and scale parameters,  $\alpha$  and  $\beta$ , in the gamma prior for parameter  $\sigma^2$  were 1.0 and 10.0, respectively. 95% HPD is shown in parentheses. Rate ( $\times 10^{-8}$  substitutions/nucleotide/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*. doi:10.1371/journal.pone.0005297.t003

encoding genes. Because the elevation of the rate occurred not only in non-synonymous substitutions but also in synonymous substitutions and because the elevated substitution rates were accompanied also by an elevated rate of insertions/deletions of nucleotides [9], the elevation of the mutation rate of chl genomes might have acted as a trigger of the adaptive evolution in the ancestral grasses, which might have facilitated the successful radiation and diversification of their descendants.

Suggested positive selection of *clpP* in *Oenothera* and *Sileneae* accompanied by elevated synonymous rate [38] might be related to our finding of rate acceleration in ancestral grasses both in terms of synonymous and non-synonymous substitutions. A more extensive study of chl genomes showed highly accelerated non-synonymous rates of ribosomal protein and RNA polymerase genes in Geraniaceae accompanied with the elevation of the  $\omega$  ratio [39]. Interestingly, the 4 genes (*atpE*, *cemA*, *rpoB*, and *rps11*) detected to have positively-selected sites in our analysis (Table 6) are included in the gene group with significantly high  $\omega$  ratio in Geraniaceae relative to other angiosperms (*clpP* was not analyzed in ref. [39]).

Recently, Smith and Donoghue [40] tested evolutionary rates across five groups of angiosperms, and found that the rates are generally low in trees/shrubs compared to related herbs. This is an interesting finding which links life history of plants to their rates of molecular evolution, and their conclusion generally holds in five different groups of Angiosperm. What we have shown in this work, however, is that the pattern of rate change during evolution is more complicated than has previously been anticipated. Our finding highlights the need for paying attention to rates of internal branches rather than averaging along a lineage in addressing the rate heterogeneity problem.

## Materials and Methods

Since our main interest was on grass evolution, we used all the monocot genera (13 species) and selected 18 species from outside monocots (31 species in total) among the 64 species in ref. [6]. We used 75 chl genes among 77 protein-encoding genes in ref. [6], excluding *infA* and *ycf2* because of missing data.

**Table 4.** Estimation of non-synonymous/synonymous rate ratio ( $\omega$ ) under different models by using CODEML in PAML [19].

Model	$\omega_0$	$\omega_1$	$\omega_2$	Ln L	LRT with 1 $\omega$ -model	AIC
1 $\omega$	0.1518	–	–	–115,741.5	–	231,485.1
Simple 2 $\omega$	0.1265	0.1617	–	–115,729.5	$9.63 \times 10^{-7}$	231,463.0
3 $\omega$	0.1255	0.2189	0.1246	–115,669.4	$4.95 \times 10^{-32}$	231,344.8
Reverted 2 $\omega$	0.1250	0.2189	–	–115,669.4	$3.29 \times 10^{-33}$	<b>231,342.8</b>

1 $\omega$ -model: (*Musa*# $\omega_0$ , *Typha*# $\omega_0$ , (crown Poaceae # $\omega_0$ ) # $\omega_0$ ).

Simple 2 $\omega$ -model: (*Musa*# $\omega_0$ , *Typha*# $\omega_0$ , (crown Poaceae # $\omega_1$ ) # $\omega_1$ ).

3 $\omega$ -model: (*Musa*# $\omega_0$ , *Typha*# $\omega_0$ , (crown Poaceae # $\omega_2$ ) # $\omega_1$ ).

Reverted 2 $\omega$ -model: (*Musa*# $\omega_0$ , *Typha*# $\omega_0$ , (crown Poaceae # $\omega_0$ ) # $\omega_1$ ).

"Crown Poaceae" includes all Poaceae branches in our tree except for the common ancestral branch (stem Poaceae). The codon-substitution model with the F61 codon frequency was used. Minimum AIC (Reverted 2 $\omega$ -model) is shown in bold italic.

doi:10.1371/journal.pone.0005297.t004



**Table 5.** LRT of 1 $\omega$ -model vs reverted 2 $\omega$ -model for individual genes.

Gene	Ancestral Poaceae $\omega_1$	Other branches $\omega_0$	ln L (reverted 2 $\omega$ )	ln L (1 $\omega$ )	$\chi^2 = 2\Delta\ln L$	P
<i>rpoB</i> *	0.3469	0.1013	-8546.12	-8574.82	57.41	3.53E-14
<i>rps11</i> *	0.7406	0.0538	-1126.87	-1144.90	36.06	1.91E-09
<i>clpP</i> *	0.5943	0.1374	-1767.99	-1778.73	21.47	3.95E-06
<i>atpE</i> *	0.9642	0.1309	-1132.80	-1143.33	21.05	4.48E-06
<i>rps3</i> *	0.5767	0.1273	-1958.45	-1968.07	19.24	1.15E-05
<i>cemA</i> *	1.4337	0.3047	-2214.75	-2224.29	19.09	1.25E-05
<i>rp122</i> *	0.6319	0.1139	-1067.47	-1075.53	16.11	5.97E-05
<i>rpoC1</i>	0.2385	0.1294	-5702.89	-5708.31	10.85	0.000989
<i>atpA</i>	0.2296	0.1102	-4201.36	-4206.20	9.68	0.001860
PS13 <sup>#</sup>	0.0540	0.1663	-3542.69	-3547.23	9.08	0.002578
<i>rps2</i>	0.3623	0.1288	-1886.14	-1890.56	8.83	0.002960
<i>rpoC2</i>	0.3364	0.2179	-10553.23	-10557.46	8.45	0.003642
<i>psaC</i> <sup>#</sup>	0.0001	0.0780	-603.22	-607.35	8.26	0.004047
<i>rbcL</i> <sup>#</sup>	0.0344	0.0937	-3805.48	-3809.48	8.00	0.004669
<i>ndhH</i>	0.1405	0.0592	-3144.67	-3148.58	7.81	0.005202
<i>rps7</i>	1.2673	0.2151	-865.87	-869.76	7.78	0.005297
<i>rps12</i>	0.6567	0.1220	-697.24	-700.69	6.91	0.008589
<i>rps19</i>	0.6328	0.1836	-727.58	-730.62	6.08	0.013646
<i>rp116</i> <sup>#</sup>	0.0268	0.1060	-1131.04	-1134.08	6.07	0.013739
<i>rp14</i>	0.2858	0.0995	-972.76	-975.00	4.47	0.034457

$\omega_1$  and  $\omega_0$  refer to the estimates of non-synonymous/synonymous rate ratios based on the reverted 2 $\omega$ -model. Only 20 genes rejecting the 1 $\omega$ -model with  $P < 0.05$  are listed. In this analysis, 11 monocot species (6 Poaceae species, *Typha*, *Musa*, *Elaeis*, *Phalaenopsis*, and *Yucca*) were used. 75 genes were analyzed, but genes with related functions were concatenated if the lengths are shorter than 180 nucleotides, and therefore the number of tests was 61. PS13 refers to the concatenated sequences of *petG+petL+petN+psaI+psaJ+psbF+psbI+psbJ+psbK+psbL+psbM+psbN+psbT*.

\* refers to a gene which remains significant after the Bonferroni correction.

<sup>#</sup> refers to a gene with  $\omega_1 < \omega_0$ .

doi:10.1371/journal.pone.0005297.t005

## Estimation of divergence times

The concatenated 75 gene sequences of chl from 31 species (from ref. [6]) and the tree topology in ref. [6] were used. To estimate divergence times and molecular evolutionary rates, a Bayesian method implemented in MCMCTREE (in PAML, [19]) was applied either with the CR model [12,13] or with the IR model [15] (using the GTR model with a discrete gamma distribution with five rate categories ( $\Gamma_5$ ) for nucleotide substitutions), and multiple calibrations were incorporated through the time prior. The Gymnosperm/Angiosperm divergence time was set at 280–310 Ma [5,7]. Three

nodes were constrained with minimum ages as follows; (1) the minimum age of the *Zea/Oryza* divergence was set either to 65 Ma [10,11] or without this constraint, (2) >115 Ma constraint to the divergence of Poales from other monocots based on the earliest fossils of Poales [41,42], and (3) >125 Ma for the most basal divergence in eudicots [7]. In order to check the robustness of the time estimation on the choice of the substitution model, the codon-substitution [21,22]+ $\Gamma_5$  model was also used. The program adopts soft bounds, so that the probability that the true divergence time is outside the bounds is small but not zero [43]. In the Bayesian framework, priors are assigned not only on times, but also on the overall substitution rate parameter  $\mu$  and on the rate-drift parameter  $\sigma^2$ . So we roughly estimated the prior mean of the overall rate  $\mu$  using the strict molecular clock with 295 Ma constraint to Gymnosperm/Angiosperm divergence time, and assigned the gamma prior  $G(4, 80)$  and  $G(4, 22)$  for this prior parameter in applying the nucleotide and codon substitution models, respectively. We next examined the impact of the rate-drift parameter  $\sigma^2$  by giving various priors for  $\sigma^2$  in applying the nucleotide substitution model. Posterior distributions of parameters were approximated using two independent MCMC analyses of  $10^7$  steps each, following a discarded burn-in of  $10^6$  steps. All the analyses were repeated with different inseed values to check for convergence of the MCMC chain.

## Non-synonymous/synonymous rate ratio

To the concatenated sequences of 75 protein-encoding genes of chl from 6 Poaceae species (*Oryza*, *Triticum*, *Hordeum*, *Zea*,

**Table 6.** Branch-site test of positive selection.

gene	LRT	Positively selected sites
<i>atpE</i>	0.0484	2T→K, 17S→C, 41A→N, 64M→W, 132V→W
<i>cemA</i>	0.0021	55N→R, 76Y→K, 161W→F, 190I→F, 204I→C
<i>clpP</i>	0.0081	26R→V, 48V→T, 86F→T, 112I→P, 134E→R, 182T→D
<i>rpoB</i>	0.0352	90R→F, 338G→K, 1026G→N
<i>rps11</i>	0.0082	54V→P, 62A→S, 82A→R, 105L→S, 115R→A, 120L→R

The numberings of amino acids are those of *Zea mays* [45]. Positively selected sites were inferred at  $P_b = 95\%$  with those reaching 99% shown in bold italic. The analyses were carried out for the 16 genes selected in Table 5, and only genes with positively selected sites and with  $P < 0.05$  (LRT) are listed.

doi:10.1371/journal.pone.0005297.t006

*Saccharum*, and *Sorghum*), *Typha*, and *Musa*, we applied the codon-based likelihood models that allow for variable  $\omega$  ratios among different lineages [44]. We used the likelihood ratio test (LRT) to compare the likelihood of one- $\omega$  ratio model, which assumes the same  $\omega$  for all branch in the tree, with the two- $\omega$  ratio models, which assumes two different  $\omega$  ratios. One of the two-ratio models (named “Simple 2 $\omega$ -model”) assumes that Poaceae (including the common ancestral branch) has different  $\omega$  from other parts of the tree as is represented by

$$(Musa\#\omega_0, Typha\#\omega_0, (\text{crown Poaceae } \#\omega_1)\#\omega_1),$$

while the other (named “Reverted 2 $\omega$ -model”) assumes that only the ancestral branch of Poaceae has a different  $\omega$  ratio than all the other branches in the tree as is represented by

$$(Musa\#\omega_0, Typha\#\omega_0, (\text{crown Poaceae } \#\omega_0)\#\omega_1).$$

All the analyses were carried out with the CODEML program in PAML [19] using the codon-substitution model with the F61 codon frequency.

### Branch-site test of positive selection

The branch-site test was applied to the dataset of 11 monocot species in our dataset excluding the two most basal monocots; *i.e.*, *Dioscorea* and *Acorus*. The branch preceding the common ancestor of the core Poaceae was specified as a foreground branch, and all the others as background branches. LRT is constructed to compare an alternative model that allows for some codons under positive selection on the foreground branch with a null model that does not. The null model restricts codons on the foreground lineage to be undergoing neutral evolution ( $\omega = 1$ ). The specific codons which evolved under positive selection were identified on the foreground branch using a Bayes empirical Bayes procedure [25,26].

### Supporting Information

**Table S1** Impact of the shape and scale parameters ( $\alpha$  and  $\beta$ ) in the gamma prior for parameter  $\sigma^2$  using IR model with the >65 Ma constraint to the *Zea/Oryza* separation. 95% HPD is shown in parentheses. Times and rates are represented in 100 Ma ( $10^8$  years ago) and  $10^{-8}$  substitutions/site/years, respectively. Found at: doi:10.1371/journal.pone.0005297.s001 (0.04 MB DOC)

**Table S2** Impact of the shape and scale parameters ( $\alpha$  and  $\beta$ ) in the gamma prior for parameter  $\sigma^2$  using CR model with the >65 Ma constraint to the *Zea/Oryza* separation. 95% HPD is shown in parentheses. Times and rates are represented in 100 Ma ( $10^8$  years ago) and  $10^{-8}$  substitutions/site/years, respectively. Found at: doi:10.1371/journal.pone.0005297.s002 (0.04 MB DOC)

**Table S3** Impact of the shape and scale parameters ( $\alpha$  and  $\beta$ ) in the gamma prior for parameter  $\sigma^2$  using IR model without the

constraint to the *Zea/Oryza* separation. 95% HPD is shown in parentheses. Times and rates are represented in 100 Ma ( $10^8$  years ago) and  $10^{-8}$  substitutions/site/years, respectively.

Found at: doi:10.1371/journal.pone.0005297.s003 (0.04 MB DOC)

**Table S4** Impact of the shape and scale parameters ( $\alpha$  and  $\beta$ ) in the gamma prior for parameter  $\sigma^2$  using CR model without the constraint to the *Zea/Oryza* separation. 95% HPD is shown in parentheses. Times and rates are represented in 100 Ma ( $10^8$  years ago) and  $10^{-8}$  substitutions/site/years, respectively.

Found at: doi:10.1371/journal.pone.0005297.s004 (0.04 MB DOC)

**Table S5** Posterior estimates of divergence times by MCMCTREE in PAML [19] using the codon-substitution+ $\Gamma_5$  model with the >65 Ma constraint to the *Zea/Oryza* separation. Shape and scale parameters,  $\alpha$  and  $\beta$ , in the gamma prior for parameter  $\sigma^2$  were 1.0 and 10.0, respectively. 95% HPD is shown in parentheses. Rate ( $10^{-8}$  substitutions/codon/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*.

Found at: doi:10.1371/journal.pone.0005297.s005 (0.04 MB DOC)

**Table S6** Posterior estimates of divergence times by MCMCTREE in PAML [19] using the codon-substitution+ $\Gamma_5$  model without constraint to the *Zea/Oryza* separation. Shape and scale parameters,  $\alpha$  and  $\beta$ , in the gamma prior for parameter  $\sigma^2$  were 1.0 and 10.0, respectively. 95% HPD is shown in parentheses. Rate ( $10^{-8}$  substitutions/codon/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*.

Found at: doi:10.1371/journal.pone.0005297.s006 (0.04 MB DOC)

**Table S7** Posterior estimates of divergence times by MCMCTREE in PAML [19] using the codon-substitution+ $\Gamma_5$  model excluding Poaceae. Shape and scale parameters,  $\alpha$  and  $\beta$ , in the gamma prior for parameter  $\sigma^2$  were 1.0 and 10.0, respectively. 95% HPD is shown in parentheses. Rate ( $10^{-8}$  substitutions/codon/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*.

Found at: doi:10.1371/journal.pone.0005297.s007 (0.04 MB DOC)

### Acknowledgments

We thank James Crabbe for improving the manuscript. Simon Joly provided thoughtful comments that helped improve the manuscript.

### Author Contributions

Conceived and designed the experiments: BZ MH. Analyzed the data: BZ TY. Wrote the paper: BZ TY YZ MH.

### References

- Jacobs BF, Kingston JD, Jacobs LL (1999) The origin of grass-dominated ecosystems. *Ann Mo Bot Gard* 86: 590–643.
- Gaut BS, Muse SV, Clark WD, Clegg MT (1992) Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *J Mol Evol* 35: 292–303.
- Bousquet J, Strauss SH, Doerksen AH, Price RA (1992) Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proc Natl Acad Sci USA* 89: 7844–7848.
- Muse SV, Gaut BS (1997) Interlocus comparisons of the nucleotide substitution process in the chloroplast genome. *Genetics* 146: 393–399.
- Chaw S-M, Chang C-C, Chen H-L, Li W-H (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58: 424–441.
- Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in



- angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104: 19369–19374.
7. Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA* 49: 19363–19368.
  8. Martin W, Deusch O, Stawski N, Grunheit N, Goremykin V (2005) Chloroplast genome phylogenies: why we need independent approaches to plant molecular evolution. *TRENDS in Plant Science* 10: 203–209.
  9. Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, et al. (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol Biol Evol* 22: 1948–1963.
  10. Prasad V, Stroemberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. *Science* 310: 1177–1180.
  11. Piperno DR, Sues HD (2005) Dinosaurs dined on grass. *Science* 310: 1126–1128.
  12. Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15: 1647–1657.
  13. Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol* 18: 352–361.
  14. Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol* 14: 1218–1232.
  15. Rannala B, Yang Z (2007) Inferring speciation times under an episodic molecular clock. *Syst Biol* 56: 453–466.
  16. Huelsenbeck JP, Larget B, Swofford D (2000) A compound Poisson process for relaxing the molecular clock. *Genetics* 154: 1879–1892.
  17. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88.
  18. Wolfe KH, Gouy MY, Yang W, Sharp PM, Li W-H (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc Natl Acad Sci USA* 86: 6201–6205.
  19. Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 24: 1586–1591.
  20. Vicentini A, Barber JC, Alicioni SS, Giussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of C<sub>4</sub> photosynthesis. *Global Change Biol* 14: 2963–2977.
  21. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11: 725–736.
  22. Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11: 715–724.
  23. Yang Z (2006) *Computational Molecular Evolution*. Oxford: Oxford Univ Press.
  24. Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, eds. *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, pp 267–281.
  25. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 1–8.
  26. Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acids sites under positive selection. *Mol Biol Evol* 22: 1107–1118.
  27. Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K (2002) Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. *Mol Biol Evol* 19: 2084–2091.
  28. Bremer K (2002) Gondwanan evolution of the grass alliance of families (Poales). *Evolution* 56: 1374–1387.
  29. Sanderson MJ, Thorne JL, Wikstrom N, Bremer K (2004) Molecular evidence on plant divergence times. *Amer J Bot* 91: 1656–1665.
  30. Aris-Brosou S, Yang Z (2003) Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Mol Biol Evol* 20: 1947–1954.
  31. Nikaido M, Matsuno F, Hamilton H, Brownell Jr RL, Cao Y, et al. (2001) Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proc Natl Acad Sci USA* 98: 7384–7389.
  32. Hasegawa M, Thorne J, Kishino H (2003) Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet Syst* 78: 267–283.
  33. Yoder AD, Yang Z (2004) Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol Ecol* 13: 757–773.
  34. Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Mol Biol Evol* 24: 2669–2680.
  35. Kitazoe Y, Kishino H, Waddell PJ, Nakajima N, Okabayashi T, et al. (2007) Robust time estimation reconciles views of the antiquity of placental mammals. *PLoS ONE* 2: e384.
  36. Brown JW, Rest JS, Garcia-Moreno J, Sorenson MD, Mindell DP (2008) Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biology* 2008, 6: 6.
  37. Renner SS, Grimm GW, Schneeweiss GM, Stuessy TF, Ricklefs RE (2008) Rooting and dating maples (*Acer*) with an uncorrelated-rates molecular clock: Implications for North American/Asian disjunctions. *Syst Biol* 57: 795–808.
  38. Erixon P, Oxelman B (2008) Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS ONE* 3: e1386.
  39. Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc Natl Acad Sci USA* 105: 18424–18429.
  40. Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science* 322: 86–89.
  41. Herendeen PS, Crane PR (1995) The fossil history of the monocotyledons. In: Rudall PJ, Cribb PJ, Cutler DF, Humphries CJ, eds. *Monocotyledon: Systematics and Evolution*. London: Royal Botanic Gardens, Kew, pp 1–21.
  42. Linder HP, Rudall PJ (2005) Evolutionary history of Poales. *Annu Rev Ecol Evol Syst* 36: 107–124.
  43. Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 23: 212–226.
  44. Yang Z, Bielawski B (2000) Statistical methods for detecting molecular adaptation. *TREE* 15: 496–503.
  45. Maier RM, Neckermann K, Igloi GL, Kossel H (1995) Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251: 614–628.